Ontology Driven Clinical Decision Support for Early Diagnostic Recommendations

Volume I

Gopikrishnan Mannamparambil Chandrasekharan

A thesis submitted in fulfilment of the requirements

for a PhD in Health Informatics

Department of Computer Science

School of Mathematics, Computer Science & Engineering

City University of London

Supervisors:

Dr Dympna O'Sullivan

Dr Andrew MacFarlane

October 2018

Table of contents

List of figures

List of tables

Key to abbreviations

Abbreviation	Full form
AAE	American Academy of Endodontics
AERS	Adverse Event Reporting System
AMD	Age-related macular degeneration
AOP	Association of Optometrists
APDG	Advanced Patient Data Generator
API	Application Program Interface
AT	Argumentation Theory
BFO	Basic Formal Ontology
CC	Chief Complaint
CDR	Cognitive Dispositions to Respond
CDSS	Clinical Decision Support System(s)
CET	Continuing Education and Training
CHF	Chronic Heart Failure
COPD	Chronic Obstructive Pulmonary Disease
DDS	Dental Diagnostic System
DIM	Dental Information Model
DL	Description Logics
DS Model	Disease-Symptom Model
EHR	Electronic Health Record(s)
EMR	Electronic Medical Record(s)
FDA	Federal Drug Administration
FHIR	Fast Healthcare Interoperability Resources
FN	False Negative
FOAF	Friend of a Friend
FOL	First Order Logic
FP	False Positive
GOC	General Optical Council
GP	General Practitioner(s)
GUI	Graphical User Interface
HES	Hospital Eye Service
HL7	Health Level Seven
HTTP	Hypertext Transfer Protocol
IOP	IOP
MCAD	Mast Cell Activation Disorder
NDCG	Normalized Discounted Cumulative Gain
NICE	National Institute for Health and Clinical Excellence
NTG	Normal Tension Glaucoma
ODO	Ocular Disease Ontology
OGMS	Ontology for General Medical Science
OHD	Oral Health and Disease Ontology

OHT	Ocular Hypertension
OMB	Oculomotor Balance
OWL	Web Ontology Language
PHR	Personal Health Record(s)
POAG	Primary Open Angle Glaucoma
RCT	Root Canal Treatment
RDF	Resource Description Framework
RDFS	RDF Schema
REST	Representational state transfer
RIF	Rule Interchange Format
SNODENT	Systematized Nomenclature of Dentistry
SNOMED CT	Systematized Nomenclature of Medicine - Clinical Terms
SPARQL	SPARQL Protocol and RDF Query Language
SPIN	SPARQL Inferencing Notation
SUS	System Usability Scale
SWRL	Semantic Web Rule Language
TN	True Negative
TP	True Positive
UMLS	Unified Medical Language System
URI	Uniform Resource Identifier
URL	Uniform Resource Locator
vs	Versus
W3C	World Wide Web Consortium
WWW	World Wide Web

Glossary

Term	Definition
Early diagnostic recommendations	Diagnostic recommendations that are provided by the CDSS early in the diagnostic encounter before all the data from the patient has been collected
Late diagnostic recommendations	Diagnostic recommendations that are provided by the CDSS late in the diagnostic encounter after all the data from the patient has been collected
Early guideline recommendations	Guideline recommendations that are provided by the CDSS during the diagnostic encounter before all the data from the patient has been collected
Pulpal	Pertaining to the pulp of the tooth
Apical	Pertaining to the apex (root tip) of the tooth
Joint observation	Combination of two or more clinical observations to form a single observation

Acknowledgements

I would like to thank my PhD supervisors, Dr Dympna O'Sullivan and Dr Andrew MacFarlane for their patience, encouragement and guidance during my PhD. I would like to thank Dr Peter Weller for his guidance during the early stages of my PhD.

I would like to thank Dr Ian Murdoch, Professor John Lawrenson and Dr Joy Myint for their advice and support during my research. I would especially like to express my gratitude towards Dr Ian Murdoch for all the support he has provided me during my time in the U.K. I would like to thank Mr Pak Sang Lee for all the help he has provided.

I am grateful to Professor John Fox and others at OpenClinical.net for their support during the research.

I would like to thank my fellow PhD students at City University for their constant encouragement and companionship during my time at City. I would like to thank Dr Vaisakh G. Nair, Dr Ajith K.V and Dr Ninan Philipose for their help and support in my research.

Finally, but more importantly, I would like to thank my family, especially my wife Amba, my father M.N Chandrasekharan, my mother Daya Chandrasekharan, sister Devi Sujit and family (Sujit Ravi, Anika and Ameya) for all their love, support, patience, kindness and understanding without which I wouldn't have been able to complete this thesis.

Declaration

I hereby grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without further reference to me.

Abstract

Diagnostic error is a significant problem in medicine and a major cause of concern for patients and clinicians and is associated with moderate to severe harm to patients. Diagnostic errors are a primary cause of clinical negligence and can result in malpractice claims. Cognitive errors caused by biases such as premature closure and confirmation bias have been identified as major cause of diagnostic error. Researchers have identified several strategies to reduce diagnostic error arising from cognitive factors. This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information. Clinical Decision Support Systems (CDSSs) have been shown to reduce diagnostic errors.

Clinical guidelines improve consistency of care and can potentially improve healthcare efficiency. They can alert clinicians to diagnostic tests and procedures that have the greatest evidence and provide the greatest benefit. Clinical guidelines can be used to streamline clinical decision making and provide the knowledge base for guideline based CDSSs and clinical alert systems. Clinical guidelines can potentially improve diagnostic decision making by improving information gathering.

Argumentation is an emerging area for dealing with unstructured evidence in domains such as healthcare that are characterized by uncertainty. The knowledge needed to support decision making is expressed in the form of arguments. Argumentation has certain advantages over other decision support reasoning methods. This includes the ability to function with incomplete information, the ability to capture domain knowledge in an easy manner, using non-monotonic logic to support defeasible reasoning and providing recommendations in a manner that can be easily explained to clinicians. Argumentation is therefore a suitable method for generating early diagnostic recommendations. Argumentation-based CDSSs have been developed in a wide variety of clinical domains. However, the impact of an argumentation-based diagnostic Clinical Decision Support System (CDSS) has not been evaluated yet.

The first part of this thesis evaluates the impact of guideline recommendations and an argumentation-based diagnostic CDSS on clinician information gathering and diagnostic decision making. In addition, the impact of guideline recommendations on management decision making was evaluated. The study found that argumentation is a viable method for generating diagnostic recommendations that can potentially help reduce diagnostic error. The study showed that guideline recommendations do have a positive impact on information gathering of optometrists and can potentially help optometrists in asking the right questions and performing tests as per current standards of care. Guideline recommendations were found to have a positive impact on management decision making. The CDSS is dependent on quality of data that is entered into the system. Faulty interpretation of data can lead the clinician to enter wrong data and cause the CDSS to provide wrong recommendations.

Current generation argumentation-based CDSSs and other diagnostic decision support systems have problems with semantic interoperability that prevents them from using data from the web. The clinician and CDSS is limited to information collected during a clinical encounter and cannot access information on the web that could be relevant to a patient. This is due to the distributed nature of medical information and lack of semantic interoperability between healthcare systems. Current argumentation-based decision support applications require specialized tools for modelling and execution and this prevents widespread use and adoption of these tools especially when these tools require additional training and licensing arrangements.

Semantic web and linked data technologies have been developed to overcome problems with semantic interoperability on the web. Ontology-based diagnostic CDSS applications have been developed using semantic web technology to overcome problems with semantic interoperability of healthcare data in decision support applications. However, these models have problems with expressiveness, requiring specialized software and algorithms for generating diagnostic recommendations.

The second part of this thesis describes the development of an argumentation-based ontology driven diagnostic model and CDSS that can execute this model to generate ranked diagnostic recommendations. This novel model called the Disease-Symptom Model combines strengths of argumentation with strengths of semantic web technology. The model allows the domain expert to model arguments favouring and negating a diagnosis using OWL/RDF language. The model uses a simple weighting scheme that represents the degree of support of each argument within the model. The model uses SPARQL to sum weights and produce a ranked diagnostic recommendation. The model can provide justifications for each recommendation in a manner that clinicians can easily understand. CDSS prototypes that can execute this ontology model to generate diagnostic recommendations were developed. The decision support prototypes demonstrated the ability to use a wide variety of data and access remote data sources using linked data technologies to generate recommendations. The thesis was able to demonstrate the development of an argumentation-based ontology driven diagnostic decision support model and

decision support system that can integrate information from a variety of sources to generate diagnostic recommendations. This decision support application was developed without the use of specialized software and tools for modelling and execution, while using a simple modelling method.

The third part of this thesis details evaluation of the Disease-Symptom model across all stages of a clinical encounter by comparing the performance of the model with clinicians. The evaluation showed that the Disease-Symptom Model can provide a ranked diagnostic recommendation in early stages of the clinical encounter that is comparable to clinicians. The diagnostic performance can be improved in the early stages using linked data technologies to incorporate more information into the decision making. With limited information, depending on the type of case, the performance of the Disease-Symptom Model will vary. As more information is collected during the clinical encounter the decision support application can provide recommendations that is comparable to clinicians recruited for the study. The evaluation showed that even with a simple weighting and summation method used in the Disease-Symptom Model the diagnostic ranking was comparable to dentists. With limited information in the early stages of the clinical encounter the Disease-Symptom Model was able to provide an accurately ranked diagnostic recommendation validating the model and methods used in this thesis.

1 Chapter 1 - Introduction

1.1 Motivation and outline

Medical diagnosis is a key component of the clinical encounter and forms the basis for understanding the problem(s) faced by the patient and helps the clinician implement a suitable treatment. Diagnosis is essentially constructing a hypothesis about the disease a patient is suspected of having based upon direct and indirect, observations and findings obtained from the patient during clinical examination. However diagnostic error is a significant problem and can occur in 5-20% of the cases (Sonderegger-Iseli *et al.*, 2000; Singh, Meyer and Thomas, 2014). It is a major cause of concern for patients and clinicians, and is associated with moderate and severe harm to the patient (Sevdalis *et al.*, 2010; Saber Tehrani *et al.*, 2013; Singh *et al.*, 2013a). Diagnostic errors are a primary cause of negligence on behalf of the clinician and can result in malpractice claims (Ely, Kaldjian and D'Alessandro, 2012a; Graber, 2013; Saber Tehrani *et al.*, 2013).

There are several reasons for diagnostic errors and Graber et al (Graber, Gordon and Franklin, 2002a) classified diagnostic errors based on aetiology into:

- No-fault errors
    - Masked or unusual presentation of disease
    - Patient-related error (uncooperative, deceptive or inconsistent patients)
    - Limited medical knowledge about condition
- System-related errors
    - Technical failure and equipment problems
        - E.g. Lack of appropriate equipment
    - Organizational failures
        - E.g. Policy failures
- Cognitive errors
    - Faulty knowledge
    - Faulty data gathering
    - Faulty information gathering
    - Faulty information synthesis

Diagnostic errors are considered multifactorial and several factors can influence the decision-making process of the clinician. Graber's later work (Graber, 2005) showed that cognitive errors are far more common than errors caused by knowledge gaps. It showed that the main type of cognitive error was premature closure bias. Premature closure bias is the tendency of the clinician to stop early and not order further tests or inquire further information to prove/disprove his/her hypothesis. Premature closure bias will prevent the clinician from gathering more information and integrating this information to formulate a new hypothesis. Clinician accept the diagnosis before it has been verified and do not seek other information (Dailey, 1952; Wellbery, 2011). Confirmation bias is the tendency of the clinician to look for data that supports his/her hypothesis (Wellbery, 2011). Even when presented with data that is of superior quality the clinician may rely on inferior or ambiguous data. For example, the clinician may rely on a history of fever as reported by the patient. Based on this history of fever the clinician may arrive at a diagnostic conclusion and look for more information that supports this wrong conclusion. There are other types of cognitive errors and Croskerry (Croskerry, 2002) has collectively called them 'Cognitive Dispositions to Respond' (CDR). He has identified several cognitive debiasing strategies to reduce the occurrence of CDR's. These include considering alternatives, decreasing reliance on memory, providing access to clear, concise and well-organized information.

Diagnostic CDSSs have been identified by researchers as a key intervention to reduce diagnostic errors (Graber, Gordon and Franklin, 2002a; Croskerry, 2003c; Croskerry, Singhal and Mamede, 2013c). However studies have shown that CDSSs have not been adopted for widespread clinical use even though it has improved other areas of clinical practice, such as: improved adherence to guidelines, improved prescribing, and improved use of preventive healthcare measures (Garg *et al.*, 2005a). The performance of current diagnostic decision support tools like differential diagnosis generator systems has been average so far (the best system scored 3.45 out of 5) (Bond *et al.*, 2012a; Umscheid and Hanson, 2012). These

diagnostic CDSS tools have not shown a marked improvement in performance even after years of continuous development since they were evaluated nearly two decades back (Berner *et al.*, 1994a).

The reasons for this lack of impact on diagnostic decision making are not apparent from these studies, but researchers believe it is partly due to the quality of input data. As the famous maxim goes "Garbage in, garbage out". Cognitive errors play the key role during the information gathering phase of the clinical encounter. Cognitive errors (example confirmation bias) lead the clinician (medical doctor or dentist) to biased hypotheses which can in turn influence the input data leading to further diagnostic errors or an incomplete diagnosis. Clinical guidelines have been suggested as a possible solution that could positively impact diagnostic decision making by standardising the process of collecting, analysing and verifying clinical data. Clinical guidelines have several benefits and limitations (Woolf *et al.*, 1999). Clinical guidelines improve consistency of care and can potentially improve healthcare efficiency. They can alert clinicians to diagnostic tests and procedures that have the greatest evidence and provide the greatest benefit. Clinical guidelines can be used to streamline clinical decision making and provide the knowledge base for guideline based CDSSs and clinical alert systems (Garg *et al.*, 2005a). Improving the information gathering of clinicians can potentially improve diagnostic performance. Evidence based guideline recommendations can help improve the information gathering of clinicians by providing timely alerts to the clinician about abnormal values and explaining the importance of the abnormal value and why it's important. Guideline recommendations can provide the clinician with information on how to interpret tests and investigations results. Similarly, guideline recommendations can help the clinician by providing evidence-based guidance about the next steps to be taken such as tests/investigations to be ordered.

In a recent review on the use of health information technology to reduce diagnostic errors (El-Kareh, Hasan and Schiff, 2013) the authors classified existing tools into different categories:

- Tools that assist in information gathering
- Tools that improve cognition through improved organization and display of information
- Tools that assist in generating a differential diagnosis
- Tools that assist in weighing the diagnosis
- Tools that support selection of diagnostic tests and plans
- Tools that enable access to diagnostic reference material and guidelines
- Tools that enable patient follow-up and continued assessment of patient course and response to treatment
- Tools that support screening by alerting presence of disease
- Tools that assist in diagnostic collaboration
- Tools that facilitate feedback and insight into diagnostic performance

A differential diagnosis is the process of "distinguishing of a disease or condition from others presenting with similar signs and symptoms"(Merriam-Webster, 2016). As the clinician is enquiring about the patient's condition, completing the physical examination and performing diagnostic tests he/she is formulating the diagnosis in his/her mind. The clinician is creating a mental model of the patient's condition and is constantly generating a differential diagnosis in the mind. Clinicians rely on the differential diagnosis to identify presence or absence of a disease when there are multiple alternatives available (Heneghan *et al.*, 2009). A clinician examining a patient goes through a multi-step procedure to attain a differential diagnosis. According to the hypothetico-deductive model of clinical reasoning (Elstein, Schwartz and Schwarz, 2002) clinicians (especially novice clinicians) obtain a diagnosis following the steps described below (Harbison, 1991).

- At first, they obtain information by taking clinical history, complete the physical examination and conduct diagnostic tests if needed.
- A working diagnosis (hypothesis) is generated and the clinicians consider all possible diagnoses in the form of a mental list.
- Further information is sought from the patient to confirm or refute the working diagnosis (if needed).
- The diagnosis is verified by weighting and combining information from all sources and then arriving at a conclusion that confirms one of the working diagnoses.

Illness script theory provides a framework that describes how clinicians organize knowledge obtained from clinical literature and experience (Schmidt and Rikers, 2007a). Based on observations seen in the patient certain illness scripts are triggered in the mind of the clinician. The clinician will traverse the network of knowledge in their memory gained from experience, supplemented by biomedical knowledge, to find the script that best suits/explains the clinical situation in front of him/her. In this manner, the clinician will generate several different hypotheses and reject the hypothesis that does not fit his/her mental model and arrive at a diagnosis. Illness scripts are an efficient way of organizing knowledge and allows the clinician to rapidly seek information, search for the appropriate illness script, find the right script and verify the values (Charlin *et al.*, 2007).

The Dual Process Theory explains how individuals make decisions and judgments under uncertainty with limited information (Kahneman and Egan, 2011; Evans and Stanovich, 2013). It can be adapted to explain how a clinician reasons when presented with a clinical case. Making a diagnostic decision in the early stages of the clinical encounter is a classic example of this scenario as he/she clinician is faced with limited information and is working under uncertainty regarding the patient's condition. The dual process theory consists of 2 types of processes, System 1 and System 2 (Evans and Stanovich, 2013). System 1 process of reasoning is used by clinicians to make rapid judgements about a patient. It is mostly activated in cases with routine clinical presentations and is dependent on the clinician's experience. It is used by busy clinicians to deal with patients with familiar clinical presentations. System 1 requires some previous experience and knowledge about the patient. If the patient has atypical signs and symptoms, and the clinician is unfamiliar with the case then the clinician slows down the process of reasoning. This is called System 2 process of reasoning. In System 2 the reasoning is slow, methodical and analytical. Both System 1 and System 2 process of reasoning work in tandem to help the clinician in formulating a diagnosis.

Croskerry et al. (Croskerry, 2002, 2003a) have suggested several techniques to prevent premature closure bias and confirmation bias. This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information. Other proposed techniques include allowing more time for clinical consultation, use of clinical algorithms, making available local disease prevalence statistics, use of CDSSs. (Phua and Tan, 2013). Croskerry et al have presented a review of cognitive debiasing strategies that can be adopted to tackle cognitive related diagnostic errors (Croskerry, Singhal and Mamede, 2013a, 2013c) which includes obtaining clinical information in a more structured manner. There is evidence suggesting that early diagnostic recommendations can help reduce diagnostic error (Kostopoulou *et al.*, 2015a).

However, a problem facing current diagnostic CDSS applications is that its ability to support decision making is limited to the data collected during the clinical encounter. Most of the tools that claim to assist the clinician in diagnosis by providing a differential diagnosis or help in weighing the diagnosis at the end of a clinical encounter after collecting all data. The reason why CDSS applications have been designed in this manner is because of the nature of clinical diagnosis in a clinical encounter. The data collected during the early stages of the clinical encounter will be incomplete.

However, to reduce diagnostic errors arising from premature closure bias and confirmation bias a CDSS should be able to able to provide some recommendations early in the diagnostic encounter. CDSSs should recommend alternative diagnoses, provide justifications for making the recommendation, provide the recommendation and justifications in a format that is easy to understand by the clinician. If the CDSS makes the recommendations early in the encounter, then the clinician can look at the alternate diagnoses and reasons and compare that with the mental model created in the clinician's mind. As the clinician is obtaining information from the patient the clinician is forming a list of diagnostic hypotheses. A list of alternate diagnoses provided by the CDSS can help the clinician consider diagnoses that he/she may not have considered without the help of the CDSS. The reasons for making the recommendation can help trigger illness scripts. This can guide the clinician in asking the next set of questions and open a new line of investigation. The list of alternate diagnosis and reasons can help slow down the thinking process of the clinician and help improve System 2 process of reasoning.

Similarly, guideline recommendations have a potential impact on helping improve information gathering. Guideline recommendations can alert abnormal values and can provide a context for these values. It can provide clinical evidence suggesting clinical conditions the abnormal values can be seen. This can trigger illness scripts for a diagnosis. The guideline recommendations can provide the clinician with information on how to interpret tests and investigations results. Similarly, the guideline recommendations can help the clinician by providing evidence-based guidance about the next steps to be taken such as tests/investigations to be ordered. All these steps can potentially improve information gathering of clinicians. Improved information presentation results in the clinician having all the information needed to make a diagnosis. Therefore, improving information gathering can help reduce diagnostic errors indirectly (Thammasitboon and Cutrer, 2013). Guideline recommendations have been shown to have an impact on management decisions as well (Glasspool *et al.*, 2003).

Argumentation Theory (AT) is an emerging area for dealing with unstructured evidence in domains such as healthcare that are characterized by uncertainty (Fox *et al.*, 2007). The knowledge needed to support decision making is expressed in the form of arguments. AT explains how the arguments can be expressed, sustained and rejected during the process of decision making. For each decision a set of positive and negative arguments are created. Argumentation is a model that is based on construction and evaluation of competing arguments (Ouerdane, Maudet and Tsoukiàs, 2010). AT evolved as a sub-discipline of philosophical logic and is an important area of logic-based AI research. In recent years it has found several applications in cognitive science and information technology (Fox *et al.*, 2007).

Argumentation uses non-monotonic logic and therefore is capable of retracting existing conclusions and formulate new conclusions. Diagnostic hypotheses formulated early in the clinical encounter are always tentative. As more information is obtained a clinician confirms or rejects a hypothesis. Similarly, argumentation-based CDSSs has the capacity to support defeasible reasoning, where initial conclusions are discarded, and new ones formed considering current information. Argumentation is well suited for generating diagnostic recommendations in the presence of incomplete information. Argumentation based CDSSs model arguments using qualitative methods. It does not rely on complex numerical probabilities for modelling uncertainty. Argumentation is therefore able to provide reasons and explanations for making the diagnosis in a manner that is easy to understand by clinicians and is presented in a structured manner. Argumentation uses a process very similar to that of clinicians. Clinicians use arguments in routine decision making. The clinician looks at the list of diagnostic hypotheses generated and cognitively makes arguments for and against each diagnosis. If the arguments supporting a diagnosis are very strong then the clinician finds more information supporting his/her diagnostic hypotheses. However, the CDSS should be able to provide arguments against a diagnosis to help the clinician overcome confirmation bias. An argumentation-based CDSS can provide arguments favouring and negating a diagnosis. Due to all these factors an argumentation-based diagnostic CDSS is well suited for providing early diagnostic recommendations to reduce diagnostic error. However, it is not known if an argumentation-based CDSS can help in reducing diagnostic error. Studies so far using argumentation-based CDSS have focussed efforts in developing risk assessment and treatment planning applications in medicine. There are a few argumentation-based diagnostic CDSSs. The studies have not evaluated how well an argumentation-based CDSS can help reduce diagnostic error and if argumentation is a feasible approach for developing diagnostic CDSSs.

Therefore, the first part of this thesis investigates feasibility and impact of an argumentation-based diagnostic CDSS on diagnostic performance of primary care clinicians. The study investigates impact of guideline recommendations on information gathering. Optometrists were recruited for this study. The study used clinical vignettes simulating real life clinical scenarios and optometrists were asked to diagnose and manage these vignettes. Clinical guideline recommendations were provided at various stages of the clinical encounter. Argumentation-based diagnostic recommendations were similarly provided at the end of the simulated clinical encounter. In addition, impact of guideline recommendations on management decision making was studied.

The results of the study showed that guideline recommendations do have a role in helping improve the optometrist's information gathering especially in cases where there is a clear pathology. The guidelines can help the optometrists ask the right questions and perform the right tests and investigations. The argumentation-based diagnostic recommendations helped improve diagnostic performance of optometrists in some cases, especially in sight-threatening cases. Guideline recommendations were shown to improve management decision making. Optometrists found the CDSS easy to use and recommendations useful and easy to understand.

This study demonstrated feasibility of using argumentation-based diagnostic recommendations in diagnostic CDSSs and how the diagnostic recommendations can help reduce diagnostic error. However current generation argumentation-based CDSSs use proprietary formats and need specialized tools for modelling and execution. CDSS applications built with current generation argumentation-based CDSS tools require extensive interface engines to be able to integrate with Electronic Health Records (EHRs) and other sources of patient data. This lack of semantic interoperability has hampered widespread adoption of argumentation-based diagnostic CDSSs and diagnostic CDSSs in general (Marco-Ruiz and Bellika, 2015).

The lack of semantic interoperability results in CDSSs not being able to integrate data from different healthcare and other potential sources of data and use them in diagnosis. The problem facing clinicians or a CDSS supporting the clinician at present is he/she/CDSS is unable to access this information as healthcare information is locked away in information silos that do not communicate with each other. Problems coordinating data from multiple sources is one reason for cognitive errors (Graber, Gordon and Franklin, 2002b). The clinician or CDSS cannot this access this additional source of information to support diagnostic decision making. The more information available to decide, the better the clinical decision making. The CDSS should be able to integrate all this information and present results in a manner that is easily understood by clinicians.

There are reference materials available online that can be used, however, the challenge is to link the CDSS to these data sources and help clinicians compare the data collected from the patient with this reference material. The data needed for decision making can reside in other clinical practices and data relevant for current diagnosis may reside in different data repositories. Similarly, patients are increasingly getting access to Personal Health Records (PHRs) (Tang *et al.*, 2006) which are controlled by the patient and more patients are recording information on online platforms such as social networking sites relevant to healthcare (Wicks *et al.*, 2010). The CDSS must traverse these disparate databases particularly on the web to find information that is relevant to the diagnosis and help clinicians in diagnostic decision making.

Semantic web and linked data technologies have been developed to help improve semantic interoperability. Semantic web tools and standards such as RDF, OWL, SPARQL are parts of a stack of technology that works together to realise the goal of semantic interoperability. RDF (Resource Description Framework) is a data model that lets you capture the meaning of information in the form of triples. It is a flexible data model and is being considered as a universal language for healthcare (Booth, 2014b, 2015c). RDF enables smarter data use and automated data translations. Multiple data models and vocabularies can be easily combined and interrelated using RDF. RDF captures the information present in data and not the syntax. RDF is easy to map from one data representation to another and RDF is self-describing. OWL (Web Ontology Language) is the de-facto standard language for representing ontologies on the web and is based on RDF and Description Logics (DL). SPARQL is both a protocol and query language for RDF databases called triple stores. It can model rules and access remote data via SPARQL endpoints. All these properties make RDF a suitable candidate for building a diagnostic inference model in OWL language that can integrate information from disparate sources using SPARQL.

Ontology driven diagnostic CDSSs have been built before that tried to address the issue of semantic interoperability. Due to limitations in modelling uncertainly using the OWL language researchers have tried several methods to overcome the problem. Some researchers have captured the degree of uncertainty outside the ontology model using external probabilistic tools to generate diagnostic

recommendations (García-Crespo *et al.*, 2010a). Other methods include representing the diagnostic criteria using OWL axioms (Ongenae *et al.*, 2010; Donfack Guefack *et al.*, 2012; Bau, Chen and Huang, 2014; Sherimon and Krishnan, 2016). However due to limits in expressiveness of DL-based axioms these methods are not well suited for representing uncertainty and cannot work with incomplete data. Most ontology-based CDSSs identified in literature do not provide a ranked diagnostic recommendation. The ontology-based diagnostic CDSSs that provide a ranked recommendation rely on complex ranking algorithms and specialized software to generate the ranking (Oberkampf *et al.*, 2012). The diagnostic CDSSs do not explain the reasons for making the recommendation by making use of argumentation or similar methods.

This thesis proposes the development of an argumentation-based ontology driven diagnostic model and CDSS. The model combines the strengths of argumentation with the strengths of semantic web technology. The proposed model provides a novel method that allows the domain expert to capture arguments favoring and negating a diagnosis in OWL/RDF model. The model overcomes problems of integrating information and semantic interoperability faced by current generation argumentation-based CDSSs. The model uses argumentation and currently available semantic web and linked data technologies to provide a ranked diagnostic recommendation that can be generated using any off-the-shelf tools without the need for specialized software and algorithms. All the information needed to generate the diagnostic recommendation will be captured within the same model. This simplifies the modelling process and enables easier sharing for collaboration. The proposed model and CDSS can generate diagnostic recommendations without the need for specialized software and algorithms using widely available standards and tools. CDSS prototypes that can use the ontology model to generate diagnostic recommendations will be developed. The CDSS prototypes will demonstrate how the proposed model called Disease-Symptom Model (DS Model) can help integrate information from a wide variety of sources to provide a ranked diagnostic recommendation.

The first part of this thesis evaluates the impact of guideline recommendations and an argumentation-based diagnostic CDSS on clinician information gathering and diagnostic decision making. In addition, the impact of guideline recommendations on management decision making was evaluated. The study found that argumentation is a viable method for generating diagnostic recommendations that can potentially help reduce diagnostic error. The study showed that guideline recommendations do have a positive impact on information gathering of optometrists and can potentially help optometrists in asking the right questions and performing tests as per current standards of care. Guideline recommendations were found to have a positive impact on management decision making. The decision support system is dependent on quality of data that is entered into the system. Faulty interpretation of data can lead the clinician to enter wrong data and cause the decision support system to provide wrong recommendations.

Current generation argumentation-based clinical decision support systems and other diagnostic decision support systems have problems with semantic interoperability that prevents them from using data from the web. The clinician and clinical decision support system is limited to information collected during a clinical encounter and cannot access information on the web that could be relevant to a patient. This is due to distributed nature of medical information and lack of semantic interoperability between healthcare systems. Current argumentation-based decision support applications require specialized tools for modelling and execution and this prevents widespread use and adoption of these tools especially when these tools require additional training and licensing arrangements.

Semantic web and linked data technologies have been developed to overcome problems with semantic interoperability on the web. Ontology-based diagnostic clinical decision support applications have been developed using semantic web technology to overcome problems with semantic interoperability of healthcare data in decision support applications. However, these models have problems with expressiveness, requiring specialized software and algorithms for generating diagnostic recommendations.

In the second part of this thesis, the DS Model was developed using semantic web and linked data technologies and can integrate information from disparate sources such as web-based patient data and provide a recommendation across all stages of the clinical encounter. The domain expert can use this

model to capture arguments favouring and negating a diagnosis. The weights that represent a degree of support of each argument can be represented in the same model. Using SPARQL rules a CDSS can access patient data and update the model for observations that are present or absent. SPARQL can sum the weights for each argument and generate a ranked diagnostic recommendation. The reasons for making a recommendation can be provided. CDSS architectures that utilize this model and generate the recommendations for the clinician have been developed and described in separate case studies. Case studies demonstrate how the DS Model can be used with a wide variety of patient information and can access information from remote sources with help of SPARQL. An extension to DS model was developed that demonstrated how the model can be extended to generate sub-diagnostic recommendations.

In the third part of this thesis, the DS Model was evaluated across different stages of the clinical encounter. An implementation of the DS Model in dentistry was developed for the purposes of evaluation. An existing information model currently used in dentistry was used as the basis to define different stages of the clinical encounter, for example, Clinical Tests, Radiographic Examination.

The evaluation measured the effectiveness of the diagnostic recommendation in early and late stages of the clinical encounter using dentistry as a test domain. This evaluation was performed using clinical vignettes verified by domain experts. Primary care dentists were recruited for the study and were asked to diagnose the dental clinical vignettes. The performance of the CDSS in all stages of the clinical encounter was compared to the ranked final diagnostic recommendation of the dentists. Normalized Discounted Cumulative Gain and Precision-Recall metrics were used for evaluating the performance.

The evaluation identifies sections of the diagnostic encounter where the DS Model is deficient when compared to dentists and could potentially be enhanced with additional data obtained using linked data technologies. The results show that the diagnostic performance of the dental instance of the DS Model is comparable to dentists across most stages of the clinical encounter and improves in the later stages eventually providing a ranked diagnostic recommendation that is better than the dentists recruited in this study. As more information is obtained the diagnostic performance of the DS Model improves. In the later stages of the clinical encounter when confirmatory evidence from clinical tests and radiographic examination is obtained the CDSS was able to provide diagnostic recommendations that are better than dentists recruited for this study. If the clinicians are provided with right diagnostic recommendations, then there is a potential for increased savings in time and money as the patient is not subject to unnecessary tests and radiographs to confirm or reject a diagnosis. In this study all the information needed to decide was provided to both the dentists and the CDSS. In real-life situations in the presence of incomplete information a decision support tool that can integrate information can be valuable even in the late stages.

The DS Model showed a dip in clinical performance in the intermediate stages of the clinical encounter. The clinicians using the DS Model based CDSS needs to be aware of the potential for this occurrence. The overall ranking of the diagnostic recommendation made by the DS Model was comparable to that of dentists. The study demonstrated how the DS Model built using argumentation-based methods and semantic web technology can deliver effective diagnostic recommendations that are comparable to dentists across all stages of the clinical encounter.

1.2 Aims

The aim of this thesis is a) to investigate the role of clinical guideline and argumentation-based diagnostic recommendations on primary care diagnosis and information gathering; b) to investigate and develop an argumentation-based model that can use data from disparate sources to assist in providing diagnostic recommendations across all stages of the clinical encounter; c) to develop CDSS prototypes that can use the model to deliver the diagnostic recommendations; d) to evaluate the model at different stages of the clinical encounter to see how a CDSS implementing the model will perform and identify areas of deficient performance of the model.

1.3 Research Questions

1. How do guideline recommendations and argumentation-based diagnostic recommendations affect information gathering and diagnostic decision making in primary care clinicians?
2. Can an argumentation-based diagnostic CDSS developed using semantic web and linked data technologies generate early diagnostic recommendations?
3. How does an ontology driven diagnostic CDSS perform across all stages (early and late) of a clinical encounter?

1.3.1 The specific objectives of this research are:

Objective 1: To review the literature on models of diagnostic reasoning and identify models used by diagnostic CDSSs to represent knowledge and formulate diagnostic recommendations and their role in reducing diagnostic error. Review state of the art web technologies to identify how they address the problem of semantic interoperability and their application in CDSSs.

Objective 2: Develop a diagnostic CDSS based on current argumentation-based CDSS technology and evaluate its effect on primary care clinician decision making and information gathering.

Objective 3: Develop a semantically interoperable diagnostic inference model that can integrate information from disparate sources and uses argumentation to generate diagnostic recommendations.

Objective 4: Implement a working CDSS prototype that implements the semantically interoperable diagnostic inference model and provides diagnostic recommendations.

Objective 5: Evaluate the CDSS prototype and the diagnostic inference model and its ability to support diagnostic decision making at all stages of the clinical encounter.

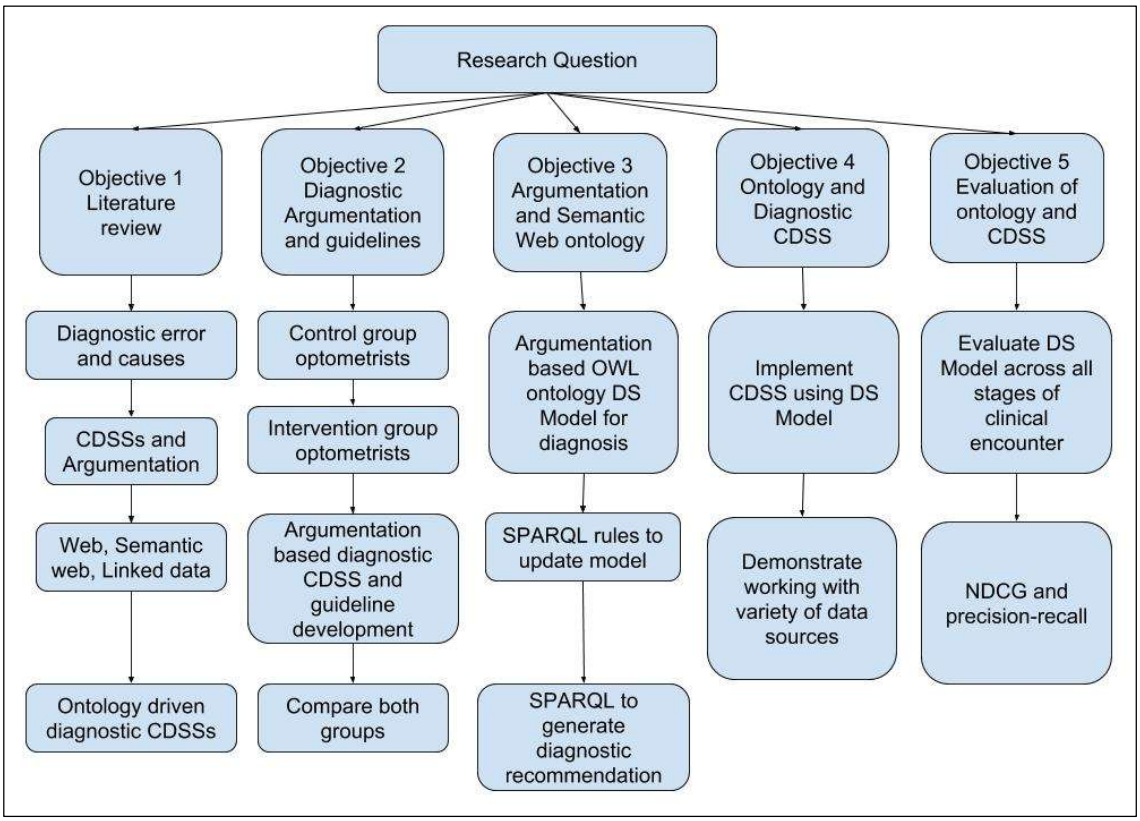


Figure 1-1 - Overall methodology

1.4 Contributions

The thesis evaluated feasibility and impact of argumentation-based CDSSs on diagnostic performance of optometrists. An off-the-shelf argumentation-based modelling tool and execution engine was used to

develop a diagnostic CDSS that can generate diagnostic recommendations. The thesis found that argumentation is a viable method for generating diagnostic recommendations that can potentially help reduce diagnostic error especially in sight-threatening cases in optometry. The thesis found that guideline recommendations do have a positive impact on information gathering of optometrists and can potentially help optometrists in asking the right questions and performing tests as per current standards of care. Guideline recommendations were found to have a positive impact on management decision making. The CDSS is dependent on the quality of data that is entered into the system. The faulty interpretation of medical images can lead the clinician to enter the wrong data and causing the CDSS to provide wrong clinical recommendations and supporting the wrong diagnostic hypothesis of the clinician. This problem can be remedied by improving the quality and quantity of arguments used for diagnosis. The optometrists found the CDSS easy to use and found its recommendations useful. This is one of the first studies looking at evaluating guideline based CDSSs amongst community optometrists. Results of this study can be used to better design and develop guideline-based and diagnostic CDSSs for optometry.

This thesis then continues from study with optometrists and develops an argumentation-based ontology driven diagnostic CDSS model and CDSS prototypes that can generate early diagnostic recommendations. The ontology model called DS Model can capture domain expertise in an ontology model using arguments in favour and against each diagnosis. This novel approach to developing ontology driven diagnostic CDSSs combines argumentation and current semantic web technologies. DS Model can be modelled using any off-the-shelf the ontology modelling tools. The thesis expands upon research in areas of argumentation-based CDSSs and ontology based CDSSs. DS Model developed for this thesis takes the strengths of argumentation (where the modelling process is a lot easier and the diagnostic recommendations can be explained to the clinician) and combines it with the strength of semantic web technologies. The current generation argumentation-based CDSS tools use proprietary software and formats and have problems with semantic interoperability. Using semantic web technologies, the thesis was able to demonstrate the ability of CDSSs using the DS Model to integrate data in a wide variety of formats to generate diagnostic recommendations. The ontology model used a simple weighting method developed for the optometry study CDSS prototype to model the degree of support of each argument favouring and negating a diagnosis. The ontology model demonstrated a novel summation method using SPARQL that can generate diagnostic recommendations without the use of specialized software and algorithms and expands upon current generation ontology driven diagnostic CDSSs. The thesis explored the development of an extension to this ontology-driven diagnostic model that can provide recommendations for sub-diagnostic states where necessary.

Lastly, an evaluation of the DS Model in all stages of a clinical encounter was performed. The evaluation showed that the DS Model can provide a ranked diagnostic recommendation in early stages of the clinical encounter that is comparable to dentists. The diagnostic performance can be improved using linked data technologies to incorporate more information into the decision making. As more information is collected, depending on the type of case, the performance of the DS Model decreases, and ranking adjusts itself to provide ranked recommendation that is better than dentists. The evaluation that showed even with a simple weighting method and summation method used in the DS Model the diagnostic recommendation was comparable to dentists. With limited information in the early stages of the clinical encounter the DS Model was able to provide an accurate ranked diagnostic recommendation validating the model and methods used in this thesis.

1.5 Summary

This chapter introduces the thesis. The next chapter of this thesis describes a review of literature in diagnostic error, diagnostic CDSSs, argumentation and ontology driven diagnostic CDSSs. In Chapter 3, the research outline has been described. Chapter 4 describes a study involving community optometrists that investigates the role of clinical guideline and diagnostic recommendations on information gathering of optometrists and diagnostic decision making. Chapter 5 details the development of an argumentation-based diagnostic inference model that is built on current semantic web standards and details development of CDSS prototypes that can access disparate datasets to support decision making. Chapter 6 discusses the evaluation of DS model and a CDSS prototype by comparing its performance with community

dentists to test ability of the CDSS using the diagnostic inference model to support diagnostic decision making across all stages of the clinical encounter.

2 Chapter 2 - Literature review

2.1 The problem of diagnostic error

The Institute of Medicine, the health division of the National Academy of Sciences published a crucial report called *'To Err is Human: Building A Safer Health System'* (Kohn, Corrigan and Donaldson, 1999). It highlighted several key problems with the healthcare system in the United States and pointed out that between 44,000 and 98,000 deaths per years can be attributed to preventable medical errors. Most these deaths are due to medication errors or surgical errors, but diagnostic errors play a key role in preventable deaths (up to 17 percent of cases). More recent studies have shown that diagnostic error rate has not shown much difference even with advances in technology and improvement in reporting of errors (Graber, 2013). Diagnostic errors are an important problem facing clinical practice as it often leads to improper and unnecessary investigations and management. It can result in delayed treatment causing significant harm or even death. Since the publication of the Institute of Medicine report several steps have been taken to address concerns highlighted by the report, including accelerated adoption of EHR systems (Leape and Berwick, 2005), however diagnostic errors have not received much attention when compared to other aspects of patient safety (Wachter, 2010). Researchers from different fields including medical cognition, medical education and health informatics have been working for the past several decades to understand and address problems in patient safety arising due to diagnostic error (Berner, 2009). However, these groups have been working in relative isolation with few instances of collaboration. Recently several independent organizations have been established with the sole purpose of bringing together various stakeholders involved in patient safety and clinical diagnosis (www.improvediagnosis.org).

One of the reasons why diagnostic errors have been side-lined in patient safety research is because of the difficulty in getting good estimates about frequency of diagnostic errors in various settings. Measuring diagnostic errors have been a challenge and several methods have been identified, including autopsy reports, self-reports, data from malpractice claims and medical records reviews, but they all have their limitations (Graber, 2013). A recent study attempted estimating the frequency of outpatient diagnostic errors in US adult population by collating data from other studies. They estimated that approximately 12 million patients (5%) in the US could potentially experience diagnostic errors each year (Singh, Meyer and Thomas, 2014). Studies have estimated that one-half of diagnostic errors have potential to cause severe harm (Singh *et al.*, 2013a).

Systematic reviews have found that diagnostic errors have been identified in a wide range of conditions and wide range of settings (Ely, Kaldjian and D'Alessandro, 2012a). Conditions like cancer are misdiagnosed due to their rarity, due to their atypical presentation or because of features common to other conditions that are less serious, but more commonly found. Comorbidity causes one condition to mask presence of another condition and can result in missed diagnosis (Kostopoulou, Delaney and Munro, 2008a; Zwaan *et al.*, 2010). The most common reason leading to error was assigning a common benign diagnosis to a less common but serious diagnosis (Ely, Kaldjian and D'Alessandro, 2012a).

In the UK, General Practitioners (GP) are first point of contact and most do not perform high risk procedures and prescribe medications with serious side-effects. Diagnosis is a key part of primary care practitioner's daily routine and therefore diagnostic errors have a greater potential for doing harm in primary care than in any other setting (Kostopoulou, Delaney and Munro, 2008a; Jones *et al.*, 2014). In some chronic conditions like Chronic Obstructive Pulmonary Disease (COPD) the opportunity for early diagnosis is being missed in vast majority of cases(up to 85%), in 5 years before diagnosis of condition (Jones *et al.*, 2014).

Many errors were related to clinical processes, especially patient-clinician encounter, during steps like eliciting medical history, performing physical examination and ordering diagnostic tests (Singh *et al.*, 2013a). This shows that clinical presentation was highly influential in affecting the accuracy of diagnosis in primary care. When compared to other types of medical errors, once the diagnostic error has been

made there is little scope for remedial measures to reduce the impact on the patient as treatment of the patient in most cases is dependent on diagnosis and a delay in treatment can adversely affect prognosis (Bradford *et al.*, 2009). Wrong diagnosis can lead to invasive treatments or incorrect diagnostic tests that can in turn cause unnecessary harm (Epner, Gans and Graber, 2013). Diagnostic errors must be prevented and any intervention to reduce these errors must be proactive. Therefore, there is a need to develop tools to improve data gathering and synthesis during the patient-clinician encounter. There is some evidence suggesting that providing clinicians with early diagnostic recommendations can help improve diagnostic accuracy (Kostopoulou *et al.*, 2015b).

When nature of medical expertise was explored by researchers for understanding causes of diagnostic errors they realised that an expert clinician reasons differently in comparison to a novice clinician. Expert clinicians form an accurate diagnosis more efficiently and effectively, using less clinical information and produce less diagnostic errors (Groves, O'Rourke and Alexander, 2003; Crespo, Torres and Recio, 2004). The main reason for this better performance is not necessarily because expert clinicians know more when compared to intermediates or novices. In fact evidence suggests that with increasing expertise there is a decreased application of biomedical knowledge (Boshuizen and Schmidt, 1992). It is argued that this increase in diagnostic performance is due to improved organisation of clinical knowledge memory by expert clinicians through several years of clinical practice (Schmidt and Rikers, 2007b). The ability of the clinician to maintain context and coherence during a diagnostic problem solving task is key to reducing diagnostic error (Sibbald, Panisko and Cavalcanti, 2011). Context for diagnostic decision making can include clinical history, clinical signs and symptoms, clinician, patient and other environmental factors (Teunissen *et al.*, 2009). Experts can quickly and efficiently build a composite picture of the patient of a series of clinical observations based on the context in which these observations are made.

2.1.1 Cognitive causes of diagnostic errors

Diagnosis is primarily a cognitive task and medical diagnosis is essentially a categorization task that allows a clinician to make sense of the patient data in front of him/her and make appropriate treatment decisions. Human error plays a significant role in diagnostic error even if other systemic factors may be primarily responsible. Clinical diagnosis is considered an example of the naturalistic decision-making model, i.e. decision-making in the real-world settings.

Factors like risk, time pressure, stress, uncertainty, distractions, ambiguity, uncertain shifting and conflicting goals, along with resource and organizational restrictions can complicate the decision making process and contribute to errors (Patel, Kaufman and Arocha, 2002; Klein, 2008a). It is difficult to identify what the exact cause of the error was due to problems in self-reporting and review. Difficulties such as hindsight bias and poor memory can affect results of self-reporting by clinicians. Diagnostic errors are not immediately noticed and there are problems with delayed feedback from secondary care regarding accurate diagnosis (Kostopoulou, 2008). While diagnostic errors are a serous cause of concern, the exact cause of diagnostic errors is still a matter of debate.

Researchers are of the opinion in a vast majority of cases, especially in domains like internal medicine (Graber, Franklin and Gordon, 2005a) and emergency medicine (Kachalia *et al.*, 2007), cognitive errors either alone or in conjunction with other factors, play a crucial role in the formulation of error. As classified by Graber (Graber, Gordon and Franklin, 2002b), cognitive errors can be due to faulty knowledge of the clinician, faulty data gathering, faulty information synthesis or faulty metacognition.

Cognitive biases and heuristics have been studied extensively as a cause of diagnostic error (Croskerry, 2002). Croskerry (Croskerry, 2003b) has given an extensive list of cognitive biases and heuristics, collectively called Cognitive Dispositions to Respond, that may lead to diagnostic error. Some key ones include (Norman, 2009):

- Availability bias: The clinician will choose a diagnosis based on the diagnostic hypothesis that can be easily retrieved from memory. If a disease has not been observed by the clinician for a long time, then chances of it being considered as a diagnosis goes down.
- Base rate neglect: Clinicians sometimes tend to ignore true prevalence rate of disease. Clinicians

can consider rare and exotic diseases when other more common and prevalent diseases may need to be considered.

- Representativeness bias: Clinicians can be misguided by the features of a disease. The prototypical features of a disease may help in diagnosing common disorders, but they can mislead the clinician when a disease with atypical features presents.
- Premature closure bias: It is the tendency of a clinician to prematurely halt the diagnostic decision-making process before all data has been obtained and diagnosis has been fully verified. The clinician does not consider all alternative diagnoses. Most diagnostic errors are due to premature closure biases (Ely, Kaldjian and D'Alessandro, 2012b).
- Confirmation bias: This can lead the clinician to look for evidence to support his/her diagnostic hypothesis, rather than looking for evidence to reject.

Graber et al. (Graber, Gordon and Franklin, 2002a; Graber, Franklin and Gordon, 2005b) has identified several reasons for cognitive errors. The key factors include:

- Faulty knowledge or skills:
  - The clinician has inadequate knowledge about the patient's condition and this predisposes to cognitive errors
  - The clinician does not have required skill in a clinical condition
- Faulty data gathering:
  - Clinician has problems coordinating data from multiple sources
  - Clinician fails to collect appropriate information during history and examination
  - Incorrect diagnostic test ordered
- Faulty information processing
  - Clinician fails to generate appropriate context
  - Clinician overestimates/underestimates the relevance of a finding
- Faulty information synthesis- Most common cause of cognitive based errors
  - Premature closure bias
  - Clinician fails to follow up on an appropriate test

It has been identified that diagnostic error is multi-factorial in nature and one error can lead to another. Cognitive errors are the most common cause of diagnostic error and faulty information synthesis is the main cause of cognitive related diagnostic error (Graber, Franklin and Gordon, 2005b). Gaps in physician knowledge or application of knowledge are a major cause of cognitive related errors (Zwaan *et al.*, 2010).

2.1.2 What can be done about cognitive errors?

Researchers have identified several possible solutions/interventions to reduce the role played by cognitive errors while acknowledging that diagnostic errors due to cognitive factors cannot be completely eliminated (Graber, Gordon and Franklin, 2002a; Thammasitboon and Cutrer, 2013).

Measures such as improving the quality of medical training (Bordage, 1999; Trowbridge, Dhaliwal and Cosby, 2013), problem based learning (Hmelo, 1998), metacognitive training (Croskerry, 2003c, 2003a) have been proposed as possible solutions to reduce cognitive errors. Other proposed solutions include improving perception, availing second opinions (Kronz, Westra and Epstein, 1999) and improving availability of expertise (Espinosa and Nolan, 2000).

Clinical guidelines are another solution that could impact diagnostic decision making by standardising the process of collecting, analysing and verifying clinical data. Clinical guidelines have several benefits and limitations (Woolf *et al.*, 1999). Clinical guidelines improve consistency of care and can potentially improve healthcare efficiency. They can alert clinicians to diagnostic tests and procedures that have the greatest evidence and provide the greatest benefit. Clinical guidelines can be used to streamline clinical decision making and provide the knowledge base for guideline based CDSSs and clinical alert systems (Garg *et al.*, 2005a).

The most important limitation of clinical guidelines is the potential for flawed recommendations (Lenzer, 2013a; Boudoulas *et al.*, 2015). Evidence for one particular sub-group may not translate into benefit for the patient in front of the clinician (Institute of Medicine, 2011). Clinicians find guidelines from different organizations conflicting and confusing (Hogeveen *et al.*, 2012). Rigid clinical guidelines and algorithms may not truly reflect the complexity of daily clinical practice and can lead to frustration on behalf of clinicians (Woolf *et al.*, 1999).

Recently there have been numerous debates concerning credibility and scientific basis of clinical guidelines, especially regarding treatment recommendation guidelines, and the presence of biased guidelines (Lenzer, 2013b). Though clinical guidelines have been shown to be effective in reducing treatment and medication errors there is very little evidence to show a similar impact on diagnostic errors (Ioannidis and Lau, 2001). Guidelines themselves are heuristics, or rules of thumb to manage complex problems in a limited time frame, and the inherent biases or preferences of a guideline development body may influence recommendations that a guideline provides thereby undermining local considerations and professional autonomy. Clinicians themselves may feel available guideline is not applicable for their patient and may reject a guideline believing their patient is somehow unique or atypical. This phenomenon is called an Aggregate bias (Saposnik, Redelmeier, Christian C Ruff, *et al.*, 2016).

Croskerry (Croskerry, 2003a) has identified several strategies to reduce diagnostic errors to 'Cognitive Dispositions to Respond' (Croskerry, 2002). This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information. Other proposed techniques include allowing more time for clinical consultation, use of clinical algorithms, making available local disease prevalence statistics and use of CDSSs (Phua and Tan, 2013). Croskerry et al have presented a review of cognitive debiasing strategies that can be adopted to tackle cognitive related diagnostic which includes obtaining clinical information in a more structured manner (Croskerry, Singhal and Mamede, 2013b, 2013d).

Researchers are of opinion that diagnostic CDSSs and other computer aided diagnostic tools will help reduce cognitive errors by reducing the cognitive burden and improving decision making (Graber *et al.*, 2012a; Croskerry, Singhal and Mamede, 2013c; El-Kareh, Hasan and Schiff, 2013; Thammasitboon and Cutrer, 2013). CDSSs help in mitigating system related factors that cause diagnostic errors as well (Schiff and Bates, 2010; Singh *et al.*, 2012a). A low-tech counterpart of diagnostic CDSS is to use checklists, but they can become too cumbersome when dealing with complex diagnostic scenarios (Ely, Graber and Croskerry, 2011; Sibbald, de Bruin and van Merrienboer, 2013). In trying to reduce diagnostic errors, there is a need to consider a problem of trade-offs. CDSSs can slow down the decision-making process if it requires clinicians to go through multiple checks and numerous data entry fields to reach a diagnosis. Therefore, CDSSs should be able to reach an optimum balance between monitoring and preventing cognitive errors and maintaining efficiency of diagnostic decision making. There are problems with existing CDSSs as most CDSSs do not impact the data gathering and synthesis phase of the patient-clinician encounter and this affects the quality of the data gathered (Wright, Phansalkar, Bloomrosen, Robert A Jenders, *et al.*, 2010).

Diagnostic reasoning models such as the hypothetico-deductive model for clinical reasoning (Elstein, Schwartz and Schwarz, 2002) explain how clinicians (especially novice clinicians) obtain a diagnosis following steps described below (Harbison, 1991).

- At first, they obtain information by taking clinical history, complete physical examination and conduct diagnostic tests if needed.
- Some working hypotheses are generated, and clinician considers all possible diagnoses in the form of a mental list.
- Further information is sought from patient to confirm or refute working hypotheses (if needed).
- The diagnosis is verified by weighing and combining information from all sources and then arriving at a conclusion that confirms one working hypotheses.

The working hypotheses for a clinical case are generated early in the diagnostic encounter based on a few clinical patterns that are observed in a patient. These working hypotheses may not always be accurate and

are tested by acquiring additional information from a patient. Some hypotheses are rejected, and other hypotheses are added as a clinical encounter progresses. The working hypotheses are generally limited in number as human memory only allow a few diagnoses to be stored simultaneously (Miller, 1956). Patterns used to trigger diagnostic hypotheses are more developed in experts than in novice clinicians. According to Evans and Patel (Patel, Evans and Groen, 1989), experts rarely use biomedical knowledge and causal reasoning to arrive at diagnoses. Early in career a novice clinician is most likely to primarily rely on biomedical knowledge for diagnostic reasoning and this biomedical knowledge is subsequently replaced by cognitive heuristics to help identify the diagnostic hypotheses (Patel, Arocha and Kaufman, 1994). The 'Illness script' is a widely accepted concept that tries to provide a framework for understanding how clinicians organize the knowledge that they gain from literature and clinical experience and use it to diagnose and treat patients (Charlin, Tardif and Boshuizen, 2000). A script can be described as a set of attributes each with a set of values.

Based on observations present in the patient, the illness script(s) for that condition is triggered and clinicians looks for default value in a patient (Schmidt and Rikers, 2007a). If default value is not present or unacceptable values are found for attributes, another illness script is triggered, and clinicians looks for attributes and values in that script. The clinician will traverse the network of knowledge in their memory gained from experience, supplemented by biomedical knowledge, to find a script that best suits/explains the clinical situation in front of him/her. In this manner, the clinician will generate several different hypotheses and reject the hypothesis that does not fit his/her mental model and arrive at a diagnosis. Illness scripts are an efficient way of organizing knowledge and allows a clinician to rapidly seek information, search for an appropriate illness script, find the right script and verify values (Charlin *et al.*, 2007).

Using illness script(s) model, the working hypotheses is triggered early in the diagnostic encounter. Generating the working hypotheses early in the encounter helps the clinician to narrow the solution space. Using the working hypothesis, further information is gathered until one of the working hypotheses is confirmed. The diagnostic CDSS should help the clinician in generating the working hypotheses and help in gathering additional information about the patient to help confirm or refute working hypotheses. The diagnostic CDSS must then help in confirming the diagnosis after all information has been collected and act as an aid to suggest alternative diagnoses if needed.

The Dual Process Theory explains how individuals make decisions and judgments under uncertainty with limited information (Kahneman and Egan, 2011; Evans and Stanovich, 2013). It can be adapted to explain how a clinician reasons when presented with a clinical case. Making a diagnostic decision in the early stages of a clinical encounter is a classic example of this scenario. The clinician is faced with limited information and is working under uncertainty regarding the patient's condition. The dual process theory consists of 2 types of processes, System 1 and System 2 (Evans and Stanovich, 2013).

System 1 is the process that is associated with pattern recognition. It is intuitive, efficient and fast and is used to make most common decisions (Saposnik, Redelmeier, Christian C. Ruff, *et al.*, 2016). It is mostly activated in cases with a routine clinical presentation and is dependent on the clinician's experience. It is used by busy clinicians to deal with patients with familiar clinical presentations. Both Illness scripts and System 1 thinking require some previous experience with a case or at least knowledge of the most common signs and symptoms of the disorder. However diagnostic errors are highly associated with atypical presentations (Kostopoulou, Delaney and Munro, 2008b). When the clinical presentation of a case is atypical, and a clinician is not familiar with the case, the clinical reasoning automatically switches to System 2 process of reasoning.

System 2 is the process that is associated with slow, methodical, analytical reasoning. This process is deliberate and is most commonly activated when a patient presents with an atypical clinical presentation. An atypical presentation forces the clinician to slow down and consider all probable causes for the patient's current condition. Medical students and clinicians with less experience tend to use System 2 for reasoning and are therefore slower and more methodical when trying to diagnose a patient.

System 1 and System 2 reasoning work in tandem and support each other (Saposnik, Redelmeier, Christian C. Ruff, *et al.*, 2016). When a clinician takes time to consider all the signs and symptoms and reflect on the patient's condition, then System 2 can override System 1 thinking. System 1 helps quickly identify patterns and formulate hypothesis, while System 2 slows down the thinking process and prevents the clinician from forming a diagnostic hypothesis without considering all the clinical data.

The diagnostic CDSS should be able to support both System 1 and System 2 processes. Using checklists and considering alternatives have been proposed to improve System 1 reasoning process (Graber *et al.*, 2012b). Using evidence based-medicine and normative decision making have been proposed as measures to improve System 2 reasoning process (Graber *et al.*, 2012b). Other measures proposed include the using guidelines and algorithms to reduce reliance on memory and provided focussed decision support on a specific condition.

However, an issue with providing early diagnostic recommendations is unavailability of decent quality information to generate accurate diagnoses by the CDSS. The poor quality of existing/historical data in EHR systems is a well-known problem (Wagner and Hogan, 1996). This in turn affects the quality of the results generated by a CDSS. If poor data quality can be addressed, existing data in EHR systems can help in identifying previous symptoms, signs, clinical tests that may be relevant to the patient's current problem and help in generating early diagnostic recommendations. However, if a patient doesn't have a previous history in the EHR that is relevant to the current problem, the diagnostic CDSS will not have enough information to generate accurate diagnostic recommendations. Guideline and algorithms have been proposed by researchers are measured to help reduce diagnostic error (Graber *et al.*, 2012b). Evidence based guideline recommendations can potentially help improve information gathering of clinicians by providing timely alerts to a clinician about abnormal values and explaining importance of an abnormal value and why it's important. The guideline recommendations can provide clinicians with information on how to interpret tests and investigations results. Similarly, guideline recommendations can help a clinician by providing evidence-based guidance about next steps to be taken including tests/investigations to be ordered.

When the clinician asks a patient a question regarding health and patient gives an answer, this data is recorded in EHR. A guideline-based CDSS can look up the data in the EHR and provide alerts. The alerts can indicate the type of diagnoses where the abnormal values can be found. The alerts can help trigger illness scripts which may not be triggered without the help of CDSSs. Guideline recommendations can then guide a clinician towards appropriate tests to be taken. Illness scripts triggered by abnormal values and corresponding guidance can help a clinician formulate diagnostic hypotheses. The clinician forms a mental model of next steps to be taken. The clinician can then compare his/her mental model with that of the guideline recommendations. If the mental model matches that of the CDSS recommendation the clinician can confidently continue along line of questioning or investigations. Guideline recommendations can help a clinician interpret test results and potentially reduce errors in interpretation. All these steps improve the quality of data collected during the clinical encounter. Guideline recommendations can help in improving management decision made after a diagnosis by providing evidence-based guidance of next steps to be taken after a diagnosis.

Guideline recommendations can therefore potentially improve information gathering of clinicians by reducing reliance on memory by acting as memory aids and providing context specific and patient specific recommendations in a structured format. If information gathering of clinicians improves, then it ensures that a clinician has all the information needed to make a diagnosis. If all information is available to a clinician, then it could correspondingly help reduce diagnostic errors caused by biases such as premature closure where the clinician halts the information gathering processes and finalizes a diagnosis. However, this is reliant on clinicians having adequate knowledge and experience. Without adequate experience a clinician will not have a knowledge base of illness scripts to rely on to guide information gathering to confirm/reject diagnostic hypotheses. Novice clinicians do not have adequate experience. Novice clinicians benefit the most from CDSSs. Clinicians can compare his/her mental model of a patient to a ranked diagnostic recommendation generated by the CDSS. The ranked diagnostic recommendation should be able to give reasons for making a recommendation. The reasons should be given in a format

that is easy to understand. Diagnostic recommendations may not be completely accurate as all information has not been collected yet. However, ranked recommendations must contain the right diagnosis in the top to help a clinician focus his/her efforts on next steps of diagnostic encounter.

Researchers believe the intervention with the best scope for improving diagnostic decision making is diagnostic CDSSs. However some of the early CDSSs were plagued with problems including non-acceptance by clinicians (Sittig *et al.*, 2006). Researchers have identified several dimensions of ambulatory care (primary care) where diagnostic errors are likely to arise (Singh and Weingart, 2009). These are provider-patient encounter, performance and interpretation of diagnostic tests, follow-up of patients and diagnostic test results, subspecialty consultation and patient specific behaviours. Later studies have confirmed that the most common dimension where diagnostic errors can be found is provider-patient encounter (84.2% of cases) and the most common cause of the errors in the provider-patient encounter is problems in data gathering and synthesis (medical history, physical examination and ordering of diagnostic tests) (Singh *et al.*, 2013b).

During provider-patient (clinician-patient) encounter, a provider/clinician may not have access to accurate and reliable data. Singh and Weingart (Singh and Weingart, 2009) suggests that this may be due to inaccurate and second hand information obtained from colleagues or trainees or sometimes directly from the patients themselves. One strategy proposed by Singh and Weingart (Singh and Weingart, 2009) to reduce diagnostic errors in clinician-patient encounter is better integrated EHR systems with improved access to accurate medical history information with patient web portals playing an important role in the process of gathering information.

A review (Singh *et al.*, 2012b) of system-related interventions to reduce diagnostic error only found 2 studies that provided empirical outcomes of interventions to reduce diagnostic error in the clinician-patient encounter. System-related interventions are those interventions that address diagnostic error issues arising from systemic issues such as technical failure of equipment or organizational problems like faulty healthcare policy or protocols implemented in hospitals (Graber, Gordon and Franklin, 2002b). One study (Perno *et al.*, 2005) implemented a designated trauma response team that helped reduce the incidence on delayed diagnosis of injury in trauma patients. The other study (Howard *et al.*, 2006) implemented a tertiary examination procedure where a comprehensive re-examination of patient and review of all laboratory results and radiology studies is conducted within 24 hours of admission. This was found to reduce the number of missed injuries in trauma patients.

2.2 Diagnostic decision support systems

2.2.1 History of diagnostic decision support systems

Computer based CDSSs have been in development for several decades. Since the earliest days of computers researchers have indicated the possibility of using machines in clinical diagnosis (Ledley and Lusted, 1959). The main decisions CDSSs can support include diagnosis and treatment planning. Treatment planning is intrinsically linked to the diagnosis and other factors including medical history, patient status and related factors such as patient values and needs. The early applications of diagnostic CDSSs built during early 1970's have required clinicians to adapt his/her diagnostic reasoning process to the mathematical and logical reasoning suitable for computation (Miller, 1994). However, it was soon realized that clinicians were not willing to adapt to this style of reasoning. Traditional decision making models including probability theory and utility theory are being slowly replaced by more naturalistic decision making models which are more suitable for real life scenarios (Klein, 2008b). This led to the development of knowledge-based systems that model the clinician's diagnostic reasoning process. These systems used categorical reasoning and probabilistic reasoning alongside the knowledge-based models to offer better explanation of the patient's symptoms. The term knowledge-based system is broad and can denote any computer program that uses a knowledge base to solve problems.

According to Shu-Hsien Liao (Shu-Hsien Liao, 2005a), Expert System (ES) methodologies generally fall into following categories: "rule-based systems, knowledge-based systems, neural networks, fuzzy ESs, object oriented methodology, case-based reasoning, system architecture, intelligent agent systems,

database methodology, modelling and ontology" based systems. Knowledge based system and expert system are sometimes used synonymously. Expert systems are primarily applications that are attempts to solve a problem normally done by a human expert (Hayes-Roth, Waterman and Lenat, 1983). Knowledge based systems represent the knowledge explicitly in the form of knowledge representation formalisms rather than computer code. Almost all expert systems available today are knowledge based, however not all knowledge-based systems are expert systems. When applied to clinical diagnosis, these knowledge-based systems/expert systems reason in a systematic manner, generating multiple hypotheses based on clinical data and arrive at conclusions to explain the symptoms that the patient has.

CDSSs can be roughly classified into three categories based on their application (Musen, Shahar and Shortliffe, 2006).

- Tools for information management: Tools that perform information storage and retrieval and provide the clinicians with the data and knowledge required to perform a task (Oberkampf, Zillner and Bauer, 2012b). These tools do not apply the information to the task. Implementation of the information and adapting the knowledge to the particular patient is accomplished by the clinician (Sullivan, Gardner and van Rijsbergen, 1999).
- Tools for focussing attention: These tools flag abnormal values and alert the clinician to possible drug-drug interactions and other events. These tools can assist in diagnosis when they alert the clinician toward abnormal values which may indicate the need for further tests and may indicate the presence of pathology. Other uses include drug-allergy checking, basic dosing guidance, formulary decision support and duplicate therapy checking (Kuperman *et al.*, 2007).
- Tools for providing patient specific recommendations: These tools provide recommendations based on the patient's clinical conditions. Diagnostic assistants such as DXplain (Barnett *et al.*, 1987a) analyse the clinical findings such as symptoms, signs and lab results to produce a ranked list of possible diagnosis. It provides a justification of why each diagnosis is present in the list and indicates other unusual manifestations that may be present in the patient and therefore needs to be considered. Other systems include QMR (Quick Medical reference) (Miller, Masarie and Myers, 1986) which was developed from the INTERNIST-1 program (Miller, Pople and Myers, 1982). QMR provides access to more than 750 diseases to support internists in daily practice and has a knowledge base of more than 5000 clinical findings.

An early successful version of a successful knowledge based system was PUFF (Aikins *et al.*, 1983). PUFF was built using EMYCIN (van Melle, 1979), which in turn was built using an earlier system called MYCIN. MYCIN was a rule based expert system for diagnosis and treatment recommendation in infections. The clinical knowledge was represented in the form of IF-THEN rules. MYCIN provided an early demonstration of the ability of rules to represent clinical knowledge (Shortliffe *et al.*, 1975). EMYCIN (Essential MYCIN) is an expert system shell that can be used to build diagnostic expert systems. Rules contain small and independent segments of domain knowledge that are easier to modify than knowledge embedded into the computer code. However, modifications to one rule affect the behaviour of other rules and can result in errors and as the number of rules increase the application becomes more and more unmanageable. However techniques for automatic discovery using data mining methods and optimisation of rules have been developed which can address some of these issues and has led to a renewed interest in rule based systems (Kwiatkowska *et al.*, 2007). PUFF was one of the first knowledge-based systems to be routinely used in clinical practice.

MYCIN (Shortliffe *et al.*, 1975), one the best known expert systems, is built around a set of concepts such as clinical history of patient and infectious organisms. The main aim of the system was to play a similar role of a human specialist in infectious disorders. The system will interact with a clinician to collect relevant information about the patient and analyse the information to provide a diagnosis and treatment recommendation. MYCIN models the chain of reasoning used by specialist diagnosticians and encode the knowledge in the form of production rules.

INTERNIST-1 (Miller, Pople and Myers, 1982) differs from other diagnostic decision support system previously developed. It did not follow Bayesian statistical methods and pattern recognition to perform

diagnostic reasoning. It is impractical to estimate the probabilities of diseases for all the diseases within internal medicine as some of the diseases as rare and not well described in literature. Pattern matching is difficult to perform in complex cases with multiple competing hypotheses and incomplete information. It relied on symbolic reasoning which was being developed around the same time and a ranking algorithm to perform complicated processing on information to arrive at a diagnosis.

2.2.1.1 Causal models in diagnostic decision support systems

Causal models play a crucial role in medical reasoning. Knowledge of causal relations helps the clinician understand the disease process, describe temporal processes in diseases, cluster symptoms into groups based on causal relations, and understand relations between diseases.

Causal models in diagnostic reasoning were first explored in the causal-associational network (CASNET) (Weiss *et al.*, 1978). It was first developed in the glaucoma domain. It was proposed as a general model for diagnosis. The CASNET model uses different layers of knowledge modelling for diagnostic reasoning. The model consists of three main components or layers: observation of the patient, pathophysiological states and categories for diagnosis, prognosis and treatment.

When the clinician examines the patient he/she records the clinical observations which are associated with certain intermediate states. The physiological states modelled summaries of abnormal physiological events. The states are causally related to each other. These states and their causal relationships form a directed acyclic graph network. Diseases are represented as pathways within the network. The progression of the disease was modelled by following pathways within the network. This was done by following the pathway from a starting node through admissible pathway of nodes. The admissible pathway passed through either confirmed or unconfirmed nodes and excludes denied nodes.

The CASNET model is different from mathematical models as it relies on causal links and temporal progression of disease to perform the diagnosis. CASNET was among the first type of knowledge-based system that represented knowledge in the form of different levels of abstraction. However, CASNET is limited by difficulty in manipulating the causal relations and dealing with comorbidities.

The Present Illness Program (Pauker *et al.*, 1976a) uses causal relations to denote relations between different diseases and their clinical states using relations such as cause-of, caused-by etc. The domain knowledge is organized into collections of facts called Frames.

These frames are like the Illness Script concept used in medical cognition research. According to this concept the physicians organize their knowledge of disease, clinical manifestations, and other conditions in which diseases are known to occur in the form of 'scripts'. As the clinician gains more experience the causal pathophysiological knowledge becomes encapsulated into diagnostic labels. The encapsulated knowledge becomes translated into scripts called Illness scripts (Schmidt and Rikers, 2007a). Illness scripts are essentially networks of knowledge that can be adapted to achieve the goals of the clinical task. When the script is adapted for diagnostic purposes it is called a diagnostic script. When the clinician examines a patient the incoming information obtained from questioning and examinations triggers the relevant script, then directs the clinician's knowledge gathering behaviour and helps in hypothesis formation and testing (Custers, Boshuizen and Schmidt, 1996; Charlin, Tardif and Boshuizen, 2000).

Within each frame of the Present Illness Program there exists a similar structure and each frame is centred on a disease and includes typical findings of the disease, scoring parameters, causal factors (caused-by, maybe-cause-of etc.) and rules for fitting the frame to individual case. These frames are in turn linked to other frames forming an elaborate network quite like the knowledge organization structure used by expert clinicians. Other links include complications (complication-of, complicated-by) and associations between diagnoses. Each hypothesis frame has 3 states that it can have; active, semi-active and inactive state. The state is determined by the match between the input date and the information held within the frames. When there is positive match, the frame is placed in the active state. Any frames that are linked to that frame are put in semi-active mode. Reasoning under uncertain conditions was performed by calculating a likelihood estimate for each active frame. When the likelihood estimate passed a threshold, the frame is confirmed, and if it fell below a threshold it was returned to the semi-active state.

The likelihood estimate was calculated as an average of a binding score and matching score. The binding score estimated the extent to which a hypothesis explained the findings of the patient. The matching score estimated the extent to which the observed findings were appropriate for the hypothesis. The semi-active frames were designed to replicate the hypotheses/scripts that are activated by the clinician during the diagnostic reasoning process. Present Illness Program was among the first expert systems to explicitly use human cognitive models to model the diagnostic reasoning mechanism within CDSSs.

Use of causal relations in diagnostic reasoning frameworks were explored in INTERNIST-II (Pople, 1982) and then in CADUCEUS (Myers, Pople and Miller, 1982) and to remedy some of the problems associated with CASNET. INTERNIST-I (Miller, Pople and Myers, 1982) used a single hierarchy to present taxonomic knowledge while CADUCEUS tried to improve upon this by organizing the knowledge base using several different hierarchies utilizing a lattice-like structure with each hierarchy organised around a concept. The disease within each hierarchy and area related to other diseases using causal links.

CADUCEUS used links to represent relations between nodes. These include 'Causal links' to represent pathophysiological relations, 'Planning links' to represent possible causal relations and 'Spanning links' to summarize the collections of causal links. Constrictor links associate the diagnosis with key findings that give a strong indication of a diagnosis.

The pathophysiological relations between clinical states are predefined in CASNET, while CADUCEUS can hypothesise the presence or absence of such a relation. The knowledge model in CASNET only has a single level of detail while CADUCEUS can vary the causal links and granularity based on the available evidence. However both CASNET and CADUCEUS does not adequately address the problem caused by interactions between multiple disorders on diagnosis (Patil and Senyk, 1987). Some diseases can affect the clinical presentation of other diseases.

ABEL (Patil, Szolovits and Schwartz, 1981) attempted to solve the problem of interaction between multiple disorders in the domain of acid-base electrolyte disorders. In this system, the knowledge base was modelled on causal hierarchical networks. The model consists of a multi-level hierarchy of representation. At the shallowest level, it will consist of causal relations between syndromes and diseases and at the deepest level it consists of pathophysiological knowledge. The knowledge at lower levels is aggregated into upper levels of abstract levels of knowledge. The authors argue that this method of representation in multiple hierarchies enables the system to handle disorders with multiple causes and interactions between multiple disorders. Each node has several attributes and each node is linked to other nodes through causal relations. Each level contained a semantic network of nodes and causal links between these nodes. Each node represents a normal or abnormal state of a physiological condition and the link represent the relations between different states. The link between the cause and effect nodes contains a set of constraints between the attributes of these nodes. The causal links help to constrain the number of nodes evaluated during the diagnostic reasoning process.

CHECK (Molino *et al.*, 1986) combines heuristic knowledge and causal knowledge within the same system. CHECK has two-level architecture with different representations of the domain knowledge in the form of causal and heuristic knowledge interacting with each other. The heuristic representation consists of productions rules and frames to assist in diagnosis. The causal representation consists of a hierarchy of interrelated nodes. The knowledge in the causal network is used to confirm hypotheses generated from the surface level, representing heuristic knowledge.

The surface level contains knowledge frames representing classifications of liver diseases and sub-classifications of these diseases. The causal network consists of several nodes which have a set of attributes: Initial causes, states, actions, findings, hypotheses. These nodes are linked to each other with arcs: cause of, has a manifestation, defined as, loops.

Another expert system that uses causal models for reasoning is the CHF (Chronic Heart Failure) Advisor (Long, 1989). This system was developed mainly to assist the clinician in diagnosis and management of patient with a condition that can cause heart failure. The cardiovascular domain is replete with uncertain

causal mechanisms. CHF Advisor used a probabilistic causal model to model the domain. In this model, there exist relations between cause and effect with a probability of a cause producing the effect. This model has attempted to include enough intermediate nodes to represent to represent causal mechanism in considerable detail. They developed a heuristic method to generating differential diagnosis based on causal pathways, starting from causes to findings and building the hypothesis, one finding at a time.

Other systems that use probabilistic causal networks include MUNIN (Andreassen *et al.*, 1987) which has causal links between nodes with conditional probabilities and apriori probabilities of the system are updated using Bayesian inference mechanisms.

Probabilistic causal models based on directed probabilistic graphs such as Bayesian networks have been developed for diagnostic reasoning in uncertain conditions (Onisko, Druzdzel and Wasyluk, 1998). Each node in the graph represents a random variable the arcs represent a direct relationship between the 2 variables. The graphs structure represents the causal structure of the domain.

One of the earliest attempts to implement diagnostic reasoning under uncertainty was (de Dombal *et al.*, 1972) developed to diagnose abdominal pain and recommend the need for surgery. It used a naïve Bayesian approach for diagnostic inference and overall diagnostic accuracy of the system was higher than clinicians involved in evaluation. The system provides access to definitions of signs and symptoms of patients with abdominal pain, information about patients with abdominal pain and outcomes for patients with similar signs and symptoms etc. The database contains over 50,000 cases. When the patient's history, signs and symptoms are entered, the correct database is selected. Using Bayesian statistical analysis, the patient's clinical data is compared to data in the database which generates the probability of the most likely diagnosis. The system has been evaluated several times and has been found to improve diagnostic accuracy, reduction in readmissions and unnecessary surgical procedures (Gunn, 1991).

However, reasoning in causal models can be computationally expensive especially for routine cases. The knowledge about the actual causal mechanism in several diseases is unavailable and therefore implementation of CDSSs based on causal reasoning is still lacking. Causal models have problems representing temporal reasoning. Another issue with causal systems is the level of detail that is required for modelling causal relations.

2.2.2 Differential diagnosis generator systems

Differential diagnosis generator systems are software that uses clinical data (clinical history, symptoms, etc.) entered by clinicians to generate a list of possible diagnoses that clinicians need to consider when making a diagnosis. Differential diagnosis generators are one of the different types of computer based decision support tools available that can help reduce diagnostic error (El-Kareh, Hasan and Schiff, 2013).

DXplain (Barnett *et al.*, 1987b) analyses the patient's medical history, clinical examination findings and laboratory findings to produce a ranked differential diagnosis list. The database has over 2000 diseases and over 4000 clinical terms to describe these diseases. Iliad is an expert system that was originally developed as a training tool for medical students (Warner, 1989) and has been evaluated as a diagnostic recommender system as well (Bouhaddou, Lambert and Morgan, 1995). Other early differential diagnosis recommender systems include MEDITEL (Waxman and Worley, 1990) and Quick Medical Reference (Miller, Masarie and Myers, 1986). The performance of these early systems left much to be desired (Berner *et al.*, 1994b). These systems produced a large differential diagnosis list and clinicians took a long time to input the data required to produce the list (Graber and VanScoy, 2003). This limited their usefulness in real world settings. These systems used either Bayesian logic or pattern matching algorithms to make diagnostic recommendations when the clinician selected clinical findings from a menu.

To overcome some of the problems with these early systems several 2nd generation systems were introduced. These systems used free-text or natural language processing technology to process findings entered by the clinicians and used newer information retrieval and pattern matching algorithms to process the data and produce a list of diagnostic recommendations. Some of these web based systems have reduced the time taken for data entry (Graber and Mathew, 2008a), but overall diagnostic performance

has not improved substantially (Bond *et al.*, 2012b). However DxPlain performed well in comparison to other 2nd generation systems such as ISABEL (Greenough, 2002; Vardell and Moore, 2011). Other well-known 2nd generation systems include Diagnosis Pro (Diagnosis Pro®, 2014), PEPID (PEPID™, 2014), VisualDx (Vardell and Bou-Crick, 2012) etc. Current (3rd generation) differential diagnosis generators using semantic web technologies (García-Crespo *et al.*, 2010b) are being developed which use ontologies as a knowledge base (See section - 2.5).

A recent review of differential diagnosis generator systems identified 23 different systems currently in use clinical practice and for teaching (Bond *et al.*, 2012b). It identified some potential barriers to widespread adoption of these tools which includes lack of integration with EHR systems, limitation of the user interface and lack of access to linked content including evidence based and non-evidence-based content.

2.2.3 Guideline-based CDSSs

Clinical management guidelines are an important part of improving standards of care, quality and clinical outcomes in healthcare. Text-based clinical guidelines are developed by domain experts such as healthcare scientists and expert clinicians using a consensus-based approach. Guidelines are based on scientific evidence and the best guidelines are based on review of randomized controlled trials. There has been a steady move towards greater use of evidence-based medicine and clinical guidelines is one several measures that have been proposed for realizing vision of using scientific evidence where possible in clinical practice. Clinical guidelines have been developed with aim of helping clinicians with decision making for common situations encountered in clinical practice. Guidelines can help the clinician reduce errors such as diagnostic and management errors, improve efficiency in clinical practice, employ best practices, reduce variation in clinical practice and improve overall clinical outcomes.

Though clinical guidelines are based on evidence their implementation has not been universally successful (Lenzer, 2013b). Main problem associated with paper-based clinical guidelines is that guidelines are designed for general population and process of personalizing guidelines for an individual patient for a specific context has be performed manually by the clinician (Woolf *et al.*, 1999). Textual content in these paper-based guidelines are narrative in nature and is often written in a relaxed language and it is up to a clinician to interpret contents based on his/her clinical judgement. The paper-based guidelines are bulky and not easily accessible at bed-side where decisions are made. Algorithms and flow-charts have been developed from these clinical guidelines to make the content more easily accessible at the bed-side. These clinical algorithms present the logic and the flow of care for a single patient and generally focusses on a single medical condition.

The logic within these algorithms and workflows can be represented in a formal manner that can be executed by computers providing automated inference and providing patient specific guideline recommendations (de Clercq *et al.*, 2004). CDSSs that can consume these formal models to provide clinical decision support recommendations are called guideline-based CDSSs. There are several formalisms that exist which can formally represent guidelines into a format that can be computer executable (Peleg, 2013).

Many of these guideline representation formalisms have a process-flow like representation called Task-Network Models (Peleg *et al.*, 2003). The steps in guidelines are represented as a series of tasks. Tasks are then joined with each other to form a network of tasks. Tasks can comprise of actions, decisions, enquiries etc.

Asbru is a skeletal plan representation language. Asbru represents the medical guideline as skeletal plan with several sub-plans. This language represents actions as well as goals of a plan. The main components of an Asbru plan includes Preferences, Intentions, Conditions, Effect and Plan-body (Seyfang, Miksch and Marcos, 2002). AsbruView is a tool developed for visualizing and developing the model.

Arden Syntax is a standard procedural language to represent protocols and plans in clinical guidelines as knowledge modules called Medical Logic Modules (Hripcsak, 1994). Medical Logic Modules execute as a sequence of logic statements, queries and calculations. The logic in each Medical Logic Module is

sufficient for a single medical decision. Medical Logic Modules can be embedded in clinical information systems to provide alerts, messages, warnings etc. Medical Logic Modules can be written using a text-editor without knowledge of any programming language. Arden Syntax cannot represent complex medical guidelines that involve multiple decisions or process flow sequences but can be used to model clinical alerts and other warnings that have an IF-THEN-ELSE representation. The Medical Logic Modules can invoke other Medical Logic Modules. Another drawback of Medical Logic Modules is the 'curly braces problem'. In Medical Logic Modules all references to specific clinical data must be contained within curly braces. Arden's syntactic description does not allow any data modelling. Definition of data is therefore left to local implementers. Therefore, each Medical Logic Module must be adapted to the architecture of individual clinical systems thereby decreasing interoperability and reusability (Samwald *et al.*, 2012).

EON defines an architecture comprised of a set of components that cooperate with each other and automates various tasks associated with a clinical guideline (Tu and Musen, 2001). EON provides recommendations that are patient specific and provides alerts if the actions of a clinician diverges from published guideline recommendations. EON models have a declarative and a procedural component. The declarative component is modelled as an ontology. The procedural component is defined by a directed graph. EON is well suited for modelling clinical algorithm and flow-charts. However, the flow-chart like representation limits the ability of EON to model complex guidelines.

GLIF 3 (Guideline Interchange Format Version 3) is an object-oriented format (Patel *et al.*, 1998). It consists of classes, attributes and relationships between classes. Clinical knowledge is represented as a relationship between medical concepts. GLIF 3 provides a framework for modelling guidelines as flowcharts (Boxwala *et al.*, 2004). This limits ability of the model to represent complex guidelines especially where there is a temporal component to a guideline. GLEE (Guideline Execution Engine) is the only commercially available execution engine that can execute GLIF models.

PROforma is a guideline representation language that was developed by Advanced Computation Laboratory of Cancer Research UK (Fox, Patkar and Thomson, 2006a). Tallis is a toolset that domain experts can use to model pathway of a guideline. The toolset contains a graphical editor and an enactment engine to execute a guideline. A PROforma model is composed of 4 classes of tasks. These are Actions, Enquiries, Decisions and Plans (Sutton and Fox, 2003). Tasks are represented as a graph of nodes and edges represent the scheduling constraints. PROforma has precise syntax and semantics. All tasks share common attributes such goals, control-flow, preconditions and post-conditions. The Decision recommendations are generated using Argumentation Theory (AT) as basis (Fox *et al.*, 2007). The choices for a decision are represented as Candidates. Each Candidate has a set of arguments that either supports or negates a Candidate (Fox, Krause and Elvang-Gøransson, 1993). The PROforma models must be executed using proprietary tools like Tallis (Sutton, Taylor and Earle, 2006). The inability to use off-the-shelf technology for execution is a main limitation of current generation guideline-based CDSSs that prevents widespread adoption.

2.2.4 Argumentation Theory and CDSSs

Argumentation Theory (AT) is an emerging area for dealing with unstructured evidence in domains such as healthcare characterized by uncertainty (Fox *et al.*, 2007). Knowledge needed to support decision making is expressed in the form of arguments. AT explains how arguments can be expressed, sustained and rejected during process of decision making. For each decision a set of positive and negative arguments are created. Argumentation is a model that is based on construction and evaluation of competing arguments (Ouerdane, Maudet and Tsoukiàs, 2010). AT evolved as a sub-discipline of philosophical logic and is an important area of logic-based AI research. In recent years it has found several applications in cognitive science and information technology (Fox *et al.*, 2007).

Non-monotonic logic is a type of formal logic where conclusions can be invalidated by adding more knowledge. Non-monotonic logic is particularly useful in real-life situations such as diagnosis where all information about a domain is not fully available. Using available information, non-monotonic logic-based system should be able to make tentative conclusions that can be retracted later when more information is available. In classical logical reasoning methods such as propositional logic, predicate logic

etc. conclusions that were formed cannot be modified or withdrawn considering new information. This phenomenon is called Monotonicity. Real life reasoning is not always monotonic. The hypothetico-deductive model for clinical reasoning suggests that clinicians use initial information to gain additional information to support a diagnostic hypothesis. Therefore, diagnostic reasoning is a dynamic process where conclusions keep on changing as new facts are obtained. Non-monotonic logics were therefore explored by AI researchers due to limitations posed by monotonic reasoning used in classical logics. Non-monotonic logics were initially developed by McCarthy and John (McCarthy and John, 1977), McDermott and Doyle (McDermott and Doyle, 1980) and Reiter (Reiter, 1980). Defeasible reasoning is a process used by humans where existing conclusions can be discarded in the light of new information or evidence. Non-monotonic logics were developed to provide a formal framework where defeasible inference and defeasible knowledge representation can be adequately represented. Using non-monotonic logics, the reasoners can arrive at a conclusion and retract conclusions in the light of new information. Reiter (Reiter, 1987) provides an example of this type of reasoning process:

- Birds fly
- Tweet is a bird
- Therefore, Tweety flies

Using this logic if Tweety is a type of a bird then we can arrive at a conclusion that Tweety is capable of flying. However, if Tweety is a special type of bird that cannot fly, for example a penguin, then the original conclusion is false. Penguins cannot fly. However, the original statement describing 'birds fly' is still true. Tweety is a penguin, and therefore not a typical bird. Therefore, we must assume that if Tweety is a typical bird, then by default we can assume it can fly. We can safely assume that Tweety flies until we have additional information that states otherwise. Different type of non-monotonic reasoning formalisms has emerged including default logic (Reiter, 1987), auto-epistemic logic (Moore, 1985) etc. Argumentation is one type of non-monotonic logic (Prakken and Vreeswijk, 2001).

Argumentation was derived from work done in the 1950's by Toulmin and others to explain how the process of argumentation happens in day-to-day decision making. There are 3 main parts to a Toulmin model (Toulmin, 1958). There are Claims, Support and Warrants. Claims are statement or assertions an agent is aiming to prove. In a diagnostic context 'patient has dental abscess' is a claim. Support are statements that can support a claim. For example, 'presence of pain' is a statement that supports the claim of dental abscess in a patient. Warrants provide underlying reasons linking claim and support. Tests that can prove the presence of pain are types of warrants.

Argumentation provides a framework to formalize defeasible reasoning (Fox, Krause and Elvang-Gøransson, 1993). This type of reasoning is especially useful for diagnosis where an initial hypothesis is used to collect more information. Based on the new information the existing hypothesis is either reinforced or discarded completely to form a new hypothesis. Defeasible reasoning is particularly useful in situations where the information is incomplete or uncertain (Longo, Kane and Hederman, 2012). Early diagnosis is one area where the information is incomplete the diagnosis is uncertain. A clinician has not collected all information needed to make a diagnosis and meaning of symptoms and signs must be confirmed using clinical tests and investigations.

Clinicians and other healthcare practitioners prefer tools that can provide better explanation about the patient's condition. The reasoning used in clinical decision making is both non-monotonic and explanatory in nature (Fox, Krause and Elvang-Gøransson, 1993). Numerical tools that can predict probability of presence of a disease is useful, however clinicians benefit more when the CDSS is able to explain why a diagnosis has been recommended (Labrie and Schulz, 2014). Subjective probabilities are sometimes used by clinicians to perform medical decision making. The costs and benefits of a decision are weighed alongside their probabilities. However, in most real-life situations, clinicians tend to use qualitative reasoning rather than numerical calculations to solve problems such as diagnosis. Reliable probabilities are not easily available for all domains to build decision models.

Using an argumentation-based approach delivers several advantages to a clinician using a CDSS (Fox and Parsons, 1998). The clinician will be able to get a good overview of the options available to him/her. A good choice amongst these decisions can be presented and the reasons underlying this decision can be presented in natural language and in a format that is easy to understand (Ouerdane, Maudet and Tsoukiàs, 2010). Argumentation allows us to model with greater expressivity the preferences needed to decide and allows partial specification of preferences especially in decisions where there are multiple options.

In argumentation-based models such as PROforma the degree of support of each argument is assigned a numeric weight (Fox *et al.*, 2007). The argumentation process begins by formalizing arguments for a decision candidate. A decision has multiple candidates. Arguments supporting a candidate are specified alongside arguments negating a candidate. The weight can be used as the basis for ranking the candidates once the arguments have been aggregated and a decision is to be made. PROforma's argumentation-based decision model supports a natural form of explaining reasons/justifications for each decision (Fox, Johns and Rahmanzadeh, 1998). Despite simplicity of argumentation-based models such as PROforma, it is a practical way to develop CDSSs and has high acceptability amongst clinicians (Fox, Patkar and Thomson, 2006a).

Argumentation-based applications have been in developed in medicine and other areas to support decision making. Argumentation-based CDSS applications have been developed to support decision making about medical treatments in patients (Atkinson, Bench-Capon and Modgil, 2006). An agent called the Drama agent orchestrates information from multiple sources to supply a set of arguments based on which a decision regarding treatment can be taken. arguEIRA, is a CDSS and was developed as an extension of the existing CDSS called EIRA (Grando *et al.*, 2013) by providing it with an argumentation-based justification system that formalizes and communicates to clinicians reasons why a patient did not respond to treatment.

Argumentation based CDSSs have been shown to improve decision making in certain areas such as prescribing. A study by Walton et al. (Walton *et al.*, 1997) using simulated cases show that CDSSs improved prescribing patterns, resulted in faster decision making and improved compliance with prescribing guidelines. Participants in the study found the CDSS easy to use and most of them were likely to use in clinical practice.

LISA is another argumentation-based CDSS for advising clinicians about drug dose adjustment for treating children with acute lymphoblastic leukemia. A study using simulated cases, questionnaire study and semi-structured interviews shows that LISA helped improve accuracy of dosage calculations (Bury *et al.*, 2004). Another study using retrospective cases showed that LISA helped improve adherence to drug dose protocols and reduce prescribing errors. Clinicians who did not use LISA deviated from trial protocol on 37% of occasions, however when they were given decision support the number dropped to zero. The clinicians who participated in the study said they would use LISA if it were routinely available (Bury *et al.*, 2005).

Due to the absence of concrete numerical estimates of uncertainty, argumentation-based CDSSs are well suited for risk assessment. An argumentation-based CDSS called RAGs (Risk Assessment in Genetics) was used to evaluate genetic risk of cancer (Emery *et al.*, 1999; Coulson *et al.*, 2001). The doctor enters a patient's details into a user interface. The patient's family tree is built using data provided. Once pedigree has been established the RAGs application assesses a patient's genetic risk of cancer. Domain experts provided the set of rules for building arguments for a patient's increased or decreased genetic risk. A simple total risk score is generated from the arguments. Based on the score a patient is put into high, medium or low category. An evaluation of user interface showed that when compared to other software and paper-based tools the clinicians preferred RAGs 91.7% of the time (Emery *et al.*, 2000).

Argumentation-based CDSS models have been developed to help clinician develop a plan of action for the patient. REACT (Risks, Events, Action and their Consequences over Time) developed by Glasspool et al. (Glasspool *et al.*, 2003) provides argumentation based support for clinicians and patients involved in medical planning. A study evaluating REACT using actors simulating patients showed that 7/8 geneticists who used the tool were very supportive of it (Glasspool *et al.*, 2003).

CADMIUM imaging system is an argumentation-based CDSS that helps clinicians in interpreting images by combining image processing and automated interpretation of images. A study using radiographers as test subjects showed that the system increased rate of correct classifications of malignant and benign abnormalities in routine breast cancer screening (Taylor, Fox and Pokropek, 1999).

Retrogram® developed by *Infer*Med is an argumentation-based CDSS that advises clinicians on the use of anti-retroviral drugs for HIV positive patients. Around 250 clinicians around the world are using this CDSS. The CDSS provided genotype interpretation alongside other decision recommendations. A multi-centre clinical trial showed that the CDSS helped improve patient outcomes by reducing viral loads to target levels (Tural *et al.*, 2002).

An argumentation-based guideline driven CDSS was built using PROforma to support triple assessment of cancer. A balanced crossover evaluation of the CDSS using breast cancer clinicians showed that there was significantly more deviation from guideline recommendations without help of decision support (Patkar *et al.*, 2006). More critical errors occurred when clinicians did not have support of decision support. Opinion of clinicians was generally positive towards the CDSS.

Argumentation-based CDSSs have been developed to support diagnostic decision making. A distributed architecture to support the diagnostic and treatment of cancer using PROforma has been proposed. Argumentation was used as the basis for developing a CDSS to help diagnosis dementia (Lindgren, 2011). Likelihood for diagnosis was displayed in the form of different types of support such as possible, unlikely, probable and excluded.

In medicine, argumentation-based CDSS tools have been developed in several areas including diagnosis, treatment planning, prescription etc. Diagnostic CDSS tools developed to date have been tools that can help interpret data and do not provide diagnostic recommendations. However, the studies to date have not shown the impact of an argumentation-based diagnostic CDSS on diagnostic decision making. There is a need to evaluate impact of an argumentation-based CDSS on the diagnostic behavior of clinicians, particularly primary care clinicians. Argumentation is a good candidate for developing a diagnostic CDSS for generating diagnostic recommendation in early stages of a clinical encounter.

2.2.5 Argumentation and other CDSS models

This section gives an outline of some of commonly used decision support models in healthcare and its limitations when compared to an argumentation-based approach.

Machine learning approaches are a major branch of AI and have been used to develop several CDSS applications. Intelligent behaviour of a CDSS is built by learning from data such as patient data from EHRs. Training data that consists of features of a diagnosis is provided to a learning algorithm alongside the correct diagnosis. Subsequently when the learning algorithm is provided a large enough dataset it can learn from the data and predict a diagnosis. The main advantage of machine-learning approaches is there is no need to develop a specific algorithm for a domain. As more data becomes available, accuracy of learning algorithms can be improved. The main disadvantage of using machine learning approaches (especially neural network-based techniques) for diagnostic CDSSs is that some of these algorithms have a black-box approach and decisions made by the CDSS cannot be explained easily. Argumentation-based methods are a simple knowledge based approach that can capture the expertise of domain experts and explain decision made by CDSSs in a manner that clinicians can understand and have high acceptability among clinicians (Fox, Patkar and Thomson, 2006a; Fox *et al.*, 2007).

Longo et al. (Longo, Kane and Hederman, 2012) have identified some advantages of Argumentation-based methods and models that machine learning techniques do not have. They include:

- The ability to deal with incomplete and missing data. Process of decision making can be completed even if all data is not available. This is important for providing diagnostic recommendations early in the diagnostic encounter.
- Argumentation-based models can capture expertise of domain experts in an organised manner.

- Argumentation is not based on statistics and probability and is therefore very close to the way in humans make decisions. If needed, statistical evidence can be incorporated into the decision model. Though subjective probabilities are used in medical decision-making, clinicians often appear to make decisions based on qualitative reasoning based on domain specific knowledge.
- Argumentation-based methods can better explain reasons for deciding. The arguments in favour and against a hypothesis can be presented and a clinician can weigh the evidence. This is important when dealing with substantial amounts of data that could potentially be available to clinicians from linked web of data.
- Argumentation is extensible and allows retraction of a decision based on new evidence. It can be updated. This is useful in providing early diagnosis where the CDSS retracts the old decision and provides a new diagnostic recommendation considering current information.

There are primarily 3 categories of human diagnostic reasoning that have been formally represented by CDSSs to provide diagnostic recommendations. These are probabilistic, causal and deterministic. Probabilistic models include models such as Brunswik's lens model (BRUNSWIK, 1955), Bayesian methods (WARNER *et al.*, 1961) and decision analysis (Thornton, Lilford and Johnson, 1992). These probabilistic models use statistical association between clinical variables such as observations and complex mathematical models to generate diagnostic recommendations. Clinicians consider prevalence and other probabilistic data about probability of a clinical observation within a population during diagnostic reasoning (Elstein, Shulman and Sprafka, 1978). For example, when clinicians are faced with a patient that complains of a lesion on the skin they are trained to look for common causes of skin lesions including infections. Once they have ruled out all common causes of lesion they will look for relatively rare causes like skin cancer. However, studies have shown that clinicians are not very good at incorporating real statistical data into their daily diagnostic decision making especially when dealing with uncertainty (Dawes, Faust and Meehl, 1989). Clinicians have problem reasoning with probabilities and studies have shown wide variability in assigning diagnostic probabilities (Cahan *et al.*, 2003).

Bayesian networks have been used by CDSS applications for making decisions under uncertainty. The Bayesian network can represent relation between diseases and its symptoms. Given a set of symptoms Bayesian network-based CDSSs can provide probability of a disease being present. Bayesian network represents uncertainties using probabilities. Bayesian network is a directed acyclic graph over which is a probability distribution is represented. The nodes of a Bayesian network graph represent random variables or events. Directed links between variables represent causal relationships. A link from one variable to another indicates that one variable can cause other. The earliest example of a Bayesian network CDSS was a program developed for diagnosis of acute abdominal pain (de Dombal *et al.*, 1972). In domains where causal knowledge is well known and can be captured easily the Bayesian network can be modelled easily and can provide accurate diagnostic recommendations (Wasyluk, Oniśko and Druzdzel, 2001; Chang, Hung and Juang, 2013). Bayesian networks perform well when incomplete or noisy data are available (Sesen *et al.*, 2013).

However, in domains where causal knowledge is not well defined it is very difficult to model a Bayesian network. Translating domain expert's clinical knowledge is difficult especially when there are several signs, symptoms, clinical test results and associated diagnoses to consider. The argumentation-based model simplifies this process of capturing the domain knowledge by using a simple technique to generate diagnostic recommendation. Presence/absence of an observation is used to provide supporting/negating arguments. The weights are a subjective representation of degree of support of each argument places on each diagnosis. The domain expert therefore does not need to know statistical probability of an observation being present and probability of a diagnosis given presence of an observation. The argumentation-based model is therefore much simpler to use to capture domain expertise. A machine-readable and machine-understandable representation of a CDSS model that uses argumentation can automatically derive the arguments from other ontologies such as SNOMED-CT. Bayesian Network Classifiers uses Bayes formal rule to classify patients. This approach involves representing each decision as a class and computing probability that a patient will fall within a class. CDSSs can recommend an option (e.g. a diagnosis) that has the highest probability given a set of data such as symptoms or signs

(Langarizadeh and Moghbeli, 2016). However, the amount of evidence needed to build a robust Bayesian Network Classifier is very high (Maria Yaneli *et al.*, 2013). It is very difficult to translate a clinician's domain knowledge into a Bayesian network.

Deterministic models such as production rules used in Expert systems based on IF-THEN rules are used to provide decision recommendation to clinicians (Horvitz, Breese and Henrion, 1988). The expert system is typically composed of 2 parts: inference engine and knowledge base. The knowledge base is composed of rules that have an IF-THEN representation. The use of natural language to model the rules make it appealing to clinicians. However, the main problem with expert systems is limitation associated with collection of knowledge from experts. They have limitations dealing with uncertainty, in translating domain expertise into the rules and they do not perform well in dynamic conditions where data keeps changing. Arguments in an argumentation-based system can be represented in natural language and argumentation supports defeasible reasoning where the decision can be retracted in the light of new information.

Fuzzy logic and fuzzy rules were developed to deal to uncertainties in decision making when using production rules. It was based on Fuzzy theory and has an advantage of modelling human reasoning that was vague. Fuzzy logic is better suited for handling different types of uncertainties that arise in clinical decision making (Warren, Beliakov and van der Zwaag, 2000). However, there are problems in translating domain expertise into fuzzy rules and difficulties capturing the subjective perception of what the fuzzy values mean to the clinicians. Argumentation-based methods are more intuitive and can be used to easily capture the expertise of domain experts.

Case-based reasoning techniques are another that can provide intuitive decision recommendation to clinicians (Begum *et al.*, 2011). In case-based reasoning the knowledge base is generated from previous clinical cases and evidence. This is similar in nature to clinicians using their memory to recall similar cases to solve the case in front of them. In theory this it is a natural technique for generating diagnostic recommendations. However, when the corpus of previous cases is limited case-based approaches can give faulty advice (HOLT *et al.*, 2005). Argumentation-based methods are not reliant on a corpus of previous cases for decision making. However, existing cases can be used to help formulate arguments for an argumentation-based CDSS.

2.3 The World Wide Web and CDSSs

We are living in an era of open healthcare data (Kayyali, Knott and Van Kuiken, 2013). As more healthcare providers switch to EHRs, availability of healthcare data increases. Easy availability of healthcare data raises some challenges. Healthcare data by its very nature comes in different shapes and sizes representing source of data (Raghupathi *et al.*, 2014).

The World Wide Web (WWW) has been a convenient and natural place for people to publish their data (Rodolfo *et al.*, 2016). This system is not owned by any government, individual, group or corporation. With dissemination of mobile phones and personal computers everyone has power to publish his/her data, including healthcare data on the web. However, with explosion of personal data available on the web it has become extremely difficult to find this data and use it to support clinical decision making. The different data sources that are available on the web can be structured, semi-structured or unstructured. Data can exist in many different syntaxes and data formats. This makes the task of integrating the diverse data sources for supporting clinical decision making very difficult.

The web allows anyone to add information via web pages that are linked to each other via hyperlinks and share this information with each other. It allows anyone to contribute to the world-wide knowledge of web and communicate with each other. From a diagnosis point of view, the utility of the web in helping clinicians diagnose and manage diseases can be illustrated with the example of rare diseases. The information that can potentially help a clinician diagnose a rare disease includes knowledge repositories on the web that contain information about all sorts of diseases that includes rare diseases (Pauer *et al.*, 2016). Since patients with a rare disease do not frequently present to a clinic, signs and symptoms of the disease will be unfamiliar to a clinician. Some diseases are so rare that clinicians, especially primary care

clinicians such as General Practitioners (GP) and Dentists, are unlikely to see even one such rare case in their whole career (DoH, 2013). Even researchers and scientists studying certain rare diseases do not have a complete picture of the most common signs and symptoms. Therefore, patients and researchers are increasing using web to share information about a rare disease of interest. The patients are using web resources to find more information about a disease, find other people that may have a similar disease, share information about their own condition or find latest treatments and clinical groups involved in diagnosing and managing the disease (Powell, Darvell and Gray, 2003). Clinicians and researchers are using these resources to find more information about a disease and recruit patients for cutting edge medical treatments (Raredisease.org, 2013; NORD, 2016). However due to time constraints and increasing workload clinicians do not have time to search these information repositories even though the incidence of diagnostic error in the community is increasing and rare diseases play an important role in this increase (Kostopoulou, Delaney and Munro, 2008b). Failure to diagnose causes considerable stress and wastage of time and resources for the patient and healthcare system and delays treatment.

2.3.1 History of knowledge representation on the web

2.3.1.1 Semantic Networks

Semantic Networks are a type of knowledge representation method which consists of vertices which represent concepts and edges representing relations between concepts. The edges/arcs are directed i.e. they are directed from one concept to another and they are labelled. Semantic networks are inspired by the semantic network models used in cognitive psychology to represent human memory (Smith, Shoben and Rips, 1974).

Semantic networks are commonly used to represent taxonomic relations. Nodes can represent concepts, objects, actions, events etc. It can represent individuals or classes of individuals. Arcs represent relations and associations between these nodes. The labels of the arcs denote the meaning of a relation. Some commonly used types of arcs include:

- IS-A (instance of)
- PART-OF
- AKO (a-kind-of)
- COMPOSED_OF

The information about a domain is stored with help of interconnecting nodes and their arcs. One of the main types of reasoning mechanism over semantic networks is inheritance of values where the 'is-a' and other instance relationships can be used to find subclass and instance links. A node can have several super-classes and child nodes can inherit properties of its parent nodes. Semantic networks can be used to query if two concepts are related or not. For example, what are the relations between these concepts, and what is the nearest concept to another pair of concepts? Figure 2-1 is an example of a simple sematic network representing relationships between people. This network can be used to find relationships between people. CASNET, ABEL and CHECK use a limited set of arcs to represent knowledge between nodes of different layers of knowledge. The Semantic Web (Web 3.0) and Web of data (Linked data) are essentially a large semantic network with interconnected databases (Shadbolt, Berners-Lee and Hall, 2006).

Semantic networks cannot represent certain types of knowledge easily, for example Non-taxonomic knowledge. Lack of formal semantics means that there is no agreed upon definition of concepts and their relations. This results in ambiguities when representing knowledge. The system will be limited by a user's understanding of links between concepts.

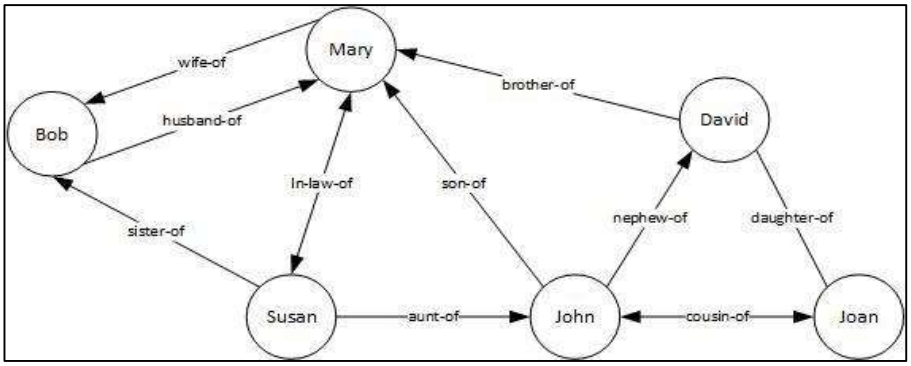


Figure 2-1 - Semantic networks

2.3.1.2 Frames

Semantic networks developed into the Frame representation languages during the '70s and '80s. The 'Frame' concept was first proposed in 1970's (Minsky, 1975) and subsequently developed into several tools for representing knowledge. A frame is represented as having a set of slots. A slot represents a relation to another frame or a value. A slot can have a range of possible values and a default value. Frames are used to present knowledge in narrow domains while sematic networks are used for knowledge representation of broad domains. Frames allow nodes to have a structure while semantic networks do not have that feature. The frame is like a record and fields and values of a record correspond to slots of a frame. A group of slots together represent an object within the domain. From semantic network shown in Figure 2-1, the following frame representing *Mary* can be created.

Mary:
Sex: Female
Spouse: Bob
Child: (John)

Sex, *Spouse* and *Child* are slots. The slots can hold multiple values.

Frames can inherit features of parent node. So, a feature that is common to all members of a class need only be represented once in the superclass. Like semantic networks, frame languages do not have standards for representation frames. Present Illness Program (Pauker *et al.*, 1976b) uses a frame based representation to model its knowledge. Knowledge base can be organized as a hierarchy of different frames. Frames can model real-world objects by using generic knowledge for representing majority of an object's attributes and use specific knowledge only where needed. Frames representing more generic concepts are at the top of a hierarchy and individual/instance level representation lies at bottom levels of a hierarchy. This is like class hierarchies in object-oriented languages. Frames can be classified based on their applications into:

- Situational frame: Contains knowledge of what to expect in each situation
- Action frame: Contains knowledge of actions to perform under different conditions
- Combination: A combination of situational and action frames can be used to represent causal knowledge and cause-effect relationships between concepts.

Frame based representation techniques have advantage of simplicity and ease of understanding and therefore makes it an ideal candidate for knowledge representation in modern applications such as semantic web (Lassila and McGuinness, 2001).

2.3.1.3 Description Logics

Description logics (DL) are a family of logic-based knowledge representation languages that allows the representation of knowledge within a domain in a structured manner (Baader *et al.*, 2003). DL evolved from a combination of Frames, Predicate logic, Semantic networks and KL-ONE (Brachman and Schmolze, 1985) another knowledge representation language developed from semantic networks. Frames and semantic networks lacked formal logic-based semantics and DL were introduced to overcome this

limitation. DL enables knowledge representation alongside logic-based semantics and reasoning capability. This limits expressiveness of the language when compared to First Order Logic based systems but improves computability. DL differ from semantic networks and frames because it is equipped with formal logic-based semantics.

First Order Logic (FOL) has very high expressivity. It is possible to model any domain knowledge using FOL. However, it is too complex for modelling and domain experts will have difficulty arriving at a consensus because of the different methods used to represent the same concept in FOL. It has problems with decidability. Therefore, DL were developed as a fragment of FOL, to overcome some of the problems with FOL. DL are decidable and have enough expressivity to model real world domains.

DL consists of a set of axioms called a Terminological Box (TBox) and an Assertional Box (ABox) (Baader *et al.*, 2003). The knowledge base is the combination of TBox and ABox. TBox contains knowledge about concepts of a domain whereas ABox contains knowledge about individuals within a domain. DL have 'Constructors' which can be used to create complex concepts and roles from simpler concepts and roles. Some of these constructors include Conjunction, Disjunction, Negation, etc.

With the rise of Semantic Web there was a requirement for a web ontology language to represent the domain knowledge comprising the web. This led to development of (Web Ontology Language) (Harmelen and McGuinness, 2004). OWL is based on DL and exploits around 15 years of DL research.

2.3.2 Limitations of the current web in supporting diagnosis

The main problem with information on the web in its current form is that information on the web has been primarily designed for human consumption. How can information in these web repositories be used for supporting diagnosis in clinical practice? For example, if a clinician wants to search for Mast Cell Activation Disorder (MCAD-a rare disease) he/she must open the browser, enter the term of interest into a search engine, browse through the list of resources retrieved by the search engine and enter the resource, in this case a website for MCAD. Within the resource, he/she must search for appropriate type of information, for example, most common signs of MCAD. Data on the resource may be outdated and may not have all the information a clinician need. The clinician will then have to open another resource and look for additional information that he/she needs. This process takes time and effort. The CDSS application in the clinician's clinic is not able to access information on the web in its current form.

Tools like ISABEL (Ramnarayan *et al.*, 2003) are web based and include a large repository of knowledge that acts as knowledge base. It is however a standalone differential diagnostic generator application that cannot exchange data with a patient's EHR and cannot find other information about a patient and other information on the web that may relate to that patient. ISABEL uses web-based technology to deliver recommendations though tools like ISABEL do not fully exploit the full potential of the web. One of the reasons why the tools are not able to fully exploit the data on the web is the data is not machine readable or sometimes only partially machine readable (metadata). The content described on the website cannot be understood by the machine i.e. CDSS (machine understandable) and synthesized to support the clinician in clinical decision making.

Another issue with data on the web is that data is frequently outdated and resources that depend on this data are therefore outdated (Purcell, Wilson and Delamothe, 2002). Data needs to be synchronized so that when one resource is updated all the other resources that depend on this updated resource are updated with the up-to-date information. This is particularly crucial for rare diseases as new research findings pertaining to the rare disease should be incorporated into the web resources. At present this process is done manually and organizations must spend a lot of money in maintaining the content of these resources. The other major problem with information on web is that information is locked away in data islands or data silos. Information about drugs and treatment for MCAD will be hosted and maintained by one organization while information about clinical trials and latest research in field of MCAD will be maintained by another organization. The task of synthesizing information from all these disparate sources is a challenge and ensuring information is up-to-date is a major challenge.

Information on the web has a diverse range of representation formats and data structures (Bizer *et al.*, 2012). The information in one resource will be represented using a format and model that is specific to an organization maintaining the information. For example, a website for patients describing MCAD signs may describe symptoms and signs using terminology commonly used by patients to describe their signs and symptoms. However, another website for researchers may describe these same signs and symptoms using technical terminology favoured by researchers and clinicians. Traditional search engines can find the right information, if you know and can enter the right search terms (keywords). However, if a clinician is conducting an exploratory search to find all information about MCAD, he/she may not know right terms being used at outset. Therefore, a clinician must spend a lot of time going through disparate sources to find right terms and then query other sources using search engine.

One way of ensuring that different web resources are linked up and share data is to have a centralized repository of data that collects and stores the data from several different sources (Weber *et al.*, 2009). Linking data from these sources requires writing purpose-built code that can then update all dependant resources appropriately when a central database record is changed or updated. This type of centralized management of data is not scalable beyond a point and discourages people from putting their data and information on the web. Having a centralized repository for all information on the web requires the different stakeholders interested in information about a rare disease to come together and agree on terminology to be used to describe terms used in their resources, which again is difficult to achieve for rare diseases as funding sources are limited (Griggs *et al.*, 2009).

So, the goal is to have a decentralized web which allows people to add data to the web in any format and intelligent software agents including CDSS applications will be able to search for the right up-to-date information and present this information to the clinician in a format that can be readily consumed by a clinician. The current web infrastructure allows the linking disparate web pages with each other with help of Uniform Resource Locaters (URL). The same concept of URLs can be extended to point one data item to another with the help of Uniform Resource Identifiers (URI). This is where the concept of the Semantic Web can help solve some of the problems with the current web. The semantic web supports distributed web at the level of data (Berners-Lee, Hendler and Lassila, 2001). The information about a resource can be located anywhere on the web and linked to each other with help of URIs. Information about MCAD drugs can be in one web resource. Information about MCAD drug indications and contraindications can be held in another resource, information about MCAD and its signs and symptoms can be in another resource. Semantic web technology allows us to link these datasets and query them for right information as if they are located on the same database. The different organizations involved in maintaining information can release their information in both a machine readable and human readable format (Rakhmawati *et al.*, 2013). The semantic web infrastructure allows smart applications and intelligent agents to use the machine readable data to drive applications such as primary care CDSS applications (Hendler, 2001).

2.4 The Semantic Web

According to Berners Lee et al. (Berners-Lee, Hendler and Lassila, 2001), the semantic web has been intended as an extension of the current web (Web 2.0). The semantic web provides a framework for sharing data across disparate applications or organizational boundaries. It is commonly referred to as Web 3.0 and is primarily concerned with meaning (semantics) of the data, rather than the structure. One of the core principles of the semantic web is the concept of AAA (Anyone can say Anything about Any topic) (Allemang and Hendler, 2011). According to this principle anyone can publish whatever they like about any topic on the web. This means that an MCAD patient can publish information about his/her signs and symptoms on a web site in a manner that is not just human readable but machine readable and can be combined with other resources. A patient may decide to use any terms that he/she prefers that best describes the condition. As more and more patients add information to the web the data from all these patients can be linked to each other. Gradually a web of data emerges that can then be linked to other similar webs of data. Researchers can add latest information about MCAD obtained from latest clinical trials to a repository that can then be linked to the data added by the patients. A semantic web driven CDSS application can then find this information based on just signs and symptoms of a patient

presenting to the clinic. The CDSS application can query the patient's PHR (once authenticated and authorized) for additional information about the patient's experience with MCAD and link them to current data being collected in the clinic. This additional information present on the web can be used for supporting clinical decision making and driving CDSS applications. However, the information on the web can have many challenges that prevent routine use in clinics. The data on the web could be wrong and sometimes information can be added to intentionally deceive others. The semantic web driven CDSS application must navigate these challenges when providing decision support recommendations. There are concerns about patient privacy and ethical concerns of publishing data on the semantic web.

The semantic web is primarily concerned with 3 types of standards that can help resolve some of these challenges. They are: RDF, RDFS/OWL and SPARQL. Other standards and technology that are part of the 'Semantic Web Layer Cake' (W3C®, 2007) can be seen in the image below (Figure 2-2).

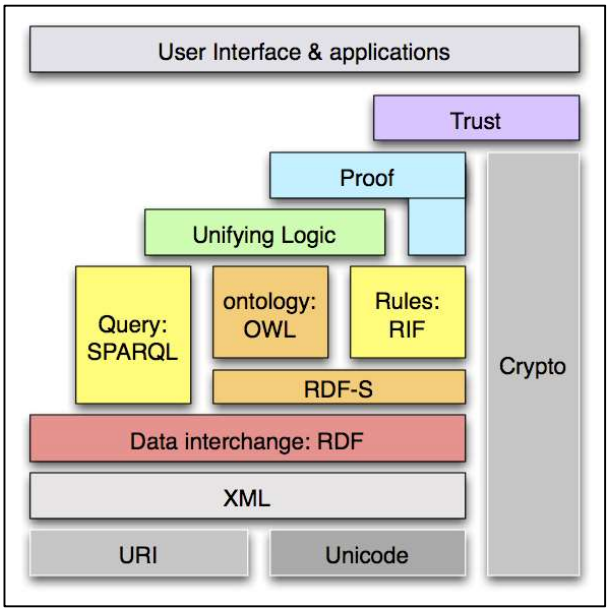


Figure 2-2 - Semantic Web Layer Cake

2.4.1 RDF

RDF stands for "Resource Description Framework". RDF is a W3C (World Wide Web Consortium) (W3C®, 2014) international standard and was developed by W3C for the purpose of representing information and has been used in a wide variety of areas including biomedical and pharmaceutical domains (Samwald *et al.*, 2013). RDF, RDFS and OWL are the basic representation languages of the semantic web (Allemang and Hendler, 2011). The 3 languages allow varying levels of expressivity. RDF is the simplest data representation model. It allows a person modelling a world around him/her to make certain basic statements about the world. For example, a MCAD patient named "Bob" describing pain can represent a fact/assertion that he has pain in following format.

PREFIX ex: <http://../PatientData>
PREFIX ont: <http://../ontology>

ex:patient124 ont:fullName "Bob Dillon".
ex:patient124 ont:value ont:PainPresent.

Each assertion has a Subject, Predicate/Property and an Object/Value. A Subject, Predicate and Object together form a "triple". The above triple states that a patient with record number 124 has a full name of "Bob Dillon". It stated that the patient has a value called "PainPresent". A set of triples can be taken to form a graph structure. The graph structures can be extremely complex and can comprise of billions of triples when scaled to the World Wide Web (Corcoglioniti *et al.*, 2015).

RDF uses URIs to uniquely identify resources on the web. This makes it easy to link existing resources.

For example, if an existing terminology/ontology used in biomedical informatics such as SNOMED CT (Spackman, Campbell and Côté, 1997; Cimino and Zhu, 2006a) has a term indicating "Abdominal Pain" then that can be used to annotate a field within an EHR. This reduces the ambiguity in defining terms and helps in finding the definition of terms. This is useful for clinical decision support, as different EHRs can have different terms to describe the same concept. If they are linked to the same vocabulary/terminology/ontology definitions and terminology codes, then a CDSS can operate across multiple EHR systems without the need for adapting a CDSS for each EHR system. If 2 web-based applications want to refer to the same resource, then they agree on a common URI for that resource. This way a CDSS application can uniquely identify a patient using URI and identification number used to denote a patient. It can possible uniquely identify the symptom described by the patient by linking it to a reference terminology.

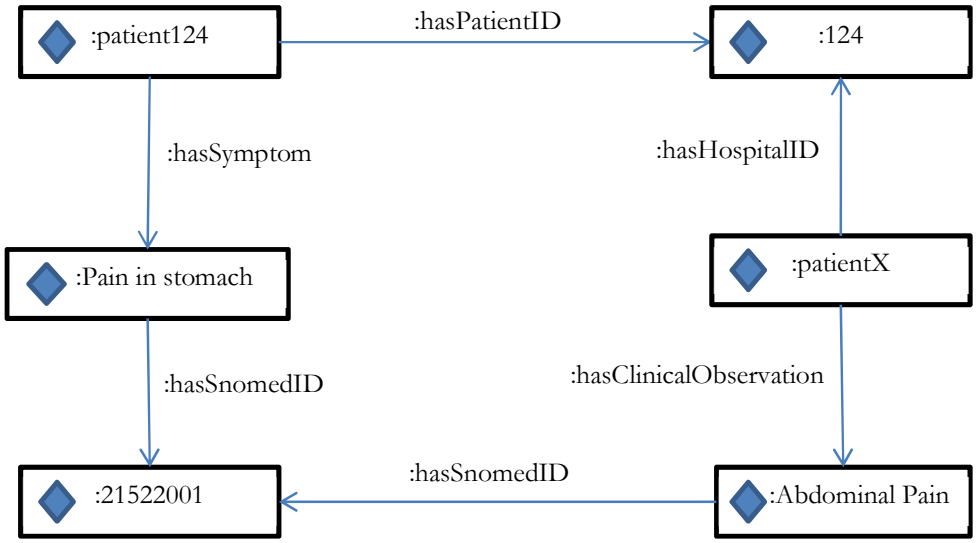


Figure 2-3 - Patient data RDF representation

In the figure above (Figure 2-3) a patient has described his/her symptoms on a web resource (for example, a website for rare disease patients) and has described his/her symptoms using a standardised terminology. Another patient data has described a patient's clinical findings in more detail and has described the pain as abdominal pain. Using the common URI with a value of ":124" we can link the data of both these resources and infer that they are from the same patient. The CDSS will be able to use the data from both these resources and use them for diagnosis even though the data was described in different ways in the different systems.

RDF is a highly flexible data model and can incorporate various other data models including relational models and other graph models. The relational data model can be translated to an RDF data model and can be combined with an existing graph. It is format independent. RDF can be represented as RDF/XML, Turtle, N-triples, JSON (W3C®, 2013a) and various other serialization formats. So, data from multiple sources using multiple data models can be integrated with ease (if the data is mapped to a RDF data model). The emphasis of representing information using RDF is on the meaning (semantics) of the information.

2.4.1.1 qnames

For ease of representation of URI, a simplified version of URI called 'qnames' are used. 'qnames' have 2 parts: a namespaces and the identifier (Allemang and Hendler, 2011). In the figure above (Figure 2-3), ":124" is the identifier and the namespace can be any model that has been used to describe that instance (for example, a hospital patient data representation model).

2.4.2 Semantic Interoperability

Different healthcare systems use different vocabularies and formats to represent their data, even though from a point of view of semantics the data may have the similar meaning. In order to achieve semantic interoperability the different systems should be able to exchange data in such a way that both the systems are able to "automatically interpret the information exchanged meaningfully and accurately in order to produce useful results as defined by the end users of both systems" (EN 13606 Association, 2016). The lack of semantic interoperability is a significant problem in healthcare and according to some estimates proper semantic interoperability can create savings up to $30 Billion (West Health Institute, 2016).

2.4.2.1 RDF as a Universal Healthcare Language

The Yosemite Project (Yosemiteproject.org, 2016) aims to achieve semantic interoperability of all structured healthcare information. They have released a roadmap for achieving this goal and one of the key steps includes using RDF as a universal healthcare information representation. Using RDF as a universal information representation format will provide health informaticians a common way to represent the meaning of the data. In medical diagnosis RDF gives us a standardised way to represent the relations between the diagnosis and observations that either support or negate a diagnosis.

Yosemite Project has identified several reasons why they think RDF is the best candidate to act as a universal healthcare information representation standard (Booth, 2014b, 2015c) :

- **RDF enables smarter data use and automated data translation:** RDF supports inference which allows you to derive new assertions from old assertions. For example, if you are querying for family history of diabetes and if the patient has a grandparent that has a history of diabetes, you can automatically infer that the patient has a family history of diabetes without asserting this fact directly. So even if the patient's EHR does not contain any record of family history but a social worker's record has details of the patient's diabetic grandparent, you can still infer the fact that a patient has a family history of diabetes by simply asserting the fact that the patient has the same grandparent as the one mentioned in the social worker's record.
- **Multiple data models and vocabularies can be easily combined and interrelated:** RDF allows multiple schemas and data models to co-exist and be semantically interconnected. There are several different data models and vocabularies in healthcare. The decision logic that drives one decision support system and utilizes data from an EHR that uses one information model and vocabulary cannot be reused in another decision support without considerable effort and cost. RDF allows different data models, vocabularies and decision models to co-exist by assisting in translations from one data model to another. For example, different organizations and researchers have different interpretations of which diagnostic criteria are accurate. Some researchers might insist upon a more comprehensive diagnostic criterion for diagnosis; however, others may have a simpler diagnostic criterion for the same diagnosis. Both these models will be able to co-exist when represented in RDF and can be translated easily from one model to another.
- **RDF captures information-not syntax:** RDF is format independent and can be represented in different syntaxes i.e. Turtle, RDF/XML, N3 etc. All these formats represent the same RDF information. Data from diverse sources having different syntaxes but the same meaning can be exchanged easily. The emphasis is placed on the meaning of the data rather than syntax. Diagnostic criteria are generally published as text-based representation in clinical guidelines or text books. The machine-readable formats can include several different formats including RDF, XML, CSV, JSON etc.
- **RDF is easy to map from other data representations:** RDF is composed of atomic assertions or Triples. A triple comprises of a Subject-Object-Predicate structure. This makes it easy to represent any form of data and incorporate any data model. Diagnostic CDSSs must use information from diverse sources that may represent data in different formats. For example, family history of a disease may be recorded in primary care, but not necessarily in other places. Blood pressure recorded in a wearable computing device will be a different format from that recorded in primary care. However, the decision support systems should be able to integrate all this information at the right place and time and assist the clinician in decision making. As mentioned previously one of the main reasons for cognitive error is faulty information gathering and synthesis. Modelling knowledge in RDF will help facilitate the process by capturing the information in a domain neutral syntax-independent manner. Semantic web tools can then integrate this information for the decision support to make diagnostic recommendations.
- **RDF is self-describing:** RDF uses URI (Uniform Resource Identifiers) to identify resources. This makes it possible to link the identifier to authoritative definitions or vocabularies. This helps reduce ambiguity when describing resources. The observations and diagnoses in a diagnostic decision support can be linked to ontologies/taxonomies/vocabularies that define these concepts.

2.4.3 RDF Schema (RDFS)

To make sense of the RDF data we use a data definition specification called RDF Schema (RDFS). It is officially called the "RDF Vocabulary Description Language". RDFS contains the definition of the classes, properties and restrictions (Brickley and Guha, 2014). It contains definitions of subclasses and super classes, and sub-properties and super-properties that link these different classes to each other and to literal values. RDFS is the schema definition language for RDF datasets (Allemang and Hendler, 2011). RDFS can be used to define a data model that can be used to create simple ontologies and was an early attempt at producing an ontology representation language.

2.4.3.1 Ontologies

Ontologies are one of the different types of knowledge representation formalisms currently available and the most popular. Ontologies are considered one the key pillars of the Semantic web (Klaczynski *et al.*, 2012).

An ontology can be defined as an "*explicit, formal specification of a shared conceptualization*" (Gruber, 1993). Conceptualization is the forming of an abstract model by identifying the classes and relations within a domain of interest. When all the concepts within the domain are defined, it is considered explicit. If there are any undefined concepts, then the ontology will fail. Formalization enables the model to be machine understandable, not just machine readable. The machine can interpret the meaning of the content. The model can be considered a shared model once there is consensus regarding the concepts and the relations within the domain.

The information within the semantic web is given a well-defined meaning which allows computers and humans to share information and work in cooperation. However, two different databases may refer the same concept using different terms. For example, one terminology may use one term to define a concept, while a different terminology may use another term. Databases that are annotated with these different terminologies will not be able to share data since they use different standards. Ontologies, by providing the explicit representation of a concept allow a program to compare the two databases and combine the data within them. The success of the semantic web is dependent on whether ontologies are widely accepted and used.

Ontologies can have Classes, Properties and Instances and is modelled using languages such as OWL. The classes represent the concepts within the domain. The classes have certain attributes that can be used to describe it. The members of a class share the attributes. The classes can be related to other classes through properties. In a clinical/biomedical ontology the classes and relations represent the concepts and their relations that are needed to represent that clinical domain. Thus, ontologies can be used to formalize the knowledge of domain experts such as dentists and other clinicians. The instances/individuals represent concrete objects such as people, drugs etc. Ontologies provide a means for classifying these instances, even though instances are not explicitly part of the ontology.

In addition, certain constraints can be placed over these classes and relations to define what classes and relations are allowed in a domain of interest. For example, in Figure 2-4 the disjunction constraint can be placed on the subclass relationship to define that both Man and Woman belong to disjoint classes. Or a member of the 'Man' class cannot belong to 'Woman' class as well.

The classes, relations and constraints can be combined to form certain 'Axioms' which describe parts of the knowledge that cannot be described by other components. A set of axioms that provide logical assertions about classes, relations and instances form an ontology representation.

DL based ontologies can exploit DL based reasoning tools to provide inference capability. Therefore, machines can infer implicit knowledge from explicitly defined knowledge bases. DL based reasoning support can be used to ensure quality control in ontologies. These reasoning tools can be used during ontology design phase and test for contradictions in concept and property representation. Reasoning tools can be used for integrating different ontologies together.

Ontology combines artificial intelligence and machine language to assist in sharing and reusing existing knowledge and combining ontologies to create new knowledge sources(Neches *et al.*, 1991). In the clinical domain, ontologies can be used to model diseases, symptoms, causes of diseases, their relationship to other aspects of the disease process and so forth. Other uses of ontologies include information retrieval, natural language processing, knowledge management, etc. (Uschold and Gruninger, 2009). Ontologies have found use in areas such as Big Data analysis and interlinking medical and dental data and performing cross-domain data analysis (Shah, Rabhi and Ray, 2014).

2.4.3.2 RDFS models

:Planet rdf:type rdfs:Class.

:Earth rdf:type :Planet.

The triples above state that Planet is a class and Earth is type of Planet that belongs to the Class "Planets". RDFS is based on formal semantics and this allows RDFS to obtain valid inferences from existing data. For example, we can state that MCAD is a type of immunological disorder using the triple:

:Immunological_disease rdf:type rdfs:Class.
:MCAD rdf:type rdfs:Class.
:MCAD rdfs:subClassOf :Immunological_disease .

We can state that a patient (:124) has a disease MCAD using the following triple:

:124 :hasDisease :MCAD.

From this triple, we can infer that the patient has an immunological disease since MCAD is a sub-class of immunological disease:

:124 :hasDisease :Immunological_disease

However, RDFS is limited by its ability to represent complex knowledge that can be read by machines. For example, in Figure 2-4, both Man and Woman are subclasses of Human. However, individuals belonging to Man and Woman cannot belong to the both classes. We cannot represent such disjunctive class relationships using RDFS.

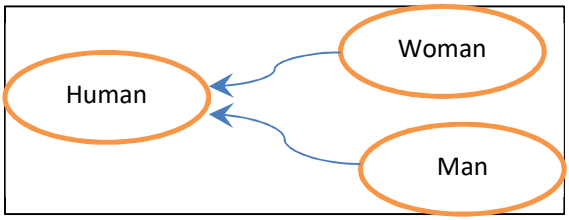


Figure 2-4 - Subclass relationship example

The other key problem of RDFS models is that RDFS cannot be used to denote negations. For example, if we denote that a clinical observation is present then it does not automatically infer that another 2nd clinical observation is absent. RDF and all other semantic web languages follow an Open World Assumption. In a closed world if one clinical observation is present, then the fact that the 2nd observation is absent is automatically inferred. In an open world unless we assert/infer the statement that the 2nd clinical observation is absent we cannot say for sure that the 2nd observation is absent or not. The open world assumption assumes that the information is incomplete. We cannot therefore automatically infer that the 2nd clinical observation is absent if the 1st observation is present. Therefore, there is need knowledge representation models that are based on mathematical logic or a formal knowledge representation that is capable of more semantic expressivity for expressing the relationships between these concepts.

2.4.4 OWL

OWL (Web Ontology Language) is the de-facto standard language for representing ontologies on the web. The current version is OWL 2.0 (W3C OWL Working Group, 2012). RDFS allows us to model simple classes, properties and the relations between these classes. OWL has vastly increased semantic expressivity and overcomes some of the shortcomings of RDFS language.

OWL comes in different types, ranging between First Order Logic based representations to plain concept hierarchies varying in semantic expressivity:

- OWL Full

- OWL DL
- OWL Lite

OWL DL is the only decidable version and is based on DL *SROIQ(D)*. OWL DL was developed so that users can have maximum expressiveness without losing computational capability and decidability since DL are a family of formal knowledge representation languages that have more semantic expressivity than propositional logic, but has higher decidability than First Order Logic (Baader *et al.*, 2003).

OWL enables the knowledge engineer to model concepts in such a way to enable maximum reuse of concepts. So concepts and their relations that are defined in ontologies such as the Oral Health and Disease (OHD) ontology (Ontobee.org, 2016) can be defined once and reused in inference models, information models and for linking existing terminology models.

2.4.5 Types of ontologies

The more knowledge about a domain that can be captured in the form of ontologies the better knowledge-based systems can the process information and communicate results. According (Noy and Mcguinness, 2001) some of the reasons for developing ontologies include:

- "To share common understanding of the structure of information among people or software agents
- To enable reuse of domain knowledge
- To make domain assumptions explicit
- To separate domain knowledge from the operational knowledge
- To analyse domain knowledge"

van Heijst et al (van Heijst, Schreiber and Wielinga, 1997)have classified ontologies based on two dimensions:

- The amount and type of conceptualization:
    - Terminological ontologies: Specify terms used within the domain (Cimino and Zhu, 2006b).
    - Information ontologies: Used to represent the data structure of databases.
    - Knowledge modelling ontologies: These ontologies are used to specify conceptualizations of the knowledge within the domain of discourse. They have far greater detail in representing the knowledge.
- The subject of conceptualization:
    - Application ontologies: Model the knowledge required for an application
    - Domain ontologies: Model the knowledge of a domain
    - Generic ontologies: Like domain ontologies but have more generic concepts that are applicable across several domains
    - Representation ontologies: Domain neutral knowledge representations

Similarly Guarino (Guarino, 1997) classified ontologies into:

- Top level ontologies/Upper-ontologies: Ontologies that describe general concepts like space, time, events, etc. and are independent of a domain. They are cross domain ontologies and are used primarily in philosophy to classify the world around them.
- Domain ontologies and Task ontologies: These ontologies are focussed on a domain or a specific problem/ task.
- Application ontologies: These are specialization of domain or task ontologies and are oriented towards application of these ontologies in an application.

Ontologies can belong to different types/categories based on their level of complexity (Lassila and

McGuinness, 2001) :

- Controlled vocabulary
- Glossary
- Thesauri
- Taxonomies
    - Informal IS-A Hierarchy
    - Formal IS-A Hierarchy
    - Formal Instance
- Frames
- Value restrictions
- General logical constraints
- Disjointness, Inverse, part-of etc.

2.4.5.1 Biomedical Ontologies

When ontologies describe the classes, relations and properties of the biomedical domain it is called a Biomedical Ontology. Biomedical terminologies are a collection of terms used within the biomedical domain, while biomedical ontologies describe the concepts and the relations of that domain as well. Several biomedical ontologies are currently available that are stored in large ontology repositories and can represent anything from the molecular level, to clinical and public health level as well. These ontologies primarily help in standardising terminologies and can be used for standardised information exchange and knowledge management. Development of the domain ontology is a crucial step in the development of a clinical knowledge management system.

Once the domain knowledge has been captured in the form of ontologies it forms a common understanding of the domain by acting as a source of well-defined terms. The ontology can then be consumed by humans and/or machines and used for communication purposes. The use of semantic web and related technologies in CDSSs, especially diagnostic systems have been increasing in recent years (Blomqvist, 2014). Ontologies are normally represented using standard machine-readable languages like RDF or OWL. RDF lacks the description of a relationship and therefore can be complemented by OWL.

GALEN (Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in medicine) was developed to provide a re-usable application independent terminology for clinical information systems (Rector and Nowlan, 1994) . The language used to describe this terminology was called GRAIL (GALEN Representation And Integration Language)(Rector *et al.*, 1997). It can be described using OWL. An open-source foundation called OpenGALEN was later developed to distribute the GALEN reference model and provide a platform for developers and vendors to use the model in developing applications (Rector *et al.*, 2003).

The UMLS (Unified Medical Language System) is collection of various controlled vocabularies within biomedical science (Lindberg, 1990). The Metathesaurus forms the foundation of UMLS and consists of 1 million concepts and 5 million concept names. Each concept in the Metathesaurus belongs to one or more semantic types and is linked to one another through semantic relationships. However the separate vocabularies used with UMLS does not have a common architecture and therefore cannot be mapped easily and combined to work in a single system (Campbell, Oliver and Shortliffe, 1998).

The Systematized Nomenclature of Medicine (SNOMED) Clinical Terms (CT) was developed by combining SNOMED Reference Terminology (RT) and Clinical Terms Version 3 (CTV3) (previously called Read Codes)(Wang *et al.*, 2001). SNOMED RT was developed by College of American Pathologists and CTV3 was developed by the National Health Service (NHS) in the United Kingdom. SNOMED CT is a multinational and multilingual terminology and is the most comprehensive biomedical terminology currently available (Lee *et al.*, 2014).

Within dentistry, SNODENT is diagnostic vocabulary designed for use in dental EHR systems (Goldberg *et al.*, 2005). However it suffers from problems in clinical coverage of terms when compared to

SNOMED CT (Torres-Urquidy and Schleyer, 2006) and will require the use of other ontologies/terminologies to overcome deficiencies. Another standardised dental terminology used in dentistry is the Dental Diagnostic System(DDS) (Kalenderian *et al.*, 2011a).

The Open Biomedical Ontologies (OBO) Foundry is an effort by several international organizations to create controlled vocabularies that can be shared across different domains and "create and maintain an evolving collection of non-overlapping interoperable ontologies that will offer unambiguous representations of the types of entities in biological and biomedical reality" (Ceusters and Smith, 2010). The aim is to develop a family of interoperable and accurate representations of biomedical reality. OGMS (Ontology for General Medical Science) is an ontology that has been developed to describe all the entities that the clinician can encounter during a patient encounter (Scheuermann, Ceusters and Smith, 2009). OGMS uses terms that are applicable across a wide range of medical disciplines and is currently available in the OBO Foundry. OGMS uses BFO (Basic Formal Ontology) as an upper level ontology to link different ontologies that use the same upper level ontology (Grenon, Smith and Goldberg, 2004). OGMS can be further elaborated to develop disease specific ontologies like OHD (Oral Health and Disease Ontology) (Schleyer *et al.*, 2013a) or ODO (Ocular Disease Ontology) (Ray and Diehl, 2013).

WordNet is a lexical database that groups English words into sets of synonyms called 'synsets' (Fellbaum, 2010). It provides definitions of these words and records sematic relations between these synsets. This general language ontology has been used in information retrieval and other clinical applications such as processing of natural language in medical records (Chankai, Prestrud and Brooks, 2005).

2.4.6 Rules on the semantic web

Rules have been used developing intelligent applications for several decades (Grosan and Abraham, 2011). Rules are more expressive than using the OWL language alone and can be used to describe relations between concepts that cannot be described using DL (Schröder and Wagner, 2003). The knowledge represented in ontologies is declarative in nature. Rules are a common type of procedural knowledge that is widely used in knowledge engineering. The knowledge of the domain expert is captured in the form of rules that have an IF->THEN format. When rules are used in conjunction with OWL or RDFS models then some of the weaknesses in expressivity of OWL or RDFS can be overcome. Rules languages that can be used with RDF data include N3 rules, Jena rules, SWRL (Semantic Web Rule Language), SPIN rules etc. (Schröder and Wagner, 2003).

A rule contains an antecedent and a consequent. The antecedent forms the body of the rule and the consequent form the head of the rule. The antecedent is composed of atomic statements and the consequent consists of conclusions that can be made from these statements. The rule can be interpreted in the following manner; if the body is true, then the head must be true as well (Sherimon and Krishnan, 2016). The rules can be represented in different formats and languages. Currently there are several different rule based systems and semantic web rule languages in use today (w3.org, 2005). One of the earliest proposals for a rules language for the semantic web was the Semantic Web Rules Language (SWRL) (Horrocks *et al.*, 2004). SWRL combines OWL with RuleML (Rule Markup Language). RuleML is a subset of Datalog rule language (Boley, 2006). For example, if we want to express the rule that a person has an uncle another person we can express this rule in the format below:

hasParent(?x,?y) ^ hasBrother(?y,?z) -> hasUncle(?x, ?z)

According to this rule if the instance represented by ?x has a parent ?y, and ?y has a brother ?z, then we can conclude that ?x has uncle ?z.

2.4.7 SPARQL (SPARQL Protocol and RDF Query Language)

Once the data has been represented in the form of RDF triples (with or without OWL/RDFS models) standardised mechanism is needed for accessing the data or querying the data (W3C®, 2013b). RDF triples are traditionally stored in a RDF store, commonly called a Triple Store. It is possible to translate existing relational data into RDF and use that as the source of RDF data (C. Bizer and Seaborne, 2004).

RDF stores are different from relational databases in its ability to allow merging of datasets regardless of the schema used. Just like SQL allows the user to access data stored in a relational database, SPARQL language allows the user to access and manipulate data present in RDF format. It allows the user to traverse a graph pattern and retrieve specific RDF patterns. The current version SPARQL 1.1 is a W3C standard, and is gaining wide adoption within the informatics community (Sahoo, Sheth and Zhang, 2012).

SPARQL query can be used to query previously unknown relations. This is quite important when handling rare cases, especially when diagnostic decisions must be made quickly. SPARQL can be used to query heterogeneous databases using a single query. Therefore, patient data and other clinical data residing in disparate databases can be queried easily. The SPARQL query will traverse the entire RDF graph and match the triple pattern within the query against RDF triples within the knowledge base. Processing of a SPARQL query is essentially matching of graph patterns that exist within the knowledge base.

Schleyer et al. (Schleyer *et al.*, 2013b) has demonstrated the use of the OHD ontology to annotate data from multiple commercial dental EMR databases and used SPARQL to query the data to answer questions such as "How long will this filling last?"

A SPARQL query is an RDF graph with variables. Consider the following example triple:

ex:patient124 ont:fullName "Bob Dillon".
ex:patient124 ont:value ont:PainPresent.

If we replace the Subject, Predicate or Object with variables (?x , ?y) we get the basic structure of a SPARQL query.

?x ont:fullName "Bob Dillon".
?y ont:value ont:PainPresent.

If we want to query a database for all Subjects that has the Predicate ont:fullName and Object of "Bob Dillon" we use the following query.

Prefix ont: http://../ontology

```
SELECT  ?person
FROM <http://example.com/dataset.rdf>
WHERE {
  ?person ont:fullName "Bob Dillon".
}
ORDER BY  ?person
```

2.4.7.1 SPARQL rules

In addition to SELECT, there are other return clauses including ASK, DESCRIBE and CONSTRUCT (Gearon, Passant and Polleres, 2013). CONSTRUCT keyword can be used to introduce a graph pattern that can be used as a template for constructing new graph patterns. It allows us to specify templates of new information based on patterns present within the existing information. This type of pattern can be considered a rule. A sample rule may have the following assertion: "If the patient data has pain present, then it supports the Reversible Pulpitis diagnosis". These rules can be logical rules or business rules.

Using the RDF data represented above we can write the following rule.

CONSTRUCT {?Patient :hasDiagnosis :Diagnosis1 }

WHERE{ :x :hasValue :PainPresent}

This rule specifies that if the data :x has value :PainPresent, create/construct a RDF statement stating that an instance of a ?patient has diagnosis Diagnosis1. The same CONSTRUCT template can be used to insert new values into the database by replacing the CONSTRUCT keyword with INSERT. It is possible to INSERT the data into a specified named graph using the INSERT INTO clause. This gives us several advantages over other rule languages. It is possible to insert data into one graph, while the patient data and the OWL model are on different graphs. It is possible to query/insert data that is present in the completely different triple store through a SPARQL endpoint. Since SPARQL is a W3C standard it is supported by all triples stores that can store data as RDF triples. There is no need for an additional rule engine for executing rule. The SPARQL rules can be executed using the SPARQL engine of the triple store. The rules can be expressed using SPARQL construct or insert statements as shown above. Therefore, there is no need to learn a different rule language to express rules for semantic web applications.

SPIN (SPARQL Inferencing Notation) is a SPARQL based rule and constraint language developed for the semantic web (Knublauch, Hendler and Idehen, 2011). It is a W3C recommendation as a rule language. Since it is based on SPARQL there is no need to learn another proprietary rule language to implement rules. SPIN and SPARQL has been used to identify data quality problems for data available for semantic web and linked data activities (Fürber and Hepp, 2010).

2.4.7.2 SPARQL endpoints

SPARQL endpoints are RDF triple stores that can be accessed via the web via HTTP protocol. SPARQL is not just a query language, but a protocol to access data and manipulate data over the web. The SPARQL query will be processed and the results will be sent back to a user. This is helpful for exploring public datasets that can be accessed via the endpoints (Buil-Aranda *et al.*, 2013). It allows transformation of data from one vocabulary to another.

It is possible to store the SPARQL queries as triples. SPARQL queries are generally written as .rq SPARQL files. A semantic web framework such as Jena can read the SPARQL query file and then query a RDF triple store. The RDF triple store can be accessed via a SPARQL server such as Jena Fuseki (Apache Software Foundation, 2015).

2.4.7.3 Federated querying of data

The RDF data model was designed to support federated querying of data (Quilitz and Leser, 2008). Federation is the process of virtually combining data but allowing the data to retain its identity (Allemang and Hendler, 2011). After translating data from different data sources into RDF and the data is published via a SPARQL endpoint the data can be queried with the help of SPARQL. The SERVICE keyword defined in the SPARQL allows the data application to merge the data from dissimilar sources distributed across the web. The information required by the CDSS application pertaining to a patient may reside in another EHR, PHR or any web application that the patient uses. The CDSS application does not need to know where the data is coming from. A single SPARQL query can therefore query multiple data sources and present the result to the CDSS. This is important for clinicians as the patient's own record may be incomplete.

Integrating data from disparate data sources remains a major challenge in health informatics (Searls, 2005; Louie *et al.*, 2007). Researchers and scientists in healthcare have to gather, merge and query data from publicly available databases such as PubMed (NCBI, 2016) to find the information that they need. At present, this is done manually or with the help of scripts that scrape the website/dataset for the right data. This process is extremely tedious even for experienced researchers. A clinician examining a patient has limited time to diagnose and manage a patient and will not have the resources to find all the information needed to decide. Patients themselves may not have accurate information about their previous history of disease. Centralized and distributed approaches have been proposed to make the process of integrating data easier. The centralized approach involves collecting and merging related data into central data repositories called data warehouses. These data warehouses can then be queries directly. The data must be carefully curated, and the sources of the data must be chosen carefully to make the queries more efficient. Bio2RDF (Belleau *et al.*, 2008) is an example of this approach in bioinformatics. Hospital networks have

used the data warehousing approach to integrate data within the organisation (Evans, Lloyd and Pierce, 2012). In the distributed approach, the data sources implement a web service providing access to the datasets. The query is split into subqueries and the client queries the data sources independently and merges the results. The distributed techniques of accessing disparate datasets have several advantages over the centralized data warehousing approach (Mandl and Kohane, 2015). There is no single point of failure as the data sources are distributed across a network. Additional web services can be added to the network without affecting other services and therefore the system become more scalable. Even data from existing warehouses can be included as part of the network. However, there is the additional cost of maintaining the endpoints through which data can be accessed and coordinating access to the endpoints. The quality of the data that can be accesses through these distributed networks is an issue as there is little or no cleaning of data (Louie *et al.*, 2007).

Linked data is a distinct type of accessing data that is distributed in nature. Linked data uses web standards and web technology as a means of sharing and connecting data or information. The Web is an ideal medium due to its inherent distributed and scalable nature and a mature technology stack (Heath and Bizer, 2011).

One of the widely-used methods for making structured data available on the web is the use of Web APIs (Application Programming Interfaces). Web APIs allows an application to perform a query on a structured data set via the HTTP protocol, for example Flickr API (flickr.com, 2016). Application called 'Mashups' combine data from several different sources. Each source of data has an API that is custom to the dataset. Significant effort is still required to integrate the data from each specialized API into a single application. Another problem with the data published and accessed through Web APIs is the use of identifiers to describe data (for example id: 21534 to identify a drug). When the data is examined outside the context of the Web API the data becomes meaningless. There is no standard way to describe the data returned from Web APIs. The data accessible through current Web APIs cannot be linked to other datasets. Therefore the data obtained from these Web API remain isolated an required significant effort to link them to each other to form a Web of data (Bizer, Heath and Berners-Lee, 2009; Heath and Bizer, 2011).

2.4.8 Linked data

In order to maximise utility of disparate data sources Linked Data approaches have been developed and it provides a framework of international standards and best practices for publishing and disseminating data on the web (Berners-Lee, 2006). Linked data tools and technologies provide a mechanism for combining data from several different source and formats from across the web. Once the data has been published on the web based on these linked data principles the data can be linked together to form a Linked Open Data Cloud that integrates data from all over the web. Linked open datasets are publicly available data set that are identified with the help of an URI and is accessible via HTTP protocol.

The data sets are linked together using URI and uses URIs for naming things on the web. URIs are HTTP URIs and therefore people can look up the names used to describe things. Linked data resources provide a means of accessing information when someone looks for information, for example SPARQL for querying the data. Linked datasets include links to other URI so people can discover more information about the thing. These 4 principles are collectively called the Linked data principles (Berners-Lee, 2006). Current web of linked data includes 9960 datasets which consists of >150 billion facts and >39 million links between these datasets (lod2.eu, 2016).

2.4.8.1 Linked data applications in healthcare

When linked data solutions are implemented for healthcare CDSSs and other healthcare applications they will be able to overcome information silos and obtain a comprehensive and holistic view of the patient's data (Bizer, 2009). The CDSS application is no longer restricted by data available at the point of care. Documents and data from disparate sources can be semantically integrated and form part of the linked web of data. In a recent review of the state of the art in semantic web application in healthcare Zenuni et al. (Zenuni *et al.*, 2015) have described some of the linked data applications in use today.

CardioSHARE (Vandervalk, McCarthy and Wilkinson, 2008) is decentralized framework based on linked data principles and provides a SPARQL endpoint for querying data related to heart diseases. Statistical data about public health is highly relevant for research and policy making. The data is currently available in the repositories such as the Global Health Observatory of the United Nations' World Health Organization (WHO). However, these datasets are only available as spreadsheets and therefore cannot be queried easily without significant effort in processing the data. Zaveri et al. (Zaveri *et al.*, 2013) converted these datasets into a RDF format from the native formats. They have setup a SPARQL endpoint to enable ease of querying the data.

Linking Open Drug Data is a project that surveys publicly available datasets on drugs and publishes them and interlink the datasets using RDF (Samwald *et al.*, 2011). The linked data is available as the linked data cloud. The information about drugs can include medicinal chemistry results, impact of drugs on gene expressions and outcomes of drug clinical trials.

LinkedCT (Hassanzadeh and Miller, 2015) is a project that publishes the results of clinical trials data using linked data principles. Existing sources of clinical trial data are transformed to RDF format and typed links between the different datasets and other information resources such as PubMed are generated.

Bio2RDF (Belleau *et al.*, 2008) is one of the largest linked data repositories that publishes biological data and provides a SPARQL endpoint for querying the data. The linked data includes data from clinical trials, PubMed and other biological data. Bio2RDF scripts convert data from formats such as flat-files, tab-delimited files, dataset specific formats, SQL, XML etc. into RDF.

2.5 Ontologies in clinical decision support systems

There are several advantages to using semantic web technologies and ontologies in diagnostic CDSSs. An ontology is a natural way of organizing information which is highly flexible and provides a multidimensional structure to information. Ontologies hold great potential for supporting CDSSs due to its ability to support semantic interoperability (Schulz and Martínez-Costa, 2013). Ontologies have therefore been used to make existing decision support application more intelligent (Blomqvist, 2014). According to Blomqvist (Blomqvist, 2014), RDF and OWL have been used to integrate data from several different sources which can then be used by the decision support applications in both clinical and non-clinical applications. Ontologies are used to model concepts that are common to these datasets and languages like SPARQL are used to query and OWL reasoning engines are used to perform reasoning across these datasets.

In addition to helping integrate data ontologies and semantic web technologies have been used to indirectly support decision making by helping retrieve information for the clinician. Ontologies have been used to improve the accuracy of search and the query string can be expanded using concepts from the domain ontology (Hasan *et al.*, 2015). Ontologies are helpful for refining the query string by reducing the use of ambiguous terms (Sendhilkumar, Mahalakshmi and Rajasekar, 2009; Rajendran and Swamynathan, 2015). Domain ontologies are used for cross-referencing and for determining related concepts (Smith *et al.*, 2007). Ontologies are helpful in exploratory search when dealing with rare and atypical cases (Sarntivijai *et al.*, 2016) especially in cases where the clinician is not sure which query string to use and which domain to look in for the right information (Roitman *et al.*, 2011) .

2.5.1 Ontology Driven Diagnostic CDSSs

Ontologies are both machine and human readable and can be used by people and intelligent software agents to share information. Thereby ontologies help makes knowledge reusable. Knowledge bases developed using ontologies can then be reused for different applications. Utilizing existing general-purpose ontology reasoners allow easy maintenance of the knowledge base. One of the major problems that expert systems in the 80's faced was the difficulty in maintaining large rule bases, and several different methodologies including ontologies were developed to overcome the shortcomings with these traditional expert systems (Shu-Hsien Liao, 2005b). Ontologies can represent the domain knowledge of large and complex domains and their ability represent this knowledge in a machine and human readable

form make them well suited for used in decision support. The availability of domain ontologies makes the process of reusing existing knowledge much easier. In this section, how ontologies and semantic web technologies have been used in diagnostic CDSSs is detailed.

Reyes-Ortiz et al (Reyes-Ortiz *et al.*, 2013) have used 3 sub ontologies to develop their ontology model called Diagnostic Ontology which include Diseases, Symptoms and Risk Factors alongside the relations between them. A similar approach has been used by Hadzic & Chang (Hadzic and Chang, 2005) to develop ontological models consisting of sub-ontologies that act as a reference point to link other existing ontologies. However, the main aim of this ontology is to provide a common language to allow multi-agent to interact and coordinate their activities in a more efficient manner. Information required for clinical diagnosis may reside in separate locations and healthcare knowledge in general is highly distributed in nature. Therefore ontologies can act as inter-lingua to coordinate the information residing in different databases and assist in communication of multi-agents systems (Hu *et al.*, 2011). The ontology model can be setup as a web service which can be accessed by intelligent mobile agents to coordinate their activities and avoid misunderstandings during communications (Wu Zhao, He Yanxiang and Jin Hui, 2005).

Romero-Tris et al (Romero-Tris, Riaño and Real, 2010; Riaño *et al.*, 2012) have developed an ontology called the Case Profile Ontology as part of the K4CARE project aimed at developing knowledge based healthcare model for providing home care to senior citizens. The application that was built on this ontology communicates with the clinician and helps him during the diagnostic procedure and helps validate the final diagnosis. The tool tailors the ontology to a patient's profile. The ontology consists of all healthcare related concepts needed for care of chronically ill patients. These concepts include diseases that can affect chronically ill patients, the signs and symptoms of these diseases etc. The relationships between concepts are formally represented with the help of an OWL ontology. The ontology is then used to personalize information about each patient. The personalization gives the users an overall idea about the patient; identify the actors involved in treating the patient, incorporate this personalized knowledge into the patients EHR and produce a knowledge structure that can then be utilized by CDSS applications. The case profile ontology and the personalized ontology of the patient are then used by the CDSS application to provide a ranked diagnostic recommendation. Each sign or symptom in the case ontology has a weight attached to it. The weight is determined by the total number of diseases (diagnoses) that are related to that sign.

Zhang et al. (Zhang *et al.*, 2014) developed and evaluated an ontology driven diagnostic CDSS for the diagnosis of mild cognitive impairment. They encoded the knowledge needed to diagnose mild cognitive impairment from Magnetic Resonance Imaging (MRI). In addition, the MRI image data set was mined using C4.5 algorithm to classify key features in the images which were then used to generate a rule set for diagnosing mild cognitive impairment. The rules were then encoded in OWL and combined with the ontology. With the help of the OWL inference engine a diagnostic recommendation was generated.

Sanchez et al. (Sanchez *et al.*, 2011) developed a framework for diagnosis of Alzheimer's disease. In this paper they outlined a technique for encoding the diagnostic criteria for Alzheimer's disease in OWL ontology with the help of supporting ontologies such as Semantic Web Application in Neuromedicine (SWAN)(Gao *et al.*, 2006) ontology and the SNOMED CT. In the framework, they have outlined several layers including a data layer, translation layer, ontology and reasoning layer and an application layer. The data layer contains the patient data obtained from EHR, Picture Archiving and Communication Systems (PACS) etc. The translation layer consists of interface applications that can read the patient data and matches them with the ontologies that contain the diagnostic criteria stored in the OWL ontology (knowledge model) contained in the ontology and reasoning layer. A reasoning model performs sematic reasoning based on rules provided by domain experts. The results are then forwarded to the application layer where a natural language processing application translates the rules and the output of the reasoner. The results are then presented to the clinician.

Rodriguez et al. (Rodriguez *et al.*, 2009) developed an OWL ontology that contains 4 classes, namely Disease, Symptoms, Laboratory tests, Medicines and Consult. Consult is a class that contains the

instances of the consultations that was made using the ontology model. There are several object properties that link the classes to each other. For example, the 'hasSymptom' property links the Disease class with the Symptoms class. The approach, called ADONIS, involved a combination of OWL and Jena rules to model the diagnostic criteria and generate diagnostic recommendations. SPARQL was not used for the diagnostic inference process as they felt the version of SPARQL at the time was not expressive enough. The current version of SPARQL (version 1.1) was released later (Gearon, Passant and Polleres, 2013). However, the method did not allow for weighting of symptoms to produce a ranked list of diagnostic recommendations. Symptoms and laboratory tests that negated a diagnosis could not be modelled and the presence and absence of a clinical observation could not be linked to the diagnosis. The software architecture that generated the recommendations was called ODDX (Rodríguez *et al.*, 2009a) and used the Jena semantic web framework. García-Crespo et al. (García-Crespo *et al.*, 2010a) further improved upon this software architecture to include probabilistic reasoning. Each symptom was given a weight of Very Low, Low, Medium, High, and Very High. Each disease has probability of its being present or not. Using the weights of the symptoms a probability for each disease is calculated by the probabilistic component of the system. Once the diagnosis is inferred using the probabilistic component of the system generates probability of the disease being present based on the weight of the symptoms. The weights and the probabilities were not modelled within the OWL ontology model and created separately and linked to the diagnosis. The ADONIS approach was not capable of multi-level diagnosis. Diseases sometimes have other disease as its findings. For example, Common Cold has Laryngitis as a finding. However, laryngitis has sore throat, cough and aphonia as its finding. CDSS based on ADONIS approach was only capable of obtaining diagnosis over a single layer. So, the finding of Laryngitis should be made explicit by the clinician. Laryngitis cannot be inferred from cough, sore throat etc. (Alejandro Rodríguez-González, Torres-Niño, Mayer, *et al.*, 2012).

Rodríguez-González et al. (Rodríguez-González *et al.*, 2011) developed another diagnostic ontology model to support diagnostic CDSS. This model called the Diagnosis ontology has several sub ontologies which includes the Clinical findings ontology and the Drugs Ontology. The Clinical Findings ontology has 3 further sub ontologies called Signs, Diagnostic Tests and Diseases ontology. Their approach involved using the SNOMED ontology (Spackman, Campbell and Côté, 1997) to extract terms to annotate the terms within the Diagnosis ontology and its sub ontologies.

SeDeLo (A Rodríguez-González *et al.*, 2012) improved upon some of the weaknesses of the ADONIS approach of modelling diagnostic criteria using DL and rules. The SeDeLo approach was capable of multi-level diagnosis. So, laryngitis can be automatically inferred from its findings, and laryngitis will then support the diagnosis of Common Cold. The system based on the SeDeLo approach was found to be faster, more efficient and more capable than the previous methods (Rodríguez-González et al. 2012). Rodríguez-González & Alor-Hernández (Rodríguez-González and Alor-Hernández, 2013) combined the SeDeLo approach with the Diagnosis ontology described by Rodríguez-González et al. (Rodríguez-González *et al.*, 2011) to develop a new system that provided diagnostic recommendations. However, these techniques still did not model the uncertainty of the clinical observations within the ontology model. The effect of the presence and absence of clinical observations and the impact on the diagnosis was not modelled in the ontology model and decision support system. These CDSS applications did not use linked data technologies to access patient data from other data sets.

Rodríguez-González et al. (Rodríguez-González et al. 2012) describes a multi-agent architecture is proposed to parallelize a reasoning system based on semantic web technologies. They found using multi-agents greatly reduced the inference times when compared to an approach that did not use multi agents and relied only on DL based inference for generating diagnostic recommendations.

Mohammed et al (Mohammed, Benlamri and Fong, 2012) have used an alignment algorithm to align two existing ontologies (Human Disease Ontology and Symptom Ontology). With the help of an Ontology Matching and Ontology Linking algorithm the 'Diseases-Symptom' ontology has been created. Oberkampf et al. (Oberkampf, Zillner and Bauer, 2012b) have proposed a similar Disease-Symptom ontology model extending the Human Disease Ontology (DOID)(Schriml *et al.*, 2012), and have developed a system that provide a ranked list of likely diseases and helps plan next clinical examinations.

In the approach for modelling diagnostic criteria described by Oberkampf et al. (Oberkampf, Zillner and Bauer, 2012a) the Disease-Symptom ontology model has 2 main classes; the Disease class and the Symptom class. These classes contain subclasses for specific diseases. The symptoms were further classified into present symptoms, absent symptoms and open symptoms. Absent symptoms are important as they are needed to rule out the presence/absence of a diagnosis. The proposed method of producing a ranked list of diagnosis involved using a ranking algorithm (Oberkampf *et al.*, 2012). The algorithm took into consideration 3 different factors that influenced the likelihood of a disease: Disease specific factors, Disease-symptom relations and Symptom specific factors. The disease specific factors involved considering factors such as the relative incidence of the disease in the population. Disease-symptom relations factored in the importance of the symptoms with respect to the disease. Leading symptoms that are indicative of a disease are given a higher weight. Symptom specific factors included all the factors that contributed to the importance of the symptom in a situation. Symptoms have a basic importance and that varies between different symptoms and is determined by the domain expert. The ranking algorithm takes into consideration all these factors and produces a ranked list of diagnosis. The weights used in this method were not encoded in the ontology model.

Researchers have used several approaches to integrate OWL ontologies and SWRL and other semantic web rule languages like Jena rules. Some have proposed using existing ontologies, but this would require creating instances for the classes within the ontology as SWRL reasons on instance data. Once the instances have been created, SWRL can be used to encode the diagnostic criteria (Bertaud Gounot *et al.*, 2011a). The combination of OWL, SWRL and other rule languages is a recognised expert system framework and has been widely used in semantic web based decision support applications (Ongenae *et al.*, 2010; Donfack Guefack *et al.*, 2012; Bau, Chen and Huang, 2014; Sherimon and Krishnan, 2016). This is largely due to the availability of ontology reasoners that are compatible with both OWL and SWRL (Sirin *et al.*, 2007a). SWRL can overcome some of the weaknesses of OWL as it can handle numerical functions better than OWL. OWL cannot express all relations and therefore by using SWRL the expressivity of OWL can be extended at the slight expense of decidability (Prcela, Gamberger and Jovic, 2008).

In real-world clinical scenarios domain knowledge contains uncertain knowledge and the difficulty in modelling real world clinical situations in OWL is compounded by the presence of incomplete/ imprecise information. To deal with partial, imprecise and vague information researchers have proposed several different formalisms including Fuzzy logic and Bayesian probability. Fuzzy ontologies are one proposed solution to the problem of vague information. Lee & Wang (Lee and Wang, 2011) proposed a five layer fuzzy ontology model to develop an ontology in the diabetes domain for diagnosis and recommend a personalised dietary plan (Chang-Shing Lee, Mei-Hui Wang and Hagras, 2010). SWRL has been used to develop a fuzzy rule base to develop a fuzzy diagnostic model (Liang, Zhai and Li, 2013). Similarly techniques for decision support using Bayesian networks have been described (Farooq *et al.*, 2012) (Bucci, Sandrucci and Vicario, 2011).

Even though there are several reasoners that are available and can perform reasoning on OWL/SWRL knowledge bases the most popular ones do not support SWRL by default. SWRL is currently not a standard and there are no standard rules languages for the semantic web. Not all reasoners have in-built support for reasoning with SWRL. For example, Jena reasoner supports Jena Rules syntax by default. Jena is a general-purpose rule-based reasoner and supports rule-based inference over RDF datasets (jena.apache.org, 2016).The Jena reasoner provides support forward chaining inference, backward chaining inference, and a hybrid execution model. In order to use SWRL rules in a Jena based triple store or semantic web application another reasoner (for example Pellet reasoner (Sirin *et al.*, 2007b)) will have to be integrated with the Jena application.

In order to overcome the problem of interoperability due to the proliferation of rules languages another standard called the Rule Interchange Format (RIF) was introduced (Kifer and Boley, 2013). The RIF standard acts as an interchange format between the different rule engines. So, the rules can be expressed in any rule language. RIF will allow different rule systems to exchange certain type of rules with other rule systems. So, knowledge base encoded in OWL and a one rule language can be integrated with another knowledge base in OWL and different rule language and it will be possible to run inference over both

datasets.

Increasingly patient records are being stored in EHR systems which make it easier to store, retrieve and analyse data. However, to extract data from data warehouses that store this data, analyse the data and use the data to create decision support tools healthcare organizations need to invest a lot of resources in manpower which includes clinicians, analysts, data miners and software developers. Haug et al. (Haug *et al.*, 2013) has demonstrated a system using real life patient data extracted from data warehouses that uses ontologies alongside data analysis applications that analyses the data to create diagnostic models. The system called the 'ontology-driven diagnostic modelling system' helps reduce the resource burden needed to develop diagnostic CDSS applications.

Corrigan (Corrigan, 2015) implemented a clinical evidence service to support diagnosis. Once a clinician has recorded signs and symptoms of a patient in a primary care EHR, decision support application accesses clinical evidence modelled as OWL ontologies and provides a ranked diagnostic recommendation.

The flexibility of ontologies allows the development of hybrid application using different methodologies. In addition to representing diagnostic criteria in the form of OWL axioms and rule the knowledge representation capability of ontologies has been used to enhance the performance of Case Based Reasoning systems (Bichindaritz, 2006; Wang and Tansel, 2013). The ontology can be used to annotate data present in clinical databases, which are excellent sources of case-based knowledge for decision support. This allows the system to solve problems from relevant experience from a knowledge base of similar cases.

Obtaining a medical history of the patient is a key part of the process of diagnostic reasoning. Improving the quality of information gathered can assist the clinician in making a better judgement regarding the condition of the patient. Ontology based information collection has been demonstrated in (Bouamrane, Rector and Hurrell, 2008) which has the potential to reduce the number of questions asked during the information gathering process by using context-sensitive adaption of the questionnaire. This is different from static questionnaires currently used in clinical practices. Once a patient's medical history has been extracted in the OWL format, the OWL reasoner can be used to provide recommendations based on clinical guidelines.

Once the diagnosis has been made, the next step in the care process includes the treatment recommendations and ontologies have been used in generating treatment recommendations as well (Douali, De Roo and Jaulent, 2012).

2.5.2 The disease-symptom pattern for modelling ontologies in diagnostic CDSSs

Ontologies, especially OWL ontologies help to represent concepts like Diagnosis, Symptoms, Signs, Clinical Test results, Radiographic examination results and the relations between these concepts etc. in a machine-readable manner that can be shared between different CDSS applications. The section above (See section - 2.5.1) details the different techniques through which OWL ontologies have been used in diagnostic CDSSs.

The literature shows that there is a common pattern for representing the relations between the concepts that represent the diagnosis and the concepts that represent the different clinical observations that either support or negate the diagnosis. Diagnosis is a label given to a disease once a clinician has confirmed the presence of the disease taking into consideration the different manifestations that the disease can take in the human body. Presence or absence of a disease is confirmed by the presence/absence of certain signs, symptom or clinical observations that are recorded by a clinician in the presence of a patient. The most important group of observations are the symptoms as they are normally reported by the patient or the patient's caregivers. The clinician then uses these symptoms to formulate certain diagnostic hypotheses. The diagnostic hypotheses will then be used by the clinician to illicit further signs and symptoms and order diagnostic tests as needed.

The concepts that are used to represent the Diagnoses in the ontology models used in diagnostic CDSSs can be collectively called the Disease concept. The different observations that comprise the signs, symptoms and clinical observations can therefore be labelled the Symptoms concept. This pattern of Disease(s) and Symptom(s) can be found in several ontology-based diagnostic CDSS models in literature. Reyes-Ortiz et al (Reyes-Ortiz *et al.*, 2013) uses 3 sub-ontologies to develop a Diagnostic ontology and the sub-ontologies include a Diseases, Symptoms and Risk factors ontologies. The complex ontology model developed by Romero-Tris et al (Romero-Tris, Riaño and Real, 2010; Riaño *et al.*, 2012) for chronically patients provides a formal representation of the all concepts needed for care of chronically ill patients at home (i.e., syndromes, diseases, social issues, signs and symptoms, problem assessments, and interventions). However, in this model signs and symptoms and all other clinical observations needed to make a diagnosis was not composed into a single concept as the aim of their model was to develop a personalized care plan based on the diagnosis.

Rodriguez et al. (Rodriguez *et al.*, 2009) and García-Crespo et al. (García-Crespo *et al.*, 2010a) developed ontology based models to that represent the Disease, Symptoms, Laboratory tests, Medicine etc. Rodríguez-González et al. (Rodríguez-González *et al.*, 2011) later developed another ontology model called the Diagnosis ontology that had a Clinical Findings ontology. The Clinical Findings ontology had several sub-ontologies such as Signs and Diagnostic Tests.

The reason for modelling the Diseases and Symptoms either within a single ontology model or as sub-ontologies is to take advantage of the presence of existing ontologies that models theses relations. For example, Rodríguez-González et al. (Rodríguez-González *et al.*, 2011) used the SNOMED-CT ontology to annotate terms within the Diagnosis ontology and its sub-ontologies.

Oberkampf et al. (Oberkampf, Zillner and Bauer, 2012a) the Disease-Symptom ontology model has 2 main classes; the Disease class and the Symptom class. These classes contain subclasses for specific diseases. The symptoms were further classified into present symptoms, absent symptoms and open symptoms.

The method used by the CDSSs using these ontology-based diagnostic models to generate diagnostic recommendations differs in different implementations. Oberkampf et al. (Oberkampf, Zillner and Bauer, 2012a) used a ranking algorithm to generate the recommendations. Others (Bertaud Gounot *et al.*, 2011a) have modelled the Disease and Symptoms in OWL and encoded in diagnostic criteria in SWRL or other semantic web rule languages such as Jena Rules (Rodriguez *et al.*, 2009). Separate tools for probabilistic reasoning have been implemented that takes into account probability of a disease being present (García-Crespo *et al.*, 2010a).

2.5.3 Linked data for clinical decision support

CDSS development is expensive and therefore to share the functionality of CDSS applications Service Oriented Architectures (SOA) have been proposed. The functionality of the CDSS application can be wrapped in a SOA application and provided as a web service which can then be accessed by different client applications. However, problem with client-service semantic interoperability remain especially around understanding the semantics of the CDSS service when shared amongst different healthcare organizations. Marco-Ruiz et al (Marco-Ruiz *et al.*, 2016) have developed ontologies to provide descriptions of CDSS services that are machine interpretable and helps in semantic interoperability and reuse of the models. This Linked Service helps expose the CDSS service and the CDSS knowledge based as Linked data.

Hoekstra et al. (Hoekstra *et al.*, 2015) translated legacy datasets of the Adverse Event Reporting System (AERS) dataset of the Federal Drug Administration (FDA) into RDF. The legacy data was in the form of dollar separated tables. After checking integrity of the data, it was imported into a relational database. The data was then converted into RDF using the D2RQ method (Christian Bizer and Seaborne, 2004). The RDF data was then imported into triple store. Using an Annotation Ontology, they annotated relevant parts of clinical guidelines and other medical literature. The user interface allows users to find relevant information about a patient using several SPARQL endpoints. When the user selects a patient, detailed

information enriched with information from the Linked Life Data SPARQL endpoint are provided. Clinical guideline annotations that match the patient's descriptions are retrieved and displayed to the user. Similar cases that are present in the AERS dataset are displayed to the user. The AERS dataset is available through a SPARQL endpoint.

2.6 Rationale for research and ontology driven CDSS for early diagnosis

This chapter gave an outline of the literature around diagnostic error and the cognitive causes of diagnostic error. Literature showcased the role played by diagnostic CDSSs in reducing diagnostic error. Faulty data gathering, faulty information processing and faulty information synthesis have been identified as some of the reasons for diagnostic error (See section - 2.1.1). Clinical management guidelines have been developed to assist a clinician in gathering information properly. Using checklists, providing alternative diagnosis and providing focussed decision support for a condition has been proposed for reducing diagnostic errors using CDSS tools.

Guideline recommendation can help improve information gathering by providing alerts, assisting in interpreting test and investigations and providing guidance on the next steps to be taken before diagnosis. Improved information gathering can potentially help in improving diagnosis by improving the quality of data collected. The guideline recommendations should be presented alongside diagnostic recommendations and both recommendations can work in tandem to help reduce diagnostic error. Literature suggests a need for early diagnostic support to reduce diagnostic error. Ranked diagnostic recommendations provide alternative diagnoses to a clinician. Clinician can compare his/her mental model with diagnostic recommendations. CDSSs will have to provide reasons for making the recommendations. The clinician can then view the reasons alongside the guideline recommendations and decide next steps to be taken in a diagnostic encounter. As more information is collected the clinician can use the information to confirm/reject diagnostic hypotheses and narrow the list of hypotheses.

The CDSS should be able to explain the patient's condition in a manner that suits the clinician's own line of reasoning. The CDSS should likewise provide alternative diagnoses and explain why it believed the patient may have a diagnosis. The model that captures the expert's domain knowledge must be able to capture the information in an easy and organised manner. It should similarly be able to perform when there is incomplete data as early diagnosis is characterised by the presence of incomplete data. The literature suggests that argumentation-based methods are a suitable solution for providing diagnostic recommendations in early diagnosis. Argumentation-based methods do not have the problem associated with other CDSS reasoning strategies such as difficult capturing the domain expertise. Argumentation using non-monotonic logic where existing conclusions has can be discarded in the light of current information and new conclusion can be formed. Early diagnosis is characterised by incomplete information. The clinician has not collected all the information needed to make a diagnosis. The clinician uses this incomplete information to form initial diagnostic hypotheses. Using this information, a clinician then obtains additional information about a patient and then uses the information to confirm or reject a hypothesis. The alternative diagnoses provided at different stages of the clinical encounter helps trigger appropriate illness scripts that can then be used by the clinician to find more information about the patient. The alternative diagnoses help in the process of generating the working hypotheses and help in gathering additional information about the patient to help confirm or refute working hypotheses. The alternate diagnostic recommendations presented in various stages of a clinical encounter alongside detailed reasons for recommendations could potentially help reduce premature closure bias by informing the clinician about alternate diagnoses. This may stop the clinician from halting the diagnostic information gathering process and consider other alternatives.

Argumentation-based methods have been used for creating guideline-based CDSS applications using the PROforma language. Though argumentation-based CDSSs have been developed for improving medical decision making in areas such as diagnosis, management, treatment planning, risk assessment etc., studies so far have not evaluated the impact of an argumentation-based diagnostic CDSSs on the diagnostic performance of clinicians. Therefore, the initial stage of the thesis evaluates the impact of guideline

recommendations on information gathering of clinicians and the feasibility and impact of argumentation-based diagnostic CDSSs on diagnostic performance of clinicians. The impact of guideline recommendations on management decision making will be studied.

Existing argumentation-based tools such as Tallis uses a proprietary format to represent the PROforma model. The model can only be modelled and executed using proprietary software such as Tallis. Integration with EHRs for collecting patient data to generate diagnostic recommendations will need expensive interface engines that can translate patient data from different formats to a format that can be understood by the Tallis inference engine. Existing argumentation-based tools and other diagnostic CDSSs tools face a semantic interoperability problem in integrating information from diverse sources.

The main problem associated with providing early diagnostic decision support is the availability of data to generate meaningful diagnostic recommendations. The clinicians have not collected all the information needed to decide. Therefore, CDSSs must access information from other sources to complement the decision making. The web is a major source of data including patient data that can be used for generating diagnostic decision recommendations. However, the variety of data representation make integrating this information difficult. One of the main challenges facing diagnostic CDSS applications even today is lack of semantic interoperability (See section - 2.4.2). Semantic web and linked data technologies have been developed to improve semantic interoperability. RDF is a strong contender for becoming the universal information representation format for healthcare interoperability. Improved semantic interoperability will help intelligent agents find the information needed for diagnostic decision support and integrate the information easily.

Even though some of the ontology driven CDSS applications do try to solve the problem of semantic interoperability by modelling the diagnostic inference models in OWL/RDF based languages there remains some deficiencies. The degree of uncertainty of diagnosis is captured outside the OWL ontology model in many ontology models outlined in this literature review. Existing ontology-based models for diagnostic CDSSs have used several techniques to deal with the uncertainty in diagnosis. Some methods have used an external probabilistic system to generate probabilities and generate the ranked diagnostic recommendations (García-Crespo *et al.*, 2010a). Other methods include representing the diagnostic criteria using OWL axioms (Ongenae *et al.*, 2010; Donfack Guefack *et al.*, 2012; Bau, Chen and Huang, 2014; Sherimon and Krishnan, 2016). However due to limits in expressiveness of DL-based axioms these methods are not well suited for representing uncertainty and cannot work with incomplete data. Most ontology-based CDSSs identified in literature do not provide a ranked diagnostic recommendation. A ranked diagnostic recommendation is important in early diagnostic CDSSs as the clinician can use the ranking to guide the next steps for gaining more information. The ontology-based CDSSs that provide a ranked recommendation rely on complex ranking algorithms and specialized software to generate the ranking (Oberkampf *et al.*, 2012). The diagnostic CDSSs do not explain reasons for making a recommendation by making use of argumentation or similar methods.

The models and the CDSS applications do not demonstrate how data in linked data repositories can be accessed to support clinical decision making (See section - 2.5.1). The review of linked data resources identified a lack of publicly available of clinical data in linked data format that can be used in CDSS applications. Translating patient data from one form to another and exposing data via SPARQL endpoints to be consumed by the ontology driven CDSS will be further explored in this thesis.

Current ontology driven diagnostic CDSS use different types of rules languages to model diagnostic criteria and generate diagnostic recommendations. Using SPARQL rules to generate diagnostic recommendations have not yet been demonstrated as a technique (See section - 2.4.6 and section - 2.5.1). Current generation of ontology driven CDSS application have other limitations as well. Use of linked data in CDSSs is still in the infancy stage though there has been numerous ontology driven CDSS demonstrations (See section - 2.5.3). However, as more and more datasets become available via linked data on the web, CDSS applications should be able to use this data via SPARQL endpoints to provide diagnostic recommendations to the clinicians. The use of SPARQL endpoints to access patient data for the CDSS will be explored in this thesis.

This thesis proposes the development of an argumentation-based ontology-driven CDSS that can provide early diagnostic decision support. The model called Disease-Symptom Model (DS Model) combines the strengths of argumentation with the strengths of semantic web technology. The proposed model provides a novel method that allows the domain expert to capture arguments favoring and negating a diagnosis in OWL/RDF model. The model overcomes problems of integrating information and semantic interoperability faced by current generation argumentation-based CDSSs. The model uses argumentation and currently available semantic web and linked data technologies to provide a ranked diagnostic recommendation that can be generated using any off-the-shelf tools without need for specialized software and algorithms. CDSS prototypes that can use the ontology model to generate diagnostic recommendations will be developed. CDSS prototypes will demonstrate how the DS Model can help integrate information from a wide variety of sources to provide a diagnostic recommendation. The DS Model can integrate this information to generate a diagnostic recommendation in the early stages of the clinical encounter. Current generation diagnostic CDSS tools are limited to information gathered during the clinical encounter. Performance of the DS Model will be evaluated across all stages of the clinical encounter.

The first part (Chapter 4) of the thesis describes the development and evaluation of an argumentation-based diagnostic CDSS using off-the-shelf technology. In the study guideline recommendations are given to study the effect of these guideline recommendations on the information gathering of community optometrists. Impact of guideline recommendations on management decision making will be evaluated. Argumentation-based diagnostic recommendations are provided. This pilot study will demonstrate the feasibility of using argumentation-based methods for diagnostic decision making and study the effect of argumentation-based diagnostic recommendations on optometry diagnostic decision making.

The next part (Chapter 5) involves development of the ontology using argumentation-based methods. The ontology model uses argumentation to model arguments supporting and negating a diagnosis. It will be built using semantic web and linked data technologies. The CDSS using this model can integrate information using semantic web and linked data technologies.

In the next part (Chapter 6) the CDSS and diagnostic recommendations generated by the ontology model in early diagnosis are evaluated by comparing its performance with community dentists.

The next chapter outlines the research questions and research objectives.

3 Chapter 3 - Research outline

3.1 Introduction

The first part of the thesis explores how clinical guideline recommendations and argumentation-based diagnostic recommendations can help clinicians in information gathering and diagnostic decision making.

The second part of the thesis explores the development of an argumentation-based ontology model for representing diagnostic criteria and CDSSs that can use this ontology model and exploit linked data and semantic web technology for generating diagnostic recommendations.

The third part of the thesis investigates the performance of the ontology based CDSS in generating diagnostic recommendations across all stages of the diagnostic encounter.

3.2 Research Questions

1. How do guideline recommendations and argumentation-based diagnostic recommendations affect information gathering and diagnostic decision making in primary care clinicians?
2. Can an argumentation-based diagnostic CDSS developed using semantic web and linked data technologies generate early diagnostic recommendations?
3. How does an ontology driven diagnostic CDSS perform across all stages (early and late) of a clinical encounter?

3.3 Outline of research

The following section outlines how objectives of this thesis was achieved and how research questions were answered in this thesis. Research methodology for each study has been described in the subsequent chapters in more detail.

To recap, specific objectives of this research are:

Objective 1: To review the literature on models of diagnostic reasoning and identify models used by diagnostic CDSS to represent knowledge and formulate diagnostic recommendations and their role in reducing diagnostic error. Review state of the art web technologies to identify how they address the problem of semantic interoperability and their application in CDSS.

Objective 2: Develop a diagnostic CDSS based on current argumentation-based CDSS technology and evaluate its effect on primary care clinician decision making and information gathering.

Objective 3: Develop a semantically interoperable diagnostic inference model that can integrate information from disparate sources and uses argumentation to generate diagnostic recommendations.

Objective 4: Implement a working CDSS prototype that implements the semantically interoperable diagnostic inference model and provides diagnostic recommendations.

Objective 5: Evaluate the CDSS prototype and the diagnostic inference model and its ability to support diagnostic decision making at all stages of the clinical encounter.

3.3.1 Objective 1

Objective 1 (Figure 3-1) has been addressed by literature search and the results of literature review have been presented in **Chapter 2**. The literature identified the common causes of diagnostic error and what role CDSSs can play in mitigating diagnostic error. The literature review provided an overview of diagnostic CDSSs and how the web can play a role in CDSSs. It identified the advantages of an argumentation-based approach when compared to other CDSS reasoning methods. Semantic web and linked data technologies such as RDF, OWL, SPARQL etc. and their applications in CDSSs was outlined.

Ontology driven diagnostic CDSS applications and some of shortcomings of the approaches takes by these systems have been identified.

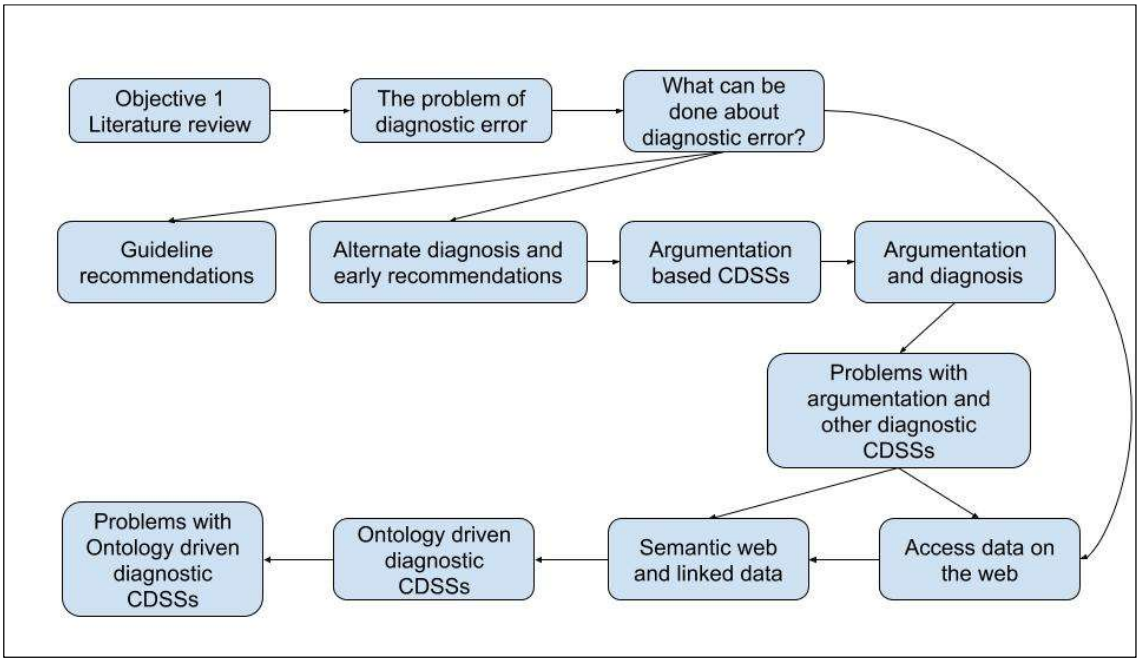


Figure 3-1 - Objective 1

3.3.2 Objective 2

From literature, it is known that quality of data that has been gathered during the patient encounter impacts CDSS recommendations (section - 2.1.1). The literature suggested that faulty information gathering, processing and synthesis combined with faulty knowledge and skills are responsible for diagnostic errors due to cognitive causes. Clinical guidelines and guideline-based CDSSs have been introduced to help in the process of collecting, analysing and verifying clinical data based on the latest clinical evidence. Guideline recommendations have been designed to help the clinician in information gathering and synthesis. Existing argumentation-based CDSS tools was identified. Argumentation provides certain advantages over other CDSS reasoning methods especially in capturing domain expertise and presenting the diagnostic recommendations to the clinician in an easy to understand manner. Argumentation works well with limited information and can support defeasible reasoning using non-monotonic logic. However, studies so far have not evaluated the role of argumentation-based CDSS in reducing diagnostic error especially in primary care.

Objective 2 (Figure 3-2) has been accomplished by developing and testing the effect of guidelines and guideline-based decision support systems on information gathering of primary care clinicians. The effect of an argumentation-based diagnostic CDSS on the diagnostic decision making of clinicians was studied. In addition, the effect of guideline recommendations on management decision making was studied.

Chapter 4 describes research methodology used and development of argumentation-based diagnostic CDSS prototype. An off-the-shelf CDSS guideline modelling and execution tool was used to create a prototype CDSS and used to study the impact of guideline recommendations and argumentation-based diagnostic recommendations. The language used to formally represent the guidelines called PROforma uses argumentation for decision making. Optometry was chosen as the test domain and the CDSS was tested using clinical vignettes (patient scenarios). Data obtained from optometrists using clinical vignettes and CDSS implemented in PROforma was compared to an earlier study where another group of optometrists was asked to diagnose and manage the same vignettes without help of any CDSS. Results of the study have been presented in the same chapter.

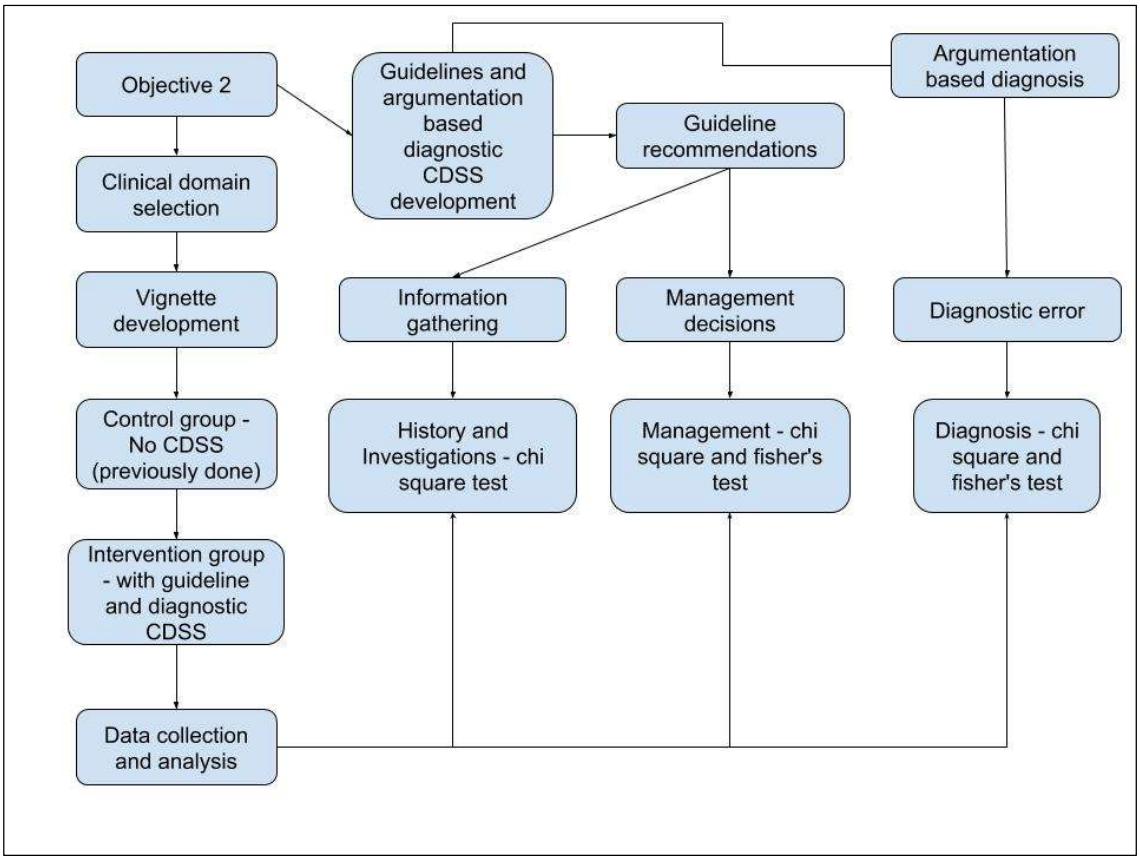


Figure 3-2 - Objective 2

3.3.3 Objectives 3 and 4

Literature identified several methods for representing diagnostic criteria for developing models to support diagnostic CDSSs. Argumentation-based CDSSs have been developed and evaluated in several domains. However current generation argumentation tools suffer from interoperability problems. These tools rely on proprietary formats and specialized tools for modelling and execution that prevents widespread use and adoption.

OWL based ontologies that can be used to represent diagnostic criteria have been identified and some of the shortcomings of CDSS applications that deploy these models have been described (See section - 2.5.1). The uncertainty involved in diagnostic decision making have been modelled outside the ontology model. Separate tools and software are needed to generate the diagnostic recommendations. Existing ontology driven CDSS applications have not demonstrated using linked data to integrate information.

Objective 2 saw the development of an argumentation-based diagnostic CDSS and demonstrated the feasibility of using argumentation as a method to generate diagnostic recommendations. In **Chapter 5**, the development of an argumentation-based OWL/RDF ontology driven model that models arguments that support or negate a diagnosis has been described.

Dentistry was chosen as the test domain to test and demonstrate the model. The model was called Disease-Symptom Model (DS Model) and **Chapter 5** details development of the model using dental examples. This chapter describes how SPARQL rules were used to map patient data that were represented in different formats to the DS Model and obtain recommendations from these diverse data representations. The literature review indicated some of the advantages of SPARQL rules when compared to other rules languages on the web (See section - 2.4.7.1). The use of SPARQL rules to generate a ranked diagnostic recommendation and generation of weights have been described in detail in Chapter 5.

In addition to DS Model (Figure 3-3) the chapter describes the development of an extension to DS Model called Multi-level DS Model. There are clinical situations where a CDSS may have to make sub-diagnostic recommendations. These sub-diagnostic states called a Finding, can be used to cluster together observations or represent concepts that may have clinical significance in certain scenarios. The Multi-level DS Model is an extension of the basic DS Model and can be used with any of the CDSS architectures described in this chapter. The chapter (**Chapter 5**) describes how the observations are linked to the findings and subsequently the findings to the diagnosis. A ranked diagnostic recommendation is generated along with weights for each diagnosis in a manner like DS Model recommendations.

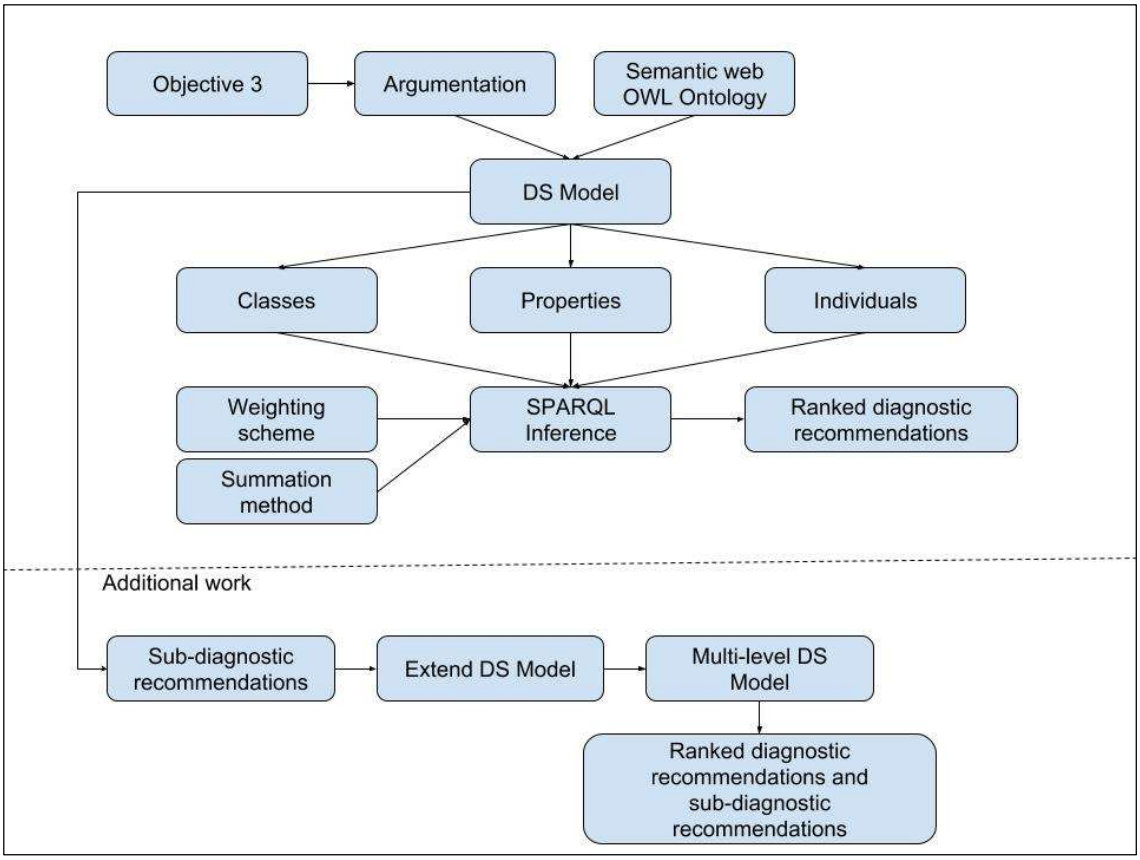


Figure 3-3 - Objective 3

Chapter 5 describes the development of several prototype CDSS applications that demonstrates the different architectures through which the DS Model can be deployed (Figure 3-4). The strengths and weaknesses of these architectures are likewise described. The prototypes show how data can be represented in different formats and how SPARQL endpoints can be used as a source of patient data to support early diagnosis. SPARQL endpoints and federated querying of data allows the data to be stored in disparate locations and allows the data to be queried and merged easily (See sections - 2.4.7.2 & 2.4.7.3). The literature suggested that providing early diagnosis could potentially help reduce diagnostic error (Kostopoulou *et al.*, 2015b). However, the data needed to make a diagnostic recommendation is stored in disparate data sources on the web. Semantic web and linked data technologies can be used to integrate this information and use them to support diagnostic CDSSs. The prototypes demonstrate how semantic web and linked data technologies can be used with the DS Model to generate diagnostic recommendations. Evaluation of DS Model was completed in Chapter 6.

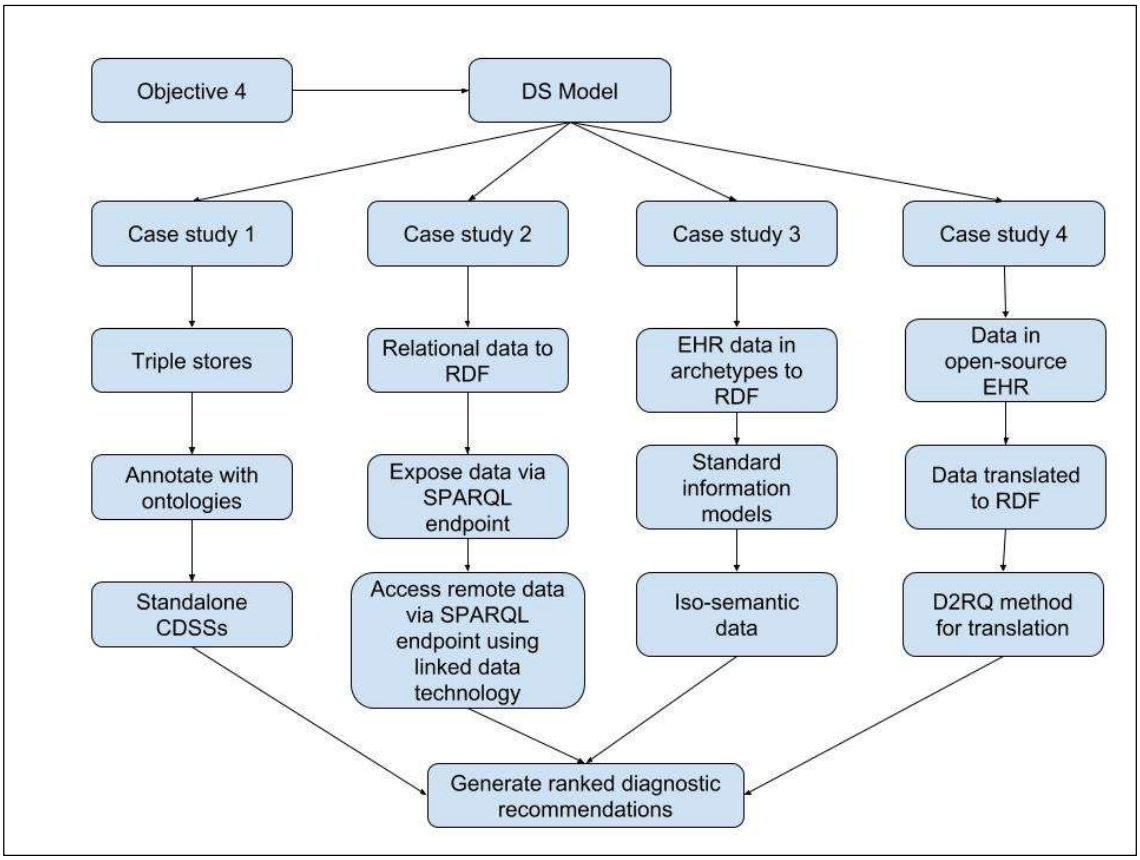


Figure 3-4 - Objective 4

3.3.4 Objective 5

The literature review (Chapter 2) outlined cognitive causes of diagnostic error and some of the cognitive biases that contribute to diagnostic error (See section - 2.1.1). Studies show that cognitive biases like premature closure bias cause the clinician to form a diagnostic hypothesis early in the clinical encounter and prevents him/her from gathering more information that can potentially negate the hypothesis (Graber, Franklin and Gordon, 2005b). There is some evidence suggesting that providing diagnostic recommendation early in the diagnostic encounter can potentially help reduce diagnostic error (Kostopoulou *et al.*, 2015b).

The DS Model developed in Objective 3 and the CDSS prototypes developed in Objective 4 showed how a ranked diagnostic recommendation can be generated. The DS Model uses a simple weighting scheme to represent weights and a simple summation method to generate diagnostic rankings. The effectiveness of this weighting scheme and summation method needs to be evaluated. In Objective 5 the ranking generated by the CDSS was evaluated. The DS Model can generate diagnostic rankings by integrating information from disparate sources using argumentation-based methods. Before linked patient data can be used to integrate patient data from disparate sources to support diagnosis, there is a need to know in which areas this additional knowledge can be used to improve diagnostic performance of the CDSS (if it needs improvement). Does the CDSS perform as good as clinicians in the early stages and the late stages of the clinical encounter? Once the areas where the DS Model is deficient are identified then the appropriate linked patient data can be discovered and more effectively used by the CDSS.

Chapter 6 describes how a dental information model was used as basis to identify the different stages of the clinical encounter (Figure 3-5). Dentistry was chosen as the test domain primarily due to the background of the author in this domain. An information model is generally used to define the information architecture of dental EHRs and define data that is collected during the clinical encounter. The information model is demarcated into different sections. These sections were used as basis for the identifying different stages of a clinical encounter. The DS Model will be evaluated at all these stages and the diagnostic performance compared to that of primary care clinicians.

A dental version of the DS Model that can provide diagnostic recommendations for dental pain (in the field of Endodontics) was developed for the purposes of evaluation. A CDSS prototype developed in Chapter 5 was used for the evaluation. Several dental clinical vignettes were developed. An expert dentist in the field of Endodontics (Endodontist) provided ranked diagnostic recommendation for these clinical vignettes. The expert's opinion was considered gold standard. The diagnostic performance of dentists was used as benchmark against which performance of the CDSS using the DS Model was compared.

Community (primary care) dentists were recruited as test subjects for the study. The participating dentists were asked to provide their Rank 1 and Rank 2 recommendations. A RDF representation of the vignettes was created and the CDSS using DS Model provided a ranked diagnostic recommendation. The performance of the CDSS was evaluated at each stage. The ability to provide a ranking was measured using Normalized Discounted Cumulative Gain (NDCG) and ability of the CDSS to accurately provide a Rank 1 and Rank 2 recommendation was measured using Precision-Recall metrics. Using these metrics, the performance of the CDSS was compared to dentists. Methodology and results of evaluation are described in more detail in **Chapter 6**.

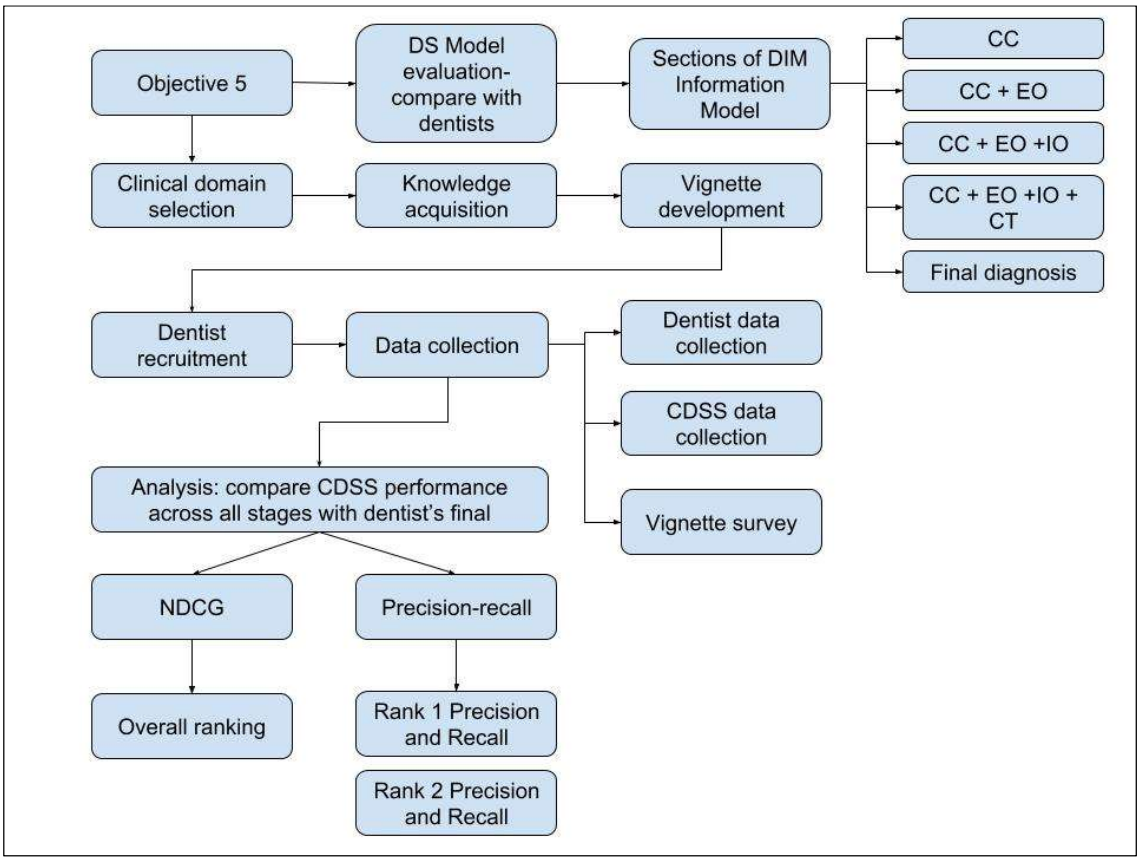


Figure 3-5 - Objective 5

4 Chapter 4 - Study investigating the role of clinical guideline recommendations and argumentation-based diagnostic decision support on clinician performance

4.1 Introduction

The main aim of this study is to implement a decision support that uses Argumentation Theory (AT) as the basis for generating diagnostic recommendations and study its effect on diagnostic performance. The literature review suggests that there is a need for decision support that can integrate information from disparate sources and provide diagnostic recommendations to the clinician at various stages of the clinical encounter. Once data has been integrated results have to be collated and presented to the clinician in a manner that is easy to understand by the clinician. AT evolved from its origins as a sub-discipline of philosophical logic and has emerged as a critical area of Artificial Intelligence (AI) research. AT and argumentation-based techniques have been used to develop decision support applications. Argumentation has been identified from literature as suitable for generating diagnostic recommendations that are easy to understand and easy for knowledge representation. PROforma is one such language that is used to develop guideline-based CDSSs. PROforma uses AT as the basis for making decision recommendations. This study used PROforma language-based tools to develop a prototype CDSS that can provide diagnostic recommendations. This study uses existing tools to study the impact an argumentation-based diagnostic CDSS has on clinical decision making. Before ontology-based model using argumentation-based approaches can be built it was decided to build a prototype based on existing tools and evaluate it with real clinicians using vignettes. This pilot study correspondingly evaluates the feasibility of providing diagnostic recommendation in the form of arguments. The subsequent studies in the remaining chapters will use argumentation-based approaches for building an ontology based CDSS and improve upon some limitations of existing argumentation-based CDSS tools.

The literature showed that guideline recommendations improve guideline adherence. The guideline alerts can indicate the type of diagnoses where the abnormal values can be found. The alerts can help trigger illness scripts which may not have been triggered without the help of a CDSS. The guideline recommendations can then guide the clinician towards the appropriate tests to be taken. The illness scripts triggered by the abnormal values and corresponding guidance can help the clinician form diagnostic hypotheses. The clinician forms a mental model of the next steps to be taken. The clinician can then compare his/her mental model with that of the guideline recommendations. If the mental model matches that of the CDSS recommendation the clinician can confidently continue along the line of questioning or investigations. Guideline recommendations can help the clinician interpret test results and potentially reduce errors in interpretation. All these steps improve the quality of data collected during the clinical encounter. Guideline recommendations can potentially help in improving management decision made after a diagnosis by providing evidence-based guidance of next steps to be taken after a diagnosis.

The study was performed using community optometrists as participants. The aim of this study is to investigate the role of clinical guideline and argumentation-based diagnostic recommendations on primary care diagnosis and information gathering. This study looks at impact of guideline recommendations on information gathering of optometrists in addition to impact of argumentation-based diagnostic recommendations on diagnostic performance of optometrists. The CDSS developed in this study provides guideline recommendations to guide/alert the optometrist about abnormal values, questions to ask the patient in the presence of abnormal values and tests to perform. Guideline recommendations are guidance from clinical guidelines that advices the clinicians to take appropriate steps during the clinical encounter. Guidance was provided on how to perform some of test or interpret the results of the tests. Guideline recommendations were given to help the optometrists in management decision making.

Once the optometrist collects all the data, the CDSS provides diagnostic recommendation using argumentation. The CDSS provides arguments in favour and against each diagnosis. The optometrist is free to choose a diagnosis based on given recommendations. These recommendations could include a ranked list of different diagnoses relevant to a patient's condition. In this study, the diagnostic recommendations were only provided after a clinician (optometrist) has collected all the information about the patient. Once the diagnosis has been provided the optometrist is asked to provide management recommendations. Guideline recommendations are provided to help the clinician in making management decisions. The guideline recommendations were provided at all stages of the clinical encounter.

4.1.1 Clinical Domain Selection

Glaucoma is one of the world's leading causes of irreversible blindness. Glaucoma is a progressive condition of the optic nerve that gradually results in loss of vision. The condition is asymptomatic in the early stages. Early detection is crucial to prevent irreversible loss of vision. In all studies to date in the developed world, only one half of all glaucoma cases in the community are diagnosed and receiving therapy (Bettin and Di Matteo, 2013).

Case detection of glaucoma patients in the community is generally by community optometrists. Diagnosis of glaucoma involves combining many sources of information including clinical and family history, examination findings and diagnostic test results. Once detected the risk of progression is dependent on a variety of risk factors. Therefore, diagnosing glaucoma is crucial for any guiding any further actions that the clinician can take.

Clinical management guidelines have been developed by the National Institute for Health and Clinical Excellence (NICE) and professional bodies including the College of Optometrists, Royal College of Ophthalmologists and European Glaucoma Society. Guidelines aim to improve care quality and decrease practice variation. However studies have shown that there are wide variations in practice and implementation of paper-based guidelines in optometry (Vernon and Ghosh, 2001; Khan, Clarke and Kotecha, 2012) may have unintended consequences. The NICE guidelines for diagnosis and management of chronic open angle glaucoma and Ocular Hypertension (OHT) did not include guidance on the detection and referral of suspected glaucoma patients by community optometrists (Shah and Murdoch, 2011; Ratnarajan *et al.*, 2013). The Association of Optometrists (AOP) however recommended that all patients with Intraocular pressure (IOP) greater than 21 mmHg should be referred to the Hospital Eye Service (HES) based on these guidelines. This resulted in a 40% increase in false positive referrals to the HES for suspected glaucoma (Vernon *et al.*, 2011; Ratnarajan *et al.*, 2013). The guideline recommendations have since been updated and other guidelines have been produced to remedy this situation and endorsed by the AOP (College of Optometrists and Royal College of Ophthalmologists, 2010). None-the-less there is a history of poor compliance with guideline recommendations and more needs to be done to reduce the problem of false-positive (Morley and Murdoch, 2006; Edgar *et al.*, 2010).

There is increasing adoption of EHR systems. Electronic referral systems can bridge communication gaps between primary care and secondary care and enable continuity of care. Guideline based CDSSs formalize the clinical guidelines into a computerised format and the latest guideline recommendations can be delivered to the clinician. They have been shown to increase the adherence of the practitioners to clinical guidelines by providing the clinician with real time, patient-specific guideline recommendations at the point of care (Garg *et al.*, 2005b; Bryan and Boren, 2008; Shalom *et al.*, 2015). The reasons given above motivated the decision to select glaucoma and community optometry as test domain for conducting this study.

4.2 Methods

The creation of a complete CDSS is a large and expensive task. To evaluate clinician performance in the presence of guideline and argumentation-based diagnostic decision support recommendations clinical vignettes were used as substitutes for real life scenarios. Clinical vignettes are a valid tool for measuring clinician performance (Peabody *et al.*, 2004; Shah, Edgar and Evans, 2010). The study is a non-

randomised controlled group study. The author was involved in setting up and running the intervention arm of this study.

The study tries to answer the following questions. Do the guideline recommendations have an impact on the information gathering and management decision making of optometrists? Guideline recommendations were given to the optometrists at various stages of the clinical encounter and during management decision making. In addition, the study tries to answer if the argumentation-based diagnostic recommendations have an impact on the diagnostic decision making of optometrists. The study looks at the overall feasibility of the argumentation-based approach and guideline recommendations with the help of a usability survey of optometrists.

H_0: The null hypothesis is that guideline recommendations have no impact on the optometrist's information gathering and management decision making.

H_0: The null hypothesis is that argumentation-based diagnostic recommendations have no impact on the optometrist's diagnostic decision making.

The following are the main steps followed in executing the study:

- Study setup
  - Vignette development
  - Description of vignette study involving optometrists (Control group study)
  - Decision support study with vignettes involving optometrists (Intervention group study)
    - PROforma modelling. PROforma is a language for formally representing clinical guidelines (See -4.2.3.2).
- Procedure
  - Recruitment
  - Data collection
- Statistical analysis

4.2.1 Vignette development

The vignettes (n=4) were developed by a group of experts including ophthalmologists, optometrists and clinical researchers with an interest in primary care optometry (Murdoch, Lawrenson and Myint, 2015). The images used in the study were obtained from real patients after data anonymization. The vignettes were used in the control group study involving community optometrists in England, U.K. Data from control group study was used to compare the results of intervention group study involving another group of community optometrists in England, U.K using same vignettes but with additional help of a CDSS.

4.2.2 Vignette study involving optometrists (Control group study)

In the control group study 4 vignettes were presented to 102 optometrists who provided a diagnosis and management strategy. The control group study was conducted by another group of researchers (Murdoch, Lawrenson and Myint, 2015).

The vignettes included in this study involved diagnosing and managing Normal, Wet Age-related macular degeneration (Wet AMD), Ocular Hypertension (OHT) and Normal Tension Glaucoma (NTG) scenarios.

Wet AMD is an ophthalmological condition that requires early diagnosis as there is an acute risk of blindness if the condition is not detected early. It is a condition affecting the macula of the eye where abnormal blood vessels form underneath the macula and causes damage to the cells (NHS Choices, 2015). Dry AMD is another type of AMD. Dry AMD is more common, however Wet AMD is responsible for majority of blindness and severe loss of vision (NHS Choices, 2015).

OHT is a condition where the pressure of the inside of the eye (IOP) is higher than normal (Boyd, 2017). In OHT the fluid of the inside of the eye will not drain properly and result in higher pressures inside the eye. When the condition is not detected and treated the increased eye pressure results in damage to the eye and this condition is called Glaucoma (Boyd, 2017).

Primary Open-Angle Glaucoma (POAG) is the most common type of glaucoma. The disease usually does not cause any symptoms until advanced vision loss has set in (Khawaja *et al.*, 2016). People with POAG have raised IOP measurements.

NTG is a type of POAG where the IOP measurements are consistently normal or lower than normal (21 mmHg) (Stein and Challa, 2007). The signs and symptoms are like POAG with some exceptions. Therefore, it is important for the community optometrist to accurately detect these conditions early to prevent onset of blindness.

The 4 vignettes are based on commonly found scenarios that the average optometrist encounters in the community clinic. The optometrists were not notified which clinical conditions the vignettes were based on. The aim of control group study was to study the diagnostic and management decision making capability of optometrists using clinical vignettes.

Each optometrist completed approximately 4 vignettes each. Some optometrists completed all 4 while others did not. Therefore, the OHT vignette was completed by 100 optometrists, Normal by 102 optometrists, NTG by 101 optometrists, and Wet AMD by 100 optometrists.

The optometrists were recruited using a list of optometrists maintained by the General Optical Council (GOC). Using a random number generator, 100 optometrists from London, 50 from Birmingham, 50 from Manchester, 50 from Glasgow and 50 from Edinburgh were invited from the list maintained by the GOC by personalised letter to participate. An incentive of £75 and 1 Continuing Education and Training (CET) point on completion for each vignette was offered. A reminder letter was sent, and due to lack of respondents a further 25 letters for participants in Manchester were sent. N=65 (64%) optometrists in the control group study did not have any post-graduate qualifications. Majority (54%) of the optometrists in the control group study graduated as optometrists in the 1995-2010 time-period (Figure 4-5).

In this control group study, the vignettes were delivered to the optometrists via a web-based application. Vignettes can be delivered in the form of paper vignettes. However, paper vignettes cannot simulate a real-life patient encounter in terms of interaction and therefore a web-based option was chosen.

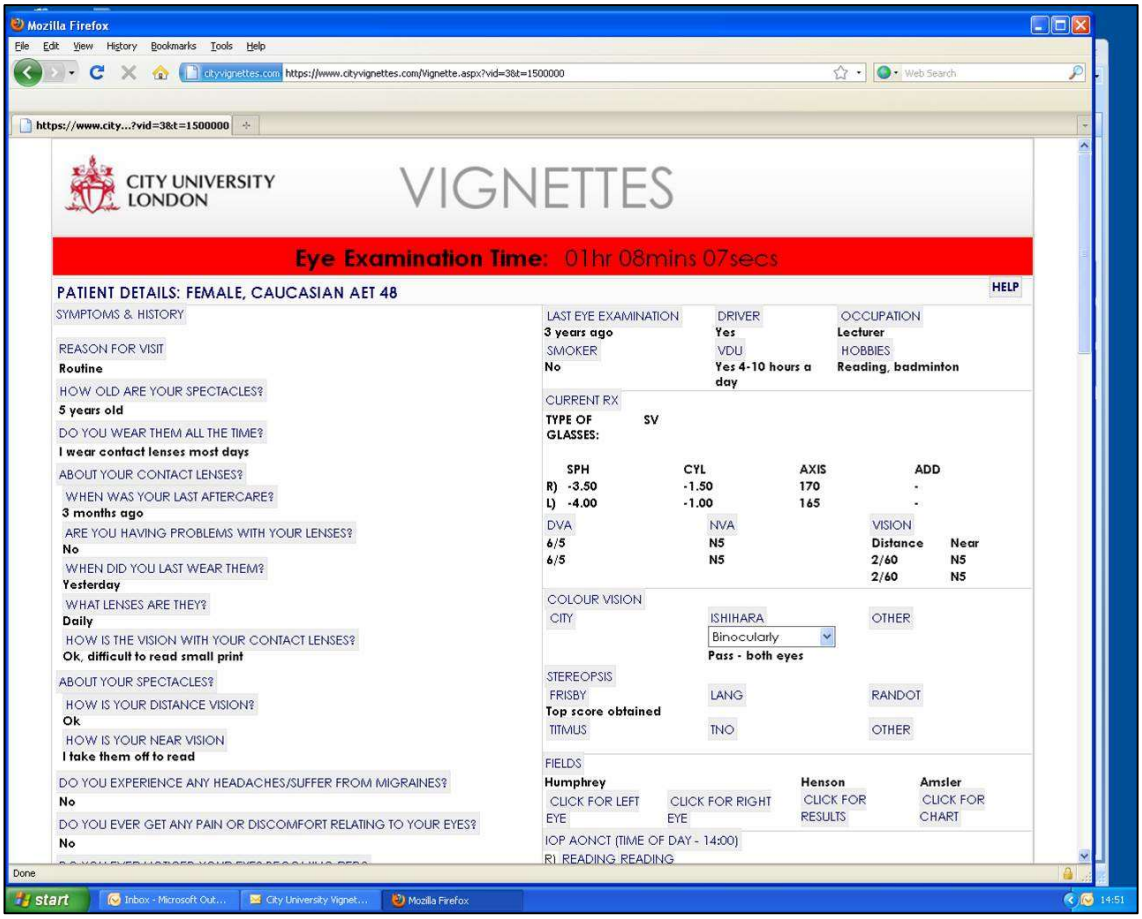


Figure 4-1 - Vignette study involving optometrists (Control group) observations

The optometrists that agreed to participate were sent a link that pointed to a website. The website contained a web-based form that contained the vignette. Each optometrist had to complete 4 vignettes. The images Figure 4-1 & Figure 4-4 shows how the vignettes were delivered via the web form.

There are a series of buttons that are related to questions normally asked by the optometrist during a routine clinical encounter. For example, in Figure 4-1 the question "Reason for visit" has been embedded into the button. When the optometrist clicks on the button, the answer to the question is displayed below the button. This process simulates a real-life encounter where the optometrist asks a series of questions to the patient and the patient responds accordingly. The optometrist is not required to ask all the questions, only those that he/she feels are relevant.

When the optometrist starts the clinical encounter by asking a question the website displays the total time completed for the clinical encounter on the top (Figure 4-1). Whenever the optometrist clicks on a question the timer automatically adds a set amount of time to the total time available for the encounter. For example, in the clinical history section the time added will be 30 seconds for each question. In the investigations section the time added will be greater and can range from 5-10 minutes depending on the type of investigation. The timer simulates the progress of time and gives the optometrist a sense of time that cannot be accurately captured using paper-based clinical vignettes. An Australian study involving optometrists showed that the average time spent on several types of consultation are: 45 minutes for a first visit, 15 minutes for subsequent visits, and 60 minutes for a contact lens consultation (Horton, Kiely and Chakman, 2006). This included time spent on administration, clinical letters etc. The total time was displayed at the end of the clinical encounter. The optometrists are trained to complete clinical examinations and investigations within 60 minutes. If all the questions and investigations are performed the overall time will exceed 60 minutes. However, optometrists in this control group study were not penalised in anyway and allowed to continue full examination and diagnosis as needed. In real-life clinical

encounters some clinical encounters may take more time than average depending on the complexity of the case.

The table below (Table 4-1) shows an extract of the questions and the answers in the symptoms and history section of the OHT vignette. All the vignettes in paper-based format can be seen in Appendix 17.

OHT Vignette question	Answer
Reason for Visit	Routine
Date of last eye examination?	2 years ago
Do you wear spectacles?	Yes, varifocals. Full time wear, 2 years old.
How is your distance vision?	Fine with spectacles
How is your near vision?	Not as good as it used to be, my arms are not long enough
Do you experience any headaches/suffer from migraines?	Once a week, associated with a heavy workload
Do you ever get any pain or discomfort relating to your eyes?	No
Do you ever notice your eyes becoming red?	No
Do you experience double vision?	No
Do you see any floaters in your vision?	No
Do you experience flashing lights?	No

Table 4-1 - OHT Vignette extract

In addition to text-based information, images were provided to the optometrists to take part in grading of these images. For example, the images from the Van Herick test are shown below (Figure 4-2)(Figure 4-3). Van Herick is a test for glaucoma that is based on grading of images of the front (anterior) of the eye. The optometrist must grade the images (0-4) based on the clinical appearance of images.

Figure 4-2 - Van Herick Left Eye image

Figure 4-3 - Van Herick Right Eye image

Similarly, the optometrist is provided with a choice of investigations in the form of buttons. When the optometrist clicks on the button the results of the investigation is provided. Based on the data in the vignette the optometrist must select a diagnosis from a list of diagnoses from a drop-down menu (Figure 4-4). He/she must provide management recommendations such as the location where the vignette patient needs to be referred, when the patient needs to be recalled for the next appointment, supplementary tests to be ordered and any eye prescriptions if needed. All the information needed to make the diagnosis was provided in the same web page and the optometrist had access to the question/investigation that he/she answered or performed.

No decision support recommendations in the form of guideline recommendations or diagnosis recommendations were provided at any stage of the patient encounter. The data was collected and analysed later once the decision support study with vignettes involving optometrists (Intervention group) was completed.

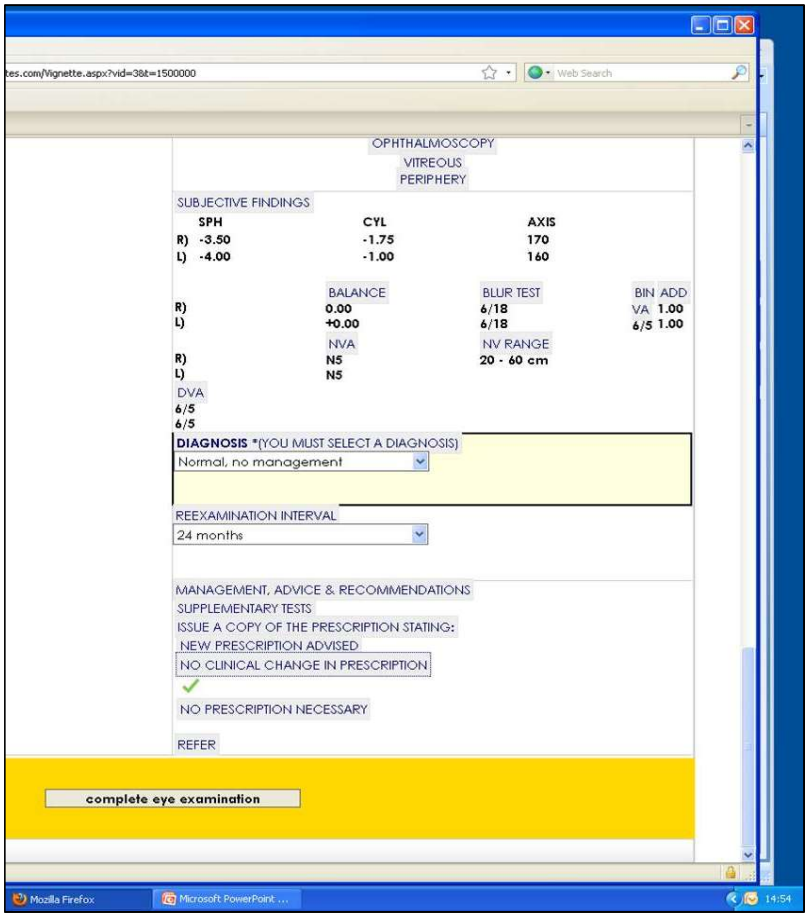


Figure 4-4 - Vignette study involving optometrists (Control group) diagnosis and management

The author was not involved in developing the vignettes, developing the web-based vignette application or data collection in the control group study.

The next section (section - 4.2.3) details the 'intervention group study' where a prototype CDSS application was created for the same 4 vignettes which were then presented to a different group of optometrists with the aim of assessing any change in diagnosis and referral outcomes. The author was involved in developing the CDSS prototypes based on the vignettes and setting up and running a web-based experiment evaluating and comparing the results of the effect of a diagnostic CDSS on the intervention group optometrists with that of the control group optometrists.

4.2.3 Decision support study with vignettes involving optometrists (Intervention group study)

4.2.3.1 Participants

In the intervention group study, the vignettes created for the control group study were presented to a group of optometrists (n=42). The community optometrists in the intervention group was a different set of optometrists from the control group study. The optometrists were recruited via email using the opportunistic sampling method (See Appendix 1- section 1.1). The optometrists belonging to a high-street optometry chain in addition to other optometrists practising in high street optometry clinics were recruited and invitations to participate were sent to 150 optometrists. 42 optometrists agreed to participate. Each optometrist was asked to complete 1 clinical vignette and provide diagnostic and management recommendations for that vignette. Some optometrists agreed to complete more than 1 vignette, therefore the vignettes were completed a total of 49 times by a total of 42 optometrists.

The **OHT** vignette was completed by **13** optometrists, **Normal** by **9** optometrists, **NTG** by **13** optometrists and **Wet AMD** by **14** optometrists.

Out of the 42 optometrists who completed the vignettes (provided diagnosis and recommendations) only 36 provided demographic information and completed the usability survey. The demographic details of the optometrists who completed the demographic information are reported in Table 4-2. Demographic information (Age group, Type of practice and Gender) for the control group optometrists was not available for comparison. The control group had a larger proportion of optometrists who did not have post-graduate qualifications (**64%**). Both groups had similar number of years of experience practising as an optometrist (Figure 4-5). The intervention group results are presented under the heading 'With CDSS' as they had the support of CDSS recommendations. The control group results are presented under the heading 'Without CDSS' as they did not have the support of a CDSS.

Participating optometrists in the intervention group study received 1 CET point for each vignette completed from the GOC for participation in the study.

Demographics	Value	With CDSS N=36 (%)
Age group (years)	<25	1 (3%)
	25-34	10 (28%)
	35-44	7 (19%)
	45-54	12 (33%)
	55-65	6 (17%)
Type of practice	Multiple / group	17 (47%)
	Independent	7 (19%)
	Locum	11 (31%)
	Joint venture	1 (3%)
Gender	Male (%)	12 (33%)
Post graduate qualification	No (%)	17 (47%)

Table 4-2 - Demographics of participants

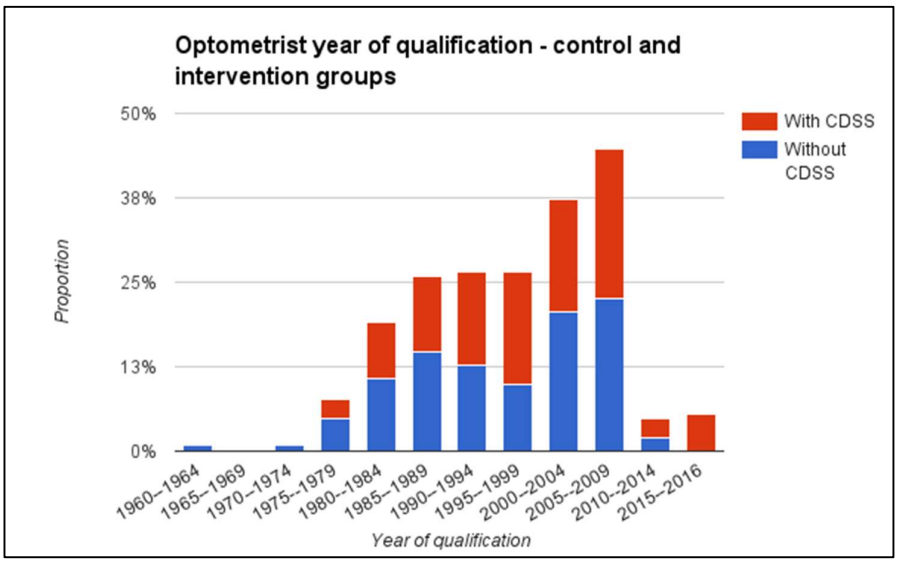


Figure 4-5 - Optometrist year of qualification

4.2.3.2 PROforma modelling

There are several different methods for modelling clinical guidelines and developing guideline based CDSS applications (Peleg, 2013). PROforma is one such language and has been evaluated several times for modelling clinical guidelines and clinical processes and develop guideline based CDSSs (de Clercq *et al.*, 2004; Fox, Patkar and Thomson, 2006a; Sutton, Taylor and Earle, 2006; Patkar *et al.*, 2012). Tallis composer (COSSAC, 2011) is the toolset that is used to model the guidelines and clinical processes into

the guideline based CDSS based on the PROforma language. PROforma was chosen as the modelling language because of the mature toolset (Tallis composer) that allowed for modelling of the guideline recommendations in the PROforma language and easily setup a prototype CDSS. PROforma allows us to map the workflow based on data selected by the optometrists and provide recommendations that are specific to each vignette. PROforma uses argumentation as the basis for generating the decision recommendations (Sutton and Fox, 2003). In this study the main decision was diagnosis decision and the diagnostic recommendations were provided. The CDSS uses the information that was provided to generate arguments in favour and against each diagnosis. Weights are provided that represent the degree of support supporting or negating each diagnosis.

In the PROforma guideline model a computerised guideline to be used in the guideline based CDSS is modelled as a set of tasks and data items (Sutton and Fox, 2003; Sutton, Taylor and Earle, 2006). PROforma captures complex clinical workflows using 4 types of tasks - Actions, Decisions, Plans and Enquiries (data collections tasks). Actions are represented as blue squares, decisions as pink circles and enquiries as green diamonds (Figure 4-6 & Figure 4-9). Plans are represented as red rounded rectangles and can be seen in Figure 4-10.

Actions are used to represent procedures that need to be carried out (for example, administering drugs). Here Actions were used to provide the guideline recommendations. Enquiries are used to model steps in the guideline where some information needs to be collected. This information can come from the patient in the form of clinical history or symptoms. The information can come from other sources of data such as lab test results from a database. In this study, the clinical data in the vignettes from control group study was used to model the enquiry nodes of the PROforma model. For example, in the image below (Figure 4-6- Top section) the different sections of the OHT vignette can be seen. These are Symptoms, Medical History, Previous Ocular History, Family History, General History. Each enquiry has several data sources (Figure 4-6- Bottom right section). The data sources correspond to the data items in the vignettes from the control group study (Table 4-1 and Figure 4-15).

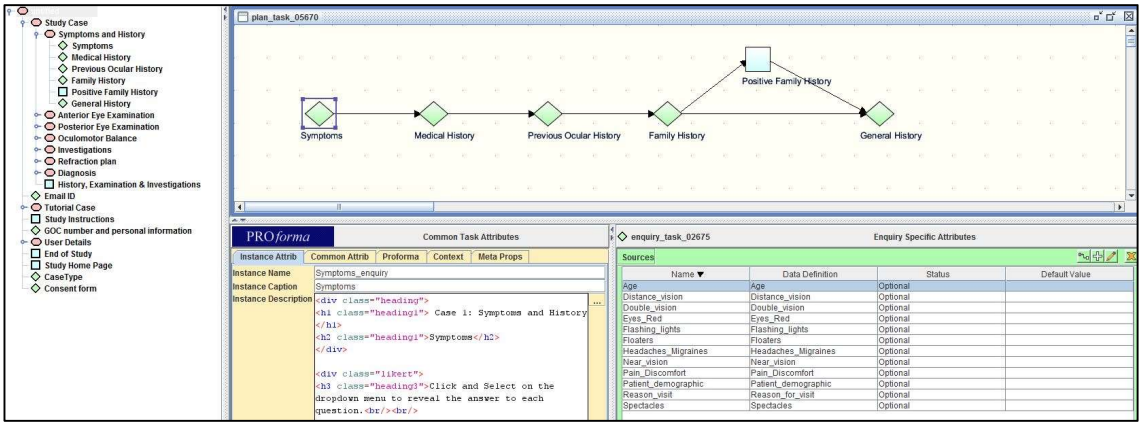


Figure 4-6 - Tallis composer – Symptoms and History as enquiry

When the PROforma computerised guideline is executed, the data sources get translated to questions for the patient to be asked by the optometrist and is enacted via a web page (Figure 4-7). In this study, the answers are hard coded into the fields of the form in the web page, as the purpose of this prototype is to simulate a real-life clinical encounter. However, in real applications the answers will not be present, and the user will be presented with blank fields or options to select based on clinical information.

The next steps taken by the CDSS application were determined by the data entered or the option selected. For example, in the image below (Figure 4-7), there is a question (data source) "Is there a family history of glaucoma?". If the optometrist clicks on the drop-down menu the answer is revealed: "Yes, my aunt and grandmother". Clinical guidance suggesting the optometrist ask this question is given above the questions with the heading "GUIDELINE ALERT", alongside the source of the guidance.

The optometrist can decide to not ask the question. However, if the optometrist selects the option and clicks submit, the clinical guidance from the College of Optometrists (College of Optometrists, 2016) is given (Figure 4-8) in the next screen. The workflow can be seen in Figure 4-6 (Positive Family History – blue square node). This guideline recommendation will only be provided if the optometrist selects the answer, i.e. simulating the process of asking a question and receiving an answer from the patient. If the optometrist does not ask the question, an answer will not be provided.

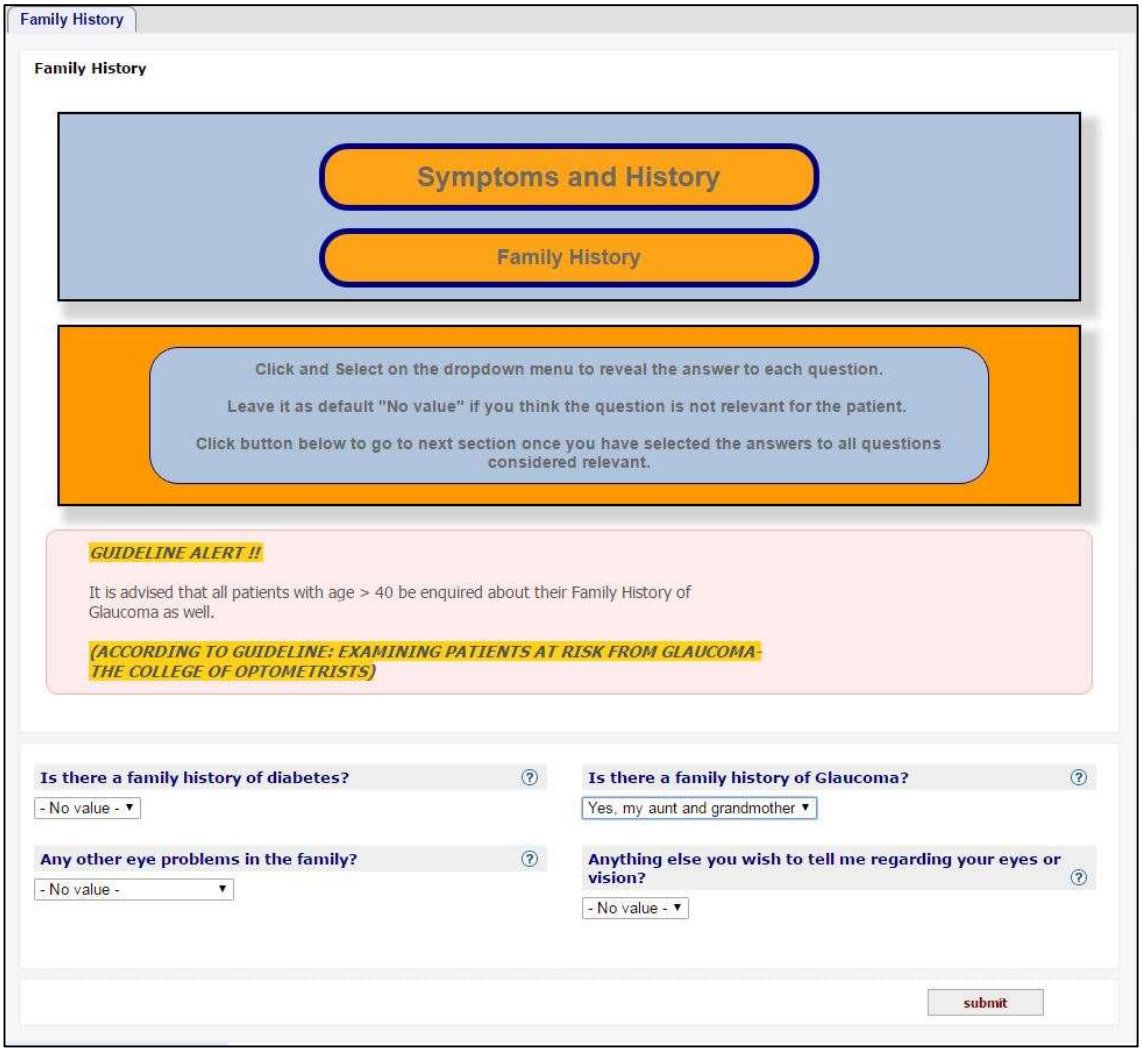


Figure 4-7 - Tallis enactment/execution of guideline via webpage

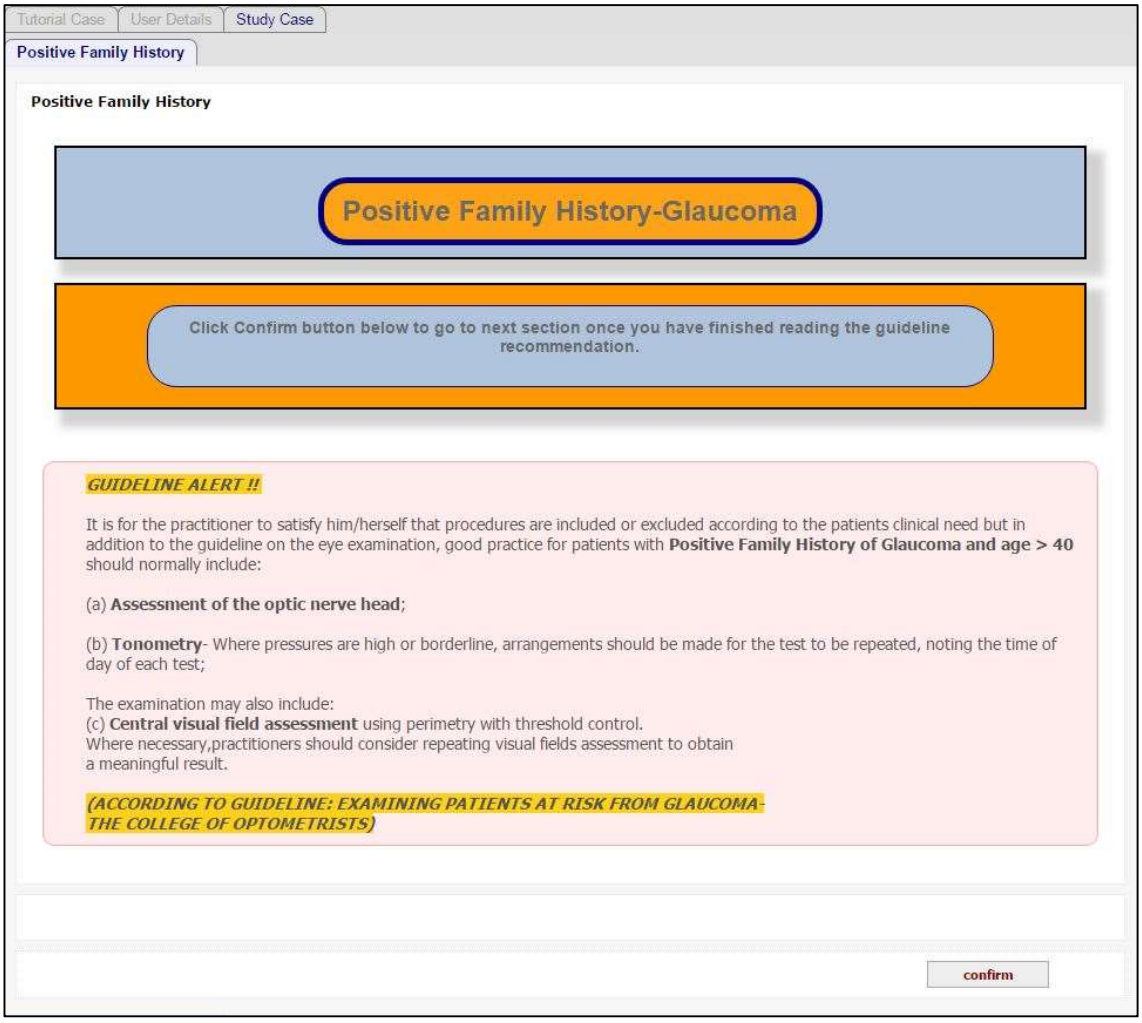


Figure 4-8 - Positive family history recommendation

Decision nodes (pink circles) are steps in the guidelines process where a decision must be made. The outcome of the decision as chosen by the optometrist, will decide how the next steps of the guideline process are executed. The decision can encompass any decision including a diagnostic decision or a risk assessment (Low, Moderate or High risk). Plans (Red rounded squares) are tasks that are used to group together other tasks that share something in common. For example, a plan can be used to execute a set of tasks together (Figure 4-10).

The clinical guidelines used as the basis for providing the guideline recommendations in this prototype CDSS based on the vignettes included the NICE Glaucoma diagnosis and management guidelines (NICE, 2009), European Glaucoma Society guidelines(European Glaucoma Society, 2008), College of Optometrists Clinical Management guidelines (College of Optometrists, 2016), College of Optometrists and the Royal College of Ophthalmologists Joint Supplementary College Guidance on Supervision in relation to Glaucoma-related Care by Optometrists (College of Optometrists and Royal College of Ophthalmologists, 2010).

In the PROforma model for this study there is only one Decision node (Decision Support Diagnosis in Figure 4-9), as diagnosis was main outcome in addition to appropriateness of management (including referral). The management recommendations in turn are based on the diagnostic decision. For example, if OHT is selected as the diagnosis by the optometrist then the recommendations for that diagnosis are displayed to the optometrist (Figure 4-9). The decision node provides recommendations to the optometrist to help choose an item from the different options available. In this CDSS application, the options are different diagnoses that the patient may have, and the optometrist must choose one.

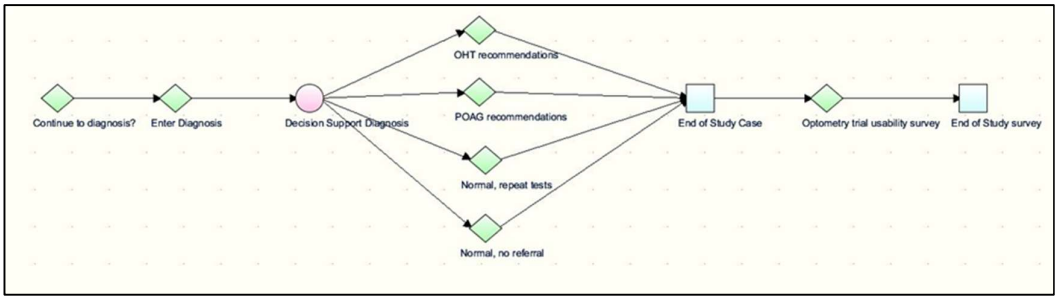


Figure 4-9 - Optometry decision support PROforma diagnosis and management workflow

4.2.3.3 Decision Support Arguments and Candidates

PROforma decisions have Candidates and Arguments for each decision (Sutton and Fox, 2003). Argumentation is a logical technique that helps in making decisions by weighing the arguments for and against a proposition (Fox, Krause and Elvang-Gøransson, 1993). The method used by PROforma to weigh in the different arguments and generate a recommendation is based on the argumentation. Each candidate can have multiple arguments that either support or negate a candidate. In a diagnostic CDSS such as the one described in this study, the candidate will be a diagnosis and the arguments will be observations that are present in the patient and either support or negate a diagnosis. Candidates and arguments can be used to model other CDSSs such as risk assessment applications where the candidates become the risk level (High, Moderate or Low) and the arguments become the clinical observations that support or negate the risk level. In this study, the candidates are diagnoses as the effect of diagnostic recommendations on optometrist decision making is being evaluated.

The degree to which the argument (clinical observations) can either support or negate is represented by Symbolic support or Numeric Weights (See support in Table 4-3). Numeric weights were used in this study to represent degree of support as the domain experts found them more intuitive and easier to understand. The weights do not signify a statistical probability of an argument. The weights are a qualitative measure of the degree of support of each argument. If needed, statistical evidence can be used to form weights by using machine learning techniques to automatically update the weights. But this requires the availability of large repositories of high quality structured clinical data and this is difficult to obtain. Therefore, in this study a much simpler method for determining weights was used.

The weight for each argument were determined by domain experts. Since there were not many studies that evaluated argumentation-based CDSSs for diagnosis decision making, no standard way for representing weights exist. Therefore, a simple weighting scheme for creating weights was devised. If a clinical observation is present/absent and it supports a diagnosis, a weight of +1 was given. If the clinical observation is present/absent and it negates a diagnosis a weight of -1 was given. If the observation strongly supports a diagnosis a weight of +2 was given. If it strongly negates a diagnosis a weight of -2 was given. Some observations are highly indicative of pathology. These observations are given a weight above 2. This can be easily explained to optometrists. The real benefit of using argumentation-based models is the argument itself. The weights help in ranking the diagnosis and help explain why the diagnosis was ranked above others in the diagnostic recommendation. The ability to better explain the patient's condition is the reason why argumentation-based CDSS tools was chosen to create the diagnostic CDSS.

In the decision support application that modelled the OHT vignette there were 4 diagnosis candidates. These candidates were "OHT", "Primary Open Angle Glaucoma (POAG)", "Normal-Repeat tests" and "Normal-No referral". The diagnosis candidates for OHT vignette (and the other vignettes) were chosen by the experts based on their similarity to the actual diagnosis. For example, POAG has signs and symptoms that are clinically like OHT and could be confused for OHT. Other diagnosis candidates (for example, Normal) are chosen to balance the diagnosis candidates and avoid giving the optometrists clues about the actual diagnosis of the vignettes determined by the experts.

The table below (Table 4-3) shows a sample of the arguments either supporting or negating the OHT candidate in OHT vignette. Table 4-3 includes a description of the arguments. For example, the argument Visual_fields_Right_Humphrey = "Normal" AND Visual_fields_Left_Humphrey = "Normal" means that if the visual field test Humphrey is normal in both left and right eye, then it supports the candidate OHT. So, if the normal value is present then the Tallis application will add a total of +2 towards the OHT candidate. The total weight for each candidate is calculated based on the arguments that are activated based on the data and the weights are attached to the candidate. The total weights are used as the basis to generate a ranked list of candidates (diagnoses) (Figure 4-20). (See Appendix 2 - section 2.1 for other candidates in the OHT vignette). Similarly, there are arguments supporting or negating all candidates for all 4 vignettes (See Appendix 2 – sections 2.1 to 2.4). The weights and arguments were validated by domain experts using test scenarios. Data was entered in the CDSS and results were validated by the domain experts using test scenarios and changes made accordingly.

OHT candidate	Support
IOP Intraocular pressure in mmHg - IOP is the fluid pressure of the eye. Non-contact tonometry (NCT) is one of the tests for measuring IOP. The other test is Goldmann applanation tonometry (GAT). If the IOP measurements is greater than 21mmHg then it is higher than normal and it indicates OHT, but not Glaucoma.	
IOP_NCT_Right_Reading_Average > 21 OR IOP_NCT_Left_Reading_Average > 21	2
IOP_GAT_Left = "24" OR IOP_GAT_Right = "23"	2
Visual Fields Humphrey Visual fields is a method for measuring the patient's entire scope of vision. Humphrey is a type of visual fields test. In OHT there is no pathology, only elevated IOP. Therefore, the visual fields tests will be normal.	
Visual_fields_Right_Humphrey = "Normal" AND Visual_fields_Left_Humphrey = "Normal"	2
Visual_fields_Right_Humphrey = "Abnormal" OR Visual_fields_Left_Humphrey = "Abnormal"	-2
(Visual_fields_Right_Humphrey = "Uncertain" AND Visual_fields_Left_Humphrey = "Uncertain") OR (Visual_fields_Right_Humphrey = "Uncertain" OR Visual_fields_Left_Humphrey = "Uncertain")	-1
Visual Fields Henson Henson is another type of visual fields test. In OHT there is no pathology, only elevated IOP. Therefore, the visual fields tests will be normal.	
Visual_fields_Right_Henson = "Normal" AND Visual_fields_Left_Henson = "Normal"	2
Visual_fields_Right_Henson = "Abnormal" OR Visual_fields_Left_Henson = "Abnormal"	-2
(Visual_fields_Right_Henson = "Uncertain" AND Visual_fields_Left_Henson = "Uncertain") OR (Visual_fields_Right_Henson = "Uncertain" OR Visual_fields_Left_Henson = "Uncertain")	-1
Optic disc status Optic disc is the part of the optic nerve that is clinically visible on examination. The optometrist must look at images of the back of the eye and grade the images (Normal/Uncertain/Abnormal) based on the status of the optic disc. In OHT there is no pathology, only elevates IOP. Therefore, the optic discs will be normal.	
Optic_Disc_Status_Right = "Abnormal" OR Optic_Disc_Status_Left = "Abnormal"	-2
Optic_Disc_Status_Right = "Normal" AND Optic_Disc_Status_Left = "Normal"	2
(Optic_Disc_Status_Right = "Uncertain" AND Optic_Disc_Status_Left = "Uncertain") OR (Optic_Disc_Status_Right = "Uncertain" OR Optic_Disc_Status_Left = "Uncertain")	-1
Van Herick Van Herick is another test for glaucoma that is based on grading of images of the front of the eye. The optometrist must grade the images (0-4) based on the clinical appearance of images. A grade of >2 means that there is no pathology and therefore supports OHT.	
Van_Herick_Right =< 2 OR Van_Herick_Left =< 2	-1
Van_Herick_Right > 2 AND Van_Herick_Left > 2	1

Table 4-3 - OHT candidate with arguments and weights for OHT vignette

4.2.3.4 CDSS development

Tallis uses PROforma (Fox, Patkar and Thomson, 2006b) guideline modelling language for modelling the different stages of clinical diagnostic encounter i.e. Symptoms and History, Refraction, Anterior Eye Examination, Posterior Eye Examination, Investigations, Oculomotor balance (OMB) and Diagnosis and executing the decision making pathway. The different stages of the clinical encounter, the questions

asked in each stage, the images used for grading tests and all the data were based on the vignettes in the control group study (Figure 4-1 & Figure 4-4). The vignettes modelled in PROforma can be seen in Appendix 3 – section 3.2.

A consensus panel comprising of a consultant ophthalmologist and an expert optometrist had independently reviewed the clinical vignettes and decided the appropriate diagnostic criteria for the CDSS and triggers for guideline recommendations (Murdoch, Lawrenson and Myint, 2015).

The CDSS application is a stand-alone web-based application and the optometrist inputs data from the vignette examination in this software. The optometrist is required to enter data relevant to make a decision, which includes symptoms, history, results of examinations and investigations etc. The program runs through a series of questions that will guide the optometrist through relevant pathways, ensuring that the optometrists follow proper guidelines and protocols when obtaining information about the patient. PROforma technology assists the optometrists in confirming a diagnosis providing a differential diagnosis and provides referral and management recommendations based on the individual characteristics of the vignette. The CDSS application modelled in Tallis allows recording of data in a database and all the data were recorded in a MySQL database for analysis.

4.2.3.5 Procedure (intervention group study)

The study was approved by City University London Department of Computer Science Research Ethics Committee (See Appendix 1 – section 1.2). Participants were given the information about the study via email (See Appendix 1 – section 1.1). The information contained details of the study purpose, duration, right to withdrawal from study, optometrist data confidentiality, CET point information, ethical committee clearances and study team contact details.

If the participants had any questions they were encouraged to contact the study team. Once the participant had read the information and agreed to its contents he/she can click on the link provided which will take them to the study application with the online consent form. Only when the participants have agreed to the terms and conditions specified in the consent form will they be taken to the study start page.

4.2.3.5.1 Step by Step process taken by the optometrists when accessing the decision support application

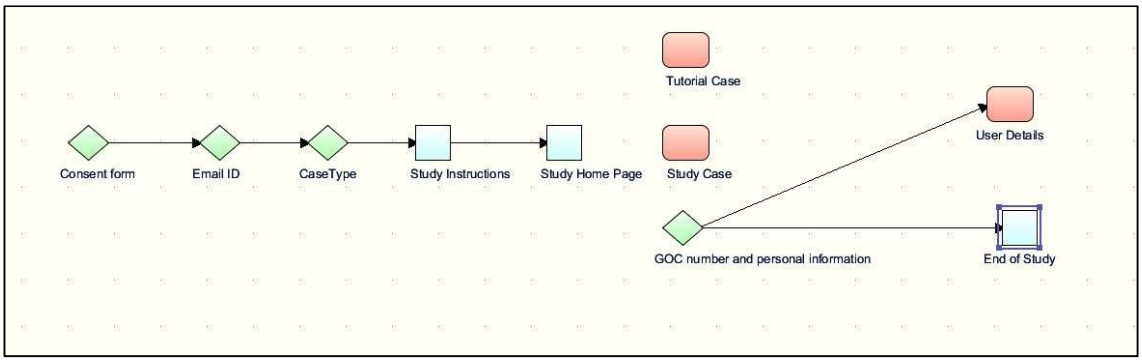


Figure 4-10 - Outline of the optometry study in PROforma

1. The optometrists were sent an email with a link to the study vignette (See Appendix 3 – section 3.1). The decision support application can be accessed via a standard web browser. When the optometrist clicked on the link s/he was forwarded to the website which ran the experiment. OpenClinical.net (OpenClinical.net, 2015) was used as the platform to conduct the experiment.

2. The optometrist is redirected to a consent form (Figure 4-11) on the via web application via a link forwarded through email. Once consent has been provided he/she is asked to enter the email ID. The vignette information is automatically recorded in a MySQL database.

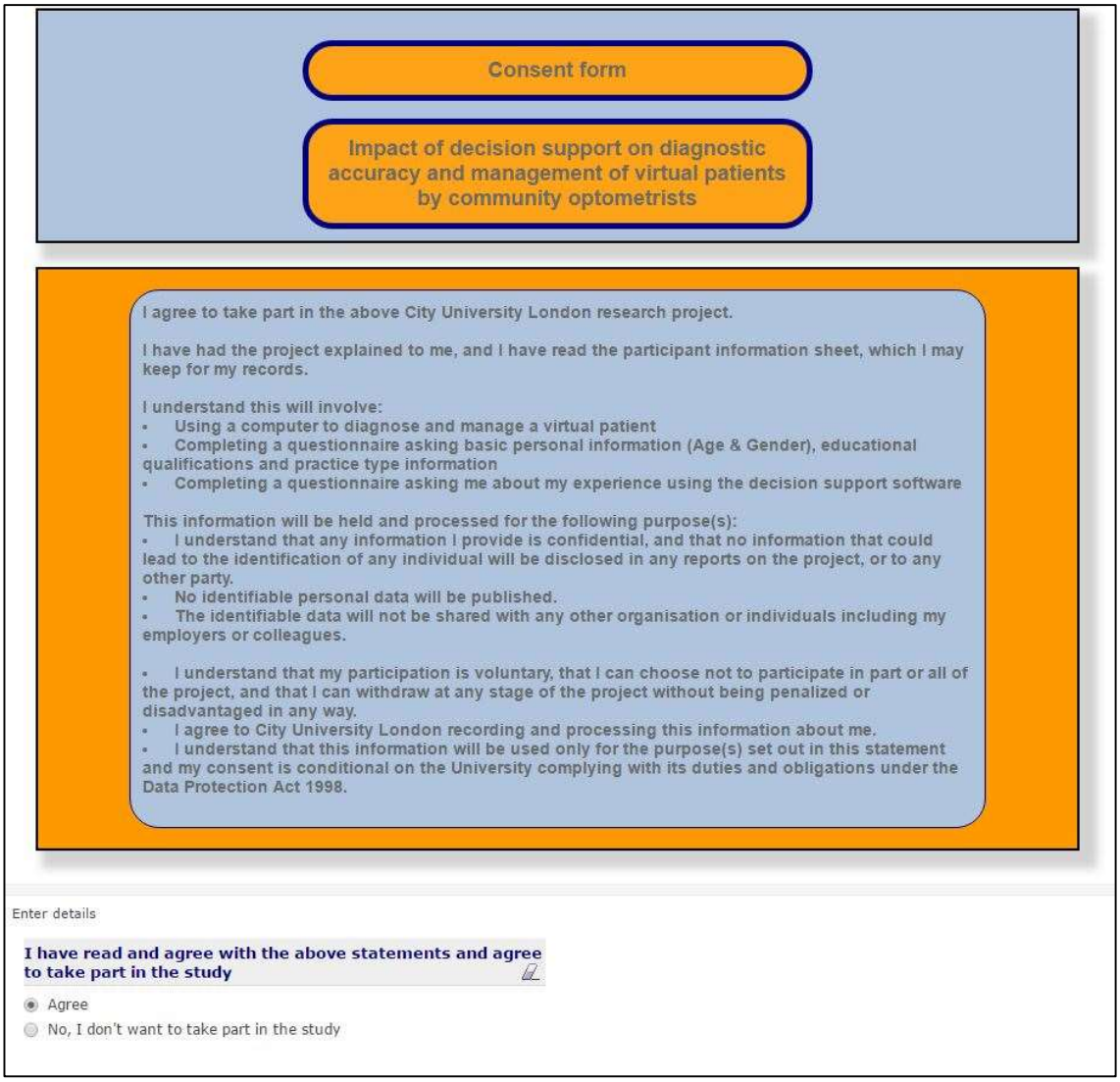


Figure 4-11 - Consent form

3. The study instructions are provided to the optometrists outlining the various steps to be taken which includes video-based instructions (Figure 4-12).

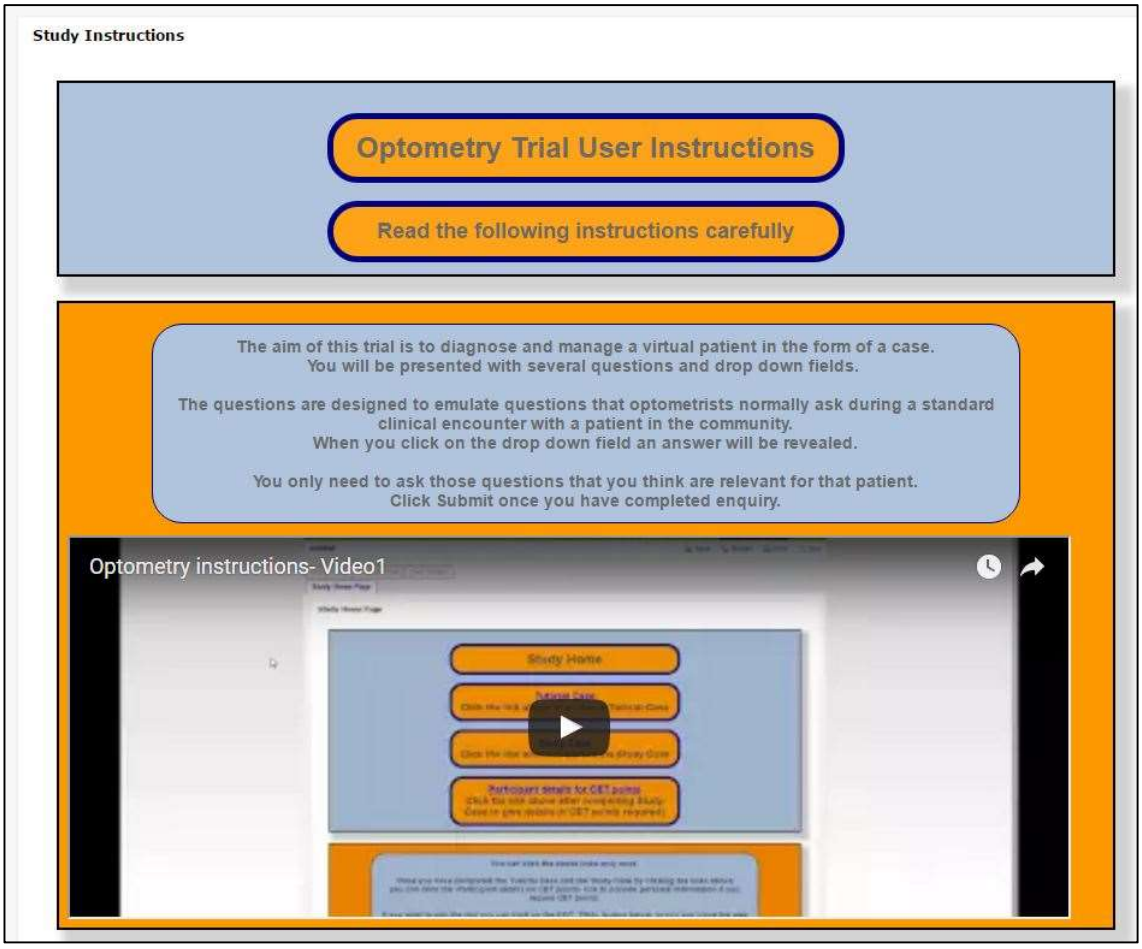


Figure 4-12 - Study Instructions

4. At the Study Home Page, they are given the option of either accessing the Tutorial Case or go directly to the Study Case (Figure 4-13). Workflow can be seen in Figure 4-10. The Tutorial Case provides an overview of the tasks the optometrist is required to perform for the study, for example grading images, diagnosis selection etc.

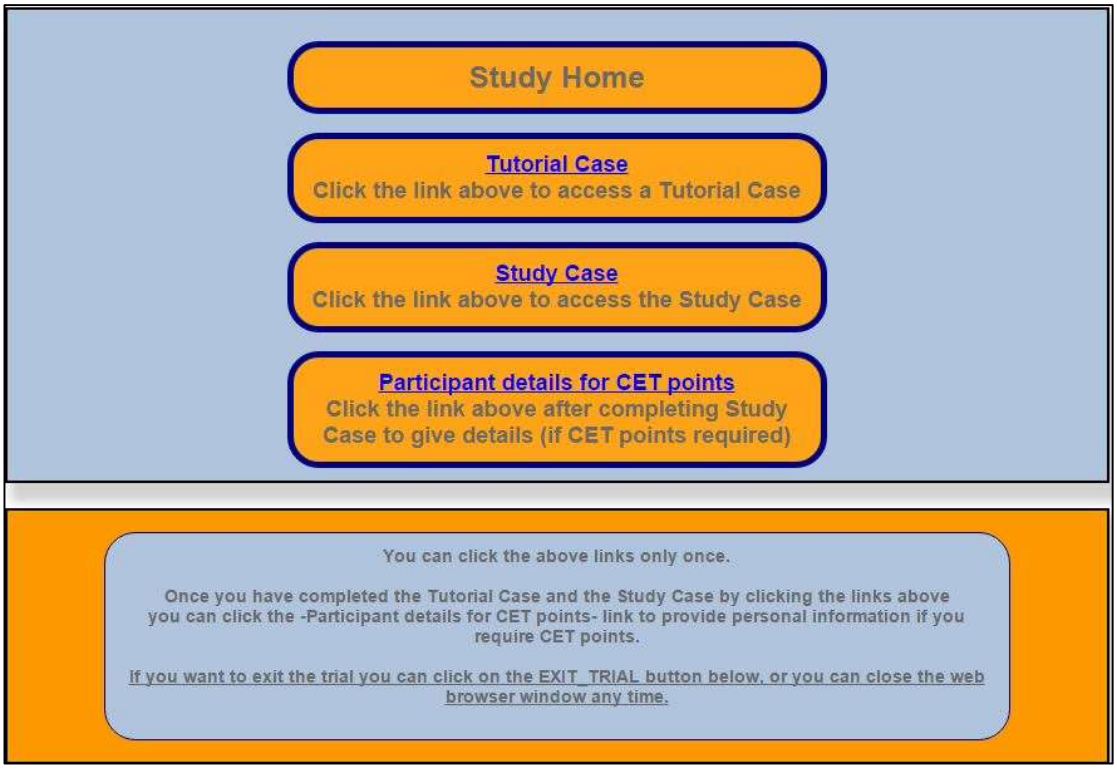


Figure 4-13 - Study Home page

5. On the Study Case page, they can access the different sections representing the different sections of the patient data including Symptom and History, Refraction, Oculomotor Balance, Anterior Eye Examination, Posterior Eye Examination, Investigations and Diagnosis and Management (Figure 4-14). In some sections, the optometrist clicks the drop-down to reveal the answer to the question (Figure 4-15). This simulates a question-answer session where the patient provides an answer to a question provided by the optometrist.

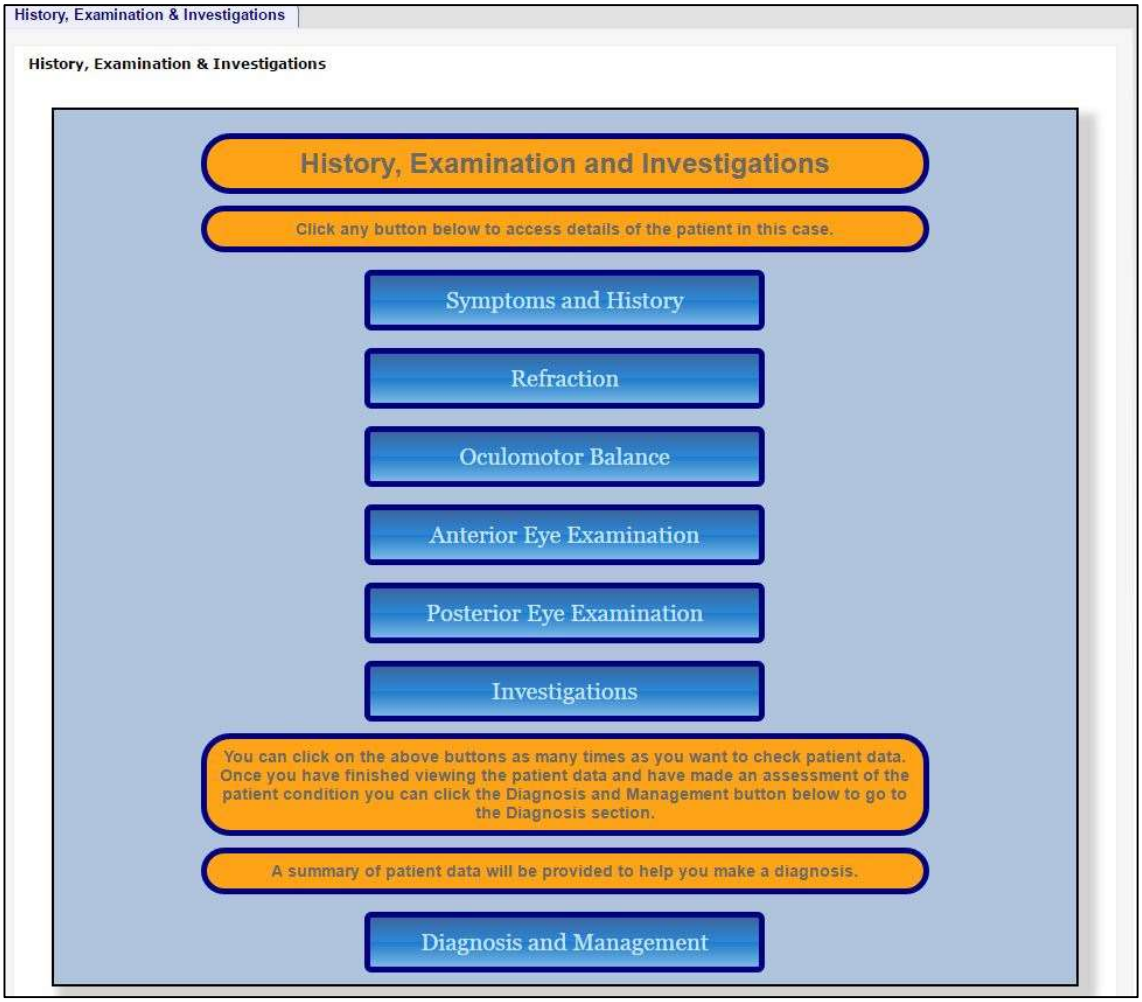


Figure 4-14 - History, Examination & Investigations

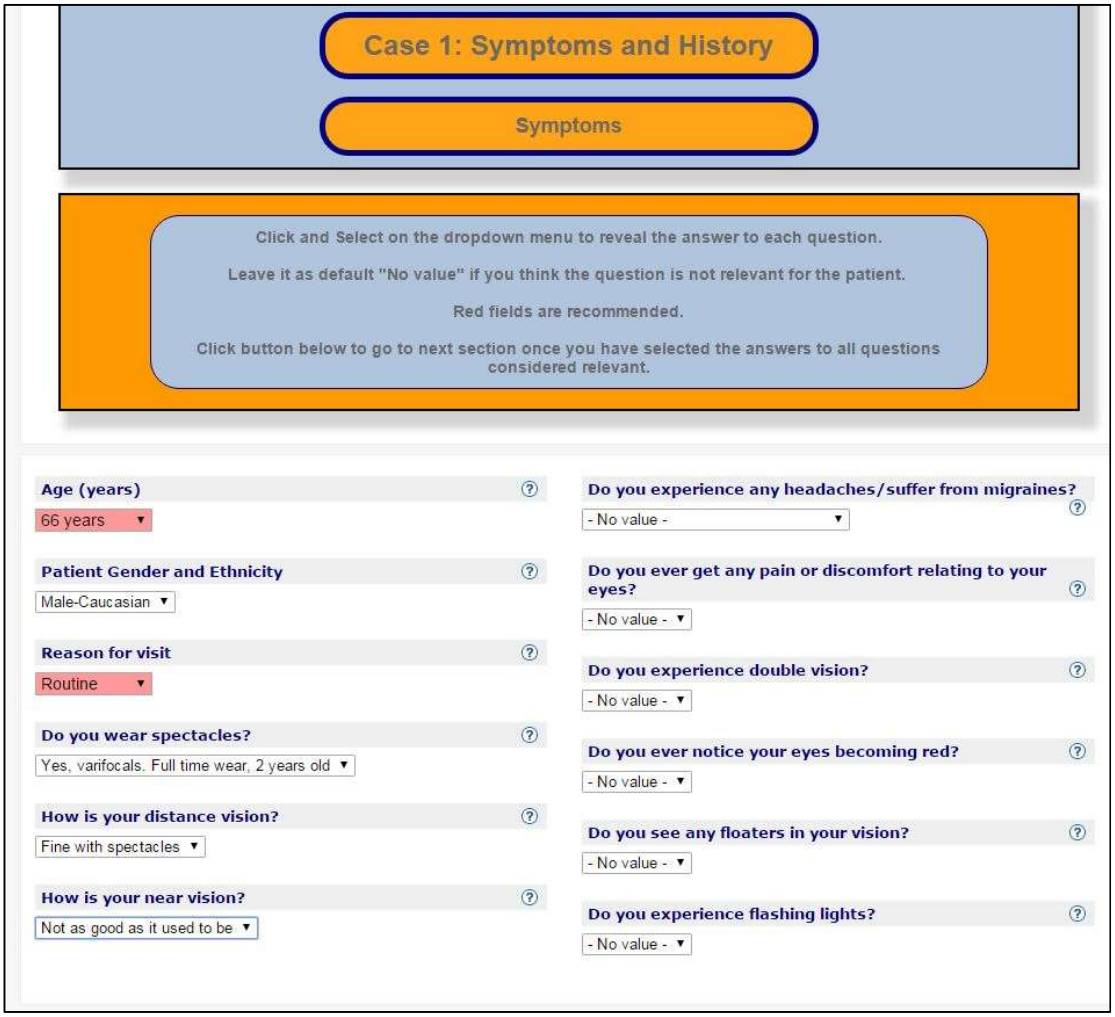


Figure 4-15 - Symptoms and History

6. The optometrists were presented with guidance from current clinical management guidelines (See GUIDELINE ALERT !! in Figure 4-16) that recommend the optometrist ask appropriate questions or select appropriate investigations. It is up to the clinician to ask these questions (Figure 4-16) or perform the tests (Figure 4-17).

Figure 4-16 - Clinical guideline recommendations

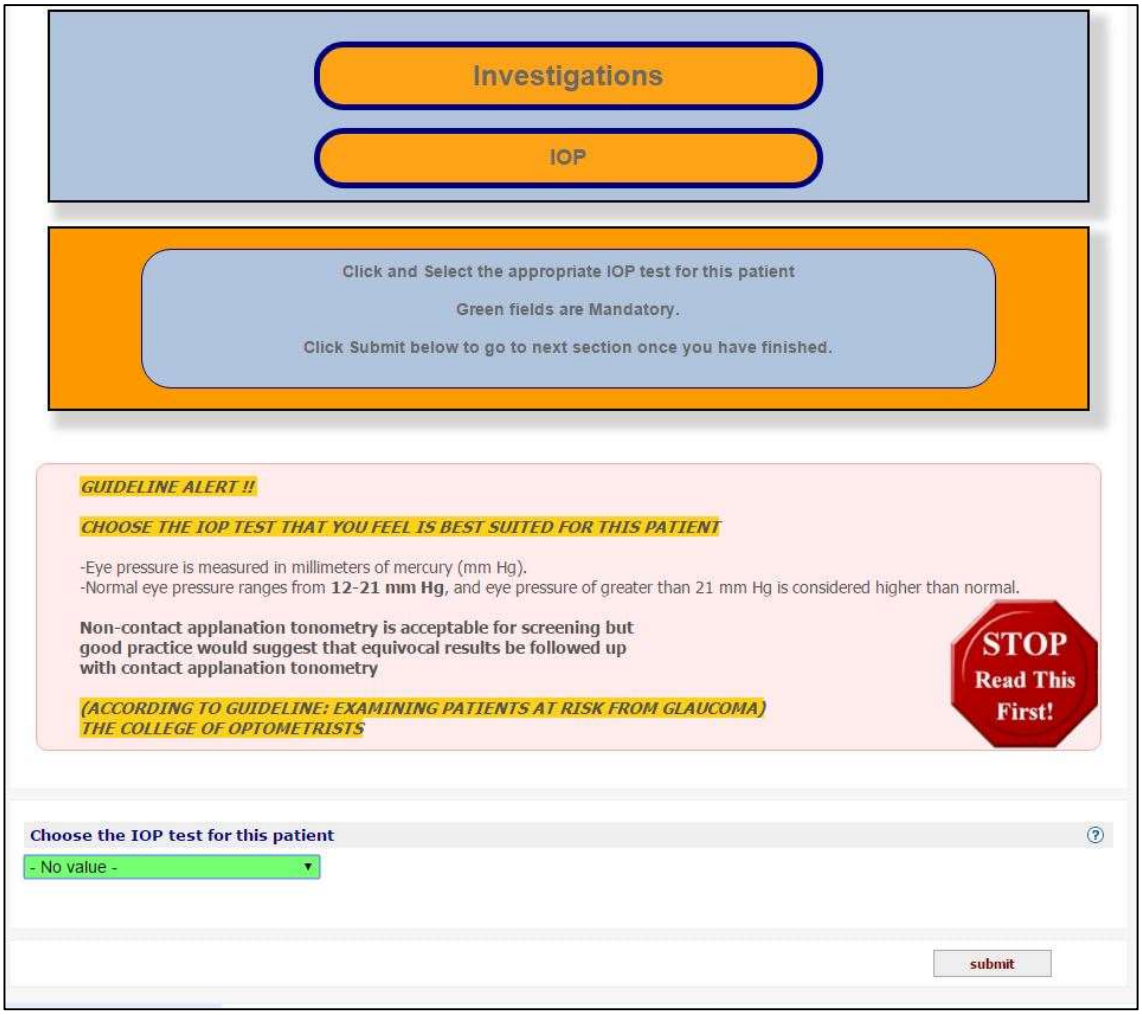


Figure 4-17 - Investigations

7. In other sections, especially Anterior Eye Examination and Posterior Eye Examination, the optometrists were given images relevant to the case and they graded the images from the options given (Figure 4-18). Clinical guidance on how to grade the images is provided. The location of these recommendations and the terminology and wording used in the recommendations was determined by the clinical experts and the guideline recommendations are specific to the questions being asked or the task the optometrist is asked to perform. For example, in Figure 4-18 the different grades that the image can take is given based on the European Glaucoma Society guideline: Terminology and Guidelines for Glaucoma – 3rd Edition (European Glaucoma Society, 2008). The guideline recommendation tells the optometrist when the Van Herick test is appropriate and how to interpret and grade the images.

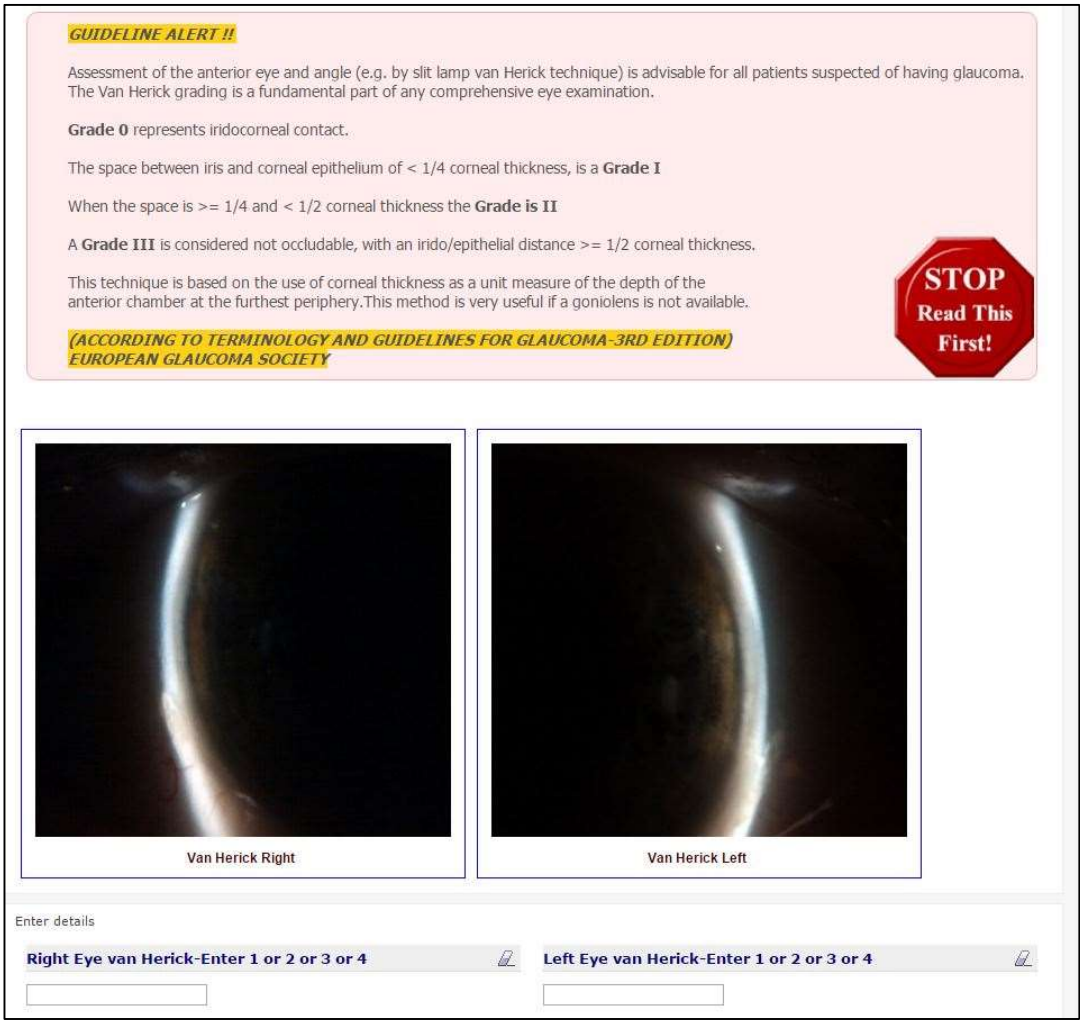


Figure 4-18 - Image grading - Van Herick

8. Once all the data that the optometrist thinks is relevant is collected, he/she can proceed to Diagnosis and Management for the case. The optometrist is given the option to go back and check any data that is needed before going to the Diagnosis and Management section. Initially the optometrist is asked to select the right diagnosis from a list. Only data that was enquired were provided to help select the diagnosis (Figure 4-19). The only decision support provided up to this point will be the different guideline recommendations provided at different stages of the process. The list of diagnoses provided in this list is the same as that of the control group study.

Figure 4-19 - Diagnosis selection by optometrist with observations

9. Once the optometrist has selected his/her diagnosis the CDSS will provide the diagnoses that it thought was appropriate (Figure 4-20). Instructions on how to interpret the recommendations provided by the decision support are given (Figure 4-21 & Figure 4-22). The weights to be attached to each argument supporting or negating a candidate was determined by the clinical experts and based on their professional judgement (See section - 4.2.3.3). The ranking of diagnostic recommendations is calculated based on the data collected from the patient, i.e. questions asked, and clinical tests and investigations performed. If the optometrist does not perform certain tests and investigations, or if the optometrist grades images incorrectly the diagnostic ranking and weights associated with the diagnosis will change. The prototype CDSS simulates a real-life CDSS which provides a list of most likely diagnoses based on data collected. The list of diagnoses provided by the CDSS is smaller than the list from which the optometrist first selects the diagnosis without diagnostic decision support. Since it was beyond the scope of this study to model the arguments and weights for all optometry diagnoses it was decided to use a smaller list which simulates a real-life CDSS that provides a list of most likely diagnosis. In the OHT vignette the candidates are OHT, POAG, Normal-no referral, No-repeat tests. The findings in the OHT vignette suggest that patient can have elevated IOP but is otherwise normal and should be managed in the community. There the 2 Normal candidates were given as recommendations. Elevated IOP can indicate possible POAG and therefore that was given as a candidate. Similarly, other candidates were chosen for the other vignettes because the findings could indicate similar conditions. The diagnostic recommendations are ranked according to the weights (Figure 4-20) with the diagnosis with the highest weight located on top of the list. The

diagnostic recommendations include the arguments in favour and against a diagnosis. In Figure 4-20 the arguments for each diagnosis is seen below the diagnosis. It shows what is the weight for each argument. In the figure Right or Left IOP-GAT greater than 21 mmHg is shown with a weight of +2. This means what the patient has been found with an IOP greater than 21mmHg during the GAT test. A weight of +2 has been attached to the diagnosis. The Tallis reasoner sums the weights of each argument based on the data that was collected. The total score of +2 was given to Primary Open Angle Glaucoma since the other arguments had a weight of +1 and -1 and therefore cancelled each other. The real benefit of the argumentation-based diagnostic CDSS is the ability to obtain a ranked diagnostic recommendation while getting the reasons for the recommendation in an easy to understand manner for the clinician.

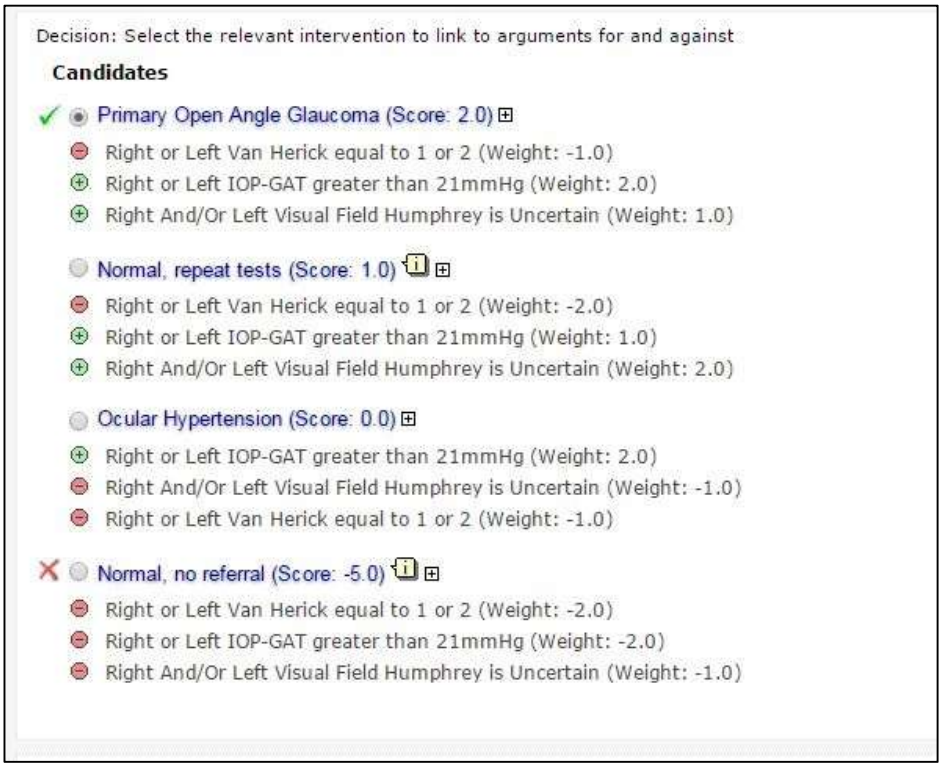


Figure 4-20 - Diagnosis candidate with arguments supporting/negating

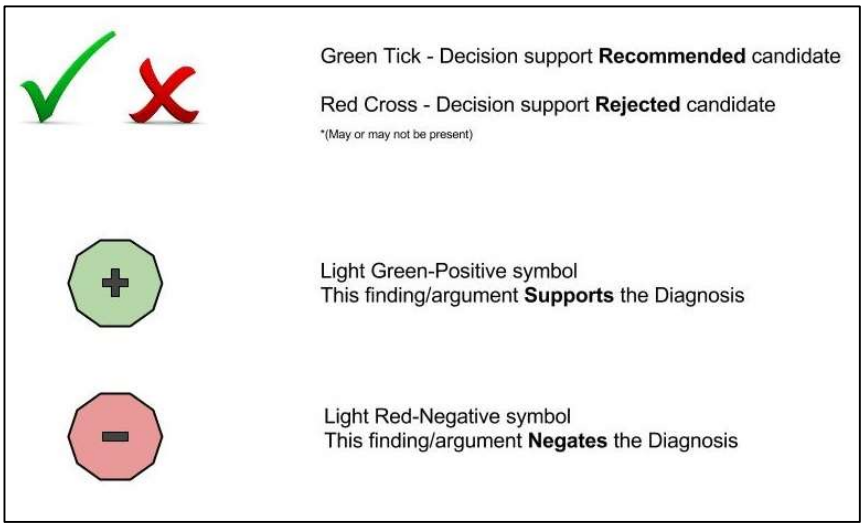


Figure 4-21 - Symbols denoting support/negation

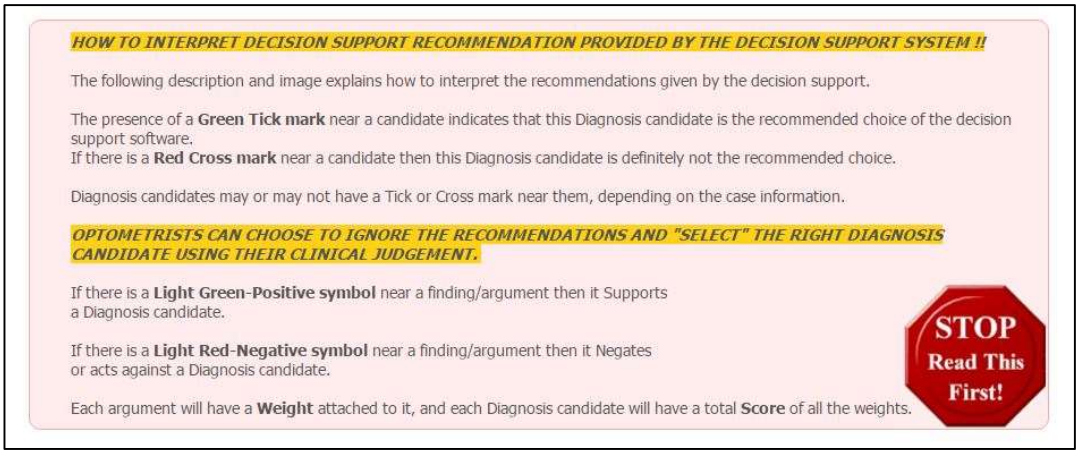


Figure 4-22 - Interpreting the symbols

10. The optometrist is free to reject the recommended decision support diagnosis and select what he/she thinks is appropriate. The optometrist is asked to select the diagnosis twice. The 1st time the optometrist will select the diagnosis without any diagnostic recommendations (Figure 4-19). The 2nd time the optometrist were given a ranked list of diagnostic recommendations and the optometrist can select from this list. The workflow can be seen in Figure 4-9.

11. Once a diagnosis provided in the diagnosis recommendation list is selected (Figure 4-20) the optometrist is then asked to provide management and referral recommendations for the patient (Figure 4-23). Clinical guidance recommendations are provided based on the diagnosis selected by the optometrists. For example, the recommendations in Figure 4-23 are shown if the optometrist selects OHT from diagnosis recommendation list.

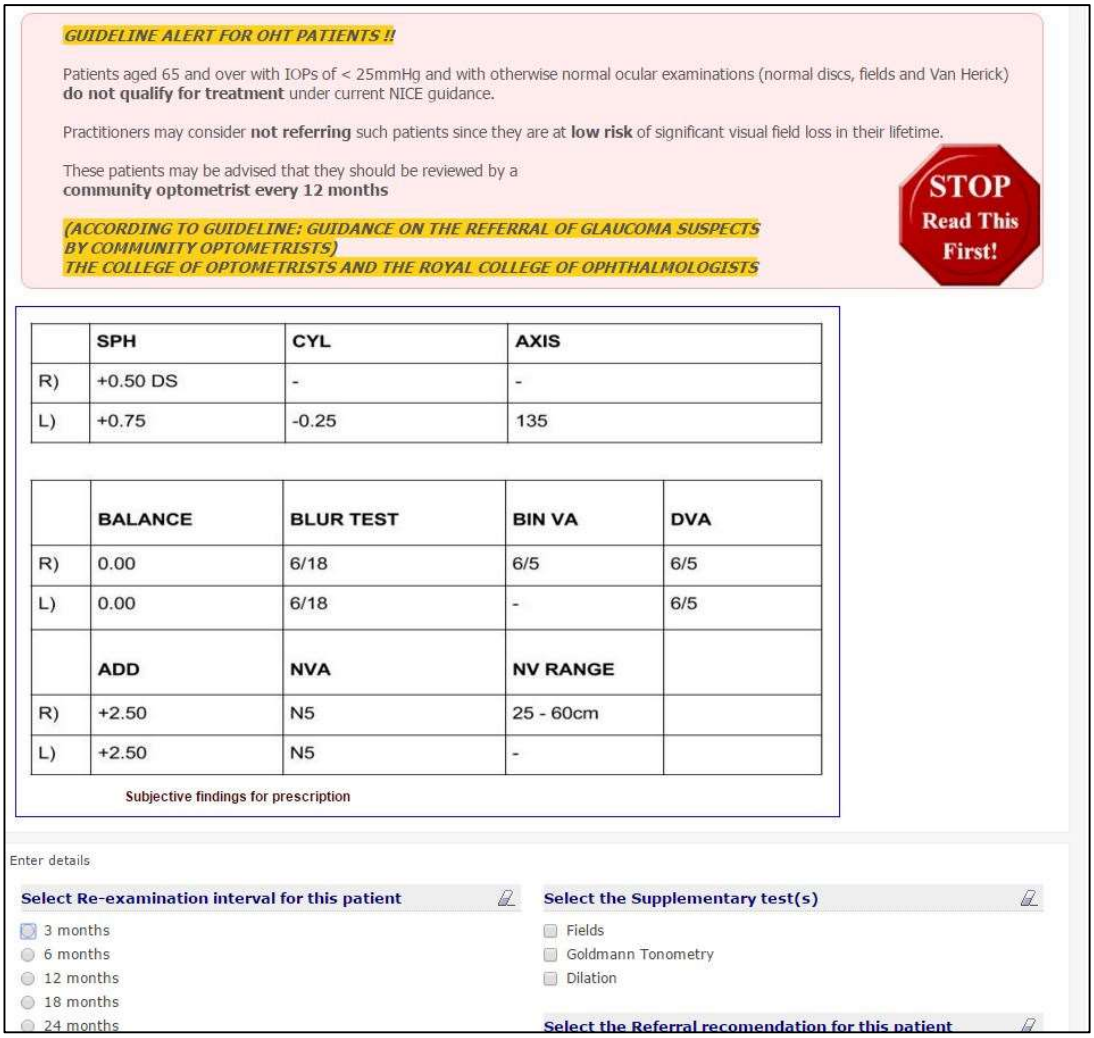


Figure 4-23 - Management advice

12. The optometrist was then asked to fill a survey form based on System Usability Scale (Brooke, 1996) (Figure 4-24) to obtain qualitative feedback about the decision support application and basic demographic and background information about the optometrist were collected (Figure 4-25).

Figure 4-24 - Survey form (1)

Figure 4-25 - Survey form (2)

4.2.4 Statistical Analysis

The data stored in the MySQL database was collected for analysis once the data collection was completed. After the data had been cleaned, the data was imported into the statistical package R (version 3.3.1) for data analysis (R Core Team, 2013). The differences between the intervention group studied in the current

study (With CDSS) and the control group (Without CDSS) were statistically analysed with the Pearson chi-square test with Yates continuity correction and Fishers exact test for categorical data. A p-value of 0.05 or less was used to indicate statistical significance.

The chi-square test is used to test if there is a possible relationship between 2 categorical variables and if the observed difference between the 2 variables has arose by chance. In this test a 2-way table is generated after the data has been counted and divided into categories.

In this study the total number of optometrists who asked the questions to the vignette patient was recorded. The total number of clinical test/investigations performed was recorded. The chi-square test was used to see if the CDSS guideline recommendations has any impact on the optometrists asking questions or performing clinical investigations. The function chisq.test in R was used to perform the test (See Appendix 18 for raw data).

The total number of optometrists from both control and intervention group who got the diagnosis and management right or wrong was calculated. The right diagnosis was determined by the experts when they developed the vignettes for the control study. The chi-square test will test if the CDSS had any impact on the decision making of the optometrists or the likelihood of the optometrists getting the diagnosis and management right/wrong is due to chance. The function chisq.test in R was used to perform the test. (See Appendix 16 for raw data).

The Pearson's chi-squared test is one the most popular types of chi-square tests and is often used in clinical research (McHugh, 2013). The Yates continuity correction is applied when in any given cell, the count is below 5 and therefore there is not enough data to conduct a chi-square test. In this study the count in the intervention group was below 5 in several instances and therefore the Yates continuity correction was applied throughout for the chi-square test (Larntz, 1978).

In addition to the chi-square test the Fisher' exact test was applied for the diagnostic and management numbers. The Fisher's test is considered more accurate when the sample size is small (Routledge, Routledge and Rick, 2005). The function fisher.test in R was used to perform the test. (See Appendix 16 for raw data). The significance level α was set at 0.05.

4.3 Results

4.3.1 History and Investigations

This section outlines the total numbers of history/investigations that were answered/performed for the Intervention group and the control group optometrists. The chi-square significance test results have been summarised in Appendix 18. This section outlines the impact of guideline recommendations on information gathering.

The vignettes presented to both control group and Intervention group optometrists were presented in different sections. The sections are:

- History
  - Symptoms
  - Medical History
  - Previous Ocular History
  - Family History
  - General History
- Clinical Tests and Investigations
  - Oculomotor balance (OMB) Habitual
  - OMB New Rx (New prescription)
  - Anterior Exam 1 (Van Herick test)
  - Anterior Exam 2 (Slit lamp investigations)
  - Posterior Exam (Ophthalmoscopy)
  - Ophthalmoscopy
  - Amsler grid
  - Colour vision
  - IOP GAT (Intraocular Pressure Goldmann Applanation Tonometry)
  - IOP NCT (Intraocular Pressure Non-Contact Tonometry)
  - Stereopsis
  - Visual field Henson
  - Visual field Humphrey
  - Refraction

4.3.1.1 OHT

The following section presents the total numbers of history/investigations answered and the statistically significant results of the history/investigations chi-square test of the OHT vignette. OHT is a condition where the pressure inside the eye (Intraocular pressure or IOP) is greater than normal. People with OHT are considered to be at risk of glaucoma and therefore needs to be regularly monitored by the optometrist for signs of glaucoma (Mowatt *et al.*, 2008).

Total number of history/investigations in the OHT vignette is **n=83**. The total number of history/investigations where there is a statistically significant difference in the answered/unanswered questions between the intervention and control group is **n=15 (18.07%)**. The significant results have been outlined in Table 4-4. The remaining results have been outlined in Appendix 18.1.1.

Section, Question/Investigation, Chi-square, p-value, Significance (α= 0.05)	**Answered with guideline rec (Intervention)**	**Not answered with guideline rec (Intervention)**	**Answered without CDSS (Control)**	**Not answered without CDSS (Control)**
Section: Symptoms, Do you ever notice your eyes becoming red?,	10 (76.92%)	3 (23.08%)	39 (39.00%)	61 (61.00%)

Chi-square: 5.281, p-value: 0.022				
Section: Symptoms, Do you see any floaters in your vision?, Chi-square: 4.105, p-value: 0.043	13 (100.00%)	0 (0.00%)	69 (69.00%)	31 (31.00%)
Section: Symptoms, Do you experience flashing lights?, Chi-square: 4.809, p-value: 0.028	13 (100.00%)	0 (0.00%)	66 (66.00%)	34 (34.00%)
Section: Medical History, Are you diabetic?, Chi-square: 3.883, p-value: 0.049	13 (100.00%)	0 (0.00%)	70 (70.00%)	30 (30.00%)
Section: Ocular History, Have you ever been told you have lazy eye?, Chi-square: 6.058, p-value: 0.014	11 (84.62%)	2 (15.38%)	44 (44.00%)	56 (56.00%)
Section: General History, Smoker, Chi-square: 4.904, p-value: 0.027	12 (92.31%)	1 (7.69%)	56 (56.00%)	44 (44.00%)
Section: OMB Habitual, Motility, Chi-square: 7.142, p-value: 0.008	11 (84.62%)	2 (15.38%)	41 (41.00%)	59 (59.00%)
Section: OMB Habitual, NPC, Chi-square: 3.898, p-value: 0.048	6 (46.15%)	7 (53.85%)	18 (18.00%)	82 (82.00%)
Section: OMB Habitual, Cover test Near, Chi-square: 6.395, p-value: 0.011	13 (100.00%)	0 (0.00%)	60 (60.00%)	40 (40.00%)
Section: OMB New Rx, Fixation Disparity Distance, Chi-square: 6.107, p-value: 0.013	6 (46.15%)	7 (53.85%)	14 (14.00%)	86 (86.00%)
Section: Anterior Exam 1, Left Eye Van Herick,	13 (100.00%)	0 (0.00%)	53 (53.00%)	47 (47.00%)

Chi-square: 8.616, p-value: 0.003				
Section: Anterior Exam 1, Right Eye Van Herick, Chi-square: 8.616, p-value: 0.003	13 (100.00%)	0 (0.00%)	53 (53.00%)	47 (47.00%)
Section: Anterior Exam 2, Tears, Chi-square: 6.756, p-value: 0.009	8 (61.54%)	5 (38.46%)	23 (23.00%)	77 (77.00%)
Section: IOP NCT, IOP NCT Left Eye Reading 1, Chi-square: 6.041, p-value: 0.014	6 (46.15%)	7 (53.85%)	81 (81.00%)	19 (19.00%)
Section: IOP NCT, IOP NCT Right Eye Reading 1, Chi-square: 6.041, p-value: 0.014	6 (46.15%)	7 (53.85%)	81 (81.00%)	19 (19.00%)
Section: Refraction, Subjective Findings, Chi-square: 6.395, p-value: 0.011	13 (100.00%)	0 (0.00%)	60 (60.00%)	40 (40.00%)

Table 4-4 - OHT vignette significant chi-square results

The guideline recommendations in the OHT vignette are in Appendix 19. The results of comparing the questions/investigations answered/not answered show that there is a significant difference in the proportions in **n=15 (18.07%)** questions/investigations. Some questions in the Symptoms section were answered more in the intervention group than the control group. These questions are part of the routine clinical history taking and needed to formulate early diagnostic hypotheses. In the Family History section, the optometrists were given the guidance to check Family History of Glaucoma but there was no significant difference between the intervention group **(n=13 (100.00%))** and the control group **(n=94 (94.00%))** in the asking the patient for an answer. Greater proportion of optometrists performed the Anterior Exam 1(Van Herick test) in the intervention group **(n=13 (100.00%))** compared to the control group **(n=53 (53.00%))**. The Van Herick test is an important test for identifying OHT patients. Since OHT patient only have elevated IOP and no other clinical pathology is present the Van Herick test can identify the presence of pathology. However other tests such as Visual Fields Humphrey that provides more detailed information about the visual field of the patient did not see a difference in the numbers between the intervention group and control group optometrists.

OMB tests are not needed for the OHT patient but have been performed more in the intervention group than the control group. Guidance was provided to the intervention group optometrists to perform 4 readings for the IOP NCT test and the average taken as the final reading. However, there was no difference in the control and intervention group for the 4th reading. Overall the use of the guideline recommendations only showed a difference in the performing on Anterior Exam 1(Van Herick test).

4.3.1.2 Normal

The following section presents the total numbers of history/investigations answered and the statistically significant results of the history/investigations chi-square test of the Normal vignette. Total number of history/investigations in the Normal vignette is **n=90**. The total number of history/investigations where there is a statistically significant difference in the answered/unanswered questions between the intervention and control group is **n=19(21.1%)**. The significant results have been outlined in Table 4-5. The remaining results have been outlined in Appendix 18.1.2.

Section, Question/Investigation, Chi-square, p-value, Significance (α= 0.05)	**Answered with guideline rec (Intervention)**	**Not answered with guideline rec (Intervention)**	**Answered without CDSS (Control)**	**Not answered without CDSS (Control)**
Section: Symptoms, Do you ever get any pain or discomfort relating to your eyes?, Chi-square: 4.141, p-value: 0.042	9 (100.00%)	0 (0.00%)	61 (59.80%)	41 (40.20%)
Section: Medical History, Do you have any allergies?, Chi-square: 5.915, p-value: 0.015	9 (100.00%)	0 (0.00%)	53 (51.96%)	49 (48.04%)
Section: Medical History, Do you have high blood pressure?, Chi-square: 4.716, p-value: 0.030	8 (88.89%)	1 (11.11%)	46 (45.10%)	56 (54.90%)
Section: General History, Smoker, Chi-square: 5.429, p-value: 0.020	9 (100.00%)	0 (0.00%)	55 (53.92%)	47 (46.08%)
Section: OMB Habitual, Accommodation BIN, Chi-square: 14.529, p-value: <0.000	6 (66.67%)	3 (33.33%)	12 (11.76%)	90 (88.24%)
Section: OMB Habitual, Accommodation Left Eye, Chi-square: 12.313, p-value: <0.000	6 (66.67%)	3 (33.33%)	14 (13.73%)	88 (86.27%)
Section: OMB Habitual, Accommodation Right Eye, Chi-square: 12.313, p-value: <0.000	6 (66.67%)	3 (33.33%)	14 (13.73%)	88 (86.27%)

Section: OMB Habitual, Confrontation, Chi-square: 4.755, p-value: 0.029	4 (44.44%)	5 (55.56%)	12 (11.76%)	90 (88.24%)
Section: OMB Habitual, PD, Chi-square: 10.487, p-value: 0.001	8 (88.89%)	1 (11.11%)	30 (29.41%)	72 (70.59%)
Section: OMB New Rx, Fixation Disparity Distance, Chi-square: 9.085, p-value: 0.003	5 (55.56%)	4 (44.44%)	12 (11.76%)	90 (88.24%)
Section: OMB New Rx, Fixation Disparity Near, Chi-square: 9.085, p-value: 0.003	5 (55.56%)	4 (44.44%)	12 (11.76%)	90 (88.24%)
Section: Anterior Exam 1, Right Eye Van Herick, Chi-square: 10.890, p-value: 0.001	9 (100.00%)	0 (0.00%)	38 (37.25%)	64 (62.75%)
Section: Anterior Exam 1, Left Eye Van Herick, Chi-square: 10.890, p-value: 0.001	9 (100.00%)	0 (0.00%)	38 (37.25%)	64 (62.75%)
Section: Anterior Exam 2, Shafers Sign, Chi-square: 6.665, p-value: 0.010	7 (77.78%)	2 (22.22%)	30 (29.41%)	72 (70.59%)
Section: Anterior Exam 2, Tears, Chi-square: 5.800, p-value: 0.016	8 (88.89%)	1 (11.11%)	42 (41.18%)	60 (58.82%)
Section: IOP NCT, IOP NCT Left Eye Reading 1, Chi-square: 9.012, p-value: 0.003	3 (33.33%)	6 (66.67%)	84 (82.35%)	18 (17.65%)
Section: IOP NCT, IOP NCT Right Eye Reading 2, Chi-square: 3.969, p-value: 0.046	3 (33.33%)	6 (66.67%)	73 (71.57%)	29 (28.43%)
Section: IOP NCT, IOP NCT Right Eye Reading 1,	3 (33.33%)	6 (66.67%)	84 (82.35%)	18 (17.65%)

Chi-square: 9.012, p-value: 0.003				
Section: Refraction, Current Rx, Chi-square: 23.418, p-value: <0.000	6 (66.67%)	3 (33.33%)	102 (100.00%)	0 (0.00%)
Section: Refraction, Subjective Findings, Chi-square: 4.340, p-value: 0.037	9 (100.00%)	0 (0.00%)	60 (58.82%)	42 (41.18%)

Table 4-5 - Normal vignette significant chi-square results

The guideline recommendations in the Normal vignette are in Appendix 19. In the Normal vignette there were no abnormal findings and general guideline recommendations were given. Guidance regarding asking Family History of Glaucoma was given but there was no significant difference between the groups in asking the patient. Like the OHT vignette, greater proportion of optometrists asked the Symptoms and Medical History questions. The OMB tests were not needed for the patient and greater proportion of control group optometrists did not perform the tests. More intervention group optometrists performed the Anterior Exam 1(Van Herick test) when compared to the control group **(n=9 (100.00%) vs n=38 (37.25%))**. The Current Rx-Refraction tests were performed more by the control group (n=102 (100.00%)) when compared to the intervention group **(n=6 (66.67%))**. However, the intervention group optometrists **(n=9 (100.00%))** performed the Subjective-Refraction tests more when compared to the control group **(n=60 (58.82%))**. Overall the guideline recommendation had a minimal impact on influencing the optometrist's decision to ask certain questions or perform investigations.

4.3.1.3 Normal Tension Glaucoma

The following section presents the total numbers of history/investigations answered and the statistically significant results of the history/investigations chi-square test of the Normal Tension Glaucoma (NTG) vignette. NTG is a type of glaucoma where the pressure inside the eye (IOP) is not raised and the lies within the normal range. However, there are signs of damage to the nerves of the eye that can be found during examination of the eye.

Total number of history/investigations in the NTG vignette is **n=84**. The total number of history/investigations where there is a statistically significant difference in the answered/unanswered questions between the intervention and control group is **n=24(28.6%)**. The significant results have been outlined in Table 4-6. The remaining results have been outlined in Appendix 18.1.3.

Section, Question/Investigation, Chi-square, p-value, Significance (α= 0.05)	**Answered with guideline rec (Intervention)**	**Not answered with guideline rec (Intervention)**	**Answered without CDSS (Control)**	**Not answered without CDSS (Control)**
Section: Symptoms, Do you ever notice your eyes becoming red?, Chi-square: 5.880, p-value: 0.015	11 (84.62%)	2 (15.38%)	45 (44.55%)	56 (55.45%)
Section: Medical History, Do you have high blood pressure?,	12 (92.31%)	1 (7.69%)	54 (53.47%)	47 (46.53%)

Chi-square: 5.624, p-value: 0.018				
Section: Medical History, Are you diabetic?, Chi-square: 4.264, p-value: 0.039	13 (100.00%)	0 (0.00%)	69 (68.32%)	32 (31.68%)
Section: General History, Smoker, Chi-square: 5.061, p-value: 0.024	12 (92.31%)	1 (7.69%)	56 (55.45%)	45 (44.55%)
Section: OMB Habitual, Accommodation BIN, Chi-square: 4.543, p-value: 0.033	2 (15.38%)	11 (84.62%)	1 (0.99%)	100 (99.01%)
Section: OMB Habitual, Confrontation, Chi-square: 6.946, p-value: 0.008	6 (46.15%)	7 (53.85%)	13 (12.87%)	88 (87.13%)
Section: OMB Habitual, NPC, Chi-square: 4.245, p-value: 0.039	7 (53.85%)	6 (46.15%)	23 (22.77%)	78 (77.23%)
Section: OMB Habitual, PD, Chi-square: 5.137, p-value: 0.023	9 (69.23%)	4 (30.77%)	33 (32.67%)	68 (67.33%)
Section: OMB New Rx, Cover test Distance, Chi-square: 5.774, p-value: 0.016	10 (76.92%)	3 (23.08%)	38 (37.62%)	63 (62.38%)
Section: OMB New Rx, Cover test Near, Chi-square: 5.774, p-value: 0.016	10 (76.92%)	3 (23.08%)	38 (37.62%)	63 (62.38%)
Section: OMB New Rx, Fixation Disparity Distance, Chi-square: 8.881, p-value: 0.003	7 (53.85%)	6 (46.15%)	15 (14.85%)	86 (85.15%)
Section: OMB New Rx, Fixation Disparity Near, Chi-square: 12.824, p-value: <0.000	8 (61.54%)	5 (38.46%)	15 (14.85%)	86 (85.15%)
Section: Anterior Exam 1, Right Eye Van Herick,	12 (92.31%)	1 (7.69%)	41 (40.59%)	60 (59.41%)

Chi-square: 10.390, p-value: 0.001				
Section: Anterior Exam 1, Left Eye Van Herick, Chi-square: 9.939, p-value: 0.002	12 (92.31%)	1 (7.69%)	42 (41.58%)	59 (58.42%)
Section: Anterior Exam 2, Shafers Sign, Chi-square: 4.631, p-value: 0.031	8 (61.54%)	5 (38.46%)	28 (27.72%)	73 (72.28%)
Section: IOP GAT, IOP Contact, Chi-square: 8.549, p-value: 0.003	11 (84.62%)	2 (15.38%)	38 (37.62%)	63 (62.38%)
Section: IOP NCT, IOP NCT Right Eye Reading 2, Chi-square: 9.959, p-value: 0.002	2 (15.38%)	11 (84.62%)	66 (65.35%)	35 (34.65%)
Section: IOP NCT, IOP NCT Left Eye Reading 3, Chi-square: 4.383, p-value: 0.036	2 (15.38%)	11 (84.62%)	51 (50.50%)	50 (49.50%)
Section: IOP NCT, IOP NCT Left Eye Reading 2, Chi-square: 9.959, p-value: 0.002	2 (15.38%)	11 (84.62%)	66 (65.35%)	35 (34.65%)
Section: IOP NCT, IOP NCT Left Eye Reading 1, Chi-square: 18.196, p-value: <0.000	2 (15.38%)	11 (84.62%)	78 (77.23%)	23 (22.77%)
Section: IOP NCT, IOP NCT Right Eye Reading 1, Chi-square: 18.196, p-value: <0.000	2 (15.38%)	11 (84.62%)	78 (77.23%)	23 (22.77%)
Section: IOP NCT, IOP NCT Right Eye Reading 3, Chi-square: 4.383, p-value: 0.036	2 (15.38%)	11 (84.62%)	51 (50.50%)	50 (49.50%)
Section: Visual Field Humphrey,	12 (92.31%)	1 (7.69%)	60 (59.41%)	41 (40.59%)

Fields Humphrey Right, Chi-square: 4.038, p-value: 0.044				
Section: Refraction, Current Rx, Chi-square: 8.149, p-value: 0.004	11 (84.62%)	2 (15.38%)	101 (100.00%)	0 (0.00%)
Section: Refraction, Subjective Findings, Chi-square: 5.624, p-value: 0.018	12 (92.31%)	1 (7.69%)	54 (53.47%)	47 (46.53%)

Table 4-6 - NTG vignette significant chi-square results

Guideline recommendations in the Normal Tension Glaucoma (NTG) vignette are in Appendix 19. The OMB tests were not relevant to the NTG vignette and are part of routine examination. More control group optometrists did not perform the test when compared to the intervention group. The Anterior Exam 1(Van Herick test) is an important test for this vignette and guidance on how to interpret the images were given. Greater proportion of intervention group optometrists performed this when compared to the control group **(n=12 (92.31%) vs n=42 (41.58%))**. IOP GAT is another test that is important for this vignette. Guidance was given on selecting the right IOP test. The optometrists were asked to select either IOP GAT or IOP NCT test and the guidance recommended that if the patient is to be referred based on IOP readings alone then IOP GAT offers better accuracy and therefore it is recommended. More number of intervention group optometrists **(n=11 (84.62%))** selected the IOP GAT test when compared to the control group optometrists **(n=38 (37.62%))**. Visual field Humphrey is another test that is recommended for NTG patients to determine the field of vision and guidance was provided on how to read the visual field reports. Greater proportion of intervention group optometrists **(n=12 (92.31%))** performed the Visual Field Humphrey test when compared to the control group **(n=60 (59.41%))**. Overall the guideline recommendation had some impact on guiding the optometrists towards performing certain key tests for the NTG vignette.

4.3.1.4 Wet AMD

The following section presents the total numbers of history/investigations answered and the statistically significant results of the history/investigations chi-square test of the AMD vignette. Wet AMD is a condition of the eye that causes the patient to lose the central vision. It is different from glaucoma where the peripheral vision is progressively reduced. Wet AMD is an acute condition and can quickly lead to loss of vision if not detected early.

Total number of history/investigations in the AMD vignette is **n=83**. The total number of history/investigations where there is a statistically significant difference in the answered/unanswered questions between the intervention and control group is **n=24(28.9%)**. The significant results have been outlined in Table 4-7. The remaining results have been outlined in Appendix 18.1.4.

Section, Question/Investigation, Chi-square, p-value, Significance (α= 0.05)	**Answered with guideline rec (Intervention)**	**Not answered with guideline rec (Intervention)**	**Answered without CDSS (Control)**	**Not answered without CDSS (Control)**
Section: Symptoms, Do you ever get any pain or discomfort relating to your eyes?,	13 (92.86%)	1 (7.14%)	54 (54.00%)	46 (46.00%)

Chi-square: 6.133, p-value: 0.013				
Section: Medical History, Are you diabetic?, Chi-square: 6.657, p-value: 0.010	14 (100.00%)	0 (0.00%)	61 (61.00%)	39 (39.00%)
Section: Medical History, Do you have high blood pressure?, Chi-square: 12.757, p-value: <0.000	14 (100.00%)	0 (0.00%)	45 (45.00%)	55 (55.00%)
Section: Ocular History, Have you ever had any other eye injuries/infections/surgeries?, Chi-square: 5.525, p-value: 0.019	13 (92.86%)	1 (7.14%)	56 (56.00%)	44 (44.00%)
Section: General History, Smoker, Chi-square: 5.522, p-value: 0.019	14 (100.00%)	0 (0.00%)	65 (65.00%)	35 (35.00%)
Section: General History, Driver, Chi-square: 4.744, p-value: 0.029	14 (100.00%)	0 (0.00%)	68 (68.00%)	32 (32.00%)
Section: OMB Habitual, Confrontation, Chi-square: 7.049, p-value: 0.008	10 (71.43%)	4 (28.57%)	31 (31.00%)	69 (69.00%)
Section: OMB Habitual, Motility, Chi-square: 40.064, p-value: <0.000	13 (92.86%)	1 (7.14%)	13 (13.00%)	87 (87.00%)
Section: OMB Habitual, Cover test Near, Chi-square: 4.888, p-value: 0.027	11 (78.57%)	3 (21.43%)	43 (43.00%)	57 (57.00%)
Section: OMB Habitual, PD, Chi-square: 31.589, p-value: <0.000	9 (64.29%)	5 (35.71%)	6 (6.00%)	94 (94.00%)
Section: OMB New Rx, Cover test Distance, Chi-square: 29.787, p-value: <0.000	11 (78.57%)	3 (21.43%)	12 (12.00%)	88 (88.00%)

Section: OMB New Rx, Cover test Near, Chi-square: 29.787, p-value: <0.000	11 (78.57%)	3 (21.43%)	12 (12.00%)	88 (88.00%)
Section: OMB New Rx, Fixation Disparity Distance, Chi-square: 15.140, p-value: <0.000	4 (28.57%)	10 (71.43%)	81 (81.00%)	19 (19.00%)
Section: OMB New Rx, Fixation Disparity Near, Chi-square: 11.258, p-value: 0.001	5 (35.71%)	9 (64.29%)	81 (81.00%)	19 (19.00%)
Section: Anterior Exam 1, Right Eye Van Herick, Chi-square: 21.669, p-value: <0.000	14 (100.00%)	0 (0.00%)	31 (31.00%)	69 (69.00%)
Section: Anterior Exam 1, Left Eye Van Herick, Chi-square: 20.852, p-value: <0.000	14 (100.00%)	0 (0.00%)	32 (32.00%)	68 (68.00%)
Section: Anterior Exam 2, Tears, Chi-square: 11.030, p-value: 0.001	10 (71.43%)	4 (28.57%)	24 (24.00%)	76 (76.00%)
Section: Ophthalmoscopy, Macula Left, Chi-square: 12.757, p-value: <0.000	14 (100.00%)	0 (0.00%)	45 (45.00%)	55 (55.00%)
Section: Colour Vision, Colour Vision City, Chi-square: 29.322, p-value: <0.000	5 (35.71%)	9 (64.29%)	0 (0.00%)	100 (100.00%)
Section: Colour Vision, Colour Vision Ishihara, Chi-square: 21.773, p-value: <0.000	4 (28.57%)	10 (71.43%)	0 (0.00%)	100 (100.00%)
Section: Colour Vision, Colour Vision Other, Chi-square: 14.439, p-value: <0.000	3 (21.43%)	11 (78.57%)	0 (0.00%)	100 (100.00%)
Section: IOP NCT, IOP NCT Right Eye Reading 4, Chi-square: 8.423, p-value: 0.004	5 (35.71%)	9 (64.29%)	77 (77.00%)	23 (23.00%)

Section: IOP NCT, IOP NCT Left Eye Reading 3, Chi-square: 9.535, p-value: 0.002	6 (42.86%)	8 (57.14%)	9 (9.00%)	91 (91.00%)
Section: IOP NCT, IOP NCT Right Eye Reading 3, Chi-square: 8.434, p-value: 0.004	6 (42.86%)	8 (57.14%)	10 (10.00%)	90 (90.00%)
Section: Visual Field Henson, Fields Henson, Chi-square: 5.813, p-value: 0.016	5 (35.71%)	9 (64.29%)	72 (72.00%)	28 (28.00%)

Table 4-7 - Wet AMD vignette significant chi-square results

The guideline recommendations in the AMD vignette are in Appendix 19. Greater proportion of the intervention group optometrists asked questions in the Symptoms **(n=13 (92.86%) vs n=54 (54.00%))**, Medical History **(n=14 (100.00%) vs n=61 (61.00%))**, Ocular History **(n=13 (92.86%) vs n=56 (56.00%))** and General History **(n=14 (100.00%) vs n=65 (65.00%))** when compared to the control group optometrists. This could be partly explained by guideline alerts flagging up Blurry vision as an abnormal value in the Symptoms section. The optometrists were encouraged to ask further questions about the condition. There was no difference between the two groups in performing tests for refraction. However, when the Current Rx-Refraction tests were performed the decision support alerted the optometrist to the reduced visual acuity in one eye and recommended additional tests such as Anterior Exam 1 and 2 and OMB tests be performed.

The result show that that tests such as OMB Habitual Confrontation **(n=10 (71.43%) vs n=31 (31.00%))**, OMB Habitual Motility **(n=13 (92.86%) vs n=13 (13.00%))**, OMB Habitual Cover Test Near **(n=11 (78.57%) vs n=43 (43.00%))** a greater proportion of intervention group optometrists performed the test. **N=94 (94.00%)** control group optometrists did not perform the OMB Habitual PD test compared to only n=5 (35.71%) in the intervention group. Similarly, **n=88 (88.00%)** of the control group optometrists did not perform the OMB New Rx Cover Test Distance and Near. However, greater proportion of control group optometrists did perform the OMB New Rx Fixation Disparity Distance and Near **(n=81 (81.00%))** when compared to the intervention group optometrists (n=4 (28.57%)).

A greater proportion of the intervention group optometrists **(n=14 (100.00%))** performed the Anterior Exam 1 (Van Herick test) when compared to the control group optometrists **(n=32 (32.00%))**. A greater proportion of the intervention group optometrists **(n=10 (71.43%))** performed the Anterior Exam 2-Tears test when compared to the control group **(n=24 (24.00%))**.

The results show that the intervention group optometrists **(n=14 (100.00%))** performed the Ophthalmoscopy Macula Left test more than the control group **(n=45 (45.00%))**.

The results show that when the optometrists were alerted to an abnormal value and asked to perform certain tests a significant number of optometrists did perform tests. The intervention optometrists were asked to check the OMB, Anterior and Posterior regions of the eye and they did this more when compared to the control group optometrists. The guideline recommendations therefore did have some impact in guiding and encouraging the optometrists to perform certain tests in the AMD vignette. As the optometrists were not penalised in any manner the optometrists choose the tests entirely based on their clinical judgement with the help of the guideline recommendations.

4.3.2 Summary of results - Diagnostic accuracy and appropriateness of management

Comparison between the intervention group and control group optometrist has been done in 2 stages. In the first stage, the number of correct/incorrect diagnoses in the intervention group and control group (before the CDSS provides its diagnostic recommendation) is compared. The optometrists are required to select a diagnosis from a drop-down list of diagnoses (See Figure 4-19). This step is done before the CDSS provides its diagnostic recommendation. The correct and incorrect numbers have been summarised in Appendix 20.

The OHT vignette was completed by 13 optometrists, Normal by 9 optometrists, NTG by 13 optometrists and Wet AMD by 14 optometrists in the intervention group.

The OHT vignette was completed by 100 optometrists, Normal by 102 optometrists, NTG by 101 optometrists, and Wet AMD by 100 optometrists in the control group.

The total number of correct diagnoses in the OHT vignette is **8 (61.54% - intervention group) vs 71 (71.00% - control group)**. The total number of correct diagnoses in the Normal vignette is **7 (77.78% - intervention group) and 84 (82.35% - control group)**. The difference in numbers is explained by the smaller sample size in the intervention group study.

The total number of correct diagnoses in the NTG vignette is **10 (76.92%% - intervention group) vs 64 (63.37% - control group)**. The total number of correct diagnoses (intervention vs control) in the Wet AMD vignette is **12 (85.71%% - intervention group) vs 71 (71.00% - control group).**

The results of the chi-square and fisher's tests (See section 4.3.5) show that there is no significant difference is the proportions between the intervention group and control group. This shows that with the help of the guideline recommendations alone the results of the intervention group and control group are similar.

In the second stage of analysis the total number of correct/incorrect diagnoses in the intervention and control group are compared after the CDSS has provided its diagnostic recommendation. The CDSS provides a list of diagnoses that are most likely to be present based on the data collected in the patient vignette. Each diagnosis has arguments supporting or negating the diagnosis and the weight are given representing the degree of support for each argument. This is dynamic, and changes based on the data that is collected or entered in the CDSS (See Figure 4-19). Once the diagnosis is selected the optometrist provides management recommendations.

Diagnostic accuracy and appropriateness of management is shown in Table 4-8. The results in Table 4-8 shows the impact of diagnostic decision support and the intervention group (With CDSS) is compared to the control group (Without CDSS). The results that are significant (light red cells) have been marked with * (Chi square test) and # (Fisher's test) to denote the tests that showed significance.

The study showed that diagnostic accuracy significantly improved with the help of the CDSS in the NTG and Wet AMD vignette ($P<0.05$). In the OHT and Normal vignettes there was no change in diagnostic accuracy when compared to the control group of optometrists (See section - 4.3.3 for detailed information). There was no improvement in referral decision in any of the 4 vignettes. The optometrists who were presented with the NTG vignette showed significant improvement in giving accurate re-examination intervals and ordering supplementary tests (Fields and Goldmann Tonometry) (See section - 4.3.4 for detailed information). Similarly, optometrists who were given the Wet AMD vignette showed improvement in ordering supplementary tests (Fields, Goldmann Tonometry and Dilation). The OHT vignette only showed improvement in ordering of Fields supplementary test. There was a significant difference in the proportion of optometrists who did not order Goldmann Tonometry test for the Normal vignette and the control group did much better (**93%**) in this area.

	OHT		Normal		NTG		Wet AMD	
	Without CDSS N=100	***With CDSS N=13***	***Without CDSS N=102***	***With CDSS N=9***	***Without CDSS N=101***	***With CDSS N=13***	***Without CDSS N=100***	***With CDSS N=14***
Diagnosis - n (%) Correct	71 (71%)	8 (67%)	84 (82%)	9 (100%)	64 (63%)	13 *# (100%)	71 (71%)	13 # (93%)
Management referral - n (%) Correct	48 (48%)	4 (33%)	93 (91%)	8 (89%)	77 (76%)	13 (100%)	88 (88%)	14 (100%)
Management re-examination Interval n (%) Correct	59 (59%)	8 (67%)	90 (88%)	7 (78%)	75 (74%)	13 # (100%)	82 (82%)	14 (100%)
Supplementary tests - Fields-n (%) Correct	30 (30%)	8 *# (67%)	77 (75%)	4 (44%)	20 (20%)	9 *# (69%)	5 (5%)	4 *# (29%)
Supplementary tests - Goldmann Tonometry - n (%) Correct	36 (36%)	8 (67%)	95 *# (93%)	6 (67%)	16 (16%)	9 *# (69%)	1 (1%)	2 *# (14%)
Supplementary tests - Dilation-n (%) Correct	10 (10%)	2 (17%)	98 (96%)	8 (89%)	12 (12%)	4 (31%)	22 (22%)	8 *# (57%)

Table 4-8 - Significance tests summary (α = 0.05) (Chi square * / Fishers test #)

4.3.3 Diagnosis

This section outlines the total numbers of correct and incorrect diagnosis results for the intervention group and the control group. The significance test results have been summarised in section - 4.3.5

4.3.3.1 OHT

OHT is a condition where the pressure inside the eye (Intraocular pressure or IOP) is greater than normal. People with OHT are considered to be at risk of glaucoma and therefore needs to be regularly monitored by the optometrist for signs of glaucoma (Mowatt *et al.*, 2008). Its correct diagnosis is therefore important to reduce unwanted referrals coming into secondary care and misdiagnosis increase the risk of the patient becoming blind due to glaucoma.

OHT	Correct	Incorrect
With Diagnostic CDSS	8 (61.54%)	5 (38.46%)
Without CDSS	71 (71.00%)	29 (29.00%)

Table 4-9 - OHT diagnosis correct vs incorrect numbers

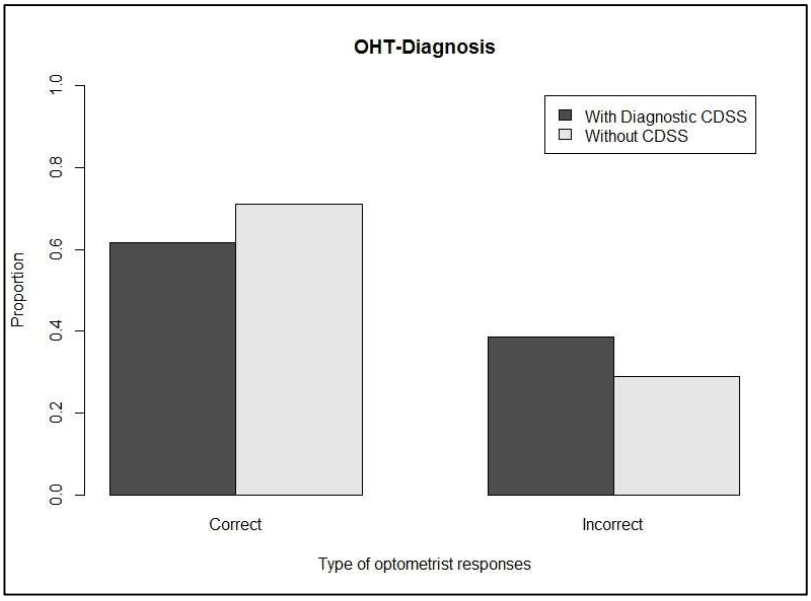


Figure 4-26 - OHT diagnosis correct vs incorrect proportion

Table 4-9 shows the total number of correct and incorrect diagnosis with the help of guideline recommendations alone and with the help of the diagnostic recommendations. These numbers were compared to the control group where the sample size was much larger (**N=100**). The different proportions are visualized in Figure 4-26 which shows the proportion of the correct and incorrect with the help of the diagnostic recommendations. The early guideline recommendations and late diagnostic recommendations in the intervention group optometrists did not have a statistically significant impact showing improvement in diagnosis over the control group according to both chi square and fisher's tests.

4.3.3.2 Normal

The optometrist should be able to correctly identify the Normal cases as sometimes certain observations can confuse the optometrist. The CDSS should be able to correctly identify the Normal cases. Misdiagnosis in Normal cases can result in unwanted clinical tests being performed, unwanted referral to hospitals and can unnecessarily increase the workload in secondary care.

Normal	Correct	Incorrect
With Diagnostic CDSS	9 (100.00%)	0 (0.00%)
Without CDSS	84 (82.35%)	18 (17.65%)

Table 4-10 - Normal diagnosis correct vs incorrect numbers

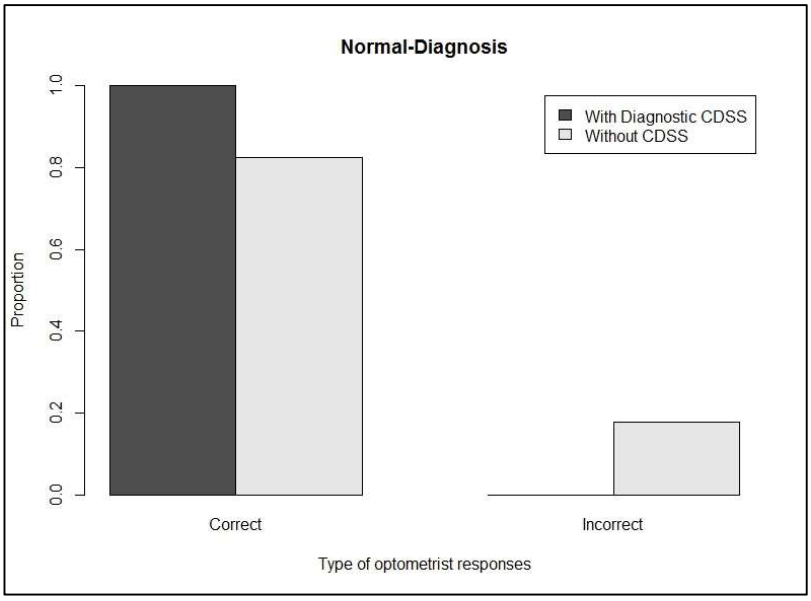


Figure 4-27 - Normal diagnosis correct vs incorrect proportion

The early guideline recommendations and late diagnostic recommendations in the intervention group optometrists did not have a statistically significant impact showing improvement in diagnosis over the control group (Table 4-10) (Figure 4-27) according to both chi square and fisher's tests.

4.3.3.3 Normal Tension Glaucoma

NTG is a type of glaucoma where the pressure inside the eye (IOP) is not raised and the lies within the normal range. However, there are signs of damage to the nerves of the eye that can be found during examination of the eye. Since the IOP is not raised NTG is one condition where is a chance for misdiagnosis which can lead to more severe forms of glaucoma.

Normal Tension Glaucoma	Correct	Incorrect
With Diagnostic CDSS	13 (100.00%)	0 (0.00%)
Without CDSS	64 (63.37%)	37 (36.63%)

Table 4-11 - NTG diagnosis correct vs incorrect numbers

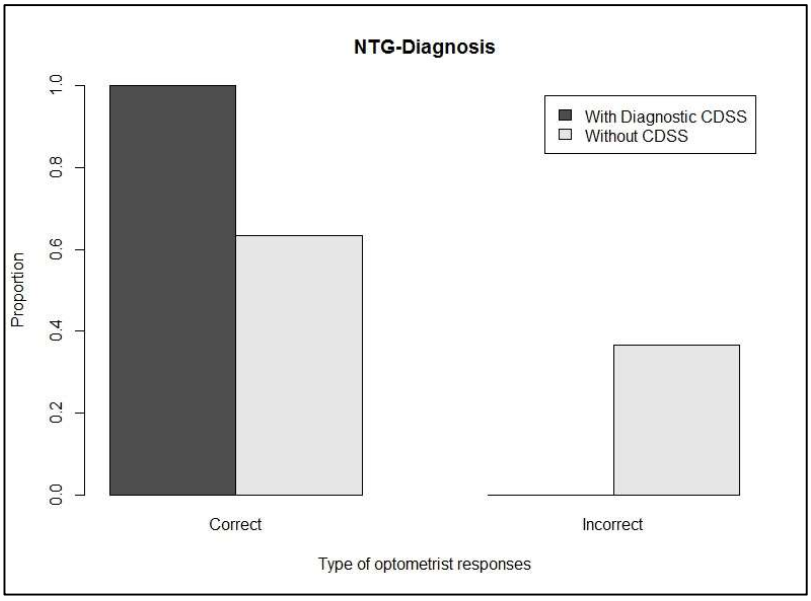


Figure 4-28 - NTG diagnosis correct vs incorrect proportion

The late diagnostic recommendations in the intervention group did have a statistically significant impact showing improvement in diagnosis over the control group (Figure 4-28) according to both chi square and fisher's tests. The early guideline recommendations did not have a statistically significant impact showing diagnosis improvement over the control group (Table 4-11) according to both chi square and fisher's tests.

4.3.3.4 Wet AMD

Wet AMD is a condition of the eye that causes the patient to lose the central vision. It is different from glaucoma where the peripheral vision is progressively reduced. Wet AMD is an acute condition and can quickly lead to loss of vision if not detected early. This condition was included in the vignettes so that there are a variety of vignettes for the optometrists to diagnose rather than all vignettes being restricted to glaucoma related conditions.

Wet AMD	Correct	Incorrect
With Diagnostic CDSS	13 (92.86%)	0 (0.00%)
Without CDSS	71 (71.00%)	29 (29.00%)

Table 4-12 - Wet AMD diagnosis correct vs incorrect numbers

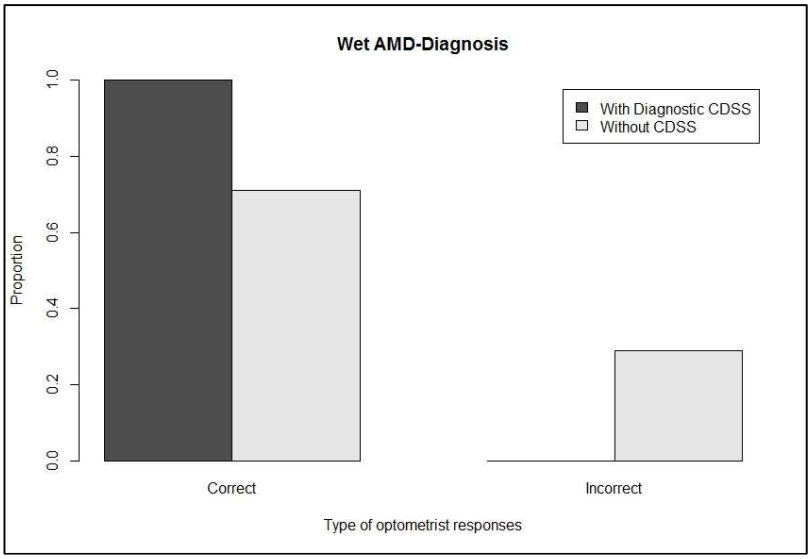


Figure 4-29 - Wet AMD diagnosis correct vs incorrect proportion

The late diagnostic recommendations in the intervention group did have a statistically significant impact showing diagnosis improvement over the control group (Figure 4-29) according to fisher's tests. The early guideline recommendations did not have a statistically significant impact showing improvement in diagnosis over the control group (Table 4-12) according to both chi square and fisher's tests

4.3.4 Management Recommendations

There are 5 types of management outcomes that were recorded as part of this study. These are:

- Referral
- Re-examination interval
- Supplementary tests
    - Fields
    - Goldmann tonometry
    - Dilation

Referral decisions are made by the optometrist and include recommendations as to where the patient needs to be referred for further management if needed or not referred and managed in the community. For example, the referral for OHT diagnosis is like the Normal diagnosis where the patient does not have to be referred to an ophthalmologist or GP by the optometrist. However, the patient will need to be recalled for re-examination at regular intervals for continuous monitoring to prevent deterioration into glaucoma.

Re-examination interval is the time-period before recalling the patient for the next review. Supplementary tests are those additional tests that need to be advised (if needed) for each diagnosis. There are 3 types of supplementary tests that can be ordered for each of the diagnosis when managing the patient. In OHT diagnosis it is recommended that Fields, Goldmann tonometry and Dilation be performed to confirm the diagnosis before making a final decision. Fields (Visual Fields) are a type of routine diagnostic eye test that test the range of vision both vertically and horizontally. Goldmann tonometry is a type of test that tests the IOP of the eye. It is recommended that the test be performed if not already done. Dilation of the eye allows the optometrist to better visualize the back of the eye.

The decisions recorded by the optometrists with the help of the CDSS guideline recommendations were recorded and compared to the control group optometrists. See Appendix 4 for the correct answers for each vignette as determined by the clinical domain experts (Murdoch, Lawrenson and Myint, 2015). The following sections (4.3.4.1 - 4.3.4.4) detail the management decisions that were found to be statistically significant for each diagnosis. All the management results for each vignette (significant and insignificant) have been detailed in Appendix 4. The significance test results have been summarised in section - 4.3.6.

4.3.4.1 OHT

4.3.4.1.1 OHT – Supplementary Tests – Fields

The CDSS was shown to have a statistically significant impact and improved the ability of the optometrists to order Fields supplementary test in the OHT vignette (Table 4-13) (Figure 4-30) according to both chi square and fisher's tests.

OHT – Supplementary Tests – Fields	Correct	Incorrect
With CDSS	8 (61.54%)	4 (30.77%)
Without CDSS	30 (30.00%)	70 (70.00%)

Table 4-13 - OHT – Supplementary Tests – Fields correct vs incorrect numbers

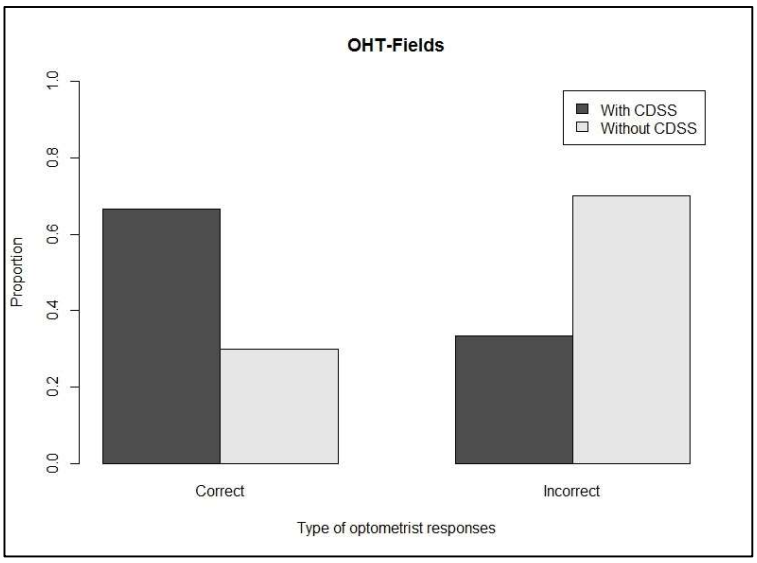


Figure 4-30 - OHT – Supplementary Tests – Fields correct vs incorrect proportion

4.3.4.2 Normal

4.3.4.2.1 Normal-Supplementary Tests – Goldmann Tonometry

In the Normal vignette, the CDSS was shown to have a statistically significant negative impact on the ability of optometrists to order Goldmann tonometry supplementary test when compared to the control group (Table 4-14)(Figure 4-31) according to both chi square and fisher's tests. The performance of the intervention group optometrists was worse than the control group optometrists.

Normal-Supplementary Tests – Goldmann Tonometry	Correct	Incorrect
With CDSS	6 (66.67%)	3 (33.33%)
Without CDSS	95 (93.14%)	7 (6.86%)

Table 4-14 - Normal-Supplementary Tests – Goldmann Tonometry correct vs incorrect numbers

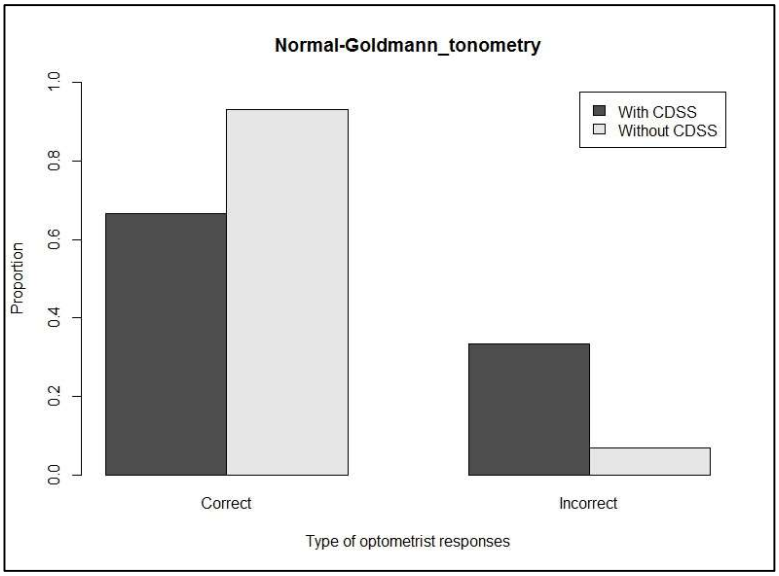


Figure 4-31 - Normal-Supplementary Tests – Goldmann Tonometry correct vs incorrect proportion

4.3.4.3 NTG

4.3.4.3.1 NTG – Re-examination Interval

The patient should be recalled every 3-12 months as needed for regular repeat eye examinations in the NTG vignette. The CDSS was shown to have a statistically significant impact and improved the ability of the optometrists in the intervention group to recall the patients when compared to the control group (Table 4-15)(Figure 4-32) according to both chi square and fisher's tests.

NTG – Re-examination Interval	Correct	Incorrect
With CDSS	13 (100.00%)	0 (0.00%)
Without CDSS	75 (74.26%)	26 (25.74%)

Table 4-15 - NTG – Re-examination Interval correct vs incorrect numbers

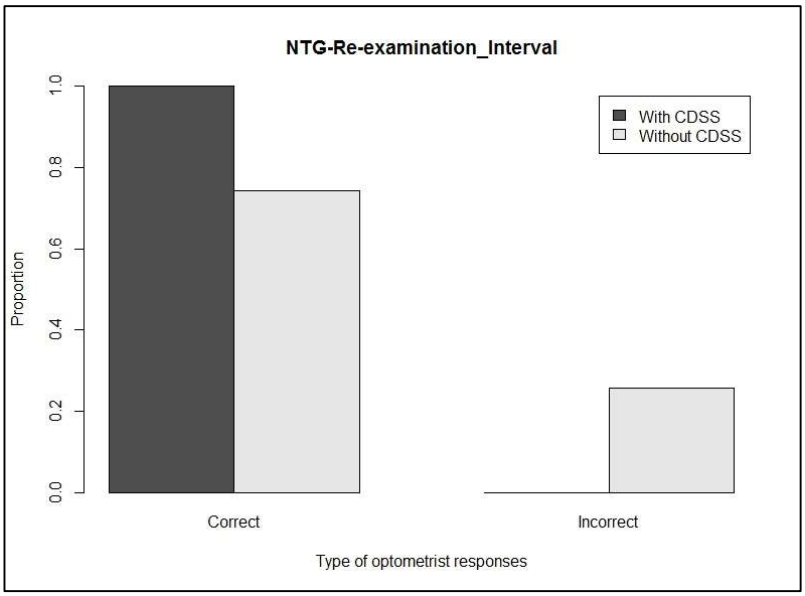


Figure 4-32 - NTG – Re-examination Interval correct vs incorrect proportion

4.3.4.3.2 NTG – Supplementary Tests – Fields

The optometrists in the intervention group performed statistically significantly better than the control group in the NTG vignette ordering Fields supplementary test for the NTG vignette (Table 4-16) (Figure 4-33) according to both chi square and fisher's tests.

NTG – Supplementary Tests – Fields	Correct	Incorrect
With CDSS	9 (69.23%)	4 (30.77%)
Without CDSS	20 (19.80%)	81 (80.20%)

Table 4-16 - NTG – Supplementary Tests – Fields correct vs incorrect numbers

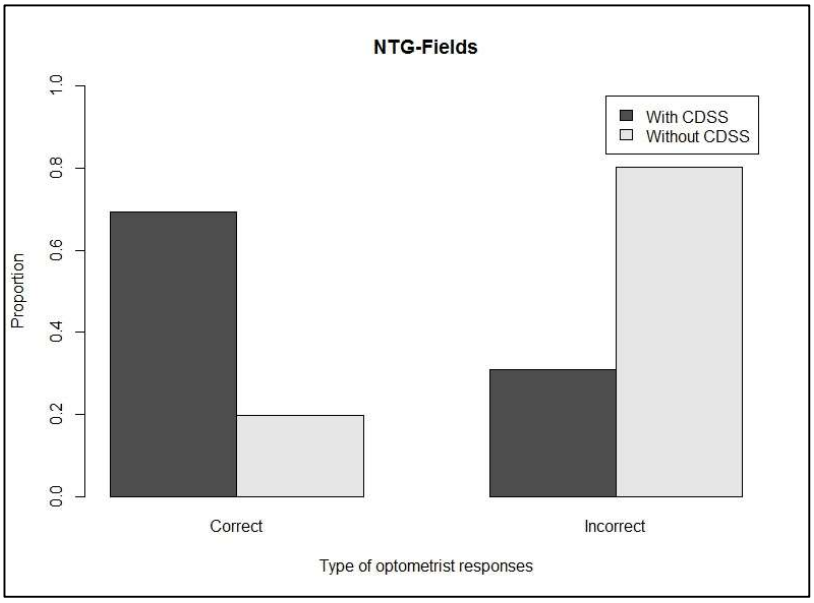


Figure 4-33 - NTG – Supplementary Tests – Fields correct vs incorrect proportion

4.3.4.3.3 NTG – Supplementary Tests – Goldmann Tonometry

The optometrists in the intervention group performed statistically significantly better than the control group in ordering Goldmann tonometry supplementary test for the NTG vignette (Table 4-17) (Figure 4-34) according to both chi square and fisher's tests.

NTG – Supplementary Tests – Goldmann Tonometry	Correct	Incorrect
With CDSS	9 (69.23%)	4 (30.77%)
Without CDSS	16 (15.84%)	85 (84.16%)

Table 4-17 - NTG – Supplementary Tests – Goldmann Tonometry correct vs incorrect numbers

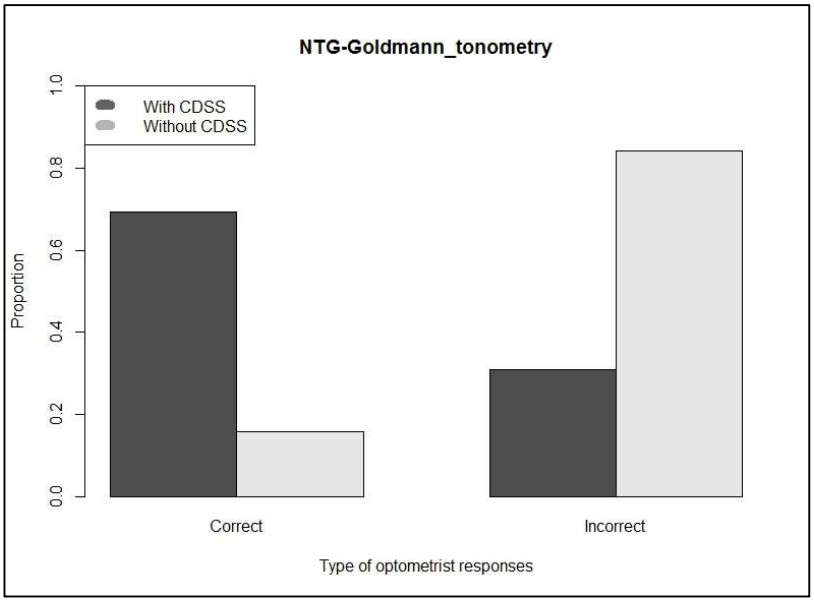


Figure 4-34 - NTG – Supplementary Tests – Goldmann Tonometry correct vs incorrect proportion

4.3.4.4 Wet AMD

4.3.4.4.1 Wet AMD – Supplementary Tests – Fields

The total number of incorrect answers was higher than the correct in the Wet AMD vignette in both intervention and control group when ordering Fields supplementary tests (Table 4-18). However, the intervention group performed better than the control group (Figure 4-35). The difference between the intervention group and the control group was found to be statistically significant according to both chi square and fisher's tests.

Wet AMD – Supplementary Tests – Fields	Correct	Incorrect
With CDSS	4 (28.57%)	10 (71.43%)
Without CDSS	5 (5.00%)	95 (95.00%)

Table 4-18 - Wet AMD – Supplementary Tests – Fields correct vs incorrect numbers

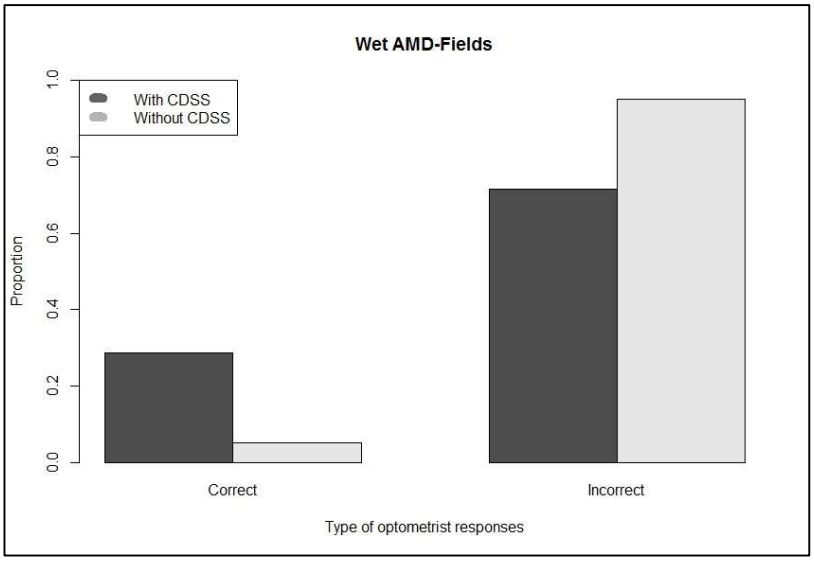


Figure 4-35 - Wet AMD – Supplementary Tests – Fields correct vs incorrect proportion

4.3.4.4.2 Wet AMD – Supplementary Tests – Goldmann Tonometry

The incorrect answers for ordering Goldmann Tonometry supplementary tests were increased in both the control and intervention groups (Table 4-19)(Figure 4-36). However, the intervention group performed statistically significantly better than the control group in ordering Goldmann tonometry supplementary tests according to both chi square and fisher's tests.

Wet AMD – Supplementary Tests – Goldmann Tonometry	Correct	Incorrect
With CDSS	2 (14.29%)	12 (85.71%)
Without CDSS	1 (1.00%)	99 (99.00%)

Table 4-19 - Wet AMD – Supplementary Tests – Goldmann Tonometry correct vs incorrect numbers

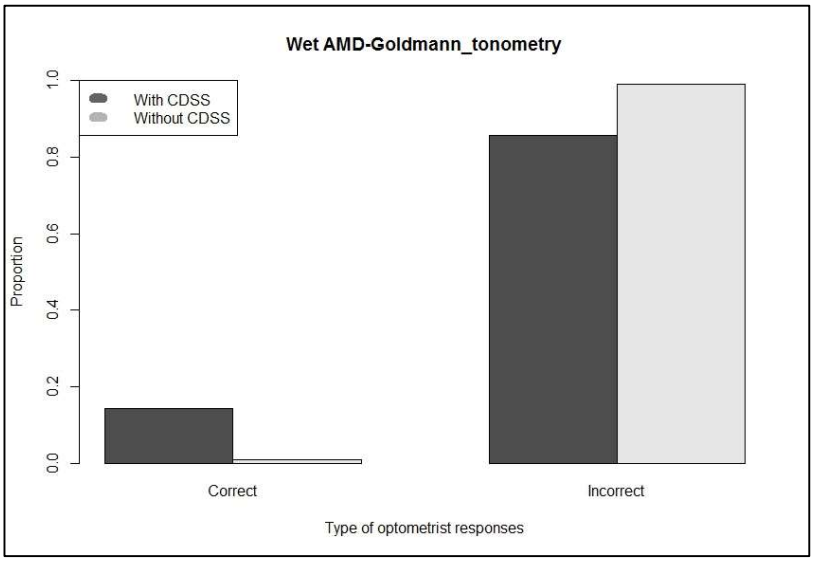


Figure 4-36 - Wet AMD – Supplementary Tests – Goldmann Tonometry correct vs incorrect proportion

4.3.4.4.3 Wet AMD – Supplementary Tests – Dilation

The intervention group optometrists performed significantly better than the control group in ordering Dilation supplementary tests for the Wet AMD vignette (Table 4-20) (Figure 4-37) according to both chi square and fisher's tests.

Wet AMD – Supplementary Tests – Dilation	Correct	Incorrect
With CDSS	8 (57.14%)	6 (42.86%)
Without CDSS	22 (22.00%)	78 (78.00%)

Table 4-20 - Wet AMD – Supplementary Tests – Dilation correct vs incorrect numbers

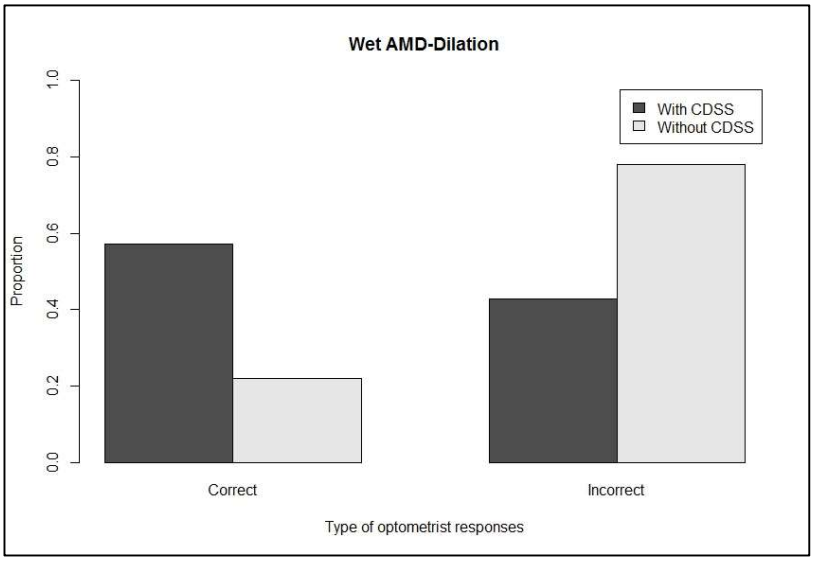


Figure 4-37 - Wet AMD – Supplementary Tests – Dilation correct vs incorrect proportion

4.3.5 Significance Tests-Diagnosis

4.3.5.1 Chi-square test

In this section, Pearson's chi-square test with Yates continuity correction is outlined. The chi-square test is performed to check if the difference in correct and incorrect numbers in the control and intervention groups has risen due to chance or not. (See Appendix 16 for raw data)

Diagnosis	Intervention group Optometrist compared to Control group	$X^2 test$ (α= 0.05) Critical value = 3.841	Significant Differences
OHT	Before Diagnostic CDSS	X^2= 0.1431, df = 1, p-value = 0.7052	✗
	With Diagnostic CDSS	X^2= 0.1431, df = 1, p-value = 0.7052	✗
Normal	Before Diagnostic CDSS	X^2=0, df = 1, p-value = 1	✗
	With Diagnostic CDSS	X^2 =0.8193, df = 1, p-value = 0.3654	✗
NTG	Before Diagnostic CDSS	X^2=0.4295, df = 1, p-value = 0.5123	✗
	With Diagnostic CDSS	$X^2$5.4787, df = 1, p-value = 0.01925	✓
Wet AMD	Before Diagnostic CDSS	X^2=0.7026, df = 1, p-value = 0.4019	✗
	With Diagnostic CDSS	X^2=3.6653, df = 1, p-value = 0.05556	✗

Table 4-21 - Diagnosis chi square test (df = Degrees of freedom)

The test shows that the CDSS only has a significant positive impact (i.e. improved the diagnosis) on the NTG diagnosis ($p<0.05$) (Table 4-21). The Wet AMD diagnosis is very close though not significant. Before diagnostic decision support there was no statistically significant difference between the control group and the intervention group. When the diagnostic recommendations were considered there was an improvement for the NTG vignette.

4.3.5.1 Fisher's test

The Fisher's test results (Table 4-22) shows that the diagnostic recommendations had a positive impact and helped the intervention group optometrist improve their diagnosis in both NTG and Wet AMD diagnosis ($p<0.05$). However, before diagnostic CDSS do not show any difference in diagnosis in any vignettes when compared to the control group. (See Appendix 16 for raw data)

Diagnosis	Intervention group Optometrist compared to Control group	95 % confidence interval	Fishers Test (α= 0.05)	Significant Differences
OHT	Before Diagnostic CDSS	0.1719065 -- 2.7712269	p-value = 0.5268	✗
	With Diagnostic CDSS	0.1719065 -- 2.7712269	p-value = 0.5268	✗
Normal	Before Diagnostic CDSS	0.1280912 -- 8.0060129	p-value = 0.6636	✗
	With Diagnostic CDSS	0.3813777 -- Inf	p-value = 0.3504	✗
NTG	Before Diagnostic CDSS	0.4544748 -- 11.5210407	p-value = 0.5381	✗
	With Diagnostic CDSS	1.630925 -- Inf	p-value = 0.008774	✓
Wet AMD	Before Diagnostic CDSS	0.4939871 -- 23.7476188	p-value = 0.3445	✗
	With Diagnostic CDSS	1.143497-- Inf	p-value = 0.02031	✓

Table 4-22 - Diagnosis Fisher's test

4.3.6 Significance Tests-Management

4.3.6.1 Chi square test

The chi-square test for significance in the management section of the decision-making shows that the CDSS has been effective in improving ordering of Fields supplementary tests in OHT, NTG and Wet AMD (Table 4-23). Even though the CDSS was not effective in improving the diagnosis of OHT if overall the CDSS can help the optometrists order the supplementary Fields test accurately then that would detect the absence/presence of any pathology during the referral stage and reduce the chance of misdiagnosis.

In the Normal diagnosis, the optometrists in the intervention group was found to significantly worse than the control group in ordering Goldmann tonometry tests. In the Normal diagnosis, there is no need for supplementary tests. The reason for this can be partially explained by the nature of the case represented in the Normal vignette. The optometrists may order the supplementary tests just in case they missed any significant finding during routine examination or if they are unsure of the findings. This could explain why the optometrists would order the supplementary test if given the option just to be on the safe-side. (See Appendix 16 for raw data)

Diagnosis	Management - Intervention group compared to Control group	$X^2 test$ (α= 0.05)	Significant Differences
OHT	Referral	X^2=0.4308, df = 1, p-value = 0.5116	✗
	Re-examination interval	X^2=0.0401, df = 1, p-value = 0.8412	✗
	Supplementary Tests – Fields	X^2=4.8942, df = 1, p-value = 0.02695	✓
	Supplementary Tests – Goldmann Tonometry	X^2=3.0366, df = 1, p-value = 0.08141	✗
	Supplementary Tests – Dilation	X^2=0.0448, df = 1, p-value = 0.8324	✗
Normal	Referral	X^2=0, df = 1, p-value = 1	✗
	Re-examination interval	X^2=0.146, df = 1, p-value = 0.7023	✗
	Supplementary Tests – Fields	X^2=2.6208, df = 1, p-value = 0.1055	✗
	Supplementary Tests – Goldmann Tonometry	X^2=4.2088, df = 1, p-value = 0.04021	✓
	Supplementary Tests – Dilation	X^2=0.025153, df = 1, p-value = 0.874	✗
NTG	Referral	X^2=2.6138, df = 1, p-value = 0.1059	✗
	Re-examination interval	X^2=2.9964, df = 1, p-value = 0.08345	✗
	Supplementary Tests – Fields	X^2=12.344, df = 1, p-value = 0.0004423	✓
	Supplementary Tests – Goldmann Tonometry	X^2=16.184, df = 1, p-value = 5.748e-05	✓
	Supplementary Tests – Dilation	X^2=2.02, df = 1, p-value = 0.1552	✗
Wet AMD	Referral	X^2=1.7919, df = 1, p-value = 0.1807	✗
	Re-examination interval	X^2=1.6462, df = 1, p-value = 0.1995	✗
	Supplementary Tests – Fields	X^2=6.422, df = 1, p-value = 0.01127	✓
	Supplementary Tests – Goldmann Tonometry	X^2=4.0692, df = 1, p-value = 0.04367	✓
	Supplementary Tests – Dilation	X^2=6.1144, df = 1, p-value = 0.01341	✓

Table 4-23 - Management chi square test

4.3.6.1 Fisher's test

The Fisher's test results (Table 4-24) were similar to the chi-square test with the additional improvement in re-examination interval in NTG diagnosis for the intervention group when compared to the control group. (See Appendix 16 for raw data)

Diagnosis	Intervention group compared to Control group	95 % confidence interval	Fishers Test (α= 0.05)	Significant Differences
OHT	Referral	0.1126035--2.1916062	p-value = 0.3766	✗
	Re-examination interval	0.3433793--6.7143945	p-value = 0.7594	✗
	Supplementary Tests – Fields	1.128325--22.489322	p-value = 0.02041	✓
	Supplementary Tests – Goldmann Tonometry	0.8687381--17.0795453	p-value = 0.05904	✗
	Supplementary Tests – Dilation	0.167855--10.379354	p-value = 0.6152	✗
Normal	Referral	0.08526982--38.12310901	p-value = 0.5867	✗
	Re-examination interval	0.07645199--5.15524514	p-value = 0.3168	✗
	Supplementary Tests – Fields	0.04833037 --1.32841955	p-value = 0.05835	✗
	Supplementary Tests – Goldmann Tonometry	0.02484175--1.13496490	p-value = 0.0336	✓
	Supplementary Tests – Dilation	0.02786504--18.06897043	p-value = 0.3501	✗
NTG	Referral	0.8652904--Inf	p-value = 0.06693	✗
	Re-examination interval	0.9663487--Inf	p-value = 0.03753	✓
	Supplementary Tests – Fields	2.208887--43.635272	p-value = 0.0005166	✓
	Supplementary Tests – Goldmann Tonometry	2.82374--57.90175	p-value = 0.0001241	✓
	Supplementary Tests – Dilation	0.6332234--14.1320047	p-value = 0.08481	✗
Wet AMD	Referral	0.6494739--Inf	p-value = 0.1213	✗
	Re-examination interval	0.520873--18.745904	p-value = 0.1055	✗
	Supplementary Tests – Fields	1.255113--41.035676	p-value = 0.01292	✓
	Supplementary Tests – Goldmann Tonometry	0.7663269--975.1985847	p-value = 0.03936	✓

	Supplementary Tests – Dilation	1.264146--18.157418	p-value = 0.009226	✓

Table 4-24 - Management Fisher's test

4.3.7 Usability survey results

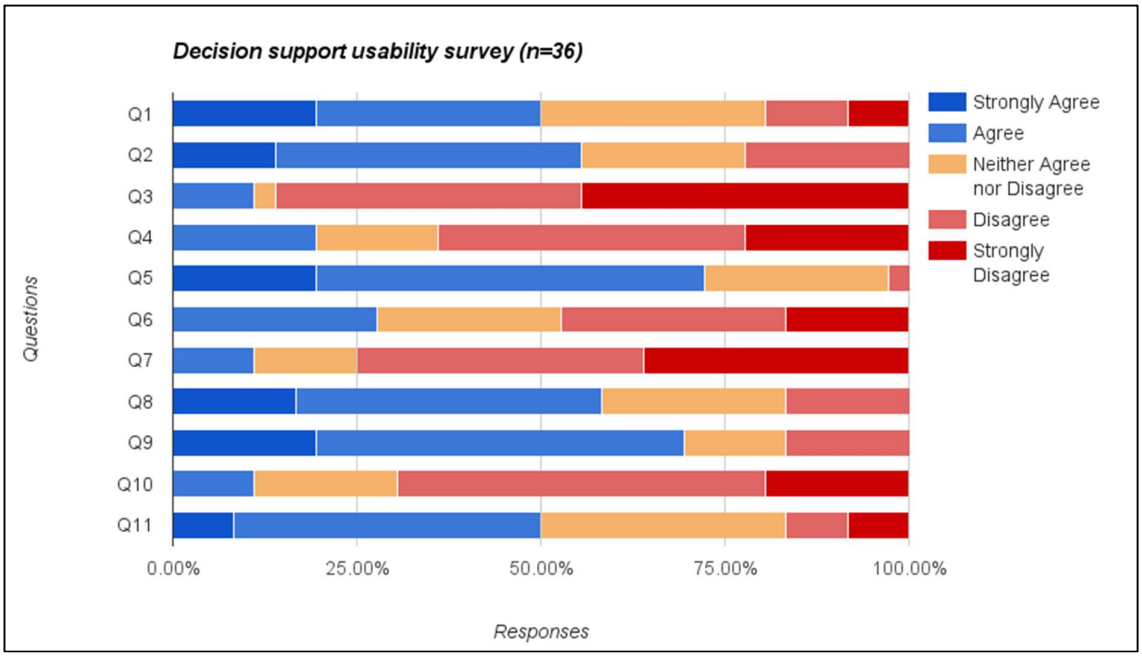


Figure 4-38 - Usability survey results

Q1. I am very likely to use the decision support system in my practice.
Q2. I thought the system was easy to use.
Q3. I think that I would need the support of a technical person to be able to use this system.
Q4. I found the system unnecessarily complex.
Q5. I feel that most people would learn to use this system very quickly.
Q6. I found the system very cumbersome to use.
Q7. I needed to learn a lot of things before I could get going with this system.
Q8. I felt very confident using the system.
Q9. I found the various clinical recommendations provided by the system relevant.
Q10. I thought there was too much inconsistency in this system.
Q11. This decision support system will help me improve my clinical decision making.

The survey results (Figure 4-38) show that most of the optometrists found the CDSS easy to use and did not need the support of a technical person to complete the tasks. They felt the CDSS in its present form was something they could learn to use quite easily. Ease of use is a key factor in helping improve adoption of CDSS applications for routine clinical use (Fossum *et al.*, 2011). Most clinicians felt that the CDSS recommendations were relevant and helped improve their clinical decision making.

4.4 Discussion

This study investigated the development of a CDSS that delivered diagnostic recommendations based on argumentation-based models. CDSSs based on other artificial intelligence model such as case-based reasoning, Bayesian probability-based reasoning, fuzzy logic reasoning and machine learning can deliver diagnostic recommendations. These systems require the availability of well-structured clinical data to

provide diagnostic recommendations. Some of these models such as machine learning-based models are developed by learning from historical clinical data. They present the diagnostic recommendation in a manner that is not always intuitive to the clinician.

The literature suggested the need for diagnostic CDSSs that can integrate information from disparate sources and provide diagnostic recommendations at different stages of the clinical encounter. Even if the information can be integrated from disparate sources there is a need to collate this information and present the information to the clinician in a format that is intuitive and easy to understand. Clinicians and other healthcare practitioners generally prefer systems that can provide better explanations for the patient's condition. They can then compare this information with their own arguments for each diagnosis and then make a final decision. The CDSS should be able to provide the arguments supporting or negating a diagnosis. The CDSS should be able to provide a diagnostic recommendation in the presence of incomplete information and should be capable of defeasible reasoning. In defeasible reasoning the system should be able to retract the decision in the light of current information and provide a new diagnostic recommendation with the updated knowledge.

Argumentation-based models have been shown to provide these capabilities and was therefore chosen as the basis for generating early diagnostic recommendations (See Chapter 5). AT based CDSSs are increasingly being used to supporting clinical decision making in areas such as referral management, risk assessment, treatment planning etc. (Emery *et al.*, 1999; Fox, Patkar and Thomson, 2006b; Sutton, Taylor and Earle, 2006). However, studies so far have not evaluated the role played by argumentation-based diagnostic CDSSs in reducing diagnostic error. This study developed a prototype diagnostic CDSS that is based on the PROforma language. CDSS applications based on PROforma used AT as the basis for generating decision recommendations. Therefore, it was decided to initially develop a CDSS based on PROforma tools and test is with real clinicians (optometrists) before developing the ontology-based model for early diagnostic recommendations.

The study looked at different aspects of the CDSS on the optometrist's performance. Guideline recommendations were provided at different stages of the clinician encounter. Some of these guideline recommendations alerted the optometrists towards abnormal values and provided guidance on what steps need to be taken next. Other guideline recommendations provided guidance on how to interpret images or clinical test results (See Appendix 19 for all guideline recommendations). The optometrists were asked to provide their diagnosis and the CDSS would then provide its diagnostic recommendations and arguments supporting and negating each diagnosis.

The results of the study looking at the effect of guideline recommendations show that the guideline recommendations have some impact on the information gathering of the optometrists. OHT and Normal vignettes are characterised by having not many abnormal values. The Normal vignette did not have any abnormal values. OHT is a condition where there is elevated IOP, but no significant pathology exists. The optometrists should not refer the patient to secondary care and should monitor the OHT patient in the community. A greater proportion of intervention group optometrists have performed tests such Van Herick test that could potentially identify cases of OHT. The guideline recommendations only had a minimal impact on the optometrists in the Normal vignette. No abnormal values were present. So, vignette specific guideline recommendations that could help orient the optometrist towards appropriate tests were not present.

In the NTG and Wet AMD vignettes a pathology was present and guideline recommendations alerted the optometrist to these abnormal values. In the NTG vignette the optometrists were given guidance that suggested that the IOP GAT test offered better accuracy for IOP readings. IOP GAT is not routinely performed in optometry practices due to issues related to costs. IOP NCT is more routinely performed however it offers less accurate readings. NTG patients have normal IOP but has other pathology that is causing problems in vision. Tests such as Visual Fields Humphrey provide a better picture of the patient's condition. Guidance was provided that recommended that the Visual Fields examination be performed as well. Greater proportion of the intervention group optometrists performed this Visual Fields test. In the

Wet AMD vignette the symptoms showed abnormal values. The optometrists were recommended to ask further questions about the patient's condition and medical history.

When the optometrist performed the refraction tests they were alerted to the reduced visual acuity in one eye and recommended tests such as OMB, Anterior Exam and Macula (Posterior Exam-Ophthalmoscopy). The intervention group optometrists performed these tests more when compared to the control group optometrists. When the optometrists are alerted to the presence of an abnormal value and provided guidance that recommended asking certain tests and investigations the optometrists responded accordingly. All vignettes were of patients aged >40. Guidance was provided that recommended that the 'Family History of Glaucoma' be enquired to check for positive family history of glaucoma. No significant difference in the proportions were seen in all the vignettes for this question. This could suggest that the guidelines may only have an impact on information gathering in non-routine questions. Optometrists are trained to ask a set of questions as part of the routine examination and guideline may not have an impact on such questions, such as asking all patients aged>40 about family history of glaucoma. More studies are needed in this area to fully understand the nature of guideline recommendations and their impact on information gathering by optometrists.

The results show that the diagnostic CDSS was not able to show significant improvement in the diagnostic and management for the OHT vignette. However, it could demonstrate improvement in 2 potentially sight threatening conditions NTG and Wet AMD. The 4 diagnosis candidates for the OHT vignette are - OHT, POAG, Normal-repeat tests and Normal-No referral. OHT and Primary Open Angle Glaucoma (POAG) had similar arguments that were either supporting or negating the diagnosis. If the optometrist made a mistake in grading the images the CDSS would rank the POAG diagnosis higher than the OHT diagnosis. Instructions on how to properly grade the images were provided to the intervention group optometrists (Figure 4-18). The arguments for OHT candidate can be seen in Table 4-3 and the corresponding arguments for the POAG candidate can be seen in Appendix 2 – section 2.1.2. The arguments for differentiating between the OHT and POAG candidate included IOP, Visual Fields (Humphrey), Visual Fields (Henson), Optic disc status and Van Herick. Excluding IOP all the arguments are based on interpretation of the images. The weights for IOP argument are the same for OHT (Table 4-3) and POAG (Appendix 2 – section 2.1.2). So, the ranking of the diagnosis in the OHT vignette will be dependent on the quality of grading done by the optometrists.

If the optometrist had selected POAG from the drop-down list (Figure 4-19 - Diagnosis selection by optometrist with observations) and the CDSS confirmed his/her decision (Figure 4-20 - Diagnosis candidate with arguments supporting/negating) then the optometrist would most likely follow the recommendations provided by the CDSS. In relatively straightforward cases this should not be a problem. However, the vignettes chosen for this study were selected for their ambiguity in observations, especially relevant clinical observations in images.

The Normal vignette requires a lot more information to discount all the other diagnosis and the optometrist may order additional tests to rule out any pathology. In vignettes where the pathology was apparent and there were clear definitions for supporting a diagnosis the CDSS tended to guide the optometrist towards the right diagnosis and therefore resulted in better management recommendations when compared to the control group optometrists. In cases like Wet AMD where there is an acute risk of blindness the CDSS could potentially assist in early diagnosis and referral of patients to secondary care for treatment. Since there was no improvement in referral decision making in any of the 4 diagnoses the CDSS will need to be improved in this area.

The CDSS is only as good as the information provided to it. Guideline based CDSSs may improve compliance with guidelines (Shalom *et al.*, 2015) as all vignettes showed improvements in prescribing supplementary tests. The guideline had some impact on guiding optometrists towards appropriate enquiry and clinical tests especially in cases where there is a definite pathology and the guidelines alert the optometrist towards abnormal values. The guideline recommendations did have positive impact on management decision making and helped optometrists in prescribing tests as shown in other studies (Fox,

Patkar and Thomson, 2006b). This could potentially lead to improved clinical outcomes in the long run, however more studies are needed to explore this area in greater depth.

The ability of the CDSS to improve diagnosis will depend on the type of case. The results of this study confirm some of the findings from previous studies on diagnostic CDSS applications and EHR systems (Grace *et al.*, 2013) where the quality of data entered affects the outcome. One of the recommended best practices for CDSS applications is to maintain high quality and complete data (Wright, Phansalkar, Bloomrosen, Robert A. Jenders, *et al.*, 2010). The data quality and correctness problems in EHR systems are well known (Wagner and Hogan, 1996; Hogan and Wagner, 1997; Curtis, Bollard and Dickson, 2002; Majeed, Car and Sheikh, 2008). If the optometrist grades the images in the wrong manner it will result in the CDSS wrongly weighting the diagnostic recommendations. So, a diagnosis that would ideally be ranked lower the diagnostic list would be ranked higher incorrectly. This problem can be remedied by increasing the quality and quantity of arguments needed to decide.

This study showed that the opinion of optometrists towards the CDSS was generally positive and even with minimal training and instructions the optometrists did not find the CDSS difficult or cumbersome to use (See section - 4.3.7). Almost half of the respondents are very likely to use the CDSS in their clinical practice. A majority felt the system was easy to use and felt that most people would find system easy to use. Most respondents felt that the CDSS was easy to learn and they didn't need the help of another individual for using the system. More than half of the respondents felt that the system helped improve their decision making and they are likely to use the CDSS in practice. This could have implications for further development and evaluation. One optometrist pointed out that the CDSS could be particularly useful to new registrants while another felt the experience was superior to flow charts, the traditional way of delivering clinical guidelines.

This confirms findings from other studies that show that the clinicians find the decision recommendations from an argumentation-based CDSSs easy to understand and more suited to their own pattern of clinical cognition (Emery *et al.*, 1999; Fox, Patkar and Thomson, 2006b; Sutton, Taylor and Earle, 2006; Longo, Kane and Hederman, 2012; Longo and Hederman, 2013). The results of this study show that the diagnostic recommendations have been shown to improve clinical diagnosis in certain sight threatening conditions and therefore there is potential for wider randomised controlled studies into the impact of argumentation-based diagnostic CDSSs on diagnosis. The guideline recommendations that were part of the CDSS had some impact on guiding the optometrists towards appropriate clinical tests and enquiry. This study demonstrated the effectiveness of an argumentation-based diagnostic CDSSs on diagnostic decision making. Argumentation-based models could therefore be used as the basis for generating early diagnostic recommendations in an ontology-based CDSS that can integrate information from disparate sources.

The survey highlighted several negative aspects of CDSSs, especially guideline based CDSSs that have been raised elsewhere. The negative comments include:

- The CDSS discouraged clinician thinking
- The CDSS could have pushed the clinician towards misdiagnosis
- More recommendations/advice needed
- Time consuming

The survey results indicate that there is a potential for greater acceptance of CDSSs if the decision support recommendations are integrated with the existing workflow of the clinicians as indicated by earlier studies (Kawamoto *et al.*, 2005).

4.4.1 Implications for further research

The results support the further development and evaluation of decision support with a larger set of vignettes and with decision support integrated with the EHR ideally with real patients in real life settings. The effect of improved guideline recommendations and instructions on the final diagnosis optometrists will need to be studied. The prototype developed in this study demonstrates the concept of using

PROforma and Tallis based CDSS applications to develop CDSSs and can be expanded to develop a full CDSS for optometry practices in future if needed.

4.4.2 Summary and Conclusions

This chapter outlined a pilot study evaluating the use of an argumentation-based diagnostic CDSS using clinical vignettes and evaluating the impact of the guideline and diagnostic recommendations on the information gathering and diagnostic performance of optometrists. The argumentation-based diagnostic CDSS was built using off-the-shelf tools (Tallis) that used PROforma as the modelling language. PROforma uses AT as the basis for generating decision recommendations. Argumentation-based models use arguments to support or negate a candidate decision. It is inherently easier to explain to clinicians and closely resembles the cognitive process involved in diagnostic decision making of clinicians (Labrie and Schulz, 2014).

The CDSS prototype developed in this study can provide guideline recommendations and diagnostic recommendations. The aim of the study was to see the impact of the CDSS on information gathering and diagnostic decision making of the optometrists. The results showed that the guideline recommendations have some impact on orienting optometrists towards appropriate history enquiry, clinical tests and investigations, especially in cases where there is a definite pathology and the CDSS can flag abnormal values. The guideline recommendations can then guide the optometrists towards the right clinical tests. The diagnostic recommendations helped the improve the diagnostic decision making of optometrists in 2 of the vignettes, especially in sight threatening cases such as NTG and Wet AMD. These conditions require care in a hospital setting and needs to be referred accordingly. The guideline recommendations helped improve management decision making. The usability survey showed that the optometrists found the recommendations useful and the system easy to use.

The CDSS prototype built in this study can use information from the clinical encounter or historical data from the EHR to make diagnostic recommendations. The results from this study show that guidelines recommendations can be provided at various stages of the clinical encounter. The guideline recommendations improve adherence to evidence-based clinical guidelines and could potentially improve clinical outcomes. Reminding the optometrist to ask the right questions and perform the right tests could similarly help the optometrist collect all the appropriate data and guideline recommendations could help interpret the clinical data from tests/investigations. This can help the optometrists form accurate diagnostic hypotheses focus their efforts on confirming the right diagnosis. The CDSS is particularly helpful for novice optometrists who may not have experience to diagnose and manage rare and complex cases. The right diagnosis can correspondingly help reduce unwanted referral to the hospital and help reduce costs. The diagnostic recommendations based on argumentation-based methods are easy to understand and help the optometrist make sense to lots of clinical data.

Integration with EHR will require connecting the Tallis CDSS application to the EHR database. This could be challenging with respect to optometry practices as not all optometry practices have access to an optometry EHR. The optometry EHRs that exist are not all designed for data exchange and interoperability. When an optometrist is entering information into EHR data is captured in a database. The PROforma pathway can be modified to provide diagnostic recommendations at all stages of the diagnostic encounter. However, the data upon which the CDSS can make the diagnostic recommendations will be restricted to the clinical encounter. The CDSS cannot access data from other clinical data repositories and provide a diagnostic recommendation at all stages of the encounter due to lack of semantic interoperability. The PROforma model is a proprietary model and the inference engine is built using proprietary software tools. The PROforma model can only be read using the Tallis toolset and therefore the diagnostic recommendations cannot be generated by openly available semantic web-based tools such as OWL/RDFS inference engines.

The PROforma model is a knowledge-based model. It is designed to capture the expertise of domain experts. The domain experts provide their subjective arguments for and against each diagnosis and determine the weights that need to be assigned to each argument. This is captured in the PROforma model and the Tallis engine then generates the decision recommendations. However, the other domain

experts and clinicians may not agree with the arguments and weights. Therefore, a website named OpenClinical.net (Fox *et al.*, 2015) was developed to provide a platform for domain experts to collaborate on developing PROforma based models including diagnostic CDSS models. Once the model has been published on the website after prolonged discussion and validation, optometry practices and hospitals can deploy the CDSS in their clinics. However, if the model needs to be updated when clinical evidence changes, the CDSS instances must be informed manually and updated manually. The users will have to be updated about the changes and how the change affects their decision making. Studies evaluating the use of machine-learning algorithms to learn from historical clinical data and automatically update the PROforma model arguments and weights are under progress. However, to automatically update the arguments and weights the intelligent agents should be able to read and understand the model. Semantic interoperability is an issue for automatically updating the CDSS instances as well.

RDF data model provides several capabilities that can address issues regarding semantic interoperability. Semantic web and linked data tools can help in integrating data from diverse sources into a single model and can generate diagnostic recommendations. The next chapter (Chapter 5) demonstrates how an ontology driven CDSS built using OWL/RDF can be developed that uses argumentation as the basis for delivering diagnostic recommendations. The ontology driven CDSS uses a model called the Disease-Symptom Model (DS Model) that uses semantic web and linked data technologies to integrate information from diverse sources to generate a diagnostic recommendation. However, once the information has been integrated there is still a need to explain the results to the clinician in an easy to understand manner.

The DS Model will have to be knowledge-based capturing the domain expertise of expert clinicians. The DS Model should be able to represent the domain expertise in an easy manner and generate a diagnostic recommendation that can be easily understood by the clinician. The model will have to generate diagnostic recommendations in the presence of incomplete data. The model should be able to recalculate and revise its decision based on current information which is called defeasible reasoning. Argumentation based models are inherently capable of doing this. Therefore, the DS Model uses the power of semantic web and linked data to integrate information and uses argumentation as the basis for generating the diagnostic recommendation. A machine-readable representation that can be processed by intelligent agents provides additional advantages when updating knowledge bases built using DS Model. Chapter 5 describes the development of the DS Model and Chapter 6 then evaluates it using clinical vignettes.

5 Chapter 5 - DS Model and decision support prototypes development

5.1 Introduction

This chapter details the development and modelling of the DS Model and demonstrates how the DS Model can help in integrating information from diverse sources and generate diagnostic recommendations using argumentation as the basis. The chapter tries to answer the following research question: Can an argumentation-based diagnostic CDSS developed using semantic web and linked data technologies generate early diagnostic recommendations?

The literature review showed the need for improved interoperability and showed how semantic web and linked data technologies can help improve semantic healthcare data interoperability. The literature review described Argumentation Theory and how argumentation-based models can help the domain experts capture domain expertise in an easy manner. The Chapter 4 optometry study demonstrated the feasibility of using an argumentation-based diagnostic CDSS to support diagnostic decision making in primary care optometrists. The diagnostic recommendations generated by the argumentation-based CDSSs can be easily understood in a manner that is like the clinician's own line of thinking.

As patients are increasingly getting access to PHRs and smartphones patients can record their data in these applications. Additional data can be captured by wearable devices and social media applications. Crucial data that can be used for clinical diagnosis is therefore locked away in these 'data islands' in a multitude of data representation formats and standards. The DS Model demonstrates the ability to integrate information from these diverse sources while generating diagnostic recommendations that is easy to understand. The diagnostic recommendation can be delivered at all stages of the clinical encounter and the CDSS is not restricted to data that is gathered during the clinical encounter.

The DS Model demonstrates how an argumentation-based ontology model can be built using currently available ontology modelling tools. The CDSS prototypes demonstrate how the argumentation-based diagnostic CDSS recommendations can be generated without the use of specialized software for inference.

The next section details DS Model and how diagnostic recommendations can help reduce diagnostic errors due to cognitive factors.

5.2 Disease-Symptom Model (DS Model)

The DS Model in this thesis is a representation of relationships between a diagnosis and its observations alongside SPARQL rules that define the presence/absence of observations and SPARQL rules that calculate weights of the observations to provide a diagnostic recommendation. **The presence of observations/absence of observations can be arguments for and against a diagnosis.**

The patient presenting to a clinic may have no disease or multiple diseases. Diagnosis is the term used to denote state of the patient. A patient having no disease can have a diagnosis called "Normal". The patient can present with a multitude of symptoms. The clinician records these symptoms, checks for further signs of the disease(s) and orders further clinical tests if needed. 'Observation' is an umbrella term used to denote all clinical observations recorded by the clinician or patient with respect to the patient's condition. This can include signs, symptoms and clinical test results. These observations are used by the clinician or the diagnostic CDSS to help make a diagnosis. The clinician uses the presence or absence of clinical observations to support or negate a diagnostic hypothesis. **The presence or absence of clinical observations are therefore arguments either supporting or negating a diagnosis and are captured in the DS Model.**

From an early diagnosis point of view, the symptoms play a greater role as they are the first observations reported by the patient. The clinician looks for further signs based on the symptoms reported by the patient and orders clinical tests accordingly. Therefore, the model was called the Disease-Symptom Model A simple representation of the link between a Diagnosis and Observations can be seen in Figure 5-1.

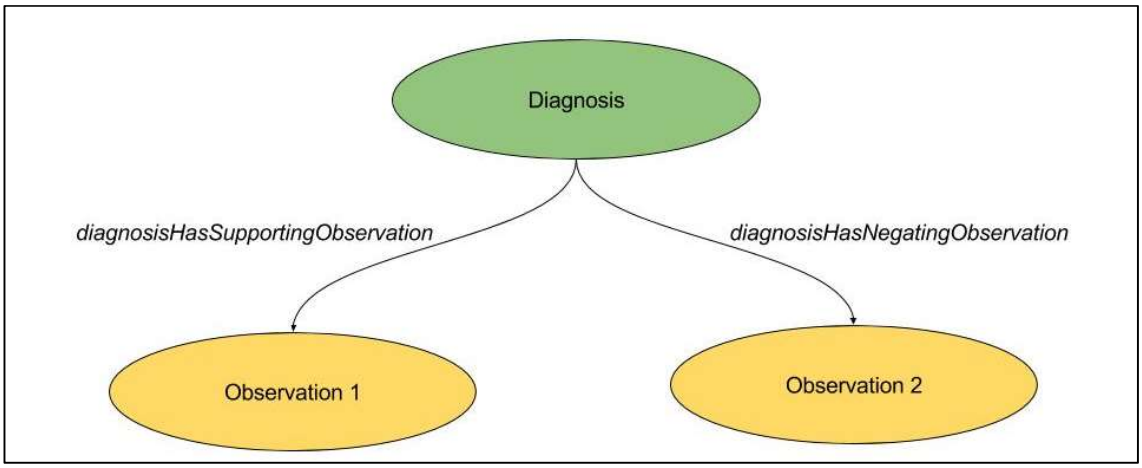


Figure 5-1 - Disease Symptom Model

The need for diagnostic triggers in a CDSS can be substantiated using the 'Illness script theory', which is a medical specialization of the more general 'Script theory' in cognitive science. Script theory tries to explain how human beings try to gather and structure information to interpret events in his/her surroundings (Schank and Abelson, 1977). When an individual encounter an environment he/she accesses memory to retrieve prior knowledge and uses that information as the basis for building a model of an object or event of interest. The individual then uses that information to predict what information he/she should be receiving from objects or events in the environment (Schank and Abelson, 1977; Frith, 2007). The brain attempts to create meaning about the environment by actively comparing the attributes of the mental model to the information he/she is receiving from the objects or events in the environment (Lubarsky *et al.*, 2015). Once the mental model of the individual 'approximately' matches the attributes of the information from the environment he/she will then take appropriate next steps if needed. Next steps can include obtaining additional information about the environment or specifically from an event or object of interest within the environment. Additional information gathered can then trigger additional mental models. When the individual is satisfied that the information coming from the environment adequately matches the mental model, then he/she will stop collecting more information. Knowledge artefacts of an individual that are structured in a particular way that makes it easy to retrieve and match to the attributes of the information obtained from the events and objects in the environment is called 'Scripts' (Schank and Abelson, 1977).

The guideline recommendations in the Chapter 4 optometry study demonstrated the role of guidelines in helping trigger the mental models by flagging abnormal values and giving additional information about the next steps. The guideline recommendations can greatly help guide the clinician in figuring out how to collect the next bit of information relevant to the patient. However, guideline recommendations cannot provide arguments favouring and negating each diagnosis. The guideline recommendations cannot give an overview of the patient's current condition. This is where argumentation-based CDSS models can help provide a better understanding of the patient's condition. The CDSS provides arguments in favour and against a diagnosis with weights explaining the level of support of each argument. The information collected in early stages of a clinical encounter are limited in nature since not all information has been collected yet. However, argumentation-based models are better suited for explaining the patient's current condition based on limited or uncertain information (Longo, Kane and Hederman, 2012). Argumentation models use non-monotonic logic and can retract existing conclusions and formulate new conclusion in the light of current information. The clinician can view the arguments for each diagnosis and match the arguments with the mental model created by the clinician from the data received from the environment i.e. the patient. The early diagnostic recommendations could in theory act as a tool (in conjunction with

the guideline recommendations) in helping create a better overall picture of the patient and help in the information gathering process. Guideline recommendations and early diagnostic recommendations work in concert with each other. Guideline recommendations can help guide the clinician towards the appropriate tests and explain how the perform the tests and interpret the test results. The data from these tests can help in formulating the diagnostic recommendation by the CDSS.

The 'Illness script' concept is a widely accepted concept that tries to provide a framework for understanding how clinicians organize the knowledge that they gain from literature and clinical experience and use it to diagnose and treat patients (Charlin, Tardif and Boshuizen, 2000). A script can be described as a set of attributes each with a set of values. The script has default values corresponding to each attribute. For example, when a dentist examines a patient with dental pain he/she first asks what the chief/primary complaint of the patient is. The patient may respond by saying that he/she has pain in lower jaw with recent onset. This triggers/activates a set of illness scripts which may include a 'dental abscess script', 'tooth fracture script' etc. The dental abscess script has a default value for nature of pain. Each attribute within the script will have different probabilities of occurring. The value with greatest chance of occurring is the default value. The dentist may use this script to further enquire about the nature of pain. If the answer provided by the patient matches the default value of the attribute in the script, then the dentist will gather further information that conforms to this illness script. The answers provided by the patient or the data collected from the patient may in turn trigger other scripts. The number of illness scripts activated will depend on the experience of the dentist and the level of prior knowledge of the dentist. If there is an unexpected symptom or sign that does not conform to the earlier activated scripts the dentist will then activate another set of scripts to explain this unexpected clinical observation. The dentist will continue to gather information, weigh the information until he/she is satisfied. If the information provided does not match the default value of the illness script, then another illness script is activated. The illness scripts are not standardised, and the scripts are formed through a mix of biomedical knowledge, clinical evidence and experience gained through years of clinical practice (Charlin, Tardif and Boshuizen, 2000). Guideline recommendations can help trigger illness scripts by providing clinicians with context specific information about the clinical encounter.

The Dual Process Theory explains how individuals make decisions and judgments under uncertainty with limited information (Kahneman and Egan, 2011; Evans and Stanovich, 2013). It can be adapted to explain how a clinician reasons when presented with a clinical case. Making a diagnostic decision in the early stages of the clinical encounter is a classic example of this scenario. The clinician is faced with limited information and is working under uncertainty regarding the patient's condition. The dual process theory consists of 2 types of processes, System 1 and System 2 (Evans and Stanovich, 2013).

Diagnostic CDSSs should be able to support both System 1 and System 2 processes. The CDSS should be able to make reasonably accurate diagnostic recommendations with limited data available from the patient during the clinical encounter. The clinician can then use the information from the patient and compare that to the information provided by the guideline recommendations and early diagnostic recommendations provided by the CDSS. The guideline recommendations and diagnostic recommendations can trigger illness scripts and trigger System 1 process of thinking. The arguments of the diagnostic recommendation can force the clinician to think about the observations and weigh the arguments in favour of each diagnosis called System 2 process of reasoning. It prevents the clinician from forming a diagnostic hypothesis without considering all clinical data. This way the arguments can potentially help reduce incidence of premature closure bias.

The CDSS should be able to use the web to help find data that is related to the patient to make this diagnostic recommendation. The clinician is therefore freed from the job of integrating the information. As the clinician is gathering data from a patient during a clinical encounter the CDSS should be able to make use of all available evidence to suggest a recommendation and help the clinician in data gathering process. This is where the use of semantic web and linked data technologies can help the CDSS in the data gathering process. The intelligent agent in a CDSS can find and integrate all relevant data in the semantic web of data and provide a diagnostic recommendation as the clinician is performing the examination of the patient.

The data needed to make a diagnosis can come from various sources including PHR, EHR of the patient, disease registries, clinical trial data, laboratory test results, pathology test results, patient surveys and other research data. There is a lot of research around the use of Big Data in healthcare (Murdoch and Detsky, 2013; Raghupathi *et al.*, 2014). Big data is generally defined using 3 V's (Volume, Velocity and Variety) (Laney, 2001). Volume describes the large volume of data generated by healthcare applications such as EHRs and from clinical trials and from biomedical and genomic sources. Velocity describes the high speed at which data especially healthcare data can be generated from applications such as wearable computing devices. However Variety is the aspect of big data that prevents its use in clinical applications such as decision support systems (Sittig *et al.*, 2008). Traditionally the data for these applications have come from EHRs or standalone applications (Osheroff *et al.*, 2007). The data that can be used for clinical applications such as diagnostic CDSSs can come from several different types (varieties) of datasets in different forms-structured, unstructured and semi-structured (Raghupathi *et al.*, 2014) including datasets residing on the web.

The amount of data available to the clinician for making decisions are increasing. While this shift in data and information availability is good in some respects it introduces additional problems. The clinician will have to spend lots of additional time trying to make sense of all this additional data. There is considerable pressure on primary care clinicians already to spend more time with patients with ever decreasing budgets and resources available to provide optimum care (Dugdale, Epstein and Pantilat, 1999). Therefore, there is a need for a framework that can help in aggregating the knowledge and help the clinician cope with the inconsistency and uncertainty associated with decision making. Argumentation can provide one such framework.

To illustrate how argumentation can be used by the CDSS to generate a diagnosis recommendation we use the example of a dentist attempting to diagnose a patient with dental pain. First the dentist tries to locate the source of pain. Then the dentist will enquire type and nature of pain and type of triggers for the pain. Once this has been ascertained the dentist will try to locate a probable cause for the pain. If the pain is a sharp type of pain, then the origin of the pain will be more likely to be pulpal in origin (originating in the pulp of the tooth). The most of common causes are dental caries or tooth fracture leading to exposure of the pulp. However, if the dentist finds no dental caries or problems with the tooth then his/her attention turns to tissues surrounding the tooth. Based on signs and symptoms reported by the patient there are arguments in favour of the pain originating in the pulp or the gums surrounding the tissue. The chances of the pain being pulpal in origin are greater. The signs and symptoms may trigger illness scripts or pattern recognition (System 1) that support this sub-diagnosis. However, once the dentist has completed examination there are no apparent problems with the tooth and the dentist's attention is focussed on the gums as probable origin of the pain. The weak argument in favour of the pain originating from the gums has now become stronger. These are knowledge-based arguments that have been formulated through years of expertise. This form of reasoning process is called defeasible reasoning (Longo, Kane and Hederman, 2012). In defeasible reasoning, the conclusions are defeasible i.e. the conclusion(s) can be retracted or modified in the light of new evidence. This is crucial for providing diagnostic recommendations in the early stages of the clinical encounter.

Argumentation evolved from its origins as a sub-discipline of philosophical logic and has emerged as a critical area of Artificial Intelligence (AI) research. Argumentation is particularly suitable for healthcare as the information available to decide is often incomplete and clinicians will have to make decisions when there is conflicting information. Argumentation based techniques have been used to develop CDSS applications in healthcare since the mid-80s (Labrie and Schulz, 2014). PROforma used argumentation-based methods to model its decision making (Sutton and Fox, 2003; Fox, Patkar and Thomson, 2006a) . A modeller can model arguments in favour and against a candidate. The candidate can be a diagnosis or some other clinical hypotheses, for example clinical risk. The Chapter 4 Optometry study uses PROforma to model the CDSS and generate diagnostic recommendations. The results of the study in Chapter 4 optometry study shows that the providing argumentation-based diagnostic recommendation to the clinician can help improve diagnostic performance in certain clinical conditions. The study showed how the argumentation-based diagnostic recommendations can help in generating diagnostic recommendations that can potentially reduce diagnostic errors while providing recommendations that are

easy to understand. Studies have shown that guideline-based CDSSs can help improve guideline adherence (Adams *et al.*, 2012; Forrest *et al.*, 2013) and has helped improve clinical outcomes (Garg *et al.*, 2005a). The results of the Chapter 4 optometry study showed that guideline recommendations can help the optometrist in information gathering before diagnosis. The qualitative survey of the Chapter 4 Optometry study CDSS showed that the optometrists found the various clinical recommendations provided by the system relevant and the CDSS easy to use and understand (See section - 4.3.7). The argumentation-based models used in the Chapter 4 optometry study CDSS to generate the diagnostic recommendations have been used to develop CDSS applications in several domains (Emery *et al.*, 1999; Fox, Patkar and Thomson, 2006b; Sutton, Taylor and Earle, 2006), however there studies so far have not studied the impact of argumentation-based CDSSs on the diagnostic performance of clinicians. Therefore, Chapter 4 Optometry study was needed.

Use of semantic web and linked data technologies help integrate patient information from various sources. However, once the data has been retrieved the CDSS that uses the proposed DS Model should be able to merge the information in such a way to provide a diagnosis that can be easily explained to the clinician. The clinician should be able to see and weigh information and understand why a diagnosis has been ranked above others in the diagnostic recommendation. This is another area where argumentation-based models have an advantage over other CDSS models. Argumentation provides a method to easily explains the arguments for each diagnosis. When this is combined with data from several sources it provides an effective mechanism for delivering diagnostic recommendations early in the clinical encounter.

To impact System 1 process of reasoning the CDSS should be able to produce a ranked list of recommendations with different weights that support each diagnosis. The clinician would then look at the ranked list of diagnoses and this triggers pattern recognition or activation of illness scripts that the clinician can use to elicit further information from the patient. Clinician enters this information into the EHR and the CDSS uses this additional information to collect more information if needed or generate an updated list of diagnostic recommendations.

If the clinician is interested in learning why a diagnosis is ranked high/low in the recommendation list, then the CDSS should be able to provide suitable arguments as to why the diagnostic decision was made. The clinician should be able to look up the arguments for and against a diagnosis and consider the weights for each argument. This process should trigger the System 2 process of reasoning and help slow down the thinking process of the clinician, help analyse all the clinical data available and consider all probable causes for the patient's current condition.

The semantic web/linked data framework helps in data integration and data interoperability and helps an intelligent agent identify and retrieve data that is related to the patient and the patient's condition. The decision model in the form of arguments helps the CDSS provide a recommendation to the clinician that can be easily understood and closely matches the clinician's own line of reasoning. The model can be easily updated in the light of current information and quickly generate a new diagnostic recommendation. The diagnostic recommendation of the CDSS in the form of ranked recommendations could in theory influence System 1 and System 2 processes of reasoning and help the clinician in reducing diagnostic error arising from cognitive factors.

Using ontologies to represent disease-symptom pattern has been done several times before (See Section - 2.5). However, these models differ from DS Model is many ways. OWL and rules such as Jena and SWRL rules have been used to model the diagnostic criteria. The combination of OWL, SWRL and other rule languages is a recognised expert system framework and has been widely used in semantic web based decision support applications (Ongenae *et al.*, 2010; Donfack Guefack *et al.*, 2012; Bau, Chen and Huang, 2014; Sherimon and Krishnan, 2016). This is largely due to the availability of ontology reasoners that are compatible with both OWL and SWRL (Sirin *et al.*, 2007a). SWRL can overcome some of the weaknesses of OWL as it can handle numerical functions better than OWL. However, the studies that evaluated the use of OWL and rule languages have not demonstrated the ability to access remote datasets to integrate patient information. The DS Model uses SPARQL to model the rules that inform the inference engine

about the presence/absence of an observation and obtains a ranking based on the weights attached to each argument. SPARQL has the additional capacity to query federated databases for data and can integrate that information into decision making.

Oberkampf et al. (Oberkampf *et al.*, 2012) demonstrated the use of a Disease-Symptom Ontology to capture the relation between the Diseases and Symptoms. Each Disease is related to the Symptoms using the 'hasSymptom' property (Oberkampf, Zillner and Bauer, 2012a). They have demonstrated use of a ranking algorithm to generate a ranked diagnosis. No specifics were given on how the ranking system was implemented.

The DS Model differs from this approach by using argumentation as the basis for generating the diagnosis. Existing diagnostic CDSS applications developed using argumentation-based languages such as PROforma can be directly mapped to the DS Model if needed. The semantics of argumentation can be captured using the DS Model. PROforma requires proprietary tools such the Tallis inference engine to generate the diagnostic recommendations. DS Model does not require any special tools to develop the ontology model.

Modelling can be performed using open-source ontology modelling tools like Protégé. The DS Model can be uploaded to any triple-store that supports simple RDFS reasoning. The SPARQL rules can then generate the ranked diagnostic recommendations. A separate diagnostic ranking algorithm was not required. Existing ontologies can be reused to capture the Diseases and the Symptoms. This vastly reduces the time taken to create the models when compared to PROforma. Argumentation reduce the time needed to capture the domain expertise linking the Diseases and their symptoms. Domain experts can easily add the weights to the same ontology model. Using semantic web technologies makes the process of sharing the model easy. Linked data technologies used in the DS Model help integrate information from diverse sources. These clinical observations can come from a variety of sources including PHR, EHR, wearable computing devices, social media website etc. Argumentation-based methods used to generate the diagnostic recommendation can help explain the recommendations to the clinicians. All these steps used in the development of the DS Model can be used to effectively deliver early diagnostic recommendations.

This section has explained the link between the DS Model and the CDSS prototype developed in Chapter 4. It explains how the DS Model and its diagnostic recommendations can help in reducing diagnostic error. The next section details the development of the DS Model. OWL, RDF, SPARQL and other languages and technologies used to develop the DS Model have been explained in Chapter 2. The full DS Model can be accessed from the following link (Further instructions in Appendix 5):

https://drive.google.com/drive/folders/0B67O3_av5-CVUzZmVzBVYktTUk0?usp=sharing

5.3 Disease-Symptom Model represented in RDF/OWL

In this section, the method thorough which the DS Model has been represented in OWL classes, properties and instances is discussed.

The degree to which each observation supports or negates a diagnosis varies from observation to observation. The same observation can support 2 different diagnoses to different extents. Some observations are more apparent than others. The clinician may not have all the tools required at his/her disposal to confirm the presence of these observations. For example, the clinician may suspect the presence of Tempero-mandibular joint (joint that connects the lower jar to the skull) problems but may not have the radiographic facility in the clinic to confirm this. This degree of support is represented in the form of weights that are attached to the observation's presence or absence.

Figure 5-2 gives an outline of the Class hierarchy that was used to model the DS Model in OWL. The diagnosis and observation classes have been expanded in the image. The top-level classes (Diagnosis, Observation, Observation Status) belong to the base DS Model. The role of Finding and Finding Status classes will be explained in section - 5.5 - Multi-level diagnosis with DS Model.

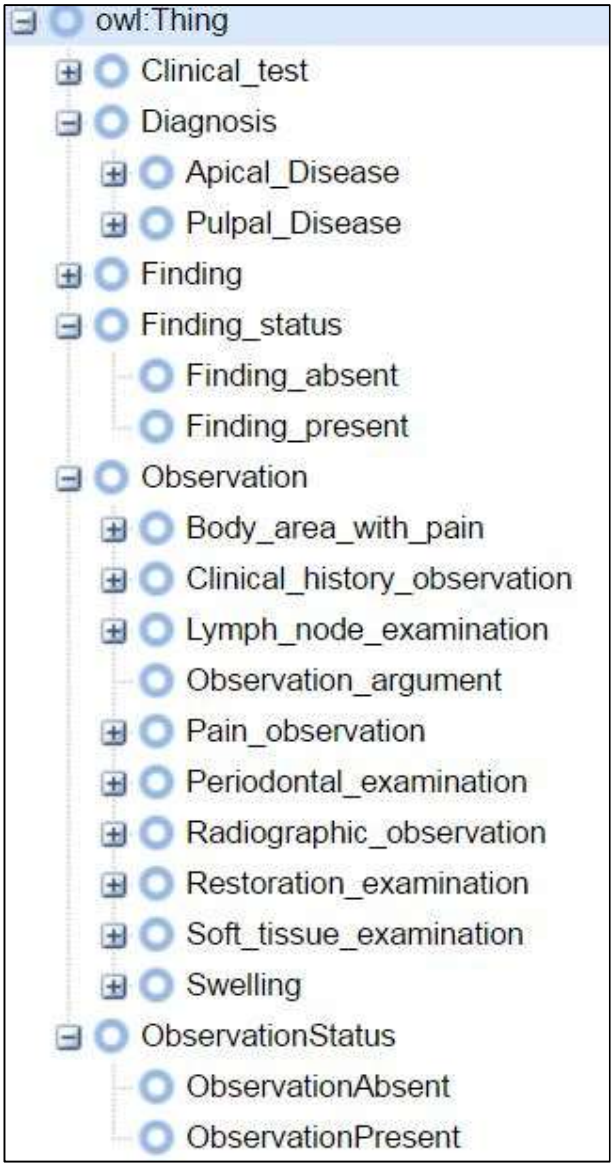


Figure 5-2 - Outline of DS Model Class hierarchy

A modeller building a DS Model for a domain will use the base classes and extend the model using classes from their domain. In this thesis, the DS Model was built using dentistry as an example. The image (Figure 5-2) shows that terminology used for the observations has come from the domain of dentistry. The main reason dentistry was chosen as a domain to develop the diagnostic inference model is because of the background of the author in dentistry and because dentistry is mainly a primary healthcare service.

Example of Reversible Pulpitis diagnosis is used to demonstrate how the DS model was developed in OWL and explain how the model is executed by the CDSS prototype to deliver diagnostic recommendations. Reversible Pulpitis is an acute condition affecting teeth. The centre of a tooth consists of living tissue called Pulp. It is a connective tissue consisting of blood vessels and nerves. When there is tooth decay and the decay extend to the pulp of the tooth it can cause inflammation of the pulp. Sometimes the inflammation is reversible, and the tooth can be saved. Other times the inflammation is irreversible (Irreversible Pulpitis) and can result in the pulp undergoing necrosis (death) later. It is therefore important to distinguish between the reversible and irreversible states of the pulp. However, symptoms in both conditions are similar and clinical history and examination is important to distinguish between both disease states. Failure to distinguish appropriately can result in the patient undergoing unnecessary expensive and invasive treatments such as Root Canal Treatment (RCT).

Turtle syntax has been used to represent the RDF triples in the DS Model (Beckett *et al.*, 2014). In this thesis, Properties are denoted in italics, Classes have been Capitalized and Instances have been underlined. The DS Model was modelled using the Protégé *5.0* ontology editing environment (Gennari *et al.*, 2003). The class structure was captured from Web-Protégé (Figure 5-2). Web-Protégé is a light weight open-source web-based ontology editor based on Protégé that is useful for collaborative ontology editing (Tudorache *et al.*, 2013) .

rp: is the prefix of the DS Model. This prefix will be used to denote classes, properties and individuals that belong to the DS Model ontology. The prefix is used to associate a label with a URI. In this thesis, the rp: prefix is used to denote the DS Model ontology model and all the classes, properties and individuals that belong to this ontology model. If a term belongs to another ontology, example RDFS, then rdfs: prefix is used to denote objects that belong to RDFS ontology (example rdfs:subClassOf). Prefixes are widely used in RDF serializations such as Turtle (Beckett *et al.*, 2014).

For example, PREFIX **rp**: <*http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#*> is the prefix used to denote the DS Model ontology in this thesis.

5.3.1 Classes

An OWL class (owl:Class) is used to represent a group of individuals that belong together and share some common properties (Harmelen and McGuinness, 2004).**The main classes in the DS model include the Diagnosis class (Figure 5-3) and the Observation class (Figure 5-4)**. The subclasses of these 2 classes may vary depending on the domain.

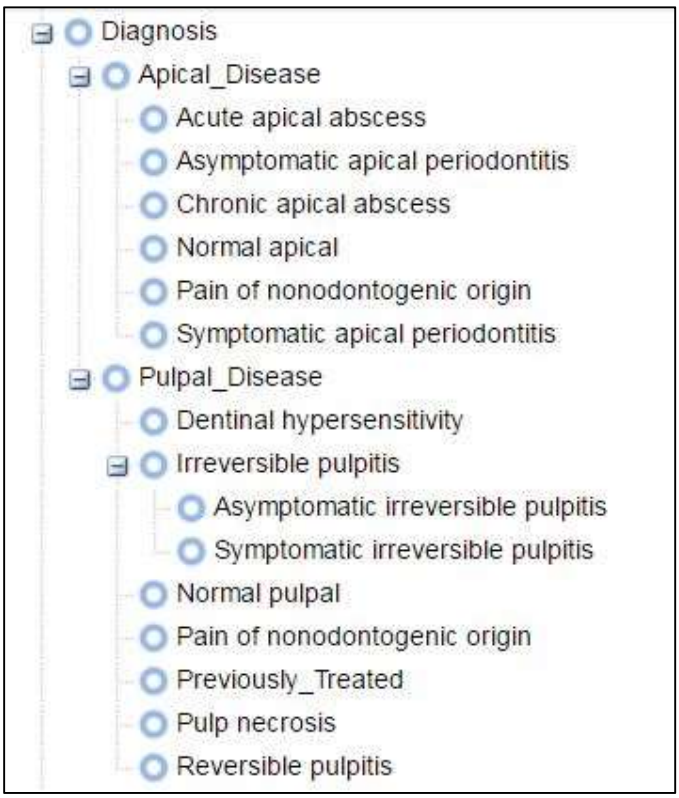


Figure 5-3 - Diagnosis class in DS Model

The diagnosis class has subclasses rp:Apical_Disease and rp:Pulpal_Disease. These classifications are relevant in dentistry as the Pulpal and Apical diagnoses are obtained separately and a composite picture of the tooth's pulpal and apical tissues is created from the clinical observations. Pulpal refers to the pulp of the tooth, while Apical refers to the apex (tip of the root) of the tooth.

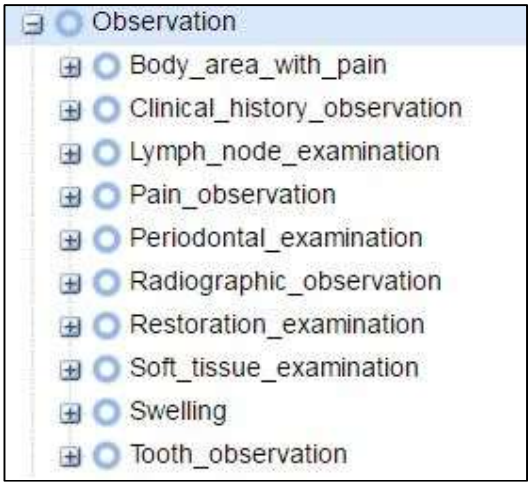
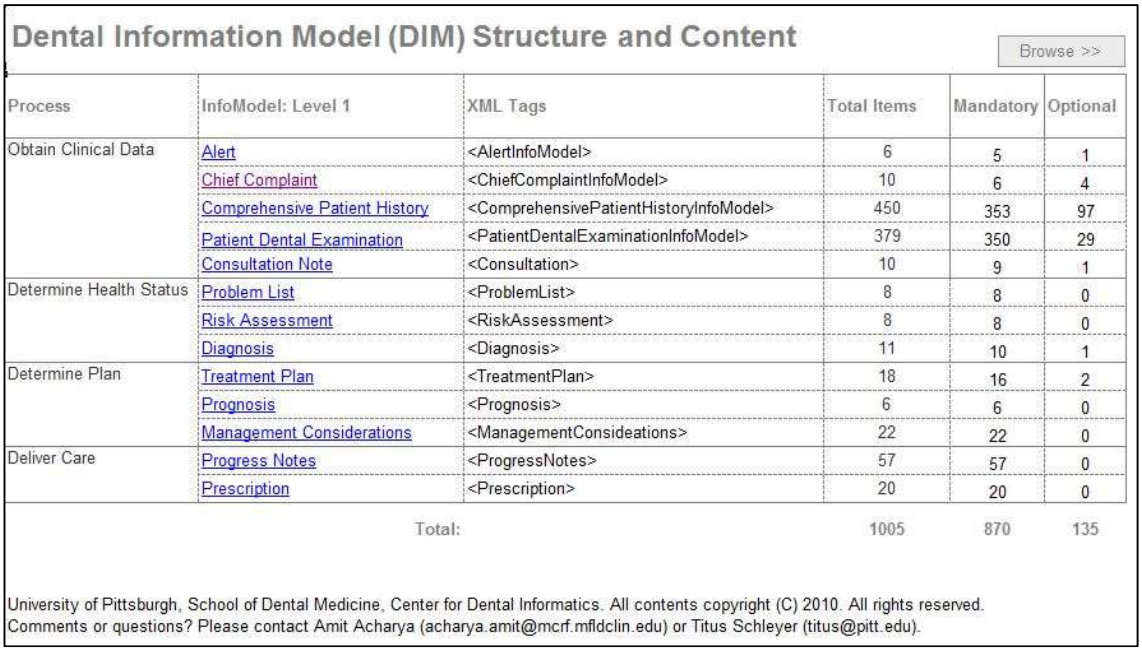


Figure 5-4 - Observation class

In addition, there is a rp:ObservationStatus class (Figure 5-5). This class has the subclasses rp:ObservationAbsent and rp:ObservationPresent.

Figure 5-5 - Observation Status class

The Observation class and instances belonging to the subclasses of the Observation class reflect the members of the information model of the EHR. It is divided in the same way the information model is classified. The Dental Information Model (DIM) which was used as a basis for developing the class structure has a similar structure (Acharya, Mital and Schleyer, 2009a). DIM is one of the few dental information models currently available and has been used as the basis for developing dental EHR and research systems (Figure 5-6) (Liu *et al.*, 2013). DIM was used as a source for some of the terms in the dental representation of the DS Model. Since the terms and the structure were derived from an existing dental information model any future mapping to a real dental EHR will be easier. The basic structure was adapted from DIM and extended where needed. Terms were included from the DDS (Dental Diagnostic System) (Kalenderian *et al.*, 2011b) dental terminology.



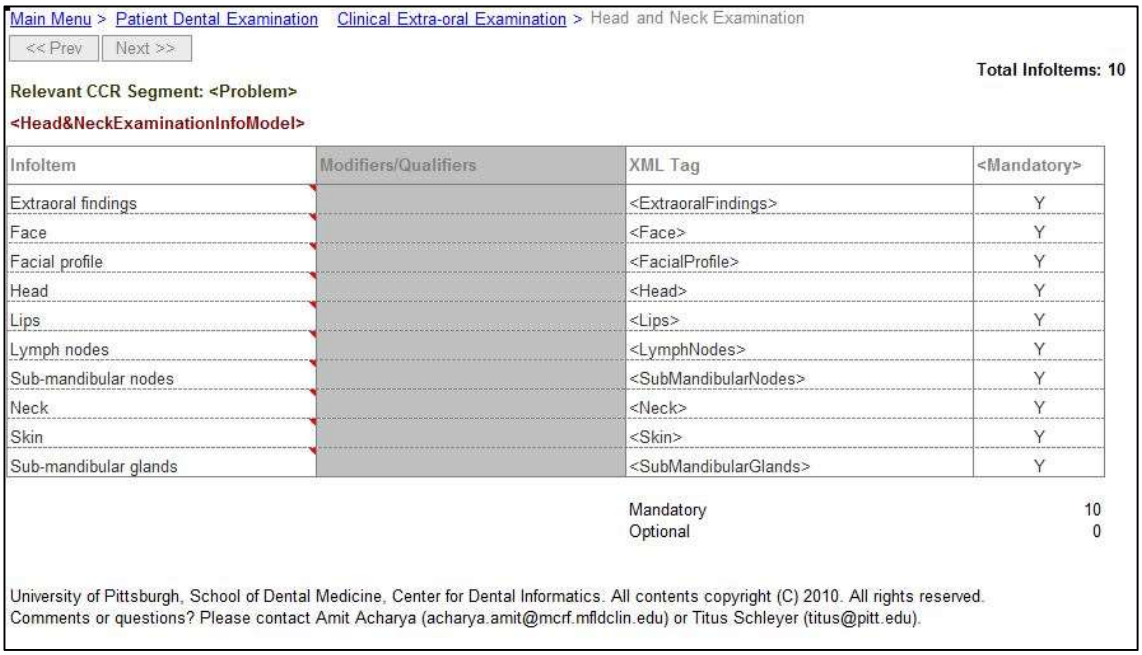
Dental Information Model (DIM) Structure and Content

Browse >>

Process	InfoModel: Level 1	XML Tags	Total Items	Mandatory	Optional
Obtain Clinical Data	Alert	<AlertInfoModel>	6	5	1
	Chief Complaint	<ChiefComplaintInfoModel>	10	6	4
	Comprehensive Patient History	<ComprehensivePatientHistoryInfoModel>	450	353	97
	Patient Dental Examination	<PatientDentalExaminationInfoModel>	379	350	29
	Consultation Note	<Consultation>	10	9	1
Determine Health Status	Problem List	<ProblemList>	8	8	0
	Risk Assessment	<RiskAssessment>	8	8	0
	Diagnosis	<Diagnosis>	11	10	1
Determine Plan	Treatment Plan	<TreatmentPlan>	18	16	2
	Prognosis	<Prognosis>	6	6	0
	Management Considerations	<ManagementConsideations>	22	22	0
Deliver Care	Progress Notes	<ProgressNotes>	57	57	0
	Prescription	<Prescription>	20	20	0
	Total:		1005	870	135

Figure 5-6 - Dental Information Model outline (DIM)

The following image (Figure 5-7) shows how DIM has organized information within Clinical Extra Oral Examination.



Main Menu > Patient Dental Examination Clinical Extra-oral Examination > Head and Neck Examination

<< Prev Next >>

Total InfoItems: 10

Relevant CCR Segment: <Problem>

<Head&NeckExaminationInfoModel>

InfoItem	Modifiers/Qualifiers	XML Tag	<Mandatory>
Extraoral findings		<ExtraoralFindings>	Y
Face		<Face>	Y
Facial profile		<FacialProfile>	Y
Head		<Head>	Y
Lips		<Lips>	Y
Lymph nodes		<LymphNodes>	Y
Sub-mandibular nodes		<SubMandibularNodes>	Y
Neck		<Neck>	Y
Skin		<Skin>	Y
Sub-mandibular glands		<SubMandibularGlands>	Y

Mandatory	10
Optional	0

Figure 5-7 - Clinical Extra Oral Examination

In the DIM model, the 'Lymph node observations' is arranged under Head and Neck Examination, which comes under Clinical Extra Oral Examination, which comes under Patient Dental Examination. The image above (Figure 5-7) shows that the information model only presents a single field for adding information about the lymph nodes, such as the condition of the lymph nodes. In the DS Model, this field was further expanded based on expert opinion and author's own clinical knowledge to include the different lymph node states, i.e. lymph node firm, lymph node swollen, and lymph node tender (Figure 5-8).

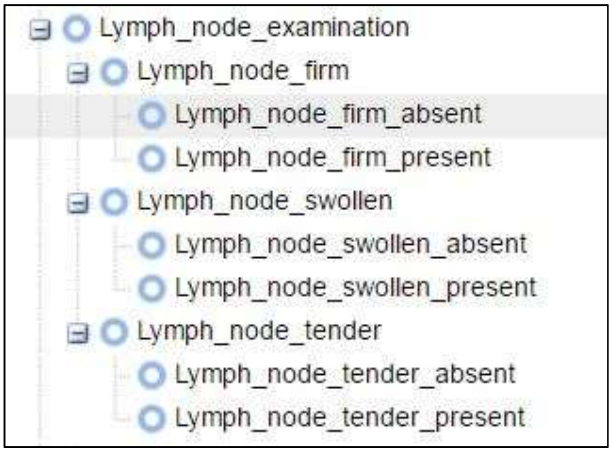


Figure 5-8 - Lymph node examination

The observations were further expanded to include the different observation states that the observation may have with respect to the patient. Having a "Present" or "Absent" observation may not be necessary in all cases. The presence of a "Firm Lymph Node" may support a Diagnosis, while the absence may support a completely different diagnosis.

The image below (Figure 5-9) shows the rp:Duration_of_pain class, and the different values that it can have in the EHR is represented as its subclasses. The presence of rp:Hours_duration_of_pain has more value than the absence, and therefore only the value itself is needed from a decision support point of view. The decision to add the present and absent states of an observation is dependent on the domain and can vary from domain to domain. However, it is advisable to create the present and absent states during modelling as different decision support applications may have different domain requirements. In this diagnostic decision support implementation, it was not considered necessary and was therefore not modelled. In the Lymph node tender example above (Figure 5-8), the presence and absence are relevant for making diagnostic decisions and was therefore included in the DS Model.

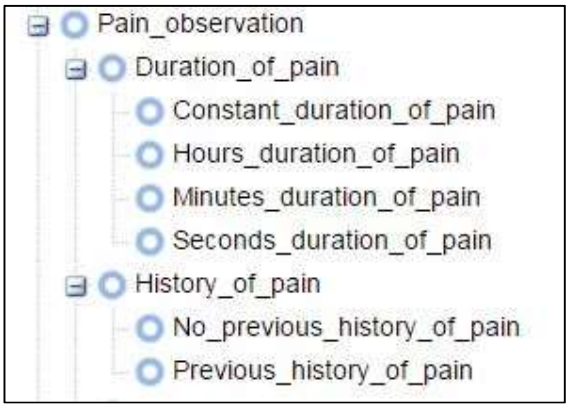


Figure 5-9 - Pain observation

5.3.2 Properties

There are 2 main types of properties (rdf:Property) in OWL. There are object properties which link individuals belonging to different classes together, and datatype properties that link individuals with datatype literals such as string or integers (Harmelen and McGuinness, 2004).

The main object property in the DS Model include the *rp:Observation_Diagnosis_properties*, and the sub-properties (Figure 5-10) of this property include:

- *rp:diagnosisHasNegatingObservation*
- *rp:diagnosisHasSupportingObservation*
- *rp:observationAbsenceNegatesDiagnosis*
- *rp:observationPresenceSupportsDiagnosis*

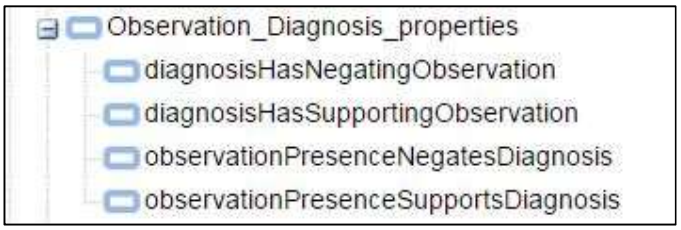


Figure 5-10 - Observation-Diagnosis properties

In addition, there is the rp:*hasObservationStatus* property that has a domain of rp:Observation and range of rp:ObservationStatus. The main datatype properties include the *rp:hasDiagnosisObsWeight* which link the diagnosis individuals to the weights represented by integers. The datatype property *rp:hasWeight* links the observation individuals to their respective weights.

5.3.3 Individuals

Individuals are instances of the classes represented in the DS Model. Individuals within the DS Model are crucial to providing the decision support capability. SPARQL inference is used to inform/update the DS Model about the presence/absence of different observations. The individuals of the DS Model and the methods through which the properties link the different classes together are explained below with the help of an example of rp:Reversible_Pulpitis diagnosis class. rp:Reversible_Pulpitis is a subclass of Pulpal_Disease and primarily denotes a state of the pulp. In Reversible Pulpitis, the inflammation of the pulp is only temporary and it does not have observations that can indicate the presence of infections, inflammation or cancer that can lead to firm/swollen lymph nodes.

rp:1_Reversible_Pulpitis_DIAG is an individual or a type of the class rp:Reversible_Pulpitis (Table 5-1). The *rdfs:label* of the individual is "Reversible Pulpitis". The label will be displayed to end-users while the name of the individual will be used for modelling. The name is a dereferenceable URI (http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#1_Reversible_Pulpitis_DIAG) and it is technically possible to access this resource with help of a browser and view the HTML page representing the resource (if hosted online), or the RDF representation of the resource with all the individuals.

subject	predicate	object
rp:1_Reversible_Pulpitis_DIAG	*rdf:type*	owl:NamedIndividual
rp:1_Reversible_Pulpitis_DIAG	*rdf:type*	rp:Reversible_Pulpitis

Table 5-1 - Reversible Pulpitis Individual

1_Reversible_Pulpitis_DIAG has the following object property assertions (Table 5-2). These assertions are made by the domain expert (Endodontist). In this case the assertions were made from clinical experience of the author in addition to referring clinical textbooks on the subject and help from the domain expert. These assertions are arguments that either favour or negate a diagnosis.

subject	predicate	object
rp:1_Reversible_Pulpitis_DIAG	*rp:diagnosisHasSupportingObservation*	rp:Lymph_node_tender_absent_Plus_1
rp:1_Reversible_Pulpitis_DIAG	*rp:diagnosisHasNegatingObservation*	rp:Lymph_node_firm_present_Minus_1

Table 5-2 - Reversible Pulpitis and Observations relationship

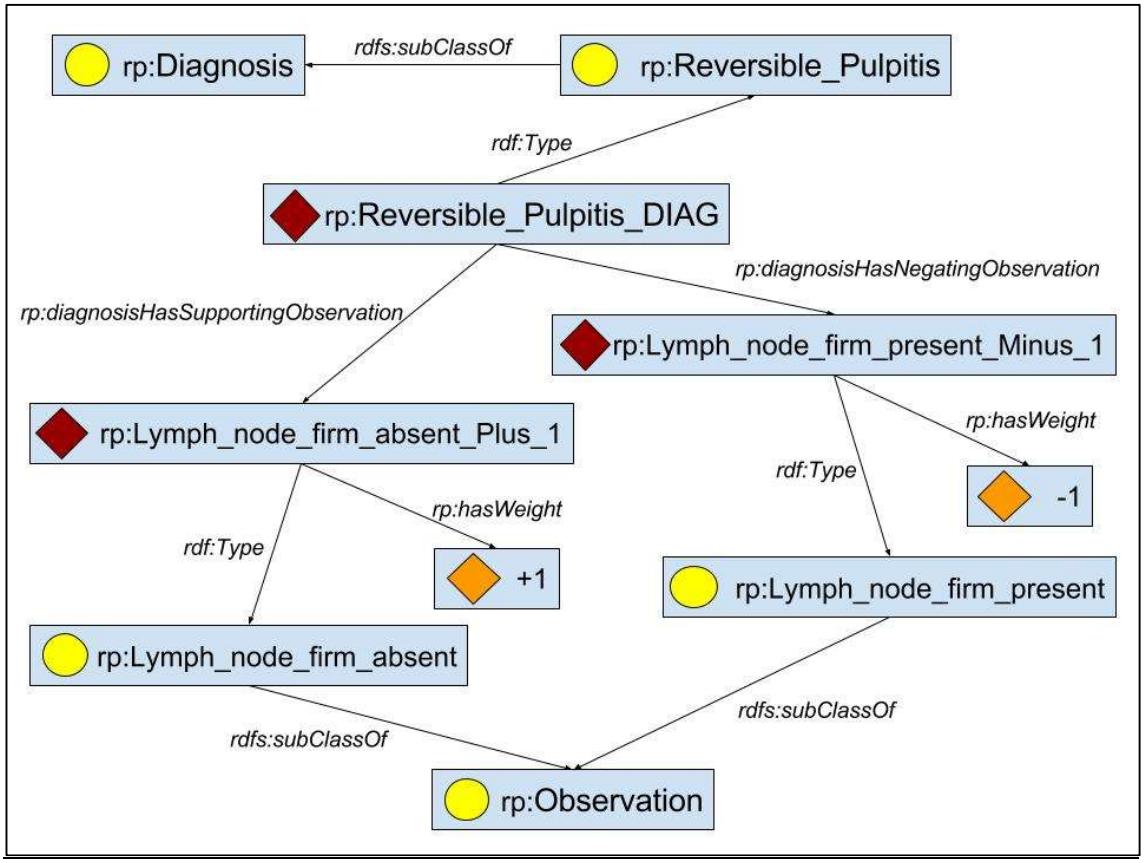


Figure 5-11 - Graph representation of Reversible Pulpitis and observations

The image above (Figure 5-11) shows a graph representation of Reversible Pulpitis and its relations to the different instances of observations that either support or negate the diagnosis. The yellow circles denote Classes, Maroon diamonds denote Individuals and Orange diamonds denote literal values (integers in this case).

rp:Lymph_node_firm_present_Minus_1 is an individual of rdf:type rp:Lymph_node_firm_present (Table 5-3). rp:Lymph_node_firm_present is a subclass of rp:Lymph_node_firm (Figure 5-13). rp:Lymph_node_firm is a subclass of rp:Lymph_node_examination which is subclass of rp:Observation (Figure 5-12).

Figure 5-12 - Observations

The weights have been placed in the name of the individual for the sake of better explaining which individuals belong to which class of observations and explaining their links to a diagnosis in this thesis. In a real-world application the names of the individual can be a coded reference. Having a coded reference will make the process of updating weights easier as the names do not have to be changed. The names visible to the domain experts will be a string literal that can be displayed using the *rdfs:label* property.

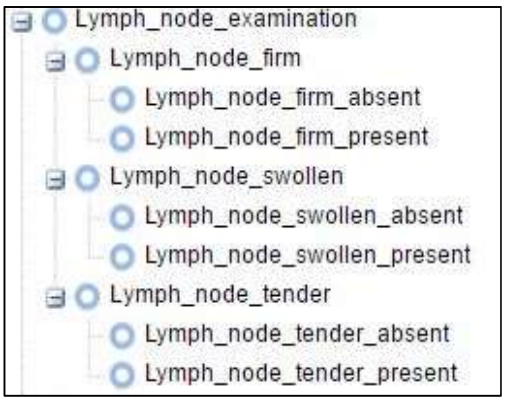


Figure 5-13 - Lymph node firm status

subject	predicate	object
rp:Lymph_node_firm_present_Minus_1	*rdf:type*	owl:NamedIndividual
rp:Lymph_node_firm_present_Minus_1	*rdf:type*	rp:Lymph_node_firm_present

Table 5-3 - Lymph node firm present

rp: Lymph_node_firm_present_Minus_1 has been given a weight of -1, (where the weight can be a positive or negative integer) (Table 5-4). Similarly, rp: Lymph_node_firm_absent_Plus_1 has weight of +1.

The decision to give the weight of +1 and -1 is based on clinical judgement and the experience of the domain expert. Since there are not many studies on argumentation-based CDSSs for diagnostic decision support the weighting scheme used in the Chapter 4 optometry study was used to create the weights for each argument (See Section - 4.2.3.3).

When developing DS Models for clinical use the weights can be determined by an expert panel. The panel can be given a diagnosis and asked to provide the arguments for each diagnosis. In argumentation theory each decision candidate is either supported or negated by arguments either in favour or against a candidate. Similarly, the domain experts in the expert panel can be asked to provide arguments in favour or against each diagnosis. They can be asked to provide a numerical value quantitatively support their argument. An integer of +1 can be provided to arguments that support a diagnosis. An integer of -1 can be given to arguments that negate a diagnosis. **In DS Model the arguments are the presence or absence of observations. If the presence of an observation supports a diagnosis, then +1 can be given to that argument. If the presence strongly supports a diagnosis, then +2 can be given.** If there is conflict among the members of the expert panel regarding the weights then the decision can be put to vote or decided through a group decision making process like the Delphi method (Van de and Delbecq, 1974) to arrive at a consensus. Web based platforms like OpenClinical.net (Fox *et al.*, 2015) have been developed to facilitate this discussion process and allow content creators and domain experts to model computerised clinical guidelines using languages like PROforma. The same platform can be used to create DS Models for each domain with domain experts creating diagnostic models for CDSS applications and using the platform to discuss and resolve conflict and disagreements about the arguments and the weights for each argument.

The argumentation-based approach used in DS Model uses a natural method for explaining how the diagnosis is support by the presence of absence of the observation. The positive integers denote the degree to which the presence or absence of the observations support the diagnosis**. If the presence/absence of an observation highly supports diagnosis, then a weight of +2 is given in the DS Model.** Depending on the domain and the type of diagnosis the weight can be increased**. If the presence/absence of an observation (argument) negates a candidate (diagnosis) then a weight using negative integers can be given.**

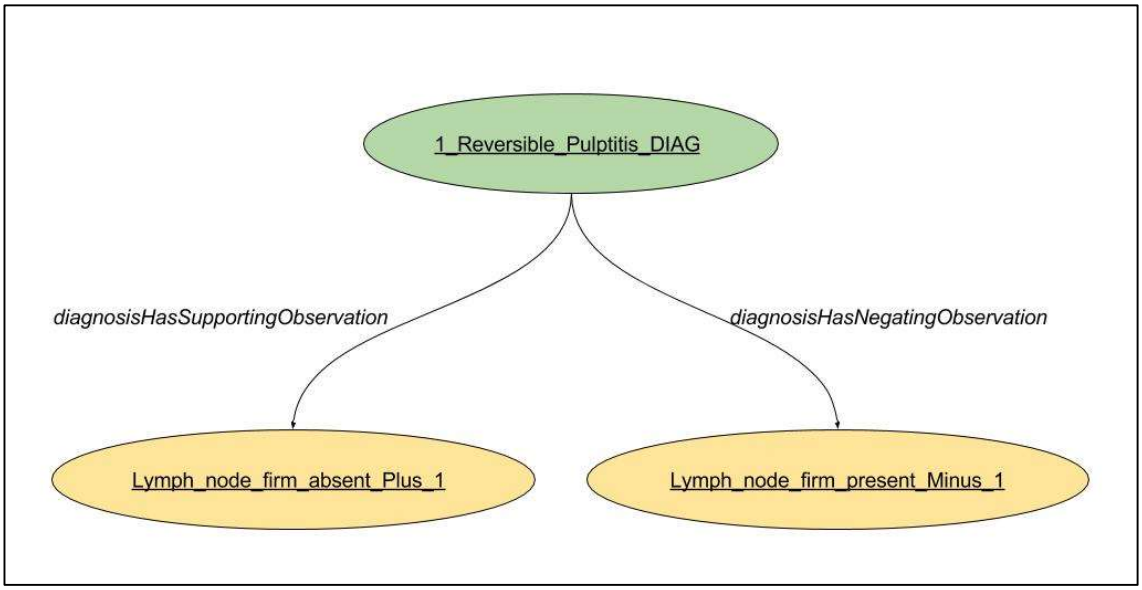


Figure 5-14 - Reversible Pulpitis and Observations relationship

The weights represent the degree to which the instance of that observation supports/negates a diagnosis. In the image above (Figure 5-14) 1_Reversible_Pulpitis_DIAG has a negating observation of Lymph_node_present_Minus_1. Since the presence of a firm lymph node negates the diagnosis of Reversible Pulpitis, if a firm lymph node is present (in the patient data) a SPARQL query (Figure 5-17) subtracts -1 from the total weight attached to OWL individual representing the Reversible Pulpitis diagnosis.

Again, from the image above (Figure 5-14), since Reversible Pulpitis has a supporting observation of rp:Lymph_node_firm_absent_Plus_1, and if a firm lymph node is absent from the record, then the state

of the observation supports the diagnosis of Reversible Pulpitis and adds +1 to the total weight of Reversible Pulpitis.

subject	predicate	object
rp:Lymph_node_firm_present_Minus_1	*rp:hasWeight*	-1
rp:Lymph_node_firm_absent_Plus_1	*rp:hasWeight*	1

Table 5-4 - Lymph node firm weight

Once all the weights are added up, factoring in the observations that are present and absent, a ranked list of diagnoses is obtained with the most probable diagnosis being the diagnosis with the highest total score. The SPARQL queries that add up the weights and produces the ranked list has been described below in the section called SPARQL inference (See section - 5.3.4). The weights are used for ranking the diagnoses. The real benefit of using the argumentation-based approach is the ability to explain the arguments using natural language to the clinician. The clinician can look at the arguments and get a good overall picture of the patient. The arguments can trigger illness scripts and System 1 process of reasoning. This is turn can help the clinician find more information about the patient.

The type of relation between the diagnosis and the observation(s) is called an N-ary relation (Noy *et al.*, 2006). In OWL and RDF, a property connecting 2 individuals or an individual to a value is a binary relation. However, when we must link individuals to more than 1 individuals or values N-ary relations are used. The 1_Reversible_Pulpitis_DIAG diagnosis individual is connected to several other individuals that are types of observations. Each of these individuals is linked to several values that form the weights.

The rp:lymph_node_firm_present or rp:lymph_node_firm_absent observations can be either present or absent depending on the patient data. The observations that are present/absent can be updated directly. It is possible to query patient data in RDF or relational format. If the clinician observes a firm lymph node in the patient then the rp:lymph_node_firm_present observation is present, and the rp:lymph_node_firm_absent observation is absent. This information needs to be updated in the model. If the clinician observes that the patient does not have a firm lymph node, then the rp:lymph_node_firm_absent observation is present (in the patient data), and the rp:lymph_node_firm_present observation is absent.

The following SPARQL insert query (Figure 5-15) will insert the Observation present or Observation absent status into the triple store.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rp: <http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#>
PREFIX wasp:<http://wasp.cs.vu.nl/apdg#>
```

```
INSERT  {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent}
WHERE
        {SELECT ?value ?Observation1
        {?Observation1 a rp:Lymph_node_firm_absent.
        ?s wasp:hasSlot ?o.
        ?o rdfs:label ?slot.
        ?o wasp:value ?value.
                FILTER (str(?slot) = "Lymph_node_firm")
                FILTER (str(?value) = "No")}
}
```

Figure 5-15 - SPARQL present/absent insert query

Using the query above (Figure 5-15), the SPARQL query engine (ARQ) will query the patient data and if the patient data has the value "No", it inserts into the default graph the triple rp:Lymph_node_absent rp:*hasObservationStatus* rp:9_ObservationPresent. The "9" in rp:9_ObservationPresent has been used to distinguish this individual from the other individuals in the DS Model. It does not have any other clinical significance.

The SPARQL insert query can be adapted to conform to any patient data model. In the current example, an archetype model of structuring information for evaluation and testing was used (see section - 5.4.3). An archetype is a model or pattern for capturing clinical information (Leslie, 2012). An archetype model was used within the Advanced Patient Data Generator (APDG) (Huang *et al.*, 2013) tool to generate RDF patient data sets (See Appendix 9). The prefix wasp: denotes the ontology used in the APDG. In the model used in wasp: ontology, information is represented in the form of several archetypes. Each archetype can contain another archetype themselves. The archetype contains a slot which represents a discrete unit of information. For example, the slot "x" contained within the archetype "y" shown in Table 5-5 has the label "Lymph node firm". This slot can have different values such as "Yes" or "No".

Consider the following RDF data:

subject	predicate	object
:y	*rdf:type*	:Archetype
:y	*wasp:hasLabel*	"Lymph node exam"
:y	*wasp:hasSlot*	:x
:x	*wasp:hasLabel*	"Lymph node firm"
:x	*wasp:hasValue*	:No

Table 5-5 - Sample RDF data

The data has been extracted from an EHR that has the following simplified information model. The text highlighted in red below (Table 5-6) has been represented in RDF graph format above (Table 5-5).

Archetype	Slot	Value
Patient Demographics		
Clinical History		
	Pain Present	Yes
		No
Clinical Examination		
Lymph node exam		
	Lymph node firm	Yes
		No
	Lymph node swelling	Yes
		No
	Lymph node tenderness	Yes
		No
Clinical Tests		
Diagnosis and Treatment Planning		

Table 5-6 - Sample information model

Lymph node firm (slot) can take the values "Yes" or "No". If the slot takes the value "No" the SPARQL query above (Figure 5-15) will insert the triple which implies "the rp:Lymph_node_firm_absent observation in the DS model is present".

```
INSERT  {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent}
WHERE
        { SELECT ?value ?Observation1
        {?Observation1 a rp:Lymph_node_firm_present.
        ?s wasp:hasSlot ?o.
        ?o rdfs:label ?slot.
        ?o wasp:value ?value.
                FILTER (str(?slot) = "Lymph_node_firm")
                FILTER (str(?value) = "Yes")}
}
```

Figure 5-16 - SPARQL present/absent conditional insert query

If the patient data has the value "Yes" in the slot for Lymph_node_firm, then the SPARQL query above (Figure 5-16) will insert the triple which implies that rp:Lymph_node_firm_present observation has observation status rp:9_ObservationPresent.

rp:9_ObservationPresent is an individual of class rp:ObservationPresent. Similarly, rp:9_ObservationAbsent is *rdf:type* rp:ObservationAbsent (Table 5-7).

This is how the CDSS using SPARQL present/absent query will update the DS Model based on the data present. The domain expert models the DS Model representing the links between the observations and the diagnosis. **The SPARQL insert query (Figure 5-15 and Figure 5-16) will inform/update the model about the observations that are present/absent in the data. Another SPARQL insert query (Figure 5-17) will sum the weights of the observations that are present and assign them to a diagnosis. Another SPARQL select query (Figure 5-18) will then produce a ranked list of diagnoses based on observations that are present/absent.**

subject	predicate	object
rp:9_ObservationPresent	*rdf:type*	rp:ObservationPresent
rp:9_ObservationAbsent	*rdf:type*	rp:ObservationAbsent

Table 5-7 - Observation present/absent

The SPARQL present/absent insert query above (Figure 5-16) roughly has a "IF..THEN.." rule like representation. This query can be considered a special type of rule called a production rule. A production rule allows us to infer new facts from the existing facts, if the new facts can be expressed in the language (in this case OWL/RDF). When the antecedent condition(s) are satisfied, the rule will fire and the consequent will result in new facts or modifications in the knowledge base.

The query above (Figure 5-16) can be represented in natural language in the following manner: "If the information model has a slot Lymph_node_firm and the slot has value 'Yes' then give the individuals belonging to rp:Lymph_node_firm_present observation in the DS Model the observation status rp:9_ObservationPresent".

The rule can be represented in semantic web rule languages such as Semantic Web Rule Language (SWRL) (Horrocks *et al.*, 2004) and Jena Rules (Apache Jena, 2016). SWRL and Jena is limited to

reasoning over the A-box, or individuals/instances present in the OWL knowledge base. In this project, we are interested in disparate datasets containing patient data and therefore these languages were not considered suitable. The DS Model requires weights to be summed and stored temporarily and retrieved later to provide a ranked list which the rules languages mentioned above (SWRL and Jena) cannot represent due to limitations in expressivity. Rules expressed in SPARQL (SPARQL Rules) were considered more appropriate for the task at hand.

SPARQL Inferencing Notation (SPIN) is language that was built on top of the capabilities that SPARQL (Knublauch, Hendler and Idehen, 2011). The syntax allows us to express business rules in SPARQL without having to implement another rules language. It is possible to store SPARQL queries as triples and store the queries (SPARQL Rules) with the RDF data. In the case studies below (See section - 5.4), the SPARQL rules were written in files or in programming functions. Storing the rules as RDF will make the process of implementing rules easier. However, SPIN was not used/evaluated further in this research project as it was not considered essential for demonstrating the concept of using SPARQL rules for generating diagnostic recommendations.

Once the query engine has queried the patient data and inserted the observation status for all observations the following query (Figure 5-17) will sum the weights and rank the diagnosis by the total score. The weights can be determined through statistical approaches. The weighting can be determined from evidence obtained from literature from clinical and epidemiological studies. However, even though some studies have shown that clinicians use prevalence and probability when making a diagnosis (Elstein, Shulman and Sprafka, 1978) other observational and experimental studies have shown that the clinicians do not routinely use statistics in diagnostic decision making (Dawes, Faust and Meehl, 1989).

5.3.4 SPARQL Inference

```
INSERT { ?Diagnosis rp:hasDiagnosisObsWeight ?DiagnosisObsWeight}
WHERE
        {SELECT (SUM (?Weight)AS ?DiagnosisObsWeight) ?Diagnosis
        { ?Diagnosis ?property ?Observation.
         ?property rdfs:subPropertyOf rp:Observation_Diagnosis_properties.
        ?Observation rp:hasWeight ?Weight.
        ?Observation rp:hasObservationStatus rp:9_ObservationPresent.
        }
GROUP BY ?Diagnosis
 }
```

Figure 5-17 - SPARQL insert query calculating weights

The query above (Figure 5-17) sums the weight of all observations present. Here the observation present refers to the observation class in the DS Model ontology, and not the observations present in the patient. The data may have the value Lymph_node_firm = "No", which means there is no firm lymph nodes in the patient. However, this means that the rp:Lymph_node_firm_absent in the DS Model will have the status rp:9_ObservationPresent .

The sum of all weights of observations either supporting or negating a diagnosis is added to a diagnosis using the insert query above (Figure 5-17) which adds the following triple to the triple store=>

?Diagnosis rp:*hasDiagnosisObsWeight* ?DiagnosisObsWeight.

The ?Diagnosis variable is any individual that belongs to the class rp:Diagnosis. Once all the instances of the Diagnosis have the weight attached to it, the following queries (Figure 5-18 and Figure 5-19) will rank them in descending order of weights. The Pulpal and Apical diagnoses are ranked separately as the clinician will be interested in the current status of the pulp and if the inflammation has spread to the apex (root end) of the tooth.

```
SELECT ?PulpalDiagnosis ?DiagnosisWeight
WHERE
        {?PulpalDiagnosis  rp:hasDiagnosisObsWeight ?DiagnosisWeight.
        ?PulpalDiagnosis a ?Diagnosis.
        ?Diagnosis rdfs:subClassOf* rp:Pulpal_Disease.
} ORDER BY desc (?DiagnosisWeight)
```

Figure 5-18 - Pulpal diagnosis ranked list SPARQL select query

```
SELECT ?ApicalDiagnosis ?DiagnosisWeight
WHERE
        {?ApicalDiagnosis  rp:hasDiagnosisObsWeight ?DiagnosisWeight.
        ?ApicalDiagnosis a ?Diagnosis.
        ?Diagnosis rdfs:subClassOf* rp:Apical_Disease.
} ORDER BY desc (?DiagnosisWeight)
```

Figure 5-19 - Apical diagnosis ranked list SPARQL select query

The following tables (Table 5-8 & Table 5-9) describe the results obtained after running the DS Model against a test scenario/vignette that has a Pulpal diagnosis of 'Symptomatic Irreversible Pulpitis' and Apical diagnosis of 'Symptomatic Apical Periodontitis'. In a real-world clinical application, the names of the individuals of the DS Model will not be displayed. In its place the rdfs:label of the instance/class can be displayed to make the results more clinician friendly (Hayes and Patel-Schneider, 2014).

Pulpal Diagnosis	**Diagnosis Weight**
rp:1_Symptomatic_Irreversible_Pulpitis_DIAG	14
rp:1_Pain_of_Nonodontogenic_Origin_DIAG	9
rp:1_Asymptomatic_Irreversible_Pulpitis_DIAG	6
rp:1_Pulp_Necrosis_DIAG	4
rp:1_Reversible_Pulpitis_DIAG	3

rp:1_Normal_Pulp_DIAG	2
rp:1_Dentinal_Hypersensitivity	-4

Table 5-8 - Pulpal Diagnosis

Apical Diagnosis	Diagnosis Weight
rp:1_Symptomatic_Apical_Periodontitis_DIAG	10
rp:1_Pain_of_Nonodontogenic_Origin_DIAG	9
rp:1_Acute_apical_abscess_DIAG	5
rp:1_Normal_Apical_Tissues_DIAG	4
rp:1_Asymptomatic_Apical_Periodontitis_DIAG	3
rp:1_Chronic_Apical_Abscess_DIAG	1

Table 5-9 - Apical Diagnosis

The DS Model needs to know which observations are present/absent to make a diagnostic decision. The observations can be 'Pain Present' observation or the 'Pain Absent' observation. The key aspect of the DS Model is that 'Pain Present' is the observation. It is easy to misrepresent 'Pain' as the observation when talking in clinical terms. However, in the DS Model 'Pain Present' observation can have the state of being present/absent. Once the clinician has completed the enquiry/examination he/she records the data in EHR or CDSS application directly.

The argumentation-based diagnostic recommendations in the Chapter 4 Optometry study provided a ranked diagnostic recommendation alongside the weights. The arguments for each diagnosis is presented below the diagnosis (See Figure 4-20). **The SPARQL query in Appendix 21 details how the DS Model can provide explanations/justifications for each diagnosis.** For example, if the patient is complaining of pain the dentist will record in the EHR that pain is present. The CDSS will look at the patient data and set the Pain_present_Plus_1 observation status to present. The ranked diagnostic recommendations are generated. The SPARQL query in **Appendix 21** can then detail how each Diagnosis is linked to the Observation via the property. The query will provide the result in a triple format such as rp:1_Symptomatic_Irreversible_Pulpitis_DIAG rp:diagnosisHasSupportingObservation rp: Pain_present_Plus_1. This means that Symptomatic Irreversible Pulpitis has a supporting observation called Pain Present with a weight of +1 and this observation is present in the patient. A natural language description of the result can be provided to the dentist. For example, the explanation can include statements like 'Since the patient has complained of pain this supports the Symptomatic Irreversible Pulpitis diagnosis'. This argumentation-based approach allows the dentist to get a good overall picture of the patient. Similar arguments against the diagnosis can be given. The dentist can then use this information to compare with his/her mental model of the patient and find more information as needed. The DS Model allows us to generate these arguments without the need for specialized software as seen in the Tallis CDSS.

5.3.5 Summary

This section details the development of the DS Model. It demonstrates how argumentation has been used to generate diagnostic recommendations. The next section details the development of CDSS prototypes that can use the DS Model to generate diagnostic recommendations.

Argumentation-based CDSS tools have been shown to improve decision making when modelled using PROforma to make guideline based CDSS tools (Fox *et al.*, 2007). The DS Model uses argumentation theory to model the arguments using semantic web technology and helps integrate information from multiple sources easily. The main advantage of using argumentation-based methods to model the arguments in DS Model to generate diagnostic recommendations is to better explain the reasons given by the CDSS for each recommendation. The arguments for each diagnosis and the weights are captured in a single model that can be easily understood by the domain expert during modelling process. When data from multiple sources are integrated the CDSS can rapidly update the recommendation considering latest information and inform the clinician about the change in data and explain in simple terms why the recommendation was made. The simple technique of summing the weights means there is no need for special software (for example Bayesian reasoning engine) to perform calculations. Any triple-store capable of simple RDFS reasoning and process SPARQL queries can perform these calculations. PROforma based models require Tallis graphical user interface to model the arguments. The Tallis reasoning engine is required to perform reasoning and generate the diagnostic recommendations as demonstrated in the Chapter 4 optometry study. The DS Model can be modelled using existing ontology modelling tools such as Protégé and Web-Protégé and can be used to make the process of creating arguments that support or negate a diagnosis easier and to help domain experts develop DS Models in their domain and help them discuss and collaboratively make changes to the model where necessary. Web-Protégé has been developed to help better support collaborative development of OWL ontologies (Tudorache *et al.*, 2013). Any triple-store capable of accepting RDF data and has a SPARQL processor can generate the diagnostic recommendations. **The simplicity of the modelling process using argumentation-based methods, the simple weighing scheme used to assign weights, the simple technique used to sum the weights and ability to use existing off-the-shelf semantic web tools to generate the diagnostic recommendations gives the DS Model and the CDSS several advantages over conventional methods for generating diagnostic recommendations.**

Clinical knowledge keeps on changing and the arguments and the weight for each diagnosis will needs updating when needed. The **Learning Healthcare System** is a proposed system where data is frequently collected from the EHR of every patient in the system, routinely analysed to answer specific clinical questions and used to continuously learn from real-world data (Friedman, Wong and Blumenthal, 2010). The results of this analysis are fed back into the system where the outcomes are routinely monitored and the data in analysed further. This results in a continuously learning healthcare system that can in theory lead to better quality, patient outcomes, effectiveness and efficiency in the healthcare system and more patient-centred care.

CDSSs are a central part of this learning healthcare ecosystem (Foley and Vale, 2017). CDSSs help deliver up-to-date patient-centred content and recommendations directly to the clinician. The results of analysis of patient data in a learning healthcare system can be used to update the logic and content of the CDSSs and the system can monitor patient data to see how the change has affected the clinical outcomes and understand where the CDSSs models can be improved. Platforms like **OpenClinical.net** (Fox *et al.*, 2015) are collaborating with organisations like the **Learning Healthcare Project** (The Learning Healthcare Project, 2017) to design and develop machine readable content for deploying in a learning healthcare system (Fox, 2015).This content can then be implemented in clinical environments and the outcomes analysed using machine learning and other data analytics methods and changes needed are studied and pushed to the clinical environment and the outcomes analysed further.

The DS Model can help in this process as the model is machine readable and understandable and built using existing semantic web technologies. The OWL ontology-based DS Model can be published on the web and CDSS applications can retrieve the DS Model using a SPARQL query to run the model on their

patient data. When changes are needed the model can be updated and all CDSS applications using the model will receive the updates. Since the DS Model has been built using a linked data framework the main DS Model can be updated using a SPARQL query and all the CDSSs using the outdated version of the DS Model can be updated using another SPARQL query. RDF uses URIs to uniquely identify resources on the web. This makes it easy to link existing resources. The resource defined in the DS Model will have a unique identification. For example, the resource rp:1_Symptomatic_Irreversible_Pulpitis_DIAG and rp:diagnosisHasSupportingObservation and rp: Pain_present_Plus_1 all have a unique resource identifier. Intelligent agents can find the resource and find all information associated with the resource. The published DS Model is inherently more suited for sharing and collaborating than the proprietary format used in current generation PROforma and other argumentation-based CDSS tools.

Healthcare data interoperability and a proper data infrastructure is central to the proper functioning of the learning healthcare system. Semantic web-based systems can help improve interoperability by providing support for semantic interoperability. Semantic web-based systems can help in machine learning and analysis. A Semantic Data Lake provides a semantic layer comprising of tools and models on top of existing systems that can integrate information from multiple systems (Archer, 2017). A SPARQL query can then find the individual bits of information needed for answering patient specific questions. **Semantic Data Lakes** have been setup to integrate data in healthcare systems such as the Montefiore Health System for data analysis (Mirhaji, 2016). Deep learning with the help of machine learning algorithms and better predictive analytics are therefore possible as the data is integrated from several different sources and includes data that is unavailable from a single source. The DS Model, patient data and all other information needed to make a clinical decision becomes part of the linked web of data when integrating data using platforms like semantic data lakes. An intelligent agent can traverse this linked web of data to find the information needed and help the CDSS using the DS Model to make a clinical decision.

To develop DS Models in other domains from scratch, the relations between the Diseases and their symptoms can be reused from existing ontologies such as **SNOMED-CT** or other ontologies developed specifically to map these relations (Bertaud-Gounot, Duvauferrier and Burgun, 2012a; Mohammed, Benlamri and Fong, 2012). When ontologies like SNOMED-CT changes the relations between the Diagnosis and Observations in the DS Model can be updated automatically as the content is machine readable and the content in the DS Model can be semantically annotated and linked with the base ontology model from which the relations were obtained. Domain experts can then evaluate the changes to see if the weights need to be changed or if the arguments need updating. Once the domain experts have approved the changes and the DS Model modified, the updated DS Model (in a domain) can be pushed to the web and all CDSS applications using (subscribing) the DS Model will receive the updated DS Model.

If the DS Model is part of a learning healthcare system, the clinicians and patients can be informed about the changes and the system can monitor outcomes (patient and clinician) to see if the changes have caused any adverse impact. This provides a system where data from the EHR and other electronic patient record sources can provide the real-world evidence needed to provide constant monitoring and upgradation of the DS Model and the real-world impact of the DS Model can be monitored and changes made as necessary.

The next section provides an overview of the CDSS prototypes development.

5.4 Decision support prototype development

The DS Model described in the earlier sections was tested using several different CDSS architectures. These architectures will be introduced in the form of 4 cases studies detailing the development of prototypes. A summary evaluation of these architectures will be provided. The following case studies details the different ways in which the clinicians can record the observations and how the CDSS application using the DS Model will be able to provide diagnostic recommendations. The most common ways in which patient data can be accessed has been demonstrated using the case studies. The DS Model

development focusses on the model and how argumentation was used as the basis to generate diagnostic recommendations. The case studies show how the DS Model can be used in real-life settings to integrate patient information and generate diagnostic recommendations.

Case Study 1 prototype uses a **triple store** as the backend to store the patient data. It demonstrates how a standalone CDSS application can use the DS Model to generate diagnostic recommendation. One of the key advantages of the DS Model when compared to the existing argumentation-based tools like Tallis is the ability to use off-the-shelf reasoning available with most triple stores. It demonstrates how data annotated with existing ontologies and terminology can be used in conjunction with the DS Model. This architecture was chosen because stand-alone CDSSs that are not connected to EHR are commonly used in today's clinical practices.

Case Study 2 demonstrates how data stored in **relational database** can be translated to RDF format using the **D2RQ method**. This case study demonstrates how linked data technology can be used in conjunction with the DS Model to integrate information from multiple sources. The D2R server exposes the data in a MySQL relational database in the form of a SPARQL endpoint. Existing applications using relational databases as their backend can expose their data via the SPARQL endpoint. Linked data technology can then identify data that belongs to the patient, provided the data is described using standard ontologies. The CDSS application can then use this data to generate early diagnostic recommendations and do not have to rely entirely on the data provided by the patient during the clinical encounter. Case Study 2 demonstrates how a CDSS application using the DS Model can use the data via a SPARQL endpoint to provide diagnostic recommendations. Linked data can be stored in triple stores or relational database and exposed via a **SPARQL endpoint**. The Case Study 2 prototype application can then access this data to provide diagnostic recommendations using the **SERVICE keyword** within the SPARQL query. The patient data can be stored in multiple SPARQL endpoints, and the CDSS application can access this data. Relational databases are a common form of storage in healthcare information systems. The architecture demonstrates how the DS Model can access this data via a SPARQL endpoint.

Case Study 3 shows how the data in a EHR can be represented in a standardised manner using **archetypes** can be used by the CDSS to generate diagnostic recommendations. A published dental information model was represented as archetypes and separated into different sections. The different sections represent the major steps in the diagnostic encounter. This cases study shows how the CDSS using the DS Model can integrate information from existing **EHRs** that used **information standards** to represent information. Case Study 3 shows how the DS Model helps the CDSS use different forms of patient data by modifying the SPARQL rules. This demonstrates how the DS Model can work with current healthcare information systems using standard information models.

Case Study 4 expands on the architecture described in Case Study 2 and used the dental information model to represent patient data in an **open-source EHR**. The data stored in the relational database of the EHR was then exported into the RDF format using software that translated relational data to RDF. The RDF data was then uploaded to a triple store alongside the DS Model and the patient data was used to generate diagnostic recommendations. Case Study 4 shows how the patient data stored in currently available and widely used open-source EHR systems can be used by the CDSS using the DS Model.

5.4.1 Case study 1-Stand-alone CDSS implementation

The first prototype (Figure 5-20) is the simplest method of deploying the DS Model and obtaining diagnostic recommendations. The prototype was built using the RDFLib library (RDFLib Team, 2013). RDFLib is a Python library that is used for working with RDF graphs.

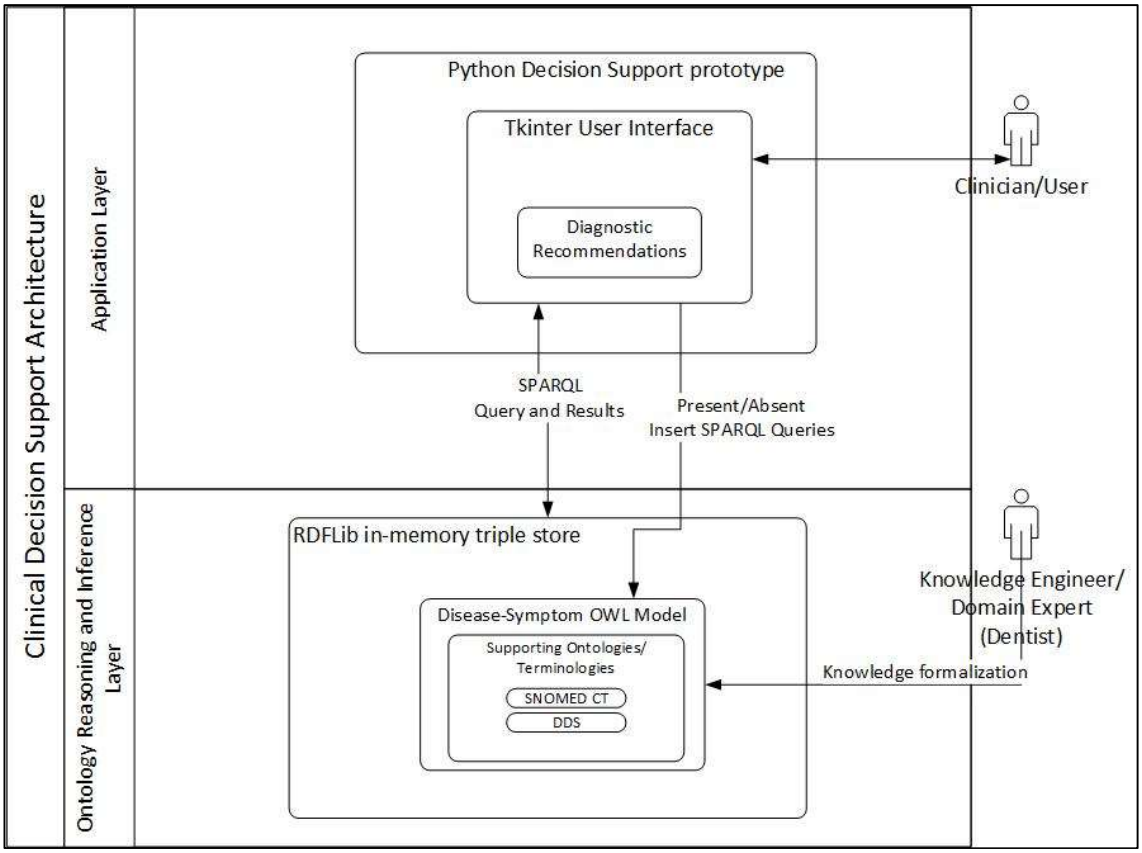


Figure 5-20 - Case Study 1- CDSS architecture

Since the DS Model is based on OWL/RDF language this library was chosen. Python was chosen because of the ability to develop simple Graphical User Interfaces (GUI). The author has more experience writing software in the Python language.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rp: <http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#>
```

Figure 5-21 - Prefixes for Select/Insert SPARQL queries

All the queries in this case study will use the prefixes defined above (Figure 5-21). The rp: prefix is the prefix used to denote the DS Model. The current dental DS Model includes the base classes of the DS Model and the dental classes required for dental decision support. All cases studies use the same dental implementation of the DS Model. The SPARQL queries may vary depending on the application. Some additional prefixes will be needed for other case studies and will be indicated accordingly. All other prefixes are standard ontologies, schemas or specifications used in OWL/RDFS/RDF based applications.

The DS Model is added to the triple store when the application is initiated. An empty graph model is created and the triples from the DS Model RDF file are added to the default graph. The present/absent insert queries are then added to this default graph that contains the DS Model.

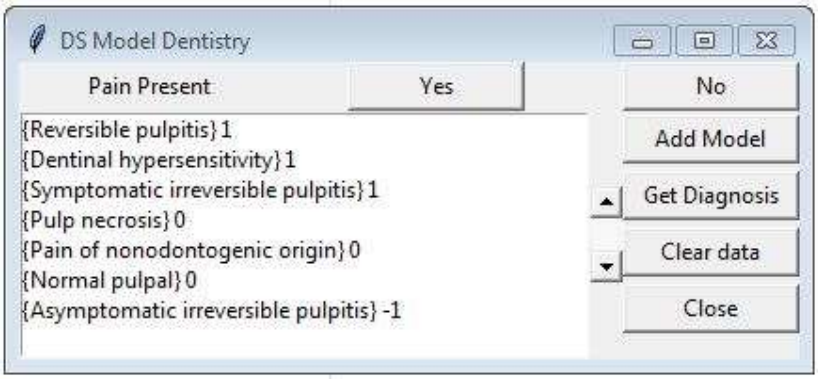


Figure 5-22 - Case Study 1 user interface diagnosis recommendations

When the clinician clicks on the Pain Present- 'Yes' or the Pain Present- 'No' button (Figure 5-22) a SPARQL insert query is executed on the in-memory triple store, like a SQL query executed on a relational database to store a data item. Figure 5-23 below shows an example of the present/absent insert query. Here the SPARQL query inserts the triple {Observation1 *rp:hasObservationStatus* rp:9_ObservationPresent} to the triple store. Observation1 variable denotes all the individuals that belong to the class rp:Pain_present.

```
INSERT  {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent}
WHERE
        {?Observation1 a rp:Pain_present.
}
```

Figure 5-23 - Pain Present insert query

The rp:Pain_present class is a subclass of the rp:Pain_status class. The sibling class of the rp:Pain_present class is the rp:Pain_absent class (Figure 5-24).

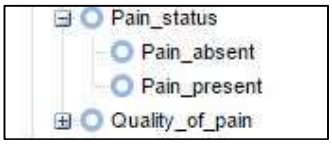


Figure 5-24 - Pain status class

The subclasses of the rp:Pain_status class indicates the different states that the rp:Pain_status observation can exist in. The Pain_present class has different individuals that belong to the class depending on the weights that the individual may have. In this example the rp:Pain_present class has 2 individuals rp:Pain_present_Plus_1 and rp: Pain_present_Minus_1 (Figure 5-25).

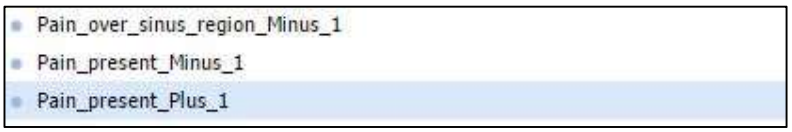


Figure 5-25 - Pain present individuals

rp:Pain_present_Plus_1 has the owl:DatatypeProperty *rp:hasWeight* with the value 1. Similarly, the rp:Pain_present_Minus_1 individual will have the weight -1. The figure below (Figure 5-26) shows the description for the individual rp:Pain_present_Plus_1. The ontologist or modeller can add SNOMED-CT

terminology codes as a property. In this example the SNOMED-CT concept code: 27355003 was used to annotate the individual. This concept code indicates a Toothache finding (SNOMED-CT, 2016) . Existing ontology definitions and terminologies can be used to link the terms in the DS Model to these terminologies. However not all classes/individuals in DS Model can be annotated as concepts like Pain_absent may not have an equivalent term.

Figure 5-26 - Pain present Plus 1 description of individual

When the decision support application inserts the triple defined in Figure 5-23 , using SPARQL inference all the individuals of the rp:Pain_present class will have the owl:ObjectProperty *rp:hasObservationStatus* with the value *rp:9_ObservationPresent*. rp:9_ObservationPresent is an individual that belongs to the rp:ObservationPresent class.

In this prototype implementation, only the Pain Present (Yes/No) functions were implemented (Figure 5-22). In a full CDSS application, all enquiries and corresponding values will be presented to the clinician in the form of buttons, drop-down menus or radio-buttons. Once the clinician has entered all the data required he/she can press the 'Get Diagnosis' button to obtain a ranked list of possible diagnoses and their weights (Figure 5-22).

When the clinician clicks on the 'Get Diagnosis' button another SPARQL insert query is sent to the triple store that sums up all the weights of the observations that are present/absent. This query has been explained previously in this chapter (Figure 5-17). The weights are only summed for observations that have the *rp:hasObservationStatus* with the value rp:9_ObservationPresent. All other observations are ignored.

The weights are inserted into the triple store and can be retrieved later if needed. It is possible to store the weights from an early diagnosis and compare it to the final diagnosis. Once the weights have been calculated and attached to each instance of a diagnosis, another SPARQL select query is sent that retrieves the ranked list of diagnosis and the attached weights. The pulpal and apical diagnoses are retrieved as separated ranked lists. The Pulpal and Apical diagnosis ranked list select query has been previously described in Figure 5-18 and Figure 5-19.

This architecture used in this case study differs from the other case studies below in that the DS Model is inserted into an empty graph and the CDSS prototype informs the model directly about observations that are present/absent. This is a much simpler architecture to implement but can only be used for stand-alone applications. The files for Case Study 1 can be accessed via Appendix 6.

5.4.2 Case Study 2-Stand-alone CDSS with relational database backend

In Case Study 1 the observations that are present/absent are recorded directly in an in-memory graph that is created during run-time. This configuration will be suitable for creating stand-alone CDSS applications. However, the most common way in which data is stored by EHR and CDSS applications is in a relational database. Relational databases have a long history of implementation (Borok, 1995), are more reliable and several clinical applications store their data in relational databases routinely. NoSQL databases, graph databases and RDF triple stores have their advantages when dealing with data in different formats but will take several years to reach a critical mass of user adoption (Leavitt, 2010).

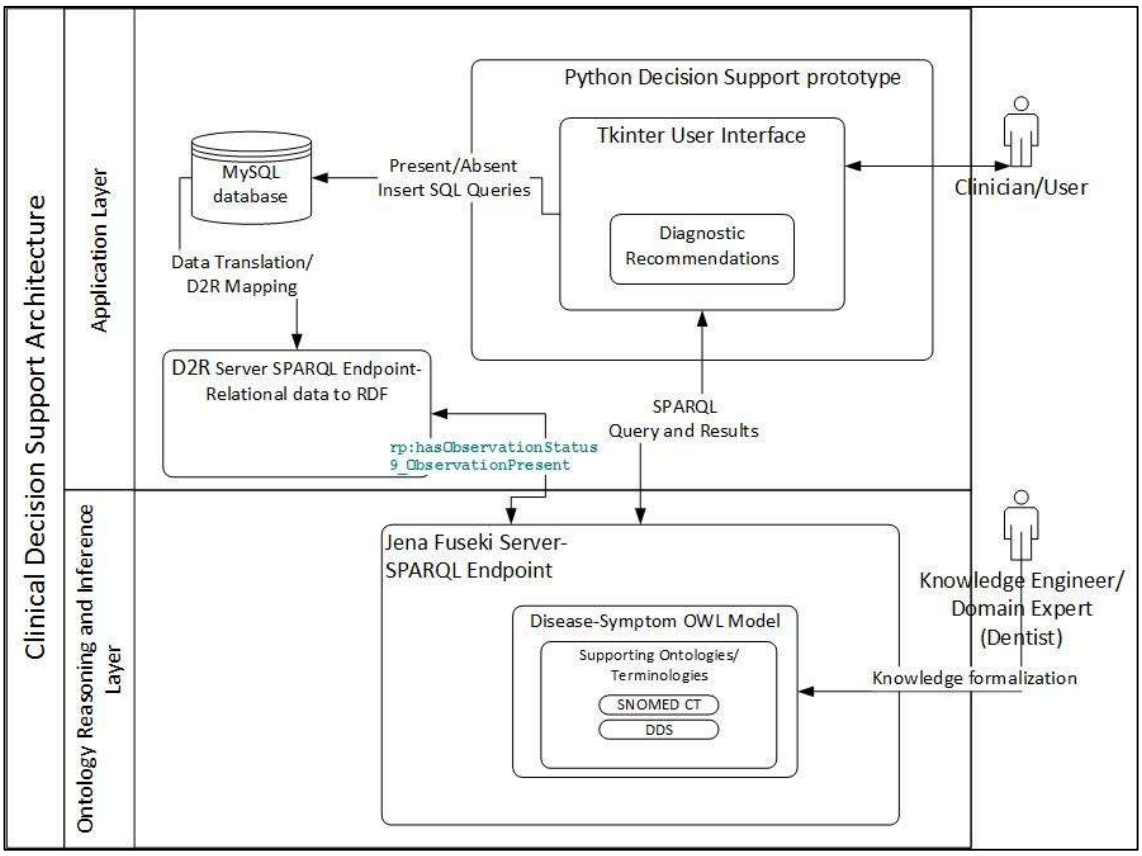


Figure 5-27 - Case Study 2- CDSS architecture

Since the DS Model is based on OWL/RDF how can the model be implemented with a CDSS application that uses a relational database as a backend to store patient data? In this case study, the application that was developed in Case Study 1 was modified to store the data in a relational database (MySQL) instead of a triple store (Figure 5-27). The clinician can complete history, physical examination, clinical tests etc. and can record which observations are present/absent in the relational database.

D2RQ was used to translate the relational data into RDF triples (C. Bizer and Seaborne, 2004). D2RQ method uses the D2RQ mapping language that describes the relationship between the database schema and OWL ontologies. In this case study, the data in the relational database was mapped to the classes/individuals in the DS Model.

Obs_id	Pain_present_name	hasWeight	hasObservationStatus
111	Pain_present_Plus_1	1	9_ObservationPresent
112	Pain_present_Plus_1	2	9_ObservationPresent
113	Pain_absent_Plus_1	1	9_ObservationAbsent

114	Pain_absent_Plus_1	2	9_ObservationAbsent

Table 5-10 - Case Study 2 prototype database pain_present table

The clinician interacts with the user interface of the prototype. When he/she selects the Pain Present 'Yes'/ 'No' button the corresponding data is entered the table in the database (Figure 5-28). The schema of the test MySQL database table 'pain_present' is shown above (Table 5-10).

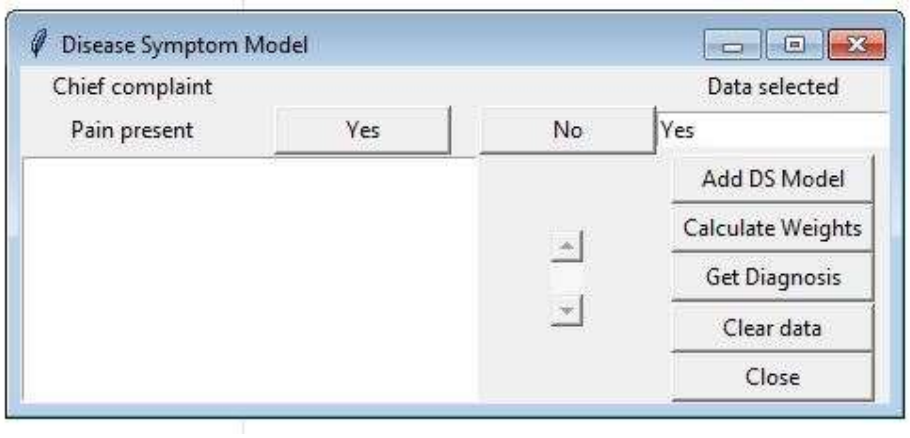


Figure 5-28 - Case Study 2 user interface 'Yes' selected

The data is translated to the virtual RDF graphs using the D2R mapping language (Cyganiak *et al.*, 2012). The mapping/translation creates individuals in the D2R virtual RDF graph that has the object property *rp:hasObservationStatus* and value "9_ObservationPresent". For example, rp:Pain_present_Plus_1 *rp:hasObservationStatus* "9_ObservationPresent" is created in the D2R server representing the same individuals from the DS Model. In this case study the translated RDF data is exposed in the form of a SPARQL endpoint. The SPARQL endpoint accepts SPARQL queries and returns the result in different formats (XML/JSON) that can be consumed by software agents. In addition, the translated data can be made available in the form of an RDF data dump that can be uploaded to a local triple store for analysis.

Once the patient data is available in RDF format the CDSS application can then query the SPARQL service for data that is present/absent either directly or via another SPARQL endpoint as demonstrated in the CDSS architecture above (Figure 5-27).

The DS Model is added to a Jena (Carroll *et al.*, 2004) triple store when the clinician clicks the 'Add DS Model' button. A Jena Fuseki 2.3.0 server was used to store the DS Model.

D2R server and endpoint is "read only" at present. We cannot insert data directly into the triple store using SPARQL insert queries. It should be done via the MySQL database.

The prototype in this case study copies the DS Model from a named graph into the default graph. The DS Model can therefore be stored on the web and a copy retrieved via HTTP web protocol and stored in the local triple store during runtime. We can then calculate weights using a similar SPARQL insert query mentioned before. This query sums the weight of all the observations that have *rp:hasObservationStatus* "9_ObservationPresent". The query instructs the Fuseki server to query the D2R server endpoint (which contains the patient data) for the right value and calculates the weights. The change from the earlier SPARQL insert query is the extra information added to the SPARQL query pointing the D2R SERVICE (See text highlighted in Red in below - Figure 5-29).

```
INSERT {?Diagnosis rp:hasDiagnosisObsWeight ?DiagnosisObsWeight}
WHERE
    {SELECT (SUM (?Weight)AS ?DiagnosisObsWeight) ?Diagnosis
        {?Diagnosis ?property ?Observation.
        ?property rdfs:subPropertyOf rp:Observation_Diagnosis_properties.
        SERVICE<http://localhost:2020/sparql>{
        ?Observation rp:hasWeight ?Weight.
        ?Observation rp:hasObservationStatus "9_ObservationPresent".
     }}
    GROUP BY ?Diagnosis
    }
```

Figure 5-29 - SPARQL insert query calculating weights for Case Study 2

In this case study a string literal was used as the object (Subject-Predicate-Object) for the property(predicate) *rp:hasObservationStatus*. rp:9_ObservationPresent can be used as the object as mentioned in Case Study 1.

The weights are then attached to the diagnosis that is either supported/negated by the presence of the observation. A SPARQL select query is then sent to the Jena Fuseki server by the Python RDFLib application which retrieves a ranked list of the diagnosis (Figure 5-30). RDFLib has a library called SPARQLWrapper (Fernández *et al.*, 2016) that provides an interface to access SPARQL endpoints. The select query is the same as mentioned before (Figure 5-18).

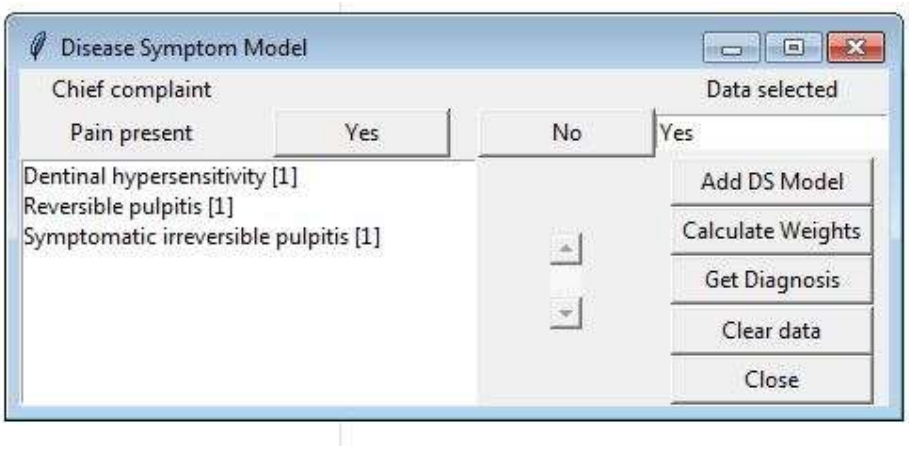


Figure 5-30 - Case Study 2 user interface diagnosis recommendations

This approach demonstrated in this case study has several advantages. In addition to the advantages of using relational databases, the patient data can reside in several different RDF data stores. This is important for early and final diagnosis. In addition to the data that is collected as part of a routine clinical encounter, additional data can be collected from other patient data repositories to help in formulating a diagnosis. Intelligent multi-agent systems can access all the data of the patient using unique patient identifiers and combine this information with information collected during the current encounter to provide a better picture of the patient's condition than relying only on the information collected during the clinical encounter. This is one of the main advantages of using SPARQL endpoints. The D2R server described in this case study can be located remotely, and the Fuseki server with the DS Model and diagnostic recommendations can provide its services remotely using HTTP protocol for any healthcare linked data applications. The approach described in Case Study 1 is limited as the in-memory triple store cannot access remote SPARQL endpoints, nevertheless is still useful for building standalone CDSS applications.

The ability to access remote endpoints is one of the distinguishing features of semantic web or linked data-based systems. The idea of the Web of Linked Data is to form a single, globally distributed dataspace (D'Aquin *et al.*, 2008; Hartig, Bizer and Freytag, 2009). This is enabled by finding resources with the help

of URIs (Uniform Resource Identifiers) and accessed via HTTP. An entity residing in one dataset can be linked to another entity via the URI. A smartphone app collecting information about patient's dental pain can annotate their data using terms from established ontologies and vocabularies or even terms defined in the DS Model (rp: 9_ObservationPresent). The data stored in the relational database of the smartphone application can be translated to RDF and exposed with the help of a SPARQL endpoint. A CDSS can then access this data and provide the clinician with a history of previous dental pain that is more accurate than what the patient will be able to remember from memory alone. Similarly, the concept of dental pain can be linked to all information available in the web of linked data and relevant facts can be retrieved and presented to the clinician in an early diagnosis situation. The clinician does not have to retrieve the information himself/herself. The CDSS application can use this information to support clinical decision making.

This approach is different from the traditional methods used to access information from disparate sources of data. The traditional approach uses proprietary Web APIs to access the data. The problem with this approach is the data is locked away in small data islands. When the underlying data structure changes, the API changes. The data representation formats are varied. The data can be stored in the form of structured, semi-structure or unstructured data. There are significant costs involved in linked these disparate sources together using APIs to produce mashups. Since the cost of linking is high, developers have not fully exploited the use of data available on the web to support clinicians. It is physically impossible for the clinician to manually query these databases to find data appropriate for clinical decision making.

Using semantic web and linked data technologies the data locked away in these data islands are represented in a common data representation format (RDF) and accessed via HTTP with the help of URIs. Linking different entities together and describing them using OWL ontologies allows the intelligent agents and software applications to discover facts that was previously not considered. For example, linking the presence of the dental pain to medication that the patient had previously taken will allow the decision support software to find out more information about this medication and find other medication that could be helpful for this patient based on profiles of similar patients. No specialized Web API is therefore needed to access the data.

Querying multiple sources is shown in Figure 5-29. The data in the sample patient database is exposed via a SPARQL endpoint. The method of using SPARQL to query disparate linked databases is called Federated SPARQL querying (Rakhmawati *et al.*, 2013). Using federated SPARQL queries we can define using the service keyword the service or SPARQL endpoint (Figure 5-29) that the client needs to access to retrieve the data. In this case study, a local endpoint was used for demonstration purposes.

Standardised representation of patient data will allow intelligent agents to access patient data in EHR systems. Current web based health informatics standards such as HL7 FHIR (Fast Healthcare Interoperability Resources) (HL7, 2016) try to solve the problem of semantic interoperability via standardised information representation accessed via RESTful APIs (Rob Zazueta, 2013). Data source using the FHIR standard can still be accessed via HTTP. However, since the entities in the data are not represented using RDF and linked using URI the intelligent agent or decision support application cannot access other data sources available on the wider web of data. There are around 600,000 query able datasets (Verborgh, 2015) currently available via the Linked Open Cloud (Cyganiak, 2014) which can be potentially accessed using linked data technologies. The files for Case Study 2 can be accessed via Appendix 7.

5.4.3 Case Study 3-Patient data vignettes in RDF format

In Case Study 1 and 2 the DS Model was informed about present/absent observations directly, either by inserting triples into the triple store or by translating relational data by mapping data to DS Model ontology. However, in real life scenarios patient data conforms to different information models. The decision support model must demonstrate interoperability with all these different information models.

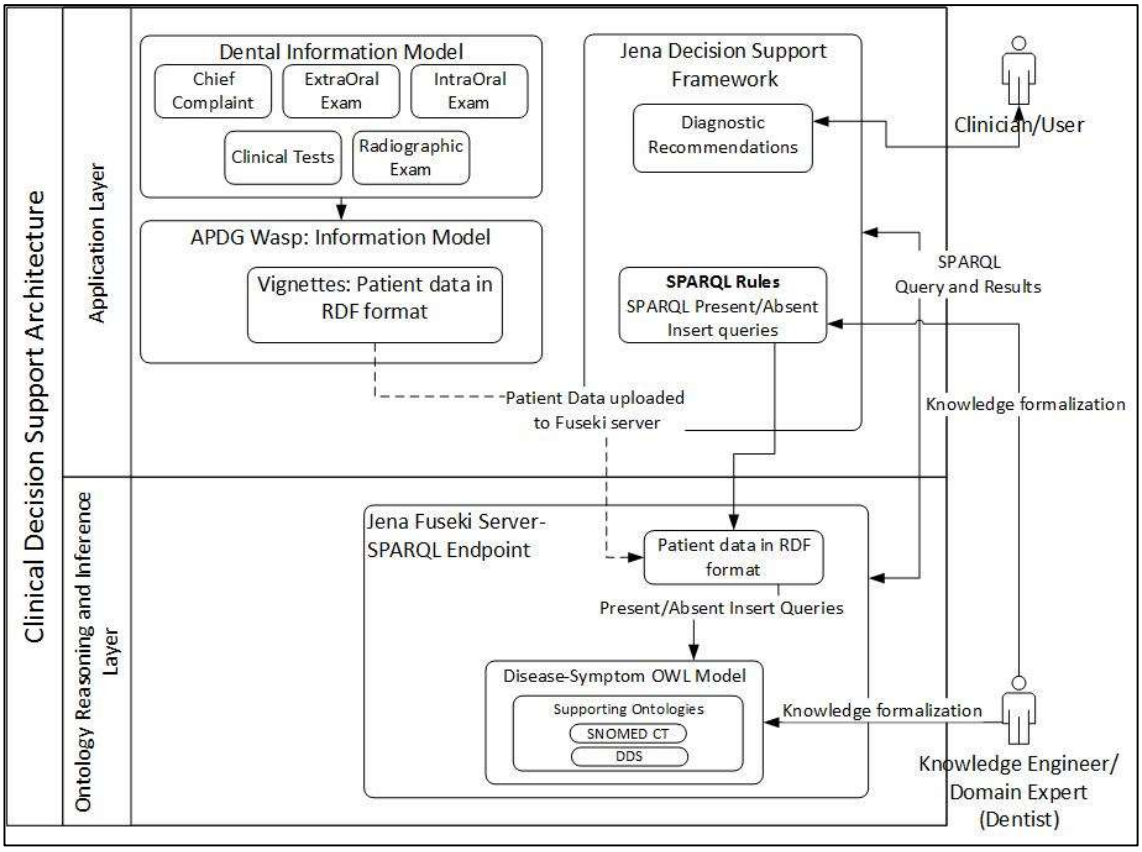


Figure 5-31 - Case Study 3 architecture of CDSS using patient data in RDF format

In this case study, the use of patient data in RDF format that is based on an archetype-based information model used in EHR systems is explored. Advanced Patient Data Generator (APDG) is a tool that allows researchers to create synthetic patient data. APDG uses the archetype model and Patient Data Definition Language (PDDL) to create patient data instances (Huang *et al.*, 2013). OpenEHR is another example of an information architecture standard used in EHR systems that is based on archetypes (openEHR Foundation, 2016).

The CDSS prototype was developed using Jena Semantic Web framework (Figure 5-31). The role of the CDSS prototype is to upload the patient data along with the DS Model into an empty default graph model in a Jena Fuseki server. Jena Fuseki server provides access to the data via a SPARQL endpoint.

In the sample patient data shown below for the archetype Clinical History there is a slot called Pain Present (highlighted in Red) (Table 5-11).

This slot can take two values "Yes" and "No". In the sample data below (Table 5-11) it has taken the value "Yes". When the clinician selects the value from the EHR or CDSS user interface the value is inserted into the database. This case study simulates this process and the process of exporting the data in RDF format with the help of the APDG generator. APDG can export in several different formats including XML and RDF. The APDG application, XML files and the RDF data can be accessed via Appendix 9.

Archetype	Slot	Value
Patient Demographics		
Clinical History		

Archetype	Slot	Value
	Pain Present	Yes
		No
Clinical Examination		
Lymph node exam		
	Lymph node firm	Yes
		No
	Lymph node swelling	Yes
		No
	Lymph node tenderness	Yes
		No
Clinical Tests		
Diagnosis and Treatment Planning		

Table 5-11 - Sample patient data showing Archetype-Slot-Value model

Once the patient data and DS Model have been uploaded to the Jena Fuseki server the CDSS prototype will insert a triple *rp:hasObservationStatus* rp:9_ObservationPresent based on the data present in the patient data (Figure 5-32).

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rp: <http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#>
PREFIX wasp:<http://wasp.cs.vu.nl/apdg#>
```

```
INSERT     {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent.
            ?Observation2 rp:hasObservationStatus rp:9_ObservationAbsent}
WHERE
        { SELECT ?value ?Observation1 ?Observation2
        {?Observation1 a rp:Pain_present.
        ?Observation2 a rp:Pain_absent.
        ?s wasp:hasSlot ?o.
        ?o rdfs:label ?slot.
        ?o wasp:value ?value.
                FILTER (str(?slot) = "Pain Present")
                FILTER (str(?value) = "Yes")}
}
```

Figure 5-32 - Case Study 3 - SPARQL present/absent insert query

If the Pain Present slot has the value "Yes", then the SPARQL insert query will insert the triple Observation1 *rp:hasObservationStatus* rp:9_ObservationPresent where Observation1 is all individuals that belong to the class rp:Pain_present in the DS Model. rp: Pain_present is class of observations that is sub-class of rp:Pain_status (Figure 5-24), which in turn is a sub-class of rp:Observation.

5.4.3.1 Iso-semantic models

The ideal way to promote semantic interoperability is to have just one model representing all medical information. However, this is not feasible in practical terms. Different healthcare systems and institutions may have their own unique way of representing the same information depending on the use case. An Iso-semantic model is the term used to define the use of multiple models to represent the same information (Huff *et al.*, 2014). This concept can be further illustrated using the example below.

In Table 5-12, the Archetype "Clinical History" has the Slot "Pain Present" which can take either "Yes" or "No" values. The same information can be represented using a different Slot called "Pain" which can take the values "Pain Present" and "Pain Absent". For the sake of simplicity most models may use the former model, but the DS Model must work with any information representation in the real world. Sometimes the information modellers must take a decision to pre-coordinate or post-coordinate the information model. For example, a Slot called "Location of pain" can have the value "Area around upper right region of cheek" (highlighted in Red in below - Table 5-12).

Archetype	Slot	Value
Patient Demographics		
Clinical History		
	Pain Present	Yes
		No

Archetype	Slot	Value
	Location of pain	Area around upper right region of cheek
		Area around lower right region of cheek

Table 5-12 - Pre-coordinated model

The table above (Table 5-12) shows an Archetype-Slot-Value that takes a pre-coordinated approach. The concept "Area around upper right region of cheek" may be linked to terminology that defines the concept/term. The same information can be represented in the post-coordinated manner seen below (Table 5-13). The concept "Area around upper right region of cheek" can be combined by joining several terms/concepts from different systems. For example, the term cheek can be defined in the Formal Model of Anatomy ontology.

Archetype	Slot	Value
Patient Demographics		
Clinical History	Pain Present	Yes
		No
Location of pain	Area	Cheek
		Lip
	Jaw affected	Upper
		Lower
	Side affected	Right
		Lower

Table 5-13 - Post-coordinated model

The post-coordinated approach is more verbose but makes it easier to analyse the data (example, an analyst may want to query instances where the pain may have started in the right side of the face). However, for simple diagnostic decision support systems the pre-coordinated approach may result in a much simpler user-interface (for example in patient facing CDSS tools). The additional data may not be required. Both these models can result in the same diagnosis. The same information has been structured differently. The DS Model must therefore operate with either of the models.

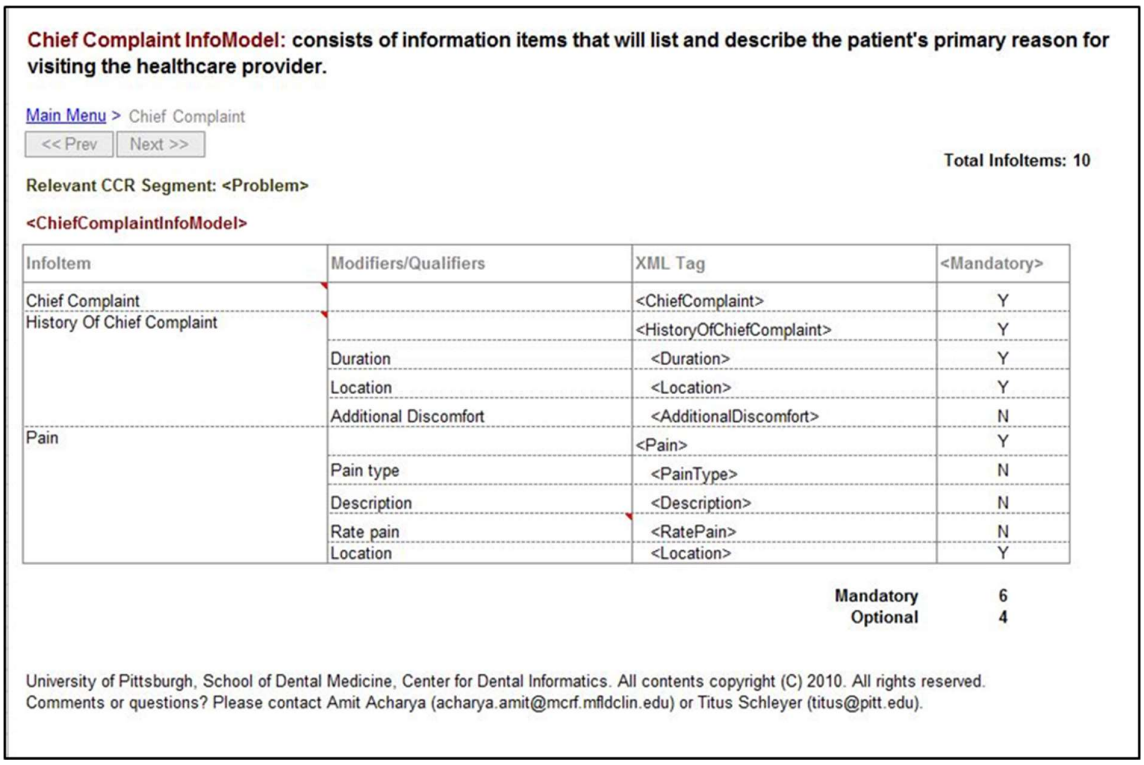
It is highly recommended that the DS Model itself be modelled using the pre-coordinated approach unless otherwise specified. For example, "pain in the area around the upper region of the cheek" is common symptom of Pulpal abscess (abscesses tooth) and maxillary sinusitis (inflammation of the sinus of the upper jaw). This observation ("pain in the area around the upper region of the cheek") when combined with a decayed tooth in the upper jaw with a deep cavity makes Pulpal abscess highly suspect. Therefore, it is theoretically possible to model "Pain in the upper jaw with deep dental caries" as an single

observation that supports Pulpal abscess if it is present (*rp:hasObservationStatus* rp:9_ObservationPresent). There are other observations that are so highly indicative of a diagnosis in their presentation that it is usually better to model these concepts as a single term, rather than trying to combine then by combining several different terms. In this current thesis, a mix of pre-coordinated and post-coordinated approaches were used for the sake of simplicity.

The advantage of using the SPARQL present/absent insert queries is that it can used to map different (iso-semantic) models to the same DS Model. By adding additional filter conditions and modifying the SPARQL insert query we can dictate the conditions under which the observation can be declared as present/absent. This is demonstrated by using different SPARQL present/absent insert queries in Case Study 3 & 4.

There are 276 individual present/absent insert queries within this CDSS prototype. The number of queries can increase with an increase in the number of items within the information model. Each query is stored in a separate query file. Multiple queries can be stored within the same query file. If the data changes for some reason, for example if the clinician feels that he has recorded the wrong observation, the inserted triple can be deleted and another inserted using the same query. The use of RDF, RDFS, OWL and SPARQL allows the flexibility to model both the information model and DS Model using different modelling paradigms.

Once the clinician has completed the patient enquiry or physical examination the patient data is exported in the RDF format. In this case study, the data was exported into a Jena Fuseki server. However, the same data can be exposed via a SPARQL endpoint and queried directly. The DS Model is uploaded by the CDSS prototype and based on the data that is present/absent the CDSS prototype inserts the appropriate present/absent triples.



Chief Complaint InfoModel: consists of information items that will list and describe the patient's primary reason for visiting the healthcare provider.

Main Menu > Chief Complaint

<< Prev | Next >>

Total InfoItems: 10

Relevant CCR Segment: <Problem>

<ChiefComplaintInfoModel>

InfoItem	Modifiers/Qualifiers	XML Tag	<Mandatory>
Chief Complaint		<ChiefComplaint>	Y
History Of Chief Complaint		<HistoryOfChiefComplaint>	Y
	Duration	<Duration>	Y
	Location	<Location>	Y
	Additional Discomfort	<AdditionalDiscomfort>	N
Pain		<Pain>	Y
	Pain type	<PainType>	N
	Description	<Description>	N
	Rate pain	<RatePain>	N
	Location	<Location>	Y

Mandatory 6
Optional 4

Figure 5-33 - DIM Chief Complaint

In this case study, the Dental Information Model (DIM)(Acharya, Mital and Schleyer, 2009b) was used to structure the EHR data (Figure 5-33). DIM is segregated into several sections such as Chief Complaint, Intra Oral Examination, Extra Oral Examination, Clinical Tests and Radiographic Examination and Diagnosis. A detailed explanation of the different sections of the information model has been given in section - 6.2 -Sections of the Information Model. The sections of DIM were used as the basis for developing the archetypes (Table 5-11) in the APDG patient data generator and generate the RDF patient

data needed for the CDSS prototype (Figure 5-31). The present/absent insert queries can be restricted to data that has been collected. For example, if we are interested in early diagnosis after the Chief Complaint section of the DIM model then we can complete the present/absent insert queries for that section which informs the model of the observations that are present/absent, calculate the weights and then obtain a ranked list of diagnosis.

However, in the architecture developed for this case study if we want to obtain a diagnosis after both Chief Complaint and Extra-Oral Examination sections of the DIM model we must delete all the data, complete inserting present/absent queries for both these sections based on the patient data, recalculate the weights and obtained another ranked list as the DS Model and the patient data was stored in the default graph model of the triple store. This architecture was found to be suitable for the purposes of evaluating the DS Model in early and late diagnosis. In real world applications, named graphs can be used to store the weights, remove the weights attached to each diagnosis when new patient data is added, recalculate the weights and obtain another diagnosis.

This architecture (Figure 5-31) was used to evaluate the DS Model (Chapter 6) because it easier to generate patient data based on the patient vignettes using the APDG patient data generator and evaluate the diagnostic accuracy. The architecture demonstrated here show how you can use different patient datasets with the same DS Model. The data can be in different SPARQL endpoints. Different sections of the DIM model can be in different endpoints if needed. The files for Case Study 3 can be accessed via Appendix 8.

5.4.4 Case Study 4-Patient data stored in open-source EHR and relational databases

Open source EHR/EMR are very popular for capturing healthcare information (Karopka, Schmuhl and Demski, 2014). Most of these EHR systems capture data in relational databases such as MySQL. However, the information models and schemas used by the open/closed source systems are not standardised now. There are ongoing efforts such as HL7 FHIR to achieve a standardised interface to the data stored by these systems using REST protocols (HL7, 2016).

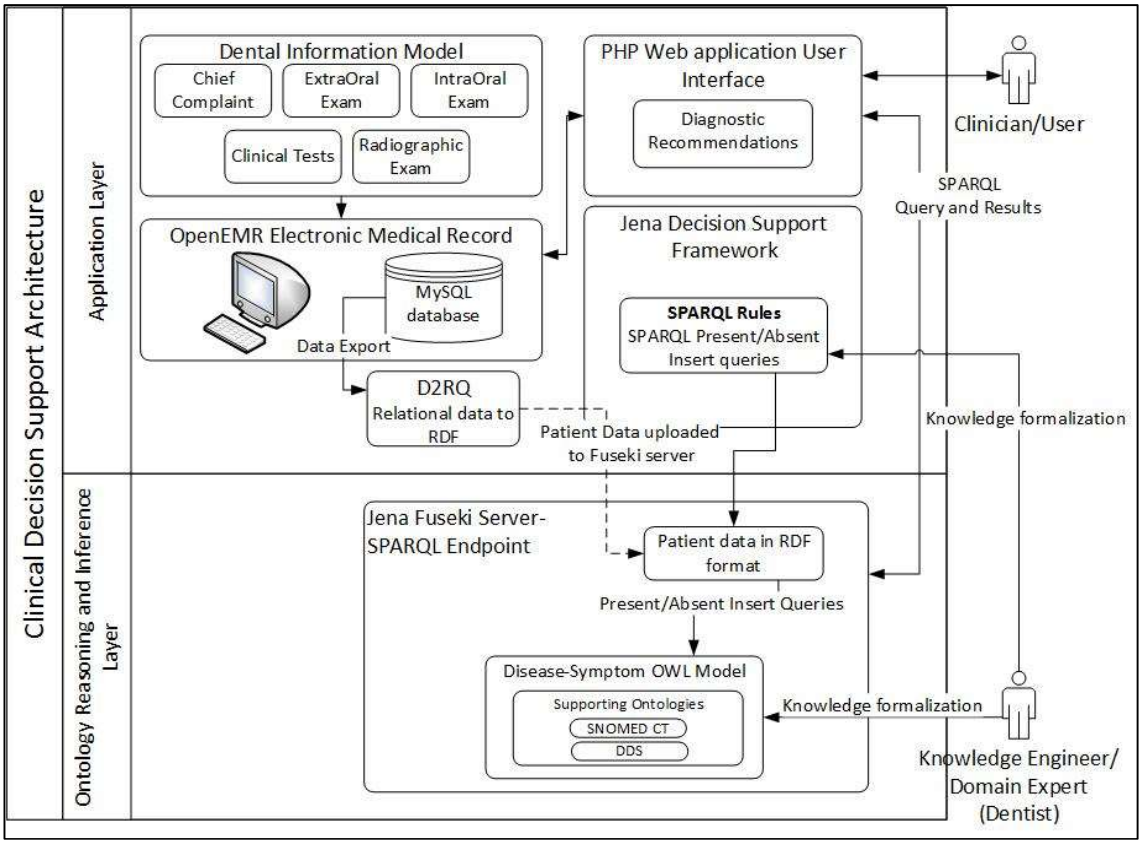


Figure 5-34 - Case Study 4- CDSS architecture of relational to RDF data

In this case study, the use of semantic web/linked data technology to convert the relational data stored in existing open source EHR systems into a RDF representation that can then be consumed by CDSS applications is demonstrated. This case study (Figure 5-34) is like Case Study 2 (Figure 5-27) in its architecture. In this study, the custom mapping used in Case Study 2 is not utilized.

OpenEMR (OpenEMR, 2016) was used as the test EHR platform (Figure 5-35). It is very popular EHR system developed using PHP language (Note: EHR and EMR has been used interchangeably here) (Noll, Beecham and Seichter, 2011). It is developed and maintained by a small group of enthusiastic programmers and clinicians. The data is stored in a MySQL database. The OpenEMR platform allows the users to create custom forms in addition to the other built-in functionality such as practice management.

DIM V.1.0 (Acharya, Mital and Schleyer, 2009a) was used as the basis for creating forms within OpenEMR. As mentioned in Case Study 3, DIM separates the information collected during a dental encounter into various processes and sub-processes. This includes Chief Complaint, Extra Oral Examination, Intra Oral Examination, Radiographic Examination followed by final Diagnosis.

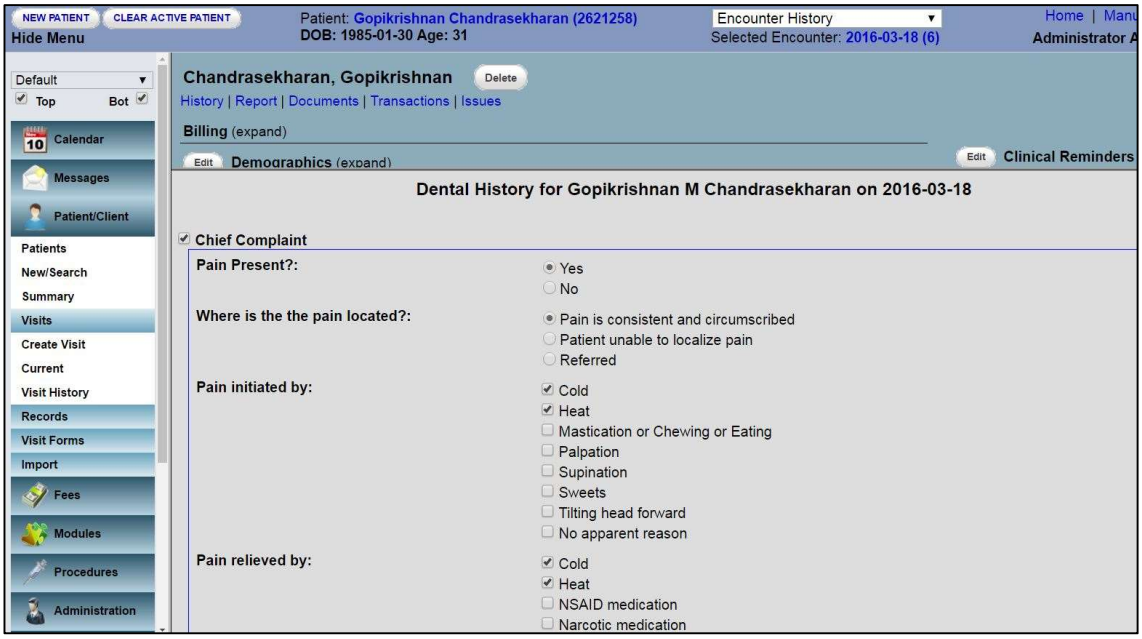


Figure 5-35 - Case Study 4- OpenEMR user interface

The data stored in the relational database was then converted to RDF using the D2RQ platform (Christian Bizer and Seaborne, 2004). However, in this case study the default mapping used by D2RQ was used to create the RDF representation of the OpenEMR patient data. The advantage of this approach is RDF representation is mapped to the default representation of data in the database and not mapped to the DS Model ontology. It is easier to setup and extract data using the default mapping. The approach used in Case Study 2 requires detailed mapping to the DS Model ontology that is difficult for healthcare applications with complex schemas. Until healthcare interoperability solutions such as HL7 FHIR becomes mainstream we may have to resort to similar techniques prescribed in this case study to access patient data in existing EHR systems.

A mapping will still need to be created to inform the DS Model which observations are present/absent. In this case study, SPARQL present/absent insert queries like the one mentioned previously (Figure 5-16) and in Case Study 3 was used.

When the clinician selects a value in the OpenEMR user interface shown above (Figure 5-35) the data is recorded in the MySQL database. D2RQ server then translates the relational data into RDF format which can then be accessed via a SPARQL endpoint. D2RQ server becomes sluggish when there is multiple FILTER or LIMIT conditions in the query. Especially since there are 276 SPARQL present/absent insert queries in this case study. Solutions for scalable SPARQL queries over relational data are available, though they were considered beyond the scope of this thesis (Zhu *et al.*, 2014).

The poor performance of D2RQ with multiple FILTER statements is a well-known problem as the virtual RDF graph is a view that resides on top of the relational model (Bizer and Cyganiak, 2007). The problem can be remedied by exporting the data in RDF format. Therefore, in this case study the translated data was exported in RDF format after experiencing difficulties obtaining diagnostic recommendations quickly when accessing the patient data via the SPARQL endpoint. Accessing the patient data via the SPARQL endpoint is not a significant problem if the application has a simple database schema like Case Study 2. Instant response times to queries are crucial for early diagnostic recommendations as a ranked differential diagnosis list will need to be generated at different stages of the clinical encounter. In this case study, a recommendation can be generated after each main section/category of the DIM information model (example Chief Complaint) since the data was exported as RDF.

The Jena semantic web framework was used to develop the CDSS prototype. The role of this CDSS prototype is to upload the patient data exported using D2RQ from the OpenEMR database into a triple

store accessed via the Jena Fuseki server. The DS Model file was uploaded onto the same default graph model in the triple store.

The CDSS prototype then queried the patient data for values using a SPARQL insert query. For example the SPARQL insert query inserted the triple => Observation1 *rp:hasObservationStatus* rp:9_ObservationPresent, where Observation1 is rdf:type rp:Pain_present.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX foaf:<http://xmlns.com/foaf/0.1/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rp: <http://www.semanticweb.org/abjb788/ontologies/2015/3/untitled-ontology-208#>
PREFIX map: <http://localhost:2020/resource/#>
PREFIX db: <http://localhost:2020/resource/>
PREFIX vocab: <http://localhost:2020/vocab/>
```

```
INSERT  {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent.
            ?Observation2 rp:hasObservationStatus rp:9_ObservationAbsent}
WHERE{
 { {
        SELECT (max(?encounter_id)as ?Recent_encounter_id)
        WHERE
                {?form vocab:forms_encounter ?encounter_id.} }
                {
                ?form vocab:forms_form_id ?form_id.
                ?form vocab:forms_deleted ?deleted.
                ?s vocab:lbf_data_form_id ?lbf_form_id.
                ?s vocab:lbf_data_field_id ?field_id.
                ?s vocab:lbf_data_field_value ?field_value.
                ?form vocab:forms_encounter ?Recent_encounter_id.
                } }
 FILTER (?deleted=0 )
```

Figure 5-36 - Case Study 4 - present/absent SPARQL insert query

The SPARQL insert query above (Figure 5-36) will inform the DS Model which observations are present/absent (*rp:hasObservationStatus* rp:9_ObservationPresent). Once all the present/absent insert queries are complete the CDSS prototype will calculate weights and another select SPARQL query returns a ranked list of recommendations (Figure 5-18). The recommendations that are returned can be modified to be more user-friendly. The rdfs: label can be displayed instead of the name of the class/individual (Figure 5-37).

```
SELECT ?Label ?DiagnosisWeight

WHERE

        {?PulpalDiagnosis  rp:hasDiagnosisObsWeight ?DiagnosisWeight.

        ?PulpalDiagnosis a ?Diagnosis.

        ?Diagnosis rdfs:subClassOf* rp:Pulpal_Disease.

        ?Diagnosis rdfs:label ?Label

} ORDER BY desc (?DiagnosisWeight)
```

Figure 5-37 - SPARQL Diagnosis recommendations select query - User friendly version

In this case study, the diagnostic recommendations were return to a webpage written using the sparqllib.php library that can query SPARQL endpoints (Fuseki server endpoint in this example). The sparqllib.php library is part of the Graphite Linked Data library built of top of another RDF PHP library called ARC2 (Gutteridge, 2012b, 2012a). The diagnostic recommendations in the form of ranked list and associated weights below (Figure 5-38). Since OpenEMR is a PHP based open-source EHR the application can be integrated into an EHR user interface without much difficulty.

Number of rows: 7 results.

Label	DiagnosisWeight
Normal pulpal	5
Pain of nonodontogenic origin	4
Pulp necrosis	4
Reversible pulpitis	2
Asymptomatic irreversible pulpitis	1
Dentinal hypersensitivity	1
Symptomatic irreversible pulpitis	1

Number of rows: 6 results.

Label	DiagnosisWeight
Acute apical abscess	17
Pain of nonodontogenic origin	4
Symptomatic apical periodontitis	4
Chronic apical abscess	3
Asymptomatic apical periodontitis	-1
Normal apical	-7

Figure 5-38 - Case Study 4-Diagnostic recommendations user interface

The main problem with the approach demonstrated in this case study is the need for an elaborate mapping SPARQL insert queries that are mapped to the schema of the open source EHR. When the schema changes, the SPARQL insert queries will need adapting as well. This problem exists even if the D2RQ mapping file approach used in Case Study 2 is utilized. However, the D2RQ mapping file-based approach is a more standardised approach.

The other major problem with the approach in this case study is the need to export the patient data as RDF data to upload to a triple store. If the diagnosis recommendations are only needed at the final stage, then the clinician can export the data after all the patient data has been collected. However, for early recommendations the clinician must export the data at each section of the DIM model described previously in Case Study 3 (Also see section - 6.2). This is not a feasible approach in the long run as it can slow down the process of obtaining a clinical diagnosis.

Another problem of querying RDF data translated from relational data obtained directly from EHR systems is that when patient data is changed the data must be exported again into the triple store (Jena

triple store with Fuseki server). Also, the patient data is stored as clinical encounters. In this case study the most recent patient encounter was used as the source of patient data. If the CDSS must access data from other encounters the queries should be modified accordingly. The query should check if the form is still active, as the data from deleted forms are recorded in the database.

Overall the implementation is more complex using the methods described in this case study when compared to the other case studies. However, the functioning of a CDSS that uses the DS Model and patient data from an existing open-source EHR to deliver diagnostic recommendations was demonstrated. Some of the problems encountered in this case study will be solved with the wide spread adoption of standard RESTful APIs such as FHIR. FHIR is both a healthcare information representation standard and an API. FHIR can be serialised in several different formats including JSON, XML and RDF.

Another recent development of interest is the RDF representation of FHIR and related FHIR ontology (HL7, 2015). The D2RQ mapping technique can be used to map the data in OpenEMR to concepts in the FHIR ontology. Having a standardised information representation will make it easier to write SPARQL present/absent insert queries that can then inform the model which observations are present. It is possible to modify the DS Model to use terms from the FHIR ontology. However, the FHIR ontology was still under development at the time of writing this thesis and could not investigated further. The files for Case Study 4 can be accessed via Appendix 8.

5.4.5 Summary

The case studies in the above section demonstrate how the DS Model can be used with different type of patient data and integrate this patient data to generate diagnostic recommendations. The next section details the development of the Multi-level DS Model which is an extension to the DS Model that can generate more detailed explanations, generate recommendations for sub-diagnostic states and generate diagnostic recommendations. **The Multi-level DS Model can be used with any architecture described in the case studies by replacing the DS Model with the Multi-level DS Model. The Multi-level DS Model has additional SPARQL rules that need to be fired to generate the recommendations. Multi-level DS Model is additional work carried out to cover certain scenarios that the DS Model cannot fulfil without modifications.**

5.5 Multi-level diagnosis with DS Model

The DS Model allows us to model the relationships between the Diagnosis and Observation class. A diagnosis can have a set of observations that are either supporting or negating the diagnosis. In the table below (Table 5-14) an archetype model with different slots and values have been represented.

Archetype	Slot	Value
Clinical History		
	Pain Present	Yes
		No
	Pain Initiated By	Cold food and drink
		Hot food and drink
		Jumping

Archetype	Slot	Value
		Mastication or Chewing or Eating
		Palpation
		Supination
		Sweets
		Tilting head forward
		No apparent reason

Table 5-14 - Multi level diagnosis sample information model

The Clinical History archetype has 2 slots "Pain Present" and "Pain Initiated By". An example patient data has been highlighted in red and the Pain Present slot takes the value "Yes" and the 'Pain Initiated By' slot takes the value "Cold food and drink". Pain Present = "Yes" has been represented in the DS Model as a class rp:Pain_present and the individuals rp:Pain_present_Plus_1 and rp:Pain_present_Minus_1 belong to this class.

rp:Pain_present_Plus_1 supports the Diagnosis rp:Dentinal_Hypersensitivity as shown in the figure (Figure 5-39) below.

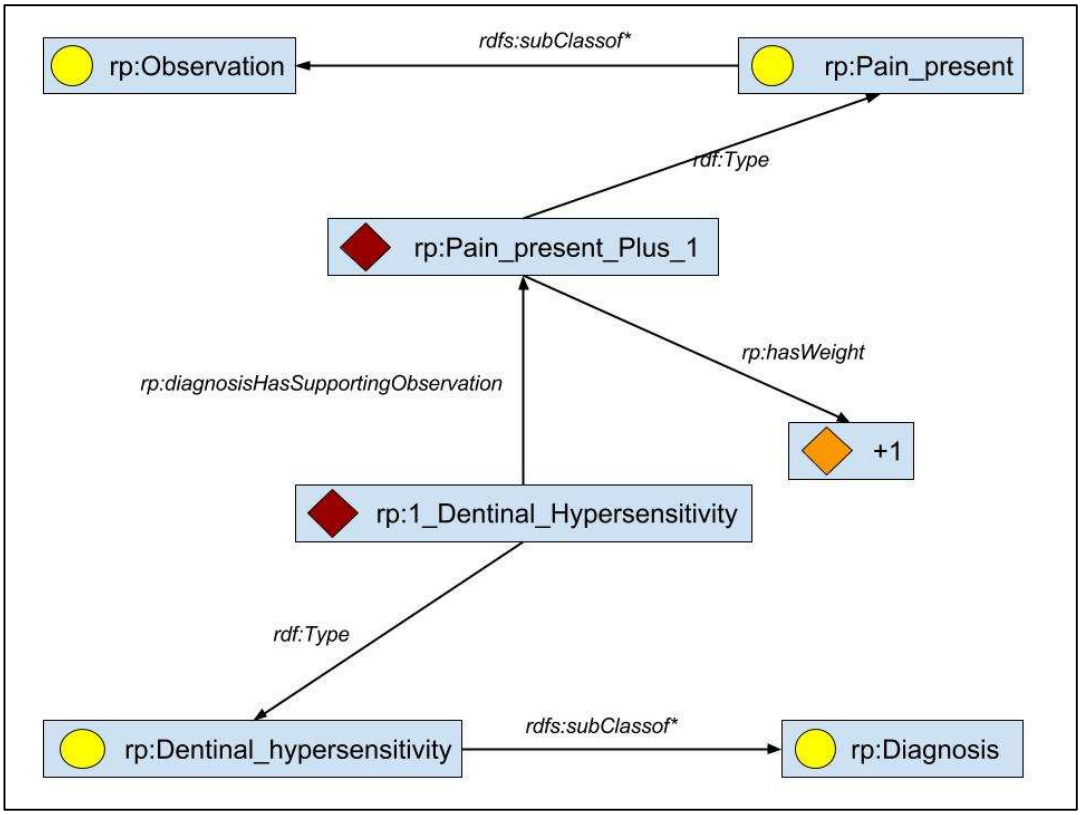


Figure 5-39 - rp:Pain_present_1_Plus_1 supports the Diagnosis rp:Dentinal_Hypersensitivity

However, the same rp:Pain_present_Plus_1 observation can negate a different diagnosis such as Normal Pulp, represented by the class rp:Normal_pulpal, which has the individual named rp:1_Normal_Pulp_DIAG as shown in the figure below (Figure 5-40).

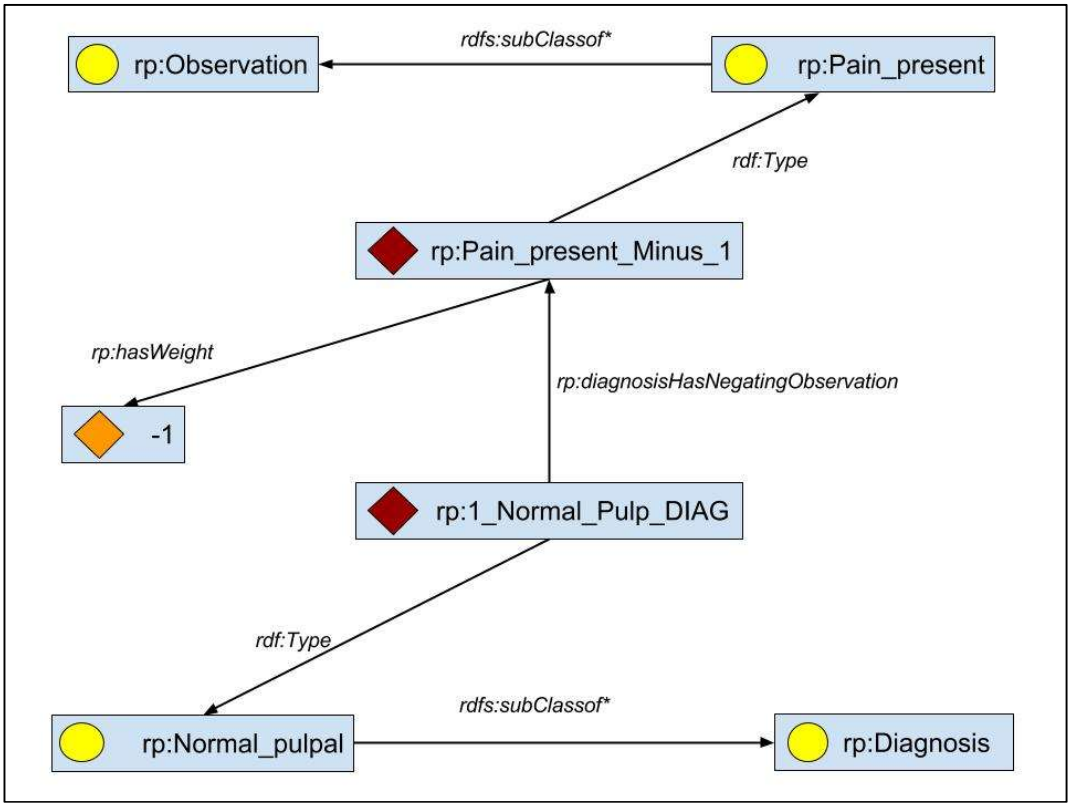


Figure 5-40 - rp:Pain_present_1_Minus_1 negates the Diagnosis rp:Normal_pulpal

So, if a patient complains of pain that supports the diagnosis Dentinal Hypersensitivity. However, if a patient complains of pain then it can negate the Normal Pulp diagnosis as the pulp is normal and no apparent pathology is present.

Similarly, the class rp:Cold_pain_initiation_factor represents the value "Cold food and drink" of the slot "Pain Initiated By" (Table 5-14). rp:Cold_pain_initiation_factor has the individual rp:Cold_pain_initiation_factor_Plus_1 which when present supports the diagnosis Dentinal Hypersensitivity (Figure 5-41).

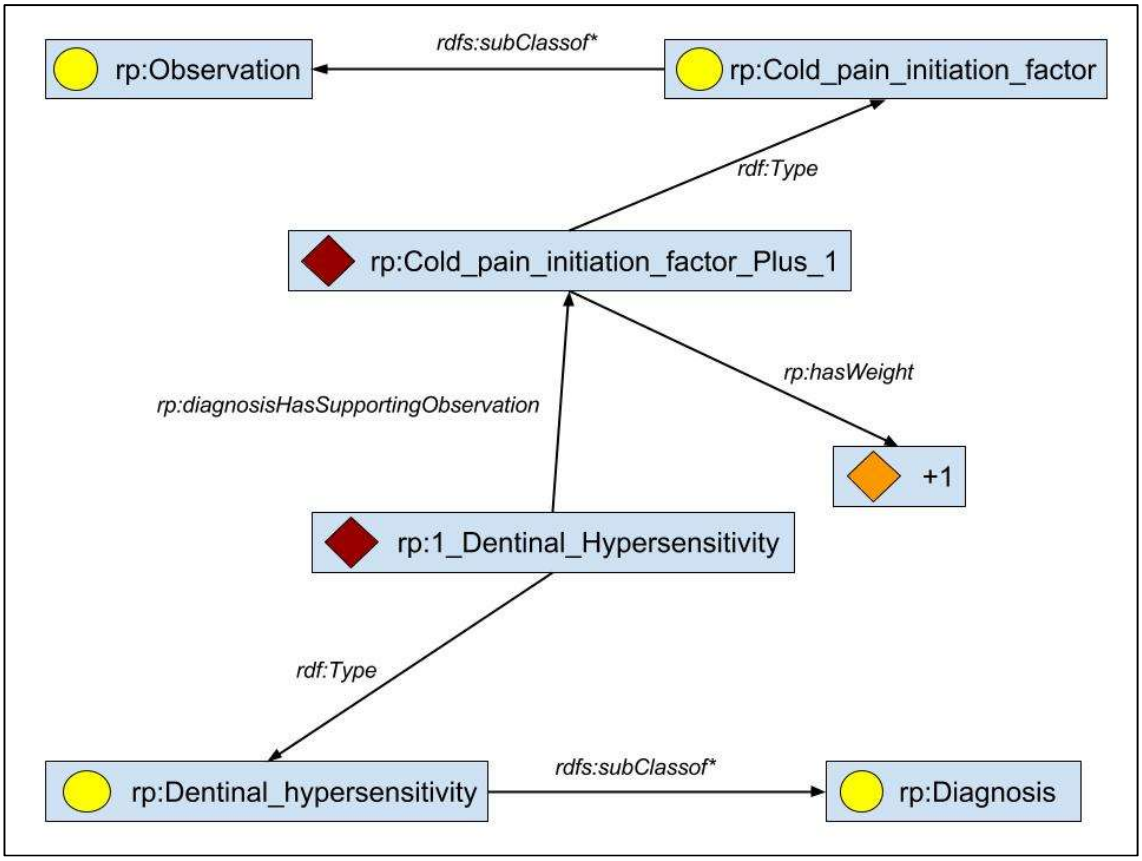


Figure 5-41 - rp:Cold_pain_initiation_factor_Plus_1 supports the Diagnosis rp:Dentinal_Hypersensitivity

5.5.1 Classes

5.5.1.1 The need for sub-diagnostic states

The 'presence of pain' and 'presence of pain initiated by cold food and drink' can have different sub-diagnostic interpretations other than diagnosis. For example, the "presence of pain that is initiated by cold food and drink" could indicate the 'presence of inflammation of pulp'. The CDSS can then trigger an alert informing the clinician that the patient may be suffering from a potential inflammation of the pulp. The need for these sub-diagnostic triggers can be substantiated using the 'Illness script theory', which is a medical specialization of the more general 'Script theory' in cognitive science. Script theory tries to explain how human beings try to gather and structure information to interpret events in his/her surroundings (Schank and Abelson, 1977). The Dual Process Theory explains how individuals make decisions and judgments under uncertainty with limited information (Kahneman and Egan, 2011; Evans and Stanovich, 2013). It can be adapted to explain how a clinician reasons when presented with a clinical case. The dual process theory consists of 2 types of processes, System 1 and System 2. System 1 is associated with pattern recognition and System 2 is associated with the slower, more analytical process of reasoning (Evans and Stanovich, 2013).

In addition to providing diagnostic recommendations during the clinical encounter the CDSS may have to provide sub-diagnostic recommendations or explanations in certain scenarios. There are clinical situations where a sub-diagnostic recommendation provides more value to the clinician than the diagnostic recommendation. Using a dental example, when there is 'presence of pain initiated by cold food and drink' this indicates a possible reversible inflammation of the pulp. The exact status of the pulp can be confirmed using clinical tests and radiographic examination at a later stage of the clinical encounter. However, if the CDSS can indicate/alert the clinician about the inflammation of the pulp, this could potentially trigger illness scripts associated with inflammation of the pulp and help the clinician ask questions or trigger a line of investigation to confirm or rule out the presence of inflammation of the

pulp. The sub-diagnostic recommendations can trigger System 1 process of reasoning. In the early stages of the clinical encounter these sub-diagnostic recommendations can help slow down the reasoning process and assist the clinician in investigating all possible causes for this 'presence of pain initiated by cold food and drink'. The aim of the Multi-level DS Model is to allow the domain experts to model the sub-diagnostic recommendations and allow the CDSS to provide sub-diagnostic recommendations/explanations to the clinician, in addition to the diagnostic recommendations. The system can be adapted to provide only the sub-diagnostic recommendations and provide the explanations for making this recommendation.

In the early stages of the clinical encounter there could be more value in providing sub-diagnostic recommendations as well as there is not enough information to fully justify a diagnostic recommendation. However, this has not been substantiated in literature. For example, to fully diagnose a Reversible Pulpitis diagnosis, the clinician will need to perform a pulp vitality test. However, in the early stages the clinician only has sufficient information to support a sub-diagnosis such as inflammation of the pulp, rather than Reversible Pulpitis. It is therefore useful for the clinician to see a diagnostic recommendation and the sub-diagnostic recommendations as well and understand why the CDSS made these recommendations and track them to the presence or absence of the clinical observations that triggered these recommendations.

The Multi-level DS Model uses the argumentation theory to provide diagnostic recommendations like the DS Model. Instead of presence/absence of observations supporting or negating a diagnosis, the presence or absence of observations can support or negate a sub-diagnostic state. The presence or absence of this sub-diagnostic state in the Multi-level DS Model can in turn support or negate a diagnosis. When the System 2 process is activated in the clinician and the clinician wants to understand why the CDSS provided a diagnostic recommendation, the clinician can obtain the arguments for the sub-diagnostic state as well. To model these sub-diagnostic states another class was introduced into the DS Model. This class is called Finding (represented as rp:Finding in the DS Model) (Figure 5-42). Finding class can be used to represent these sub-diagnostic states. The model that expands the DS Model (detailed in previous sections) and contains the Diagnosis, Findings and Observations is called the Multi-level DS Model.

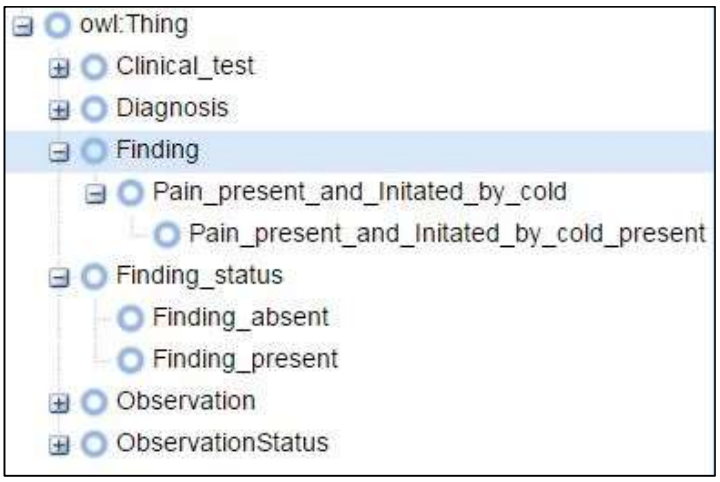


Figure 5-42 - DS Model outline with Finding class

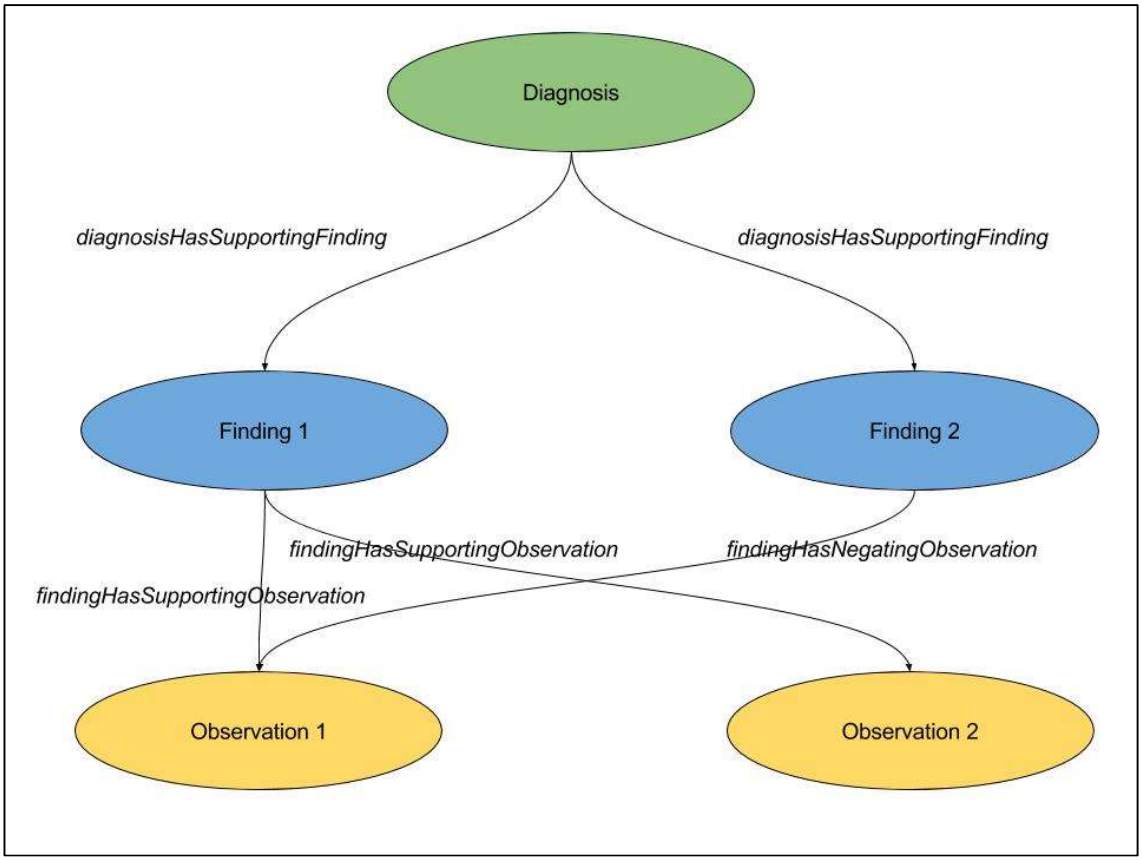


Figure 5-43 - Multi-level diagnosis with DS Model

The figure above (Figure 5-43) shows the relationships between the diagnosis, finding and observations in the Multi-level DS Model. The finding class is located between the diagnosis and observation classes and adds another layer of decision making and provides the sub-diagnostic recommendations. Each finding can have a set of observations that are either supporting or negating it. Similarly, the diagnosis may have a set of findings that are present and either supporting or negating the diagnosis.

In the coming sections the example of "Pain present and Initiated by Cold" was used to represent a Finding. During a clinical examination the presence of pain is an interesting observation and could provide crucial information about the state of the dental pulp. However, when this clinical observation is combined with the observation that the pain is initiated by cold food and drink that narrows the possibilities and could indicate the presence of inflammation. Therefore, this was used as an example to demonstrate how the observations can support the finding and how the finding can in turn provide arguments to support or negate a Diagnosis. The finding was represented in the Multi-level DS Model as a class: rp:Pain_present_and_Initated_by_cold.

This class has a subclass called rp:Pain_present_and_Initated_by_cold_present. A class called rp:Finding_status represents the different states of the findings and has the subclasses rp:Finding_absent and rp:Finding_present (Figure 5-42).

The modelling of sub-diagnostic states has relevance when using the linked web of data to support decision making. The patient's data may be distributed among different clinical systems. The clinical observations present in those systems may only support a sub-diagnostic state and there may not be sufficient information to fully support a diagnosis. If there is enough evidence to support a sub-diagnostic state, then the CDSS can make the recommendation using the data in the disparate clinical systems. **In addition, sub-diagnostic state can be used as a trigger for guideline recommendations instead of relying on abnormal values alone.**

Other ontology driven diagnostic CDSS tools have demonstrated the use of multi-level diagnosis to support decision making. SeDeLo (A Rodríguez-González *et al.*, 2012) is an ontology driven diagnostic CDSS that models diagnostic criteria using DL and rules. The SeDeLo approach was capable of multi-level diagnosis. So, laryngitis can be automatically inferred from its clinical observations, and laryngitis will then support the diagnosis of Common Cold. In SeDeLo laryngitis becomes the sub-diagnostic state like the 'pulp inflammation' example and the 'pain present and initiated by cold' example described in this section above. The system based on the SeDeLo approach was found to be faster, more efficient and more capable than the previous methods (Rodríguez-González et al. 2012). The Multi-level DS Model differs from SeDeLo by using the argumentation approach to explain how the presence or absence of observations support/negate findings (sub-diagnostic state) and how the presence or absence of findings either support/negate diagnoses. The degree to which the observations support/negate a diagnosis or finding is provided by the Multi-level DS Model and is something that is not supported by the SeDeLo approach. The linked data approach used in the DS Model and the Multi-level DS Model helps in integration of data from multiple sources with relative ease, which is something not inherently built into the SeDeLo approach.

The Multi-level DS Model can be implemented using any architecture demonstrated in the previous cases studies. The DS Model can be replaced by the Multi-level DS Model and the diagnostic recommendations can be displayed to the clinician. The Multi-level DS Model has the additional advantage of providing the recommendations in the form in intermediate states in addition to the diagnostic recommendations. Arguments for and against each finding and arguments for and against each diagnosis can be displayed to the clinician to help decide.

5.5.2 Properties

The object property linking Observations to Findings is the *rp:Observation_Finding_properties* and its sub-properties (Figure 5-44).

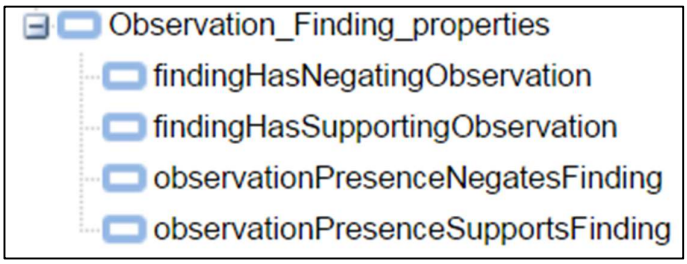


Figure 5-44 - rp:Observation_Finding_properties and its sub-properties

The object property linking Findings to Diagnosis is the *rp:Finding_Diagnosis_properties* and its sub-properties (Figure 5-45).

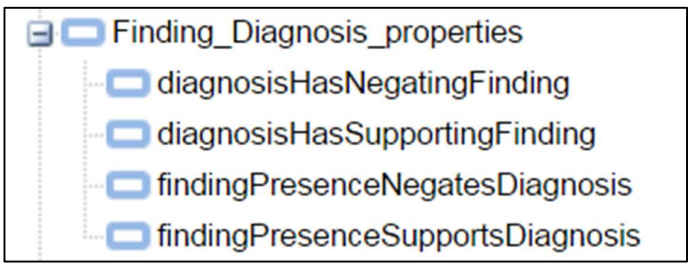


Figure 5-45 - rp:Finding_Diagnosis_properties and its sub-properties

In addition, there is the *rp:hasFindingStatus* property that has a domain of rp:Finding and range of rp:FindingStatus.

The main datatype properties include:

- *rp:hasDiagnosisFindWeight*
- *rp:hasAssignedFindingWeight*
- *rp:hasDiagnosisObsWeight*
- *rp:hasDiagnosisFindWeight*
- *rp:hasTotalFindingWeight*
- *rp:hasDiagnosisTotalWeight*
- *rp:hasFindingWeight*

These datatype properties allow the modeller to add the different weights to the findings and observations and denote the degree to which they support or negate a finding or diagnosis.

5.5.3 Individuals

The main individuals in this example that belong to the Finding class rp:Pain_present_and_Initated_by_cold_present include:

- rp:2_Pain_present_and_Initated_by_cold_present_Minus_2
- rp:2_Pain_present_and_Initated_by_cold_present_Plus_2

The following figure (Figure 5-46) illustrates the relations between the observations and the findings in a graph structure. The (*) near rdfs:subClassOf* denotes a transitive relationship.

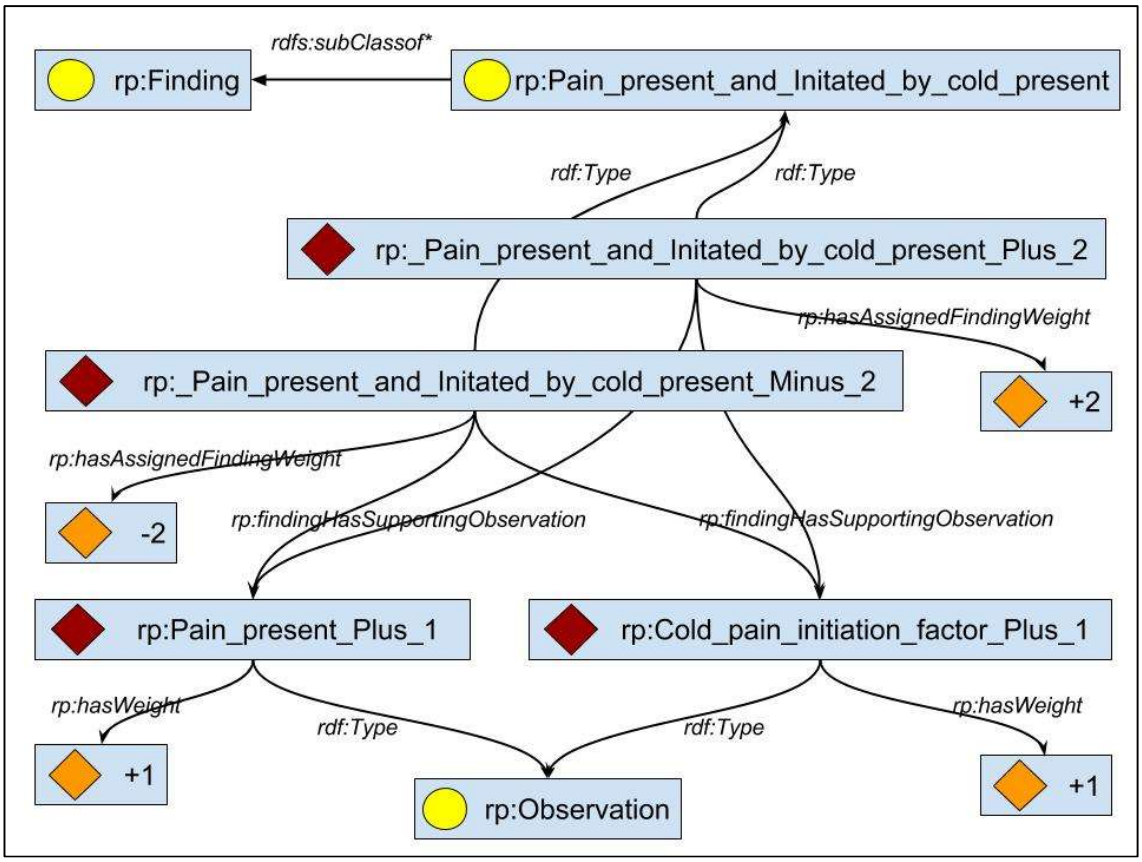


Figure 5-46 - Observation-Finding relations

The figure above (Figure 5-46) shows how the different individuals belonging to Observation classes are linked to the individuals belonging to the Finding classes. The individuals that are *rdf:type* rp:Finding has

an assigned finding weight of +2 and -2 respectively. The assigned weight is given by the domain expert and describes the degree to which the Finding will support or negate a diagnosis.

In the DS Model existing ontologies such as SNOMED-CT can be reused for obtaining the relations between diagnoses and observations. However, in the Multi-level DS Model only those relations between the diagnoses and observations can be reused. The intermediate states represented by findings and the relations between the observations and findings followed by findings and diagnoses cannot be reused as ontologies that map these relations do not exist. The domain experts will therefore have to map these relations from scratch and capture the weights for the arguments in favour and against each finding and diagnoses.

This is a major disadvantage of the Multi-level DS Model when compared to the DS Model. Whenever there is a change in knowledge the model should be updated easily, and the changes need to be pushed to CDSSs in clinical practice. The DS Model captures the relationship between the diagnoses and the observations. Ontologies like SNOMED-CT already capture this relationship in varying degrees of granularity. Other OWL based ontologies have been developed to capture this relationship (Bertaud-Gounot, Duvauferrier and Burgun, 2012a; Mohammed, Benlamri and Fong, 2012). Whenever the knowledge changes the base model (SNOMED-CT) the DS Model can be updated automatically. Domain experts will then have to review the weights to see where the DS Model needs to be updated and then the changes can be pushed to CDSSs using the DS Model in their domain.

Implementing the Multi-level DS Model will make this process much more difficult as existing ontologies that capture the intermediate states are not available and therefore domain experts will have to capture the intermediate states in separate OWL ontologies and integrate them with the simpler DS Model when needed. The Multi-level DS Model extends the DS Model and therefore this integration is possible.

5.5.4 SPARQL Inference

When the CDSS application that is using the DS Model with multi-diagnosis is run the following SPARQL insert query is run and the triple: ?Finding *rp:hasFindingWeight* ?FindingWeight is inserted into the default graph.

```
INSERT {?Finding rp:hasFindingWeight ?FindingWeight}
WHERE
        {SELECT (SUM (?Weight)AS ?FindingWeight) ?Finding
        {?Finding ?property ?Observation.
        ?property rdfs:subPropertyOf rp:Observation_Finding_properties.
        ?Observation rp:hasWeight ?Weight.
        ?Observation rp:hasObservationStatus rp:9_ObservationPresent.
 }
        GROUP BY ?Finding}
```

Figure 5-47 - SPARQL insert query calculating Finding Weights

The SPARQL insert query (Figure 5-47) calculates the "Finding Weight" from the observations that are present.

```
INSERT {?Finding rp:hasFindingStatus rp:9_FindingPresent.}
WHERE
        {?Finding rp:hasFindingWeight ?FindingWeight.
         FILTER (?FindingWeight > 0 )
}
```

Figure 5-48 - Finding present SPARQL insert

```
INSERT  {?Finding rp:hasFindingStatus rp:9_FindingAbsent.}
WHERE
        {?Finding rp:hasFindingWeight ?FindingWeight.
        FILTER (?FindingWeight <= 0 )
}
```

Figure 5-49 - Finding absent SPARQL insert

Two different SPARQL insert queries (Figure 5-48 & Figure 5-49) then sets the status for the Finding using the Finding Weight. **Any finding that has Finding Weight less than or equal to 0 is given the Finding Status of rp:9_FindingAbsent. Since only findings that has finding status of rp:9_FindingPresent are used for obtaining the diagnosis this technique ensures that findings with finding status of rp:9_FindingAbsent are discarded.** rp:9_FindingAbsent is an individual of *rdf:type* rp:FindingAbsent.

```
INSERT {?Finding rp:hasTotalFindingWeight ?TotalFindingWeight}
WHERE{
        SELECT  ?Diagnosis ?Finding ((?AsignFindingWeight * ?FindingWeight) AS
?TotalFindingWeight) ?FindingWeight ?AsignFindingWeight
        {
        ?Diagnosis ?property ?Finding.
         ?property rdfs:subPropertyOf rp:Finding_Diagnosis_properties.
          ?Finding rp:hasAssignedFindingWeight ?AsignFindingWeight.
          ?Finding rp:hasFindingWeight ?FindingWeight.
          ?Finding rp:hasFindingStatus rp:9_FindingPresent.
          }
}
```

Figure 5-50 - SPARQL insert calculating Total Finding Weight

The SPARQL insert query described above (Figure 5-50) calculates the Total Finding Weight. The Total Finding Weight is the product of the Assigned Finding Weight and the Finding Weight. The Assigned

Finding Weight has been assigned to each individual Finding by the domain expert. The Finding Weight is calculated by the observations that are present and either support or negates a Finding.

The following 2 figures (Figure 5-51 & Figure 5-52) illustrates the relationships between the Diagnosis individuals and the individuals belonging to the Finding classes. The example shown in (Figure 5-51) shows the relationship between the Dentinal Hypersensitivity (Diagnosis) and the Pain present and Initiated by Cold (Finding). The figure shows that the Finding has an individual rp:Pain_present_and_Initiated_by_cold_present_Plus_2 has an Assigned Finding Weight of +2, and its supports the individual belonging to the Dentinal Hypersensitivity Class.

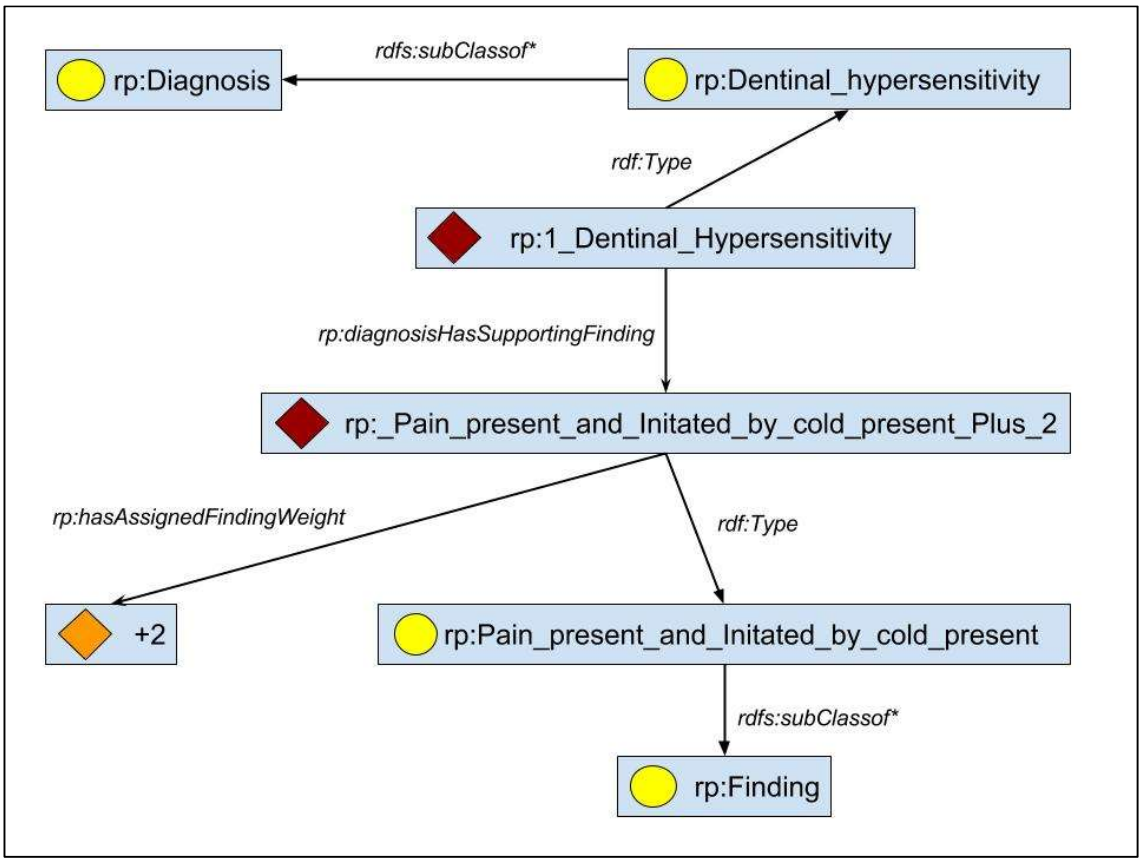


Figure 5-51 - rp:Pain_present_and_Initiated_by_cold_present_Plus_2 supports the Diagnosis rp:Normal_pulpal

The example shown in (Figure 5-52) shows the relationship between the Normal Pulp Diagnosis and the Pain present and Initiated by Cold Finding. The figure shows that the Finding has an individual rp:Pain_present_and_Initiated_by_cold_present_Minus_2 .

This individual has an Assigned Finding Weight of -2, and it supports the individual belonging to the Normal Pulp class.

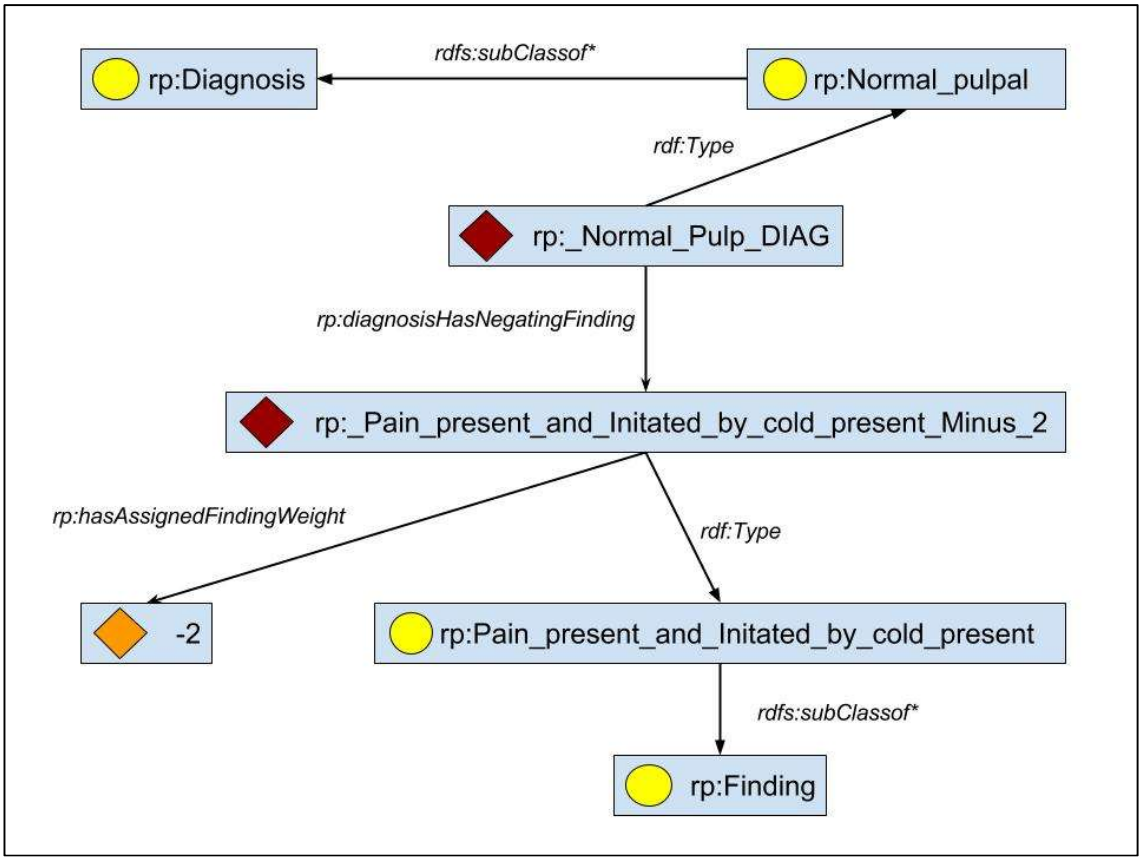


Figure 5-52 - rp:Pain_present_and_Initiated_by_cold_present_Minus_2 negates the Diagnosis rp:Normal_pulpal

The different Diagnosis candidates can have Observations and Findings supporting/negating them at the same time. Some Observations may support/negate a Diagnosis directly while other Observations may support/negate a Finding, which then supports/negate a Diagnosis. A CDSS application with a Multi-level DS Model should consider both Observations and Findings. Figures (Figure 5-39) and (Figure 5-40) show the direct relationships between the Observations and Diagnosis. A Diagnosis-Observation Weight is calculated using a SPARQL Insert query below (Figure 5-53) based on Observations that are Present. This Diagnosis-Observation Weight is attached to each Diagnosis. This process of calculating the Diagnosis-Observation Weight is like the DS Model (but without a Multi-level structure) as shown in (Figure 5-17).

```
INSERT {?Diagnosis rp:hasDiagnosisObsWeight ?DiagnosisObsWeight}
WHERE
        {SELECT (SUM (?Weight)AS ?DiagnosisObsWeight) ?Diagnosis
        {?Diagnosis ?property ?Observation.
        ?property rdfs:subPropertyOf rp:Observation_Diagnosis_properties.
        ?Observation rp:hasWeight ?Weight.
        ?Observation rp:hasObservationStatus rp:9_ObservationPresent.
        }
GROUP BY ?Diagnosis
}
```

Figure 5-53 - SPARQL Insert Diagnosis-Observation Weight

In the Multi-level diagnosis model a Diagnosis-Finding Weight is added to the weights obtained from observations. Certain observations support a diagnosis directly. However, some observations can support a finding which in turn can support the same diagnosis. The sum of weights of observations becomes total weight of the finding. The total weight of the finding will in turn affect the total weight of the diagnosis. Therefore, the final weight of the diagnosis can be influenced by the observation directly and through findings indirectly in some cases. However, the negative weights can reduce the weights of findings in certain cases.

The Diagnosis-Finding Weight is calculated from the Findings that are present and either support or negates a Diagnosis (See Figure 5-47 to Figure 5-50). This Diagnosis-Finding Weight is calculated using the SPARQL Insert query below (Figure 5-54).

```
INSERT {?Diagnosis rp:hasDiagnosisFindWeight ?DiagnosisFindWeight }
WHERE
        {SELECT ?Diagnosis (SUM (?TotalFindingWeight)AS ?DiagnosisFindWeight)
        {?Diagnosis ?property ?Finding.
        ?property rdfs:subPropertyOf rp:Finding_Diagnosis_properties.
        ?Finding rp:hasTotalFindingWeight ?TotalFindingWeight.
        ?Finding rp:hasFindingStatus rp:9_FindingPresent.
 }
 GROUP BY ?Diagnosis }
```

Figure 5-54 - SPARQL Insert Diagnosis-Finding Weight

Once the Diagnosis-Observation Weight and the Diagnosis-Finding Weight has been calculated the Total Diagnosis Weight is attained by summing => (Diagnosis-Observation Weight + Diagnosis-Finding Weight). This is done using the SPARQL Insert query below (Figure 5-55). The Total Diagnosis Weight is then attached to each Diagnosis and is used as the basis for obtaining the ranked diagnostic recommendations.

```
INSERT {?Diagnosis rp:hasDiagnosisTotalWeight ?TotalDiagnosisWeight }
WHERE
        {SELECT ?Diagnosis  (SUM (?DiagnosisFindWeight+ ?DiagnosisObsWeight)AS
        ?TotalDiagnosisWeight)
        {
        ?Diagnosis  rp:hasDiagnosisObsWeight ?DiagnosisObsWeight.
        ?Diagnosis rp:hasDiagnosisFindWeight ?DiagnosisFindWeight.
 }
GROUP BY ?Diagnosis }
```

Figure 5-55 - SPARQL Insert Total Diagnosis Weight

Once all the instances of the Diagnosis have the weight attached to it, the following queries (Figure 5-56 & Figure 5-57) will rank them in descending order of weights. The pulpal and apical diagnoses are ranked separately.

```
SELECT ?PulpalDiagnosis ?DiagnosisTotalWeight
WHERE
        {?PulpalDiagnosis  rp:hasDiagnosisTotalWeight ?DiagnosisTotalWeight.
        ?PulpalDiagnosis a ?Diagnosis.
        ?Diagnosis rdfs:subClassOf* rp:Pulpal_Disease.
} ORDER BY desc (?DiagnosisTotalWeight)
```

Figure 5-56 - Pulpal diagnosis (Multi-level) ranked list SPARQL select query

```
SELECT ?ApicalDiagnosis ?DiagnosisTotalWeight
WHERE
        {?ApicalDiagnosis  rp:hasDiagnosisTotalWeight ?DiagnosisTotalWeight.
        ?ApicalDiagnosis a ?Diagnosis.
        ?Diagnosis rdfs:subClassOf* rp:Apical_Disease.
} ORDER BY desc (?DiagnosisTotalWeight)
```

Figure 5-57 - Apical diagnosis (Multi-level) ranked list SPARQL select query

The following tables (Table 5-15 & Table 5-16) describe the results obtained after running the DS Model against a test scenario/vignette using only the observations and the findings that are present in the patient data example in Table 5-14.

Pulpal Diagnosis	Diagnosis Weight
rp:1_Dentinal_Hypersensitivity	6
rp:1_Reversible_Pulpitis_DIAG	2
rp:1_Symptomatic_Irreversible_Pulpitis_DIAG	2
rp:1_Pain_of_Nonodontogenic_Origin_DIAG	0
rp:1_Pulp_Necrosis_DIAG	0
rp:1_Asymptomatic_Irreversible_Pulpitis_DIAG	-1
rp:1_Normal_Pulp_DIAG	-4

Table 5-15 - Pulpal Diagnosis (Multi-level)

Apical Diagnosis	Diagnosis Weight
rp:1_Acute_apical_abscess_DIAG	1
rp:1_Asymptomatic_Apical_Periodontitis_DIAG	0
rp:1_Chronic_Apical_Abscess_DIAG	0
rp:1_Normal_Apical_Tissues_DIAG	0
rp:1_Pain_of_Nonodontogenic_Origin_DIAG	0
rp:1_Symptomatic_Apical_Periodontitis_DIAG	0

Table 5-16 - Apical Diagnosis (Multi-level)

5.5.5 Summary

The Multi-level DS Model can be accessed via Appendix 5 alongside the DS Model. The CDSS prototype demonstrating the generation of diagnostic recommendations can be accessed via Appendix 8.

The Multi-level DS Model can be used with any of the case studies described in Section 5.4. **The Multi-level DS Model need additional SPARQL rules that indicate which Observations are present for the Findings and which Findings are present for the Diagnosis. These additional SPARQL rules will need to be added to the CDSS alongside the Multi-level DS Model to generate the recommendations.**

The Multi-level DS Model expands upon the DS Model and gives its more granularity when modelling intermediate states/sub-diagnostic states like Findings. It is a demonstration of how the sub-diagnostic states can be modelled by reusing an existing DS Model if the requirements dictate its development.

The next section discusses the key findings and conclusions of this chapter.

5.6 Discussion

In this chapter, development of the DS Model was outlined. The different sections of the DS Model were explained using dentistry as an example. The DS Model was implemented using several CDSS architectures with each one outlining the different methods through which the DS Model can be implemented. The techniques used to model the DS Model can be used to create diagnostic inference models in domains other than dentistry.

The DS Model uses argumentation-based methods to generate diagnostic recommendations. In Argumentation Theory the decision is characterised by having one or more candidates. Each candidate has a set of arguments that support or negate the candidate. The Chapter 4 optometry demonstrated the feasibility of using argumentation-based methods in optometry. The DS Model applies argumentation to support diagnostic decision making. The candidates are the different diagnoses and the arguments are the presence of absence of observations. The DS Model demonstrates how these diagnostic arguments can be modelled using OWL/RDF using existing ontology modelling tools.

The DS Model demonstrates how the domain expert can capture his/her knowledge in an OWL model in the form of arguments. It allows the domain expert to add the degree of support for an argument in the form of weights within the same OWL ontology model. A simple weighting scheme of using +1 numeric

weight is used if the presence/absence of an observation supports a diagnosis. Similarly, +2 is used if the observation is highly indicative of a diagnosis and strongly support the diagnosis. Likewise, a negative numeric weight is used if the presence/absence of the observation negates a diagnosis. Weights above +2 or below -2 can be used for observations that can confirm or completely negate a diagnosis. This weighting scheme is easy to understand for domain experts and can be used to provide effective diagnostic recommendations for the clinicians as demonstrated in the Chapter 4 Optometry study. The DS Model uses the same weighting scheme but in the domain of dentistry. However, one of the main problems associated with using a simple weighting scheme is the potential for clashes. Many observations can share the same weight for a diagnosis. In the early stages of the clinical encounter where there is little information there is potential for many diagnoses to share the same rank. Therefore, it is critical to study how accurate the rankings are in the early stages of the clinical encounter. Chapter 6 studies this aspect in more detail. The use of argumentation and modelling diagnostic arguments using semantic web technologies makes the DS Model different from other researchers who have developed ontology driven CDSSs.

The work done by García-Crespo et al. (García-Crespo *et al.*, 2010a) demonstrates the use of a diagnostic CDSS using OWL ontologies that is based on an external probabilistic system. The use of probabilities requires the use of an external probabilistic system to process the probabilities. The literature shows how argumentation-based methods have an advantage over other reasoning techniques such Bayesian reasoning. It is easier for the domain expert to capture his/her knowledge in the form of arguments. Translating the domain expert's knowledge into Bayesian probabilities is only easy in well understood domains. The DS Model has demonstrated how diagnostic recommendations can be generated without the use of any external tools or specialized software.

Others have demonstrated the use of OWL ontologies for medical diagnosis but have relied upon DL and SWRL/Jena rules to map the relationships between diagnosis and observations and provide diagnostic recommendations (Bertaud Gounot *et al.*, 2011b; A Rodríguez-González *et al.*, 2012; Bertaud-Gounot, Duvauferrier and Burgun, 2012b; Donfack Guefack *et al.*, 2012). However, these the models cannot represent the degree of support of a diagnosis due to limitations in the expressiveness of DL.

DL were developed as a subset of First Order Logic and is based on First Order Logic (Krötzsch, Simancik and Horrocks, 2012). First Order Logic is highly expressive and can be used to express any fact. However, First Order Logic is very complex and is only semi-decidable. Therefore, it is not suitable for reasoning facts given a set of data, especially on the web. The reasoner could potentially run forever and not be able to provide a suitable inference from a set of data. DL are a family of logics that were derived from First Order Logics to remedy some of the problems with decidability. DL are more decidable and are simpler when compared to First Order Logics. There are several variations in DL and some variants are more decidable than others. OWL DL is the DL variant of the OWL language (Harmelen and McGuinness, 2004).

DL uses the Open World Assumption. In Open World Assumption what is not known to be true is considered as an unknown entity. In the alternate Closed World Assumption what is not known to be true is considered false. The Open World Assumption is useful when working with distributed data and data from the web. When working with distributed data a new fact can be discovered at any time. The Closed World Assumption works when the system has complete information. This is used by many database applications. If a patient is not registered in the hospital EHR system and if you query for the patient's details, then the system will look in the database and return a negative result if the patient is not found.

However, in the Open World Assumption, the system operates with incomplete information particularly when the information is located on the web. Open World Assumption is helpful when we use inference engines and ontologies to infer new information from existing ones. If the patient's EHR doesn't have information about allergy to a medicine, we cannot assume that the patient doesn't have allergy to the medicine. If someone queries for the patient's medication allergy a system operating with an Open World Assumption will return an 'unknown' as result.

The Open World Assumption is used in DS Model to infer if a clinical diagnosis/observation belongs to a class of diagnoses/observations in the DS Model. We can use OWL DL axioms to define the criteria with which the diagnosis/observation belongs to a class or subsumed by another class. For example, if the data has been annotated using a SNOMED-CT code representing dental pain then this information can be used to infer the fact that the patient has previously complained about dental pain. We can use OWL axioms to represent this fact in the DS Model and the OWL reasoners can infer this fact automatically.

However, for the CDSS to make a diagnostic decision the CDSS must know which observations are present or absent. It must use this information to give a definitive recommendation about the patient. DL cannot be used to represent 'negation as failure', as it uses an Open World Assumption. In 'negation as failure', the CDSS tries to prove that the observation is present. If the observation is present (i.e. it is proven) by examining the patient data, then its negation i.e. not (observation is present) fails. Therefore 'observation is absent' fails.

If the result of a query is unknown the CDSS cannot decide. It must use the information currently available to make the decision. Therefore, the Closed World reasoning must be used. DS Model uses SPARQL rules to close the world and infer which observations are present/absent based on the patient data. Other researchers have used OWL axioms to represent diagnostic criteria and use SWRL rules to overcome the limitations of DL to infer additional facts (Bertaud Gounot *et al.*, 2011b; A Rodríguez-González *et al.*, 2012; Bertaud-Gounot, Duvauferrier and Burgun, 2012b; Donfack Guefack *et al.*, 2012). However, SWRL cannot model disjunctions and negation as failure. SWRL cannot be used to model complex calculations using weights and query remote endpoints to retrieve patient information.

The DL reasoners such as Pellet (Sirin *et al.*, 2007a), can be used to infer facts about the patient and used for consistency checking during model development of the DS Model. The SPARQL rules help query remote endpoints for patient information, assert which observations are present/absent based on the patient data and calculate the weights for each diagnosis. The diagnostic recommendation is then presented using argumentation-based methods and the reasons for the recommendation are provided. The DS Model provides a simple method of modelling that takes advantage of all modelling capabilities of OWL and its DL reasoners to help in the modelling process. The use of OWL helps in reusing existing ontologies to capture the relations between Diagnoses and Observations. SPARQL rules overcome the limitation of existing semantic web rules languages and uses the weighting mechanism in the DS Model to generate a clinician friendly diagnostic recommendation.

Most ontology driven diagnostic CDSS applications identified in literature do not provide a ranked diagnostic recommendation. The ones that do provide a ranked recommendation use external algorithms that incorporate other factors such as relative incidence of the disease into the decision making (Oberkampf *et al.*, 2012; Oberkampf, Zillner and Bauer, 2012b). The CDSS applications in this study uses SPARQL rules on the DS Model to generate the ranking. There is no need for another rule language or custom ranking algorithm when using SPARQL rules to map the patient data to the DS Model and provide the diagnostic ranking. SPARQL is a W3C standard (W3C®, 2013b) and is widely implemented in several commercial and open-source triple stores. Therefore, additional software to generate the rankings are not needed. Any triple store capable of accepting RDF data and processing SPARQL queries could be used to generate diagnostic recommendations. The simple weighting scheme used and using SPARQL inference to sum the weights means the diagnostic recommendations and associated weights can be generated easily when compared to other methods. This is crucial for generating early diagnostic recommendations. The data needed to make the early diagnostic recommendation will come from the diagnostic encounter. The use of semantic web technology means that the data can be integrated easily. Using a light-weight mechanism to generate the rankings means large quantities of data can be processed to quickly generate the diagnostic recommendation in the early stages of the clinical encounter.

The use of argumentation-based methods will result in a simple explanation of the patient's condition. The clinician is presented with alternative diagnoses for the patient. The literature suggests that providing alternative diagnoses can help in System 1 reasoning process and help reduce diagnostic error. Similarly,

Croskerry (Croskerry, 2003a) has identified several strategies to reduce diagnostic errors due to cognitive causes. This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information etc. The working hypotheses for the clinical case is generated early in the diagnostic encounter based on a few clinical patterns that are observed in the patient. These working hypotheses may not be accurate always and are tested by acquiring additional information from the patient. Using the illness script model, the working hypotheses is triggered early in the diagnostic encounter. Generating the working hypotheses early in the encounter helps the clinician to narrow the solution space. Using the working hypothesis, further information is gathered until one of the working hypotheses is confirmed. The argumentation-based diagnostic recommendations generated by the DS Model can help in this process by providing a list of diagnosis at various stages of the clinical encounter. It provides the reasons for making these recommendations in the form of arguments. This helps in the process of generating the working hypotheses and help in gathering additional information about the patient to help confirm or refute working hypotheses. The argumentation-based CDSS can help organizing the expert's knowledge into arguments and provide recommendations that are easy to understand.

The use of linked data technology and use of RDF will allow the CDSS implementing the DS Model to access disparate sources of data and use the data in diagnostic decision making. Case Study 1 demonstrates a prototype CDSS application that uses the DS Model as the inference model to provide diagnostic recommendations in a standalone desktop application. Case Study 1 prototype uses a triple store as the backend to store the patient data. It demonstrates how a standalone CDSS application can use the DS Model to generate diagnostic recommendation. This case study demonstrates how the existing triple-stores and semantic web technology can be used to generate the diagnostic recommendations without the use of any other specialized tools. It cannot access linked data directly but can be modified if needed to access data from other sources, rather than relying entirely on the clinician input.

Case Study 2 demonstrates how data stored in relational database can be translated to RDF format using the D2RQ method. The D2R server exposes the data in a MySQL relational database in the form of a SPARQL endpoint. Existing applications using relational databases as their backend can expose their data via the SPARQL endpoint. Linked data technology can then identify data that belongs to the patient, provided the data is described using standard ontologies. For example, the FOAF (Friend of a Friend) ontology provides a set of terms that can be used to describe people and link people with other types of information on the web (Brickley and Miller, 2014). The patient currently being examined by the clinician can be linked to the 'foaf profile' of the patient published on the web. If a web resource such as a PHR can describe the data using FOAF and other clinical terminologies, intelligent agents and other software can find the relevant information about the patient and can then use it in CDSS applications. The CDSS application can then use this data to generate early diagnostic recommendations and do not have to rely entirely on the data provided by the patient during the clinical encounter. Case Study 2 demonstrates how a CDSS application using the DS Model can use the data via a SPARQL endpoint to provide diagnostic recommendations. Linked data can be stored in triple stores or relational database and exposed via a SPARQL endpoint. The Case Study 2 prototype application can then access this data to provide diagnostic recommendations using the SERVICE keyword within the SPARQL query. The patient data can be stored in multiple SPARQL endpoints, and the CDSS application can access this data.

Case Study 3 showed how the data in an EHR can be represented in a standardised manner using archetypes. The DIM model was represented as archetypes and separated into different sections. The different sections represent the major steps in the diagnostic encounter. The architecture developed in Case Study 3 will be used in Chapter 6 for evaluating the DS Model, and the diagnostic performance of the DS Model will be compared to that of dentists. APGD was used to generate patient data based on the archetype model in the RDF data format. The RDF data was added to a triple store alongside the DS Model and the CDSS application could use the data to generate the diagnostic recommendations. This Case Study demonstrates how patient data can be represented in different formats and using SPARQL rules the patient data can be mapped to inform/update the DS Model. To generate early diagnostic recommendations based on linked data on the web the CDSS must be able to use patient data in different formats and stored in disparate locations. Case Study 2 shows how the data can be accessed when stored

in disparate locations. Case Study 3 shows how the DS Model helps the CDSS use different forms of patient data by modifying the SPARQL rules.

Case Study 4 expands on the architecture described in Case Study 2 and used the DIM to represent patient data in an open-source EHR. The data stored in the relational database of the EHR was then exported into the RDF format using the D2RQ software. The RDF data was then uploaded to a triple store alongside the DS Model and the patient data was used to generate diagnostic recommendations. One of the challenges facing current differential diagnosis generator systems is the problem of integration with EHR systems (Bond *et al.*, 2012a) and this has affected the wide spread uptake of these systems in routine clinical practice. Case Study 4 shows how the patient data stored in currently available EHR systems can be used by the CDSS using the DS Model. However, exposing patient data via the SPARQL endpoints on EHR databases and querying them using complex SPARQL queries may result on slow performance and therefore the data should be exported for the CDSS application to use. This is not conducive for early diagnosis as the clinician will need diagnostic recommendations at each stage of the clinical encounter.

One of the main challenges faced during development of the prototypes was the availability of actual patient data via linked data repositories. This is mainly due to concerns about patient data confidentiality and issue concerning the individual's privacy on the linked web of data (Corsar, Edwards and Nelson, 2013). Relational databases and healthcare web services go to extraordinary lengths to protect the patient's privacy and providing complete protection remains a challenge (Sachdeva, Mchome and Bhalla, 2010; Harman, Flite and Bond, 2012). Linked data and semantic web technologies have been developed to increase data sharing and interoperability. Ensuring privacy and confidentiality on the linked web of data is a challenge and some research has been done in area (Aron, 2013) though more needs to be done in the area of privacy and confidentiality of linked healthcare data.

Case Study 3 and Case Study 4 prototypes show how the same DS Model can be used with different representations of patient data. The patient data in Case Study 3 is an archetype based model and patient data in OpenEMR is a model that is specific to that system. The same patient data can be represented in different forms and ability to generate diagnostic recommendations across these iso-semantic models (See 5.4.3.1) is a key strength of the DS Model. By making minor modifications to the SPARQL present/absent insert queries the same CDSS prototype in both cases studies could generate diagnostic recommendations. The SPARQL rules that mapped the patient data to the DS Model was specific to that patient data model. If a standardised patient data representation is available, the mapping will need to be created only once. Such efforts are currently being undertaken by several different healthcare standards organizations such as OpenEHR and HL7. However, until a common patient data representation is available we will need rules for mapping the patient data to inference model.

Developing a common standard that can be applicable across all healthcare environments has not been achieved due to a variety of reasons. The nature of medicine and the complexities associated with it means finding a standard that can suit a diverse set of use cases is difficult. The process of creating standards and adopting them is time consuming process. This results in a situation where there is competition amongst healthcare standards vying to become the common model (Booth, 2015d) although this does not achieve the aim of improving healthcare semantic interoperability.

The Yosemite Project (Yosemiteproject.org, 2016) tries to address the problem of semantic interoperability amongst healthcare standards by using RDF as a universal representation language. The patient data can be represented in any format (including non-RDF formats). RDF captures the semantics of the data and can then be used to transform the data from one format to another. The figure below (Figure 5-58) from yosemiteproject.org (Booth, 2015a) illustrates how the meaning of different formats (HL7 v2.X and FHIR) can be captured in a RDF graph structure.

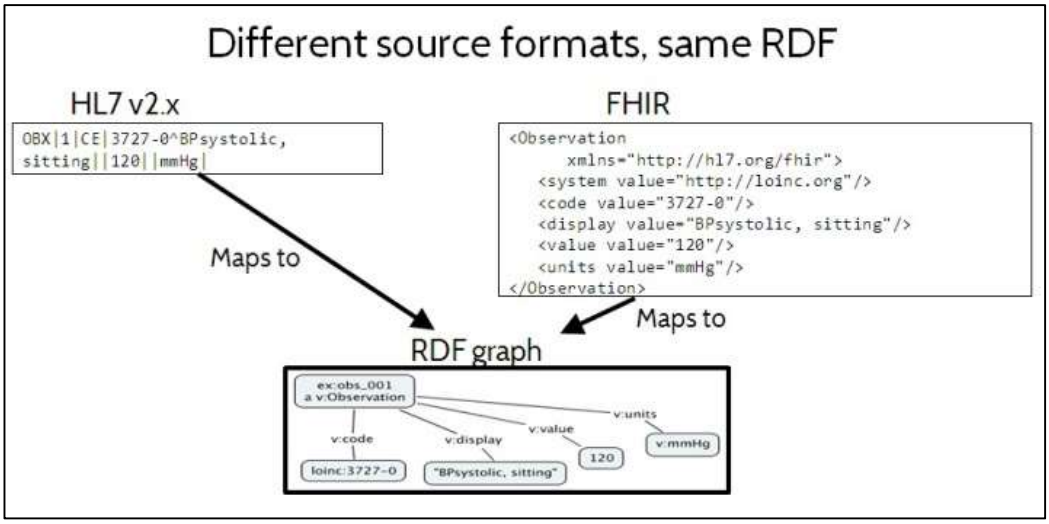


Figure 5-58 - Different source formats, same RDF

RDF enables us to capture the information content and the meaning of the content, separate from the data format and that quality makes RDF a suitable candidate for a *universal information representation* language (Booth, 2014b). One of the long-term goals of the Yosemite Project is to build a collaborative crowd-sourced translation rules hub that would hold the translation rules for translating data from one format to another (Booth, 2014a). SPARQL rules can be used alongside other rules languages or programming languages (Jena rules, SPIN, N3, Python) to perform this translation.

The figure below (Figure 5-59) from yosemiteproject.org (Booth, 2015b) shows how the rules stored in the translation hub will help perform the translation.

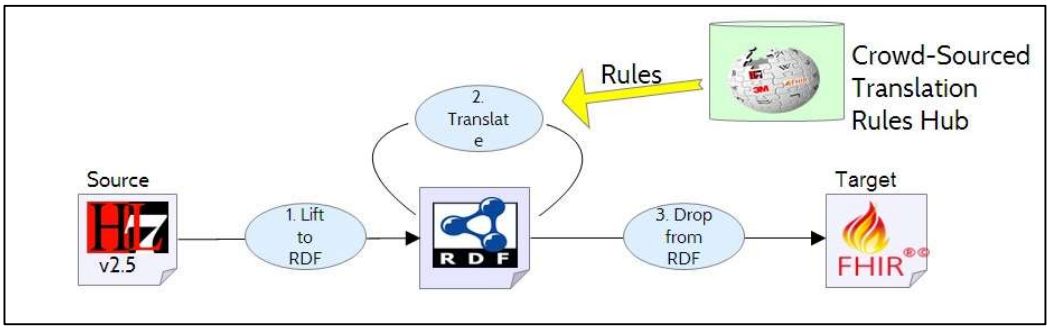


Figure 5-59 - Translating patient data

Once these translation rules are in place there is no need for a common standard or data model or vocabulary across healthcare. The SPARQL present/absent insert rules used in this study can work in conjunction with the SPARQL rules of the Yosemite project to map one data representation to another and map the patient data to the DS Model.

The Multi-level DS Model expands upon the DS Model and gives its more granularity when modelling intermediate states/sub-diagnostic states like Findings. It is a demonstration of how the sub-diagnostic states can be modelled by reusing an existing DS Model if the requirements dictate its development.

The same DS Model can be expanded to include intermediate states via the Multi-level DS Model. The ability to model multiple levels has been a challenge faced by other researchers as well (Rodríguez *et al.*, 2009b). They had attempted to solve the problem of multi-level diagnosis by using an approach similar to the Multi-level DS Model where the diagnosis is supported by a finding, which in turn can be supported by another set of findings (Alejandro Rodríguez-González, Torres-Niño, Mayer, *et al.*, 2012). They have used the term 'finding' to denote all clinical observations.

The Multi-level DS Model is an extension to the DS Model that can capture additional detail in the form in intermediate sub-diagnostic states. Argumentation-based methods used in the DS Model was used to develop arguments that supported each Finding. The presence or absence of the Finding was used as arguments to support a diagnosis. The Finding is a sub-diagnostic state and there are use-cases where additional explanation needs to be provided to the clinician. Certain diagnostic tests are indicated when the dentist needs to confirm the status of the pulp. The finding 'Inflammation of the pulp' can be triggered by the observations ("Pain present" and "Pain initiated by cold") during the clinical encounter and a ranked list of findings can be displayed to the clinician to help the clinician choose the appropriate test based on these findings or let the CDSS application choose the right test for the clinician. The findings can therefore be used as a trigger for generating guideline recommendations as well. However, Chapter 4 Optometry study showed how the guideline recommendations can be generated without the presence of findings, instead the Tallis CDSS uses abnormal values as triggers.

In this study, the example of a joint observation was used to represent a finding and to demonstrate the CDSS generating a diagnostic recommendation in the Multi-level DS Model. A joint observation is an observation that is a combination of 2 different observations that are present. The presence of 2 clinical observations can have other interpretations as well. For example, the presence of pain and the observation "pain is initiated by cold" is a possible indication of inflammation of pulp. There are areas where this additional level of explanation is needed for the clinician.

However, this additional level of explanation comes at the cost of increased complexity. The domain experts will need to model the relationships between the observations and the findings. In the DS Model the Diagnoses and the Observations can be reused from existing ontologies such as SNOMED-CT. Findings as a sub-diagnostic state are not captured in existing ontologies at present. Existing DS Models can be expanded to develop the Multi-level DS Model.

The literature review did not provide examples of models like the Multi-level DS Model in other argumentation-based languages such as PROforma. Examples from literature that demonstrated modelling multiple levels of observations using ontologies did not model sub-diagnostic states as demonstrated in the Multi-level DS Model (Alejandro Rodríguez-González, Torres-Niño, Mayer, et al. 2012). These can be adapted to create Multi-level DS Models in future if they find widespread use however, they have not found wide applicability to date.

The guideline recommendations can be generated using abnormal values as triggers as demonstrated in the Chapter 4 study. Therefore, the efforts needed to develop the more complex Multi-level DS Model is not justified for routine diagnostic CDSS applications. Chapter 4 optometry study demonstrates the use of argumentation-based methods to generate diagnostic recommendations that are easy to understand. The DS Model used argumentation as well in addition to using semantic web technologies to model the arguments and to integrate information from disparate sources. In situations where the requirements dictate a need for findings as shown in the sections above, the Multi-level DS Model can be developed.

This thesis is focussed on developing an ontology model capable of early diagnostic recommendations. The DS Model offers an elegant solution to this problem and expands on existing research in the use of argumentation-based methods and the use of ontologies for CDSSs. Developing a fully modelled Multi-level DS Model that can generate recommendations for sub-diagnostic states and diagnosis and evaluating the results with clinical vignettes and comparing the results to dentists was considered beyond the scope of this thesis. Therefore, evaluation was restricted to the DS Model and Case Study 3 prototype in Chapter 6. The Case Study 3 prototype was chosen because of the use of APDG patient data generator tool that allows us to generate RDF patient vignettes easily for the purposes of evaluation.

5.6.1 Conclusions

Health information technology have been used to reduce diagnostic errors (El-Kareh, Hasan and Schiff, 2013) and the existing tools can be classified into different categories:

- Tools that assist in information gathering

- Tools that improve cognition through improved organization and display of information
- Tools that assist in generating a differential diagnosis
- Tools that assist in weighing the diagnosis
- Tools that support selection of diagnostic tests and plans
- Tools that enable access to diagnostic reference material and guidelines
- Tools that enable patient follow-up and continued assessment of patient course and response to treatment
- Tools that support screening by alerting presence of disease
- Tools that assist in diagnostic collaboration
- Tools that facilitate feedback and insight into diagnostic performance

The DS Model CDSS belongs to the category of helping in information gathering. It helps the clinician in information in improving cognition through improved organization and display of information using argumentation. It helps in generating a differential diagnosis and the weights and arguments help in weighing the diagnosis.

The DS Model demonstrates the ability to capture the expert's knowledge in easy to understand manner using argumentation-based methods. The DS Model demonstrates how using arguments the diagnostic criteria can be modelled in OWL language. The case study prototypes show how data in various forms can be used with the DS Model and how using semantic web and linked data technologies data can be integrated to generate diagnostic recommendations. DS Model likewise demonstrates how the model can be created using existing ontology modelling tools and the diagnostic recommendations generated without the use of specialized software tools to perform inference. The ranking was generated using a simple weighting scheme that was demonstrated in Chapter 4. The diagnosis, ranking and the weights together better explain the patient's condition for early diagnosis and thereby help reduce diagnostic error.

In the next chapter, the DS Model will be evaluated with the help of the architecture developed in Case Study 3. Clinical vignettes were developed and used to evaluate the DS Model by comparing the results of ranked diagnostic recommendations with that of dentists.

6 Chapter 6 - Model evaluation and Results

6.1 Introduction

The DS Model described in the previous chapter (Chapter 5) was evaluated using an experiment described below. The CDSS prototype developed using the architecture described in Case Study 3 of the previous chapter (Chapter 5 – section - 5.4.3) was used for model evaluation. Any of the case study prototypes can be used for the evaluation. The architecture in Case Study 3 was used because of the ease of generating RDF patient data based on the vignettes using APDG tool for evaluating the performance of clinicians. APDG uses the archetype model and Patient Data Definition Language (PDDL) to create patient data instances (Huang et al. 2013). The APDG tool allows us to model different clinical vignettes in RDF format. An experimental evaluation of the impact of DS Model on clinician's performance is important; however, this is beyond the scope of this thesis.

In this chapter, the DS Model was evaluated by comparing the diagnostic results generated by the CDSS with the results generated by clinicians. The aim of this chapter is to answer the research question:

How does an ontology driven diagnostic CDSS perform across all stages (early and late) of a clinical encounter and can it support the clinician in making reliable diagnostic decisions?

The study aims to evaluate performance of the DS Model and study its performance across all stages of the clinical encounter. Literature review had identified cognitive causes such as premature closure bias as the main reason for diagnostic error. In premature closure the clinician prematurely halts the diagnostic decision-making process before all the data has been collected and the diagnosis has been fully verified. Interventions suggested to address this includes providing alternative diagnoses. Providing alternate diagnosis has been suggested as a method to improve System 1 process of reasoning. Providing early diagnostic recommendations have been suggested as a technique that could potentially reduce diagnostic error. Similarly, Croskerry (Croskerry 2003b) has identified several strategies to reduce diagnostic errors due to cognitive causes. This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information etc. Other causes for diagnostic error include problems faced by clinicians in coordinating information from multiple sources.

Clinicians rely on a differential diagnosis to identify the presence or absence of a disease when there are multiple alternatives available (Heneghan *et al.*, 2009). A clinician examining a patient goes through a multi-step procedure to attain a differential diagnosis. According to hypothetico-deductive model of clinical reasoning (Elstein, Schwartz and Schwarz, 2002) clinicians (especially novice clinicians) obtain a diagnosis following the steps described below (Harbison, 1991).

- At first, they obtain information by taking clinical history, complete the physical examination and conduct diagnostic tests if needed.
- A working diagnosis (hypothesis) is generated and the clinicians consider all possible diagnoses in the form of a mental list.
- Further information is sought from the patient to confirm or refute the working diagnosis (if needed).
- The diagnosis is verified by weighting and combining information from all sources and then arriving at a conclusion that confirms one of the working diagnoses.

The working hypotheses for the clinical case are generated early in the diagnostic encounter based on a few clinical patterns that are observed in the patient. These working hypotheses may not be accurate always and are tested by acquiring additional information from the patient. Using the illness script model, the working hypotheses is triggered early in the diagnostic encounter. Generating the working hypotheses early in the encounter helps the clinician to narrow the solution space. Using the working hypothesis, further information is gathered until one of the working hypotheses is confirmed.

The literature identified argumentation as an appropriate choice for presenting diagnostic recommendations to clinicians, especially early diagnostic recommendations. Argumentation using non-monotonic logic where existing conclusions can be discarded in the light of current information and new conclusion can be formed. Early diagnosis is characterised by incomplete information. The clinician has not collected all the information needed to make a diagnosis. The clinician uses this incomplete information to form initial diagnostic hypotheses. Using this information, the clinician then obtains additional information about the patient and then uses this information to confirm or reject a hypothesis. The alternative diagnoses provided at different stages of the clinical encounter helps trigger appropriate illness scripts that can then be used by the clinician to find more information about the patient. The alternative diagnoses help in the process of generating the working hypotheses and help in gathering additional information about the patient to help confirm or refute working hypotheses. The alternate diagnostic recommendations presented in various stages of the clinical encounter alongside detailed reasons for the recommendations could potentially help reduce premature closure bias by informing the clinician about alternate diagnoses. This may stop the clinician from halting the diagnostic information gathering process and consider other alternatives.

Providing alternate diagnosis helps in System 1 process of reasoning and helps reduce diagnostic error. The clinician would look at the ranked list of diagnoses and this triggers pattern recognition or activation of illness scripts that the clinician can use to elicit further information from the patient. The arguments for each diagnosis help explain the reasons for the recommendation. This could potentially help in System 2 process of reasoning and help slow down the process of thinking and help analyse all the information available.

The CDSS must make an inference based on the incomplete information collected by the clinicians in the early stages of the clinical encounter and suggest a diagnostic recommendation. The web is excellent source of information that can be used to access additional information about a patient. However, the information is presented in various formats and due to problems in semantic interoperability the data cannot be integrated to support inference in diagnostic CDSSs.

Though argumentation-based CDSSs have been developed for improving medical decision making in areas such as diagnosis, management, treatment planning, risk assessment etc., studies so far have not evaluated the impact of an argumentation-based diagnostic CDSS on the diagnostic performance of clinicians. Chapter 4 Optometry study evaluated impact of argumentation-based diagnostic CDSS on the diagnostic performance of optometrists. The Chapter 4 Study demonstrated the feasibility of developing an argumentation-based diagnostic CDSS using existing tools and highlighted the deficiencies of existing argumentation-based CDSS tools. The study showed that argumentation-based diagnostic CDSS could potentially help reduce diagnostic error. It demonstrated the feasibility of the argumentation approach for creating diagnostic CDSSs and the users found the system easy to use and recommendations helpful.

Existing argumentation-based tools such as Tallis uses a proprietary format to represent the PROforma model. The model can only be modelled and executed using proprietary software such as Tallis. Integration with EHRs for collecting patient data to generate diagnostic recommendations will need expensive interface engines that can translate the patient data from different formats to a format that can be understood by the Tallis inference engine. The need for a CDSS ontology model was identified that

could use argumentation as the basis to generate diagnostic recommendations. Chapter 4 study evaluated the impact of guideline recommendations on the information gathering of optometrists.

Chapter 5 detailed the development of this argumentation-based ontology driven diagnostic CDSS model called the DS Model and CDSS prototypes that can use this DS Model to generate diagnostic recommendations. The DS Model is aimed at generating diagnostic recommendations at all stages of the clinical encounter. The ontology could be modelled using any existing ontology editing tool.

The simple weighting scheme used in DS Model allows the domain expert to capture his/her knowledge in a relatively easy manner. The simple weighting scheme was derived from the weighting scheme used in the Chapter 4 Optometry study. Chapter 4 study validated the model and studies the impact of diagnostic recommendation on decision making of optometrists. However, a comprehensive test evaluating the diagnostic accuracy of the PROforma argumentation-based CDSS was not performed.

The DS Model used a simple summation method to calculate the overall weights for each diagnosis based on the presence or absence of an observation. PROforma argumentation-based decisions likewise uses a similar summation method calculate overall weights for the diagnosis. The DS Model can be used with any off-the-shelf triple store that can read RDF data and process SPARQL queries. No specialized software is needed to perform inference.

The simple weighting scheme and the simple method for calculating weights has the potential for problems when the DS Model is used in real life situations. There can be tendency for diagnoses to share the ranking especially when there is not enough information to make a diagnosis. This study evaluates the performance of the DS Model and compares the ranking with that of dentists. The study will allow us to evaluate the effectiveness of the simple weighing scheme and the summation method used to generate diagnostic ranking.

Clinical vignettes were created and ranked diagnostic recommendations were generated using the CDSS. The same vignettes were presented in a tabular format to the dentists and they were asked to provide diagnostic recommendations for each vignette presented. The performance of dentists was used as the benchmark against which the performance of the DS Model is compared.

The argumentation-based CDSS can help organizing the expert's knowledge into arguments and provide recommendations that are easy to understand. However, the accuracy of the ranking the DS Model will need to be further evaluated to understand if the DS Model can help clinicians in formulating the diagnostic hypothesis. The ranked list generated by the CDSS at the early stages of the clinical encounter may not be as accurate or precise when compared to the later stages or when compared to a clinician. How much does it differ and where can improvements be made in the diagnostic recommendations? Using linked data and semantic web technology it is possible to obtain data from several different sources. It is therefore important to identify areas where additional data needs to be obtained and areas where data obtained from the patient alone will suffice in diagnostic decision making.

The purpose of the evaluation is to compare the ranked diagnostic recommendations generated by the CDSS at all stages of the clinical encounter with the ranked diagnostic recommendations provided by the dentists. The different stages of the clinical encounter as represented in an information model have been described in section - 6.2. The term stage is used to denote the different stages of the clinical encounter. The term section is used to denote the different sections of the information model that correspond to the stage and represent the stage. The dental adaptation of the DS Model which forms the core logical component of the CDSS, the diagnostic recommendations generated by the CDSS using the DS Model and SPARQL rules from a simulated dataset will form the focus of the system evaluation that will be described in section - 6.3.5.2.

6.2 Sections of the Information Model

The information model used in a comprehensive and systematic examination of a patient presenting to the dental clinic is presented in this section. The information model forms the basis for developing data capturing forms and designing the architecture of EHR. Standardised representations of information models captured with the help of languages such as OpenEHR (openEHR Foundation, 2016) are available, or the information model can be captured in a structured format (for example, tabular) from domain experts and later represented using standardised information representation languages. The information model used here has been delineated into several sections for purposes of this study. The information model has been used to develop the dental adaptation of the DS Model and the clinical vignettes used for evaluating the DS Model. Each section contains units of information that are grouped together representing different stages of the clinical encounter. The information model used in this study is based on the DIM model (Acharya, Mital and Schleyer, 2009a) which has been adapted and extended where necessary to support dental diagnostic decision making.

- Chief Complaint (CC): In this section the patient's chief complaint, clinical history of the complaint, family history and medical history is included. For the purposes of this study the demographic information of the patient was included in this section as well. The section was called Chief Complaint (CC) because that is the main information collected by the clinician that provides the initial working diagnosis for the clinician. For example, if the patient presents with pain, then a brief description of the pain followed by the history and characteristics of the pain will form the core of the CC that is relevant to the diagnosis (See Table 6-1). The CDSS must generate a diagnosis at this section based on the data collected. In current clinical environments, such as dental clinics, the information model is restricted to collecting data that has been obtained from the patient directly. Therefore, the CDSS is limited to using the data collected directly from the patient. However, it is possible that history of the pain and other medical history relevant to the diagnosis are stored in other data sources. Using linked data technology, the CDSS software or intelligent agents can obtain information relevant to the patient's diagnosis within the context of CC and support decision making. Therefore, it is important to understand how well the CDSS performs using information models in the configurations currently used in dental environments to identify areas where additional information can be used to improve early diagnosis.

- Extra Oral Examination (EO): Once the CC section is complete the clinician then progresses to Extra Oral Examination. In this section the dentist will examine the areas of the face and neck and examine tissues and glands such as lymph nodes to check if they exhibit signs relevant to the main complaint of the patient. For example, a patient with an infected tooth may complain of pain. The information about pain is collected in CC. However even before examining the tooth directly it is possible to detect if infection has spread from the local environment of the tooth to the patient's system by examining the status of the patient's lymph nodes (Murdock, 2016). The CDSS must combine information from both CC and EO to provide a diagnostic recommendation at this section as well. If the lymph nodes are indeed swollen or tender, then the CDSS should provide a diagnostic recommendation of "Apical Abscess" early on. This may trigger or encourage the clinician to enquire if the patient had an episode of fever or malaise recently if the information is not captured in the information model. The information may remind the clinician to further probe the nature of the fever. The CDSS will be able to query data sources such as a General Practitioner's (GP) EHR about the most recent episode of fever. Sometimes patients can present to the GP's clinic or Accident and Emergency with fever and pain instead of the dentist. The CDSS should be able to access the data residing in these systems as well. The architectures described in Chapter 5 (Case Study 3 & 4) explain how the data can be used from disparate sources to support decision making.

- Intra Oral Examination (IO): The dentist will then proceed to examine the inside of the patient's mouth to examine the hard and soft tissues. The hard tissues include the teeth and surrounding tissues and soft tissue include structures such as tongue and cheek. If the dentist suspects the pain is caused by a decayed tooth then a detailed examination of the teeth, especially teeth on the side indicated by the patient in CC is done. If the teeth are found to be normal or not the cause of the pain, other tissues are examined. Sometimes if no relevant observations are found the dentist might go back to EO to check if any clinical signs have been missed during the EO examination. The ranked list provided by the CDSS should be able to highlight diagnosis that may be relevant, and the dentist has missed. For example, if the patient complains of pain during chewing and there are no relevant IO observations then the CDSS should highlight a non-odontogenic (non-dental) diagnosis as a possible working diagnosis.

- Clinical Tests (CT): The data collected from CC and the physical examination sections (EO and IO) will be used to identify clinical tests to be performed. Clinical Tests (sometimes called Diagnostic tests) can occur in several successive rounds. Sometimes tests must be repeated. The most recent value will be used by the dentist and the CDSS. In some cases, the patient is given a tentative diagnosis and the referred to a diagnostic testing centre for blood or imaging tests and the data used to formulate a final diagnosis. CT includes tests conducted to assess the condition of the pulp of the tooth. The condition of the pulp needs to be assessed first and any pathology ruled out before other potential diagnoses are considered or excluded. In the information model used in this study, diagnostic imaging in the form of radiographs has been modelled in a separate section. CT can indicate the presence/absence of pathology in a tooth. However, this should be confirmed with the help of radiographs and therefore the data items were separated into a different section. In a realistic clinical scenario at this section of the information model the CDSS should be able to provide a diagnosis recommendation as a ranked list that closely matches that of the final diagnosis of the clinician. The most likely recommendations would be located on the top of the list. The least likely recommendations would be positioned at the bottom of the list. Radiographs are only needed to confirm/refute this hypothesis.

- Radiographic Examination (RE): In this section, radiographs of the area of interest are developed. The patient should not be exposed to any additional radiation that what is required for routine clinical use. The risk of harm to the patient from routine dental radiographs is low (Abbott, 2000). Nevertheless, the data from the CT should be used to identify areas of interest and diagnostic radiographs are generated. The radiographs are examined by the dentist or a specialist radiologist and then a report of clinically relevant observations are created. At this section the CDSS should be able to provide a diagnostic recommendation that is either as good as or better than the dentist's own recommendation. Current CDSS systems use the information collected during the encounter to generate the diagnostic recommendation. Therefore, if the clinician has entered the right information into the system and the diagnostic inference model has been tested and evaluated it should provide a recommendation that is at least as good as the clinician's recommendation.

Once CT has been completed the diagnosis is confirmed therefore it is important that the diagnostic CDSS recommend the right diagnosis in the early stage, i.e. before the CT data is collected. The tests performed in CT are sometimes invasive and can be expensive for patients. Getting the right recommendations at an early stage helps the clinician choose the appropriate test(s). Once the clinician has made a mental model of the patient and framed a working diagnosis based on the data collected it is difficult for the CDSS or any suggestion to have an impact on helping the clinician change his/her earlier assertion. This premature conclusion or premature closure bias will prevent the clinician from gathering more information and integrating this information to formulate a new hypothesis (Dailey, 1952). The evidence indicates that early diagnostic recommendations can help clinicians, especially primary care clinicians (including dentists) in improving diagnostic accuracy (Kostopoulou *et al.*, 2015b) especially in

cases where the cases have an atypical clinical presentation and there is a higher likelihood of producing a diagnostic error (Kostopoulou, Delaney and Munro, 2008b). The problem of diagnostic error gets compounded because of the increasing pressures on the primary care clinician to meet patient targets and can therefore lead to medical errors (Barraclough, 2013). Assessing the performance of the DS Model at different sections of the information model can help us identify the areas where the CDSS underperforms and will help us expand the information model to include data from other sources. The DS Model is only as good as the data entered/obtained. By comparing the performance of the CDSS using the DS Model with dentists we can identify deficient areas. It will help us identify potential types of external sources of data that can be linked to the DS Model to support diagnosis.

6.3 Method

6.3.1 Clinical Domain Selection

Dental caries is a major public health problem affecting 60-90% of school children in most industrialised nations (Iheozor-Ejiofor *et al.*, 2015). If left untreated the dental caries can progress to inflammation or irritation of the dental pulp, which can then result in a dental abscess requiring Root Canal Treatment (RCT) or extraction of the infected tooth. One of the most common symptoms of deep dental caries and dental inflammation/infections is pain in the orofacial region. Dental pulp is the neuro-vascular tissue that resides within the tooth and occupies the central cavity of the tooth. The pain that the patient experiences when there is trauma or dental caries is an indication of the nervous tissue in the pulp being irritated.

There are several different diagnoses reflecting the state of the dental pulp. In some cases, the dental pulp will be vital and can be saved. In other cases, the dental infection has spread too deep into the tooth to prevent the rescue the dental pulp. In some cases, the patient does not experience any pain. There are other non-odontogenic (non-dental) causes of pain that can simulate a dental pain (Yatani *et al.*, 2014). Diagnostic error can result in many patients receiving dental treatment when in fact the cause for the dental pain was non-odontogenic. Acute and chronic dental/orofacial pain can have different clinical manifestations. It is one of the most common complaints that patients visiting the dental clinic have. As many as 1 in 6 patients have reported orofacial pain in the past year due to odontogenic and non-odontogenic causes (Horst *et al.*, 2015).

Dental pain is one of the most common diagnostic challenges that the average dental practitioner encounters in his/her clinic. One of the main problems with dental pain is the difficulty of diagnosing the cause of dental pain. The main reason for this is the subjective nature of the symptoms experienced by the patient and the diagnostic tools available to the clinician. The clinical signs and symptoms do not correlate with the pathological findings of the tooth. Once diagnosed, the management is relatively straightforward for the clinician. Therefore, there is a need for tools that can assist the clinician in identifying the most likely diagnoses the patient may be having.

Endodontics and dental pain has been chosen as the test domain within dentistry due to various factors (Bender, 2000) .

- Indirect signs and symptoms: Dental infections of the dental pulp and the surrounding tissues take place in a confined and concealed tissue that cannot be visually inspected in most cases. The clinician must rely on indirect information to assess the patient and arrive at a diagnosis. This increases the risk of the clinician making a false-positive or a false-negative diagnosis. Diagnosis is currently based on the presence and nature of pain; and the presence and type of exposure of the dental pulp.
- Lack of the discriminatory ability of patient's description or response to pain: There is no direct relationship between the amount of tissue damage and the reported pain.

- Multiple sources of diagnostic information: Nature of complaint, dental history, clinical examination and radiographic examination.
- Diagnostic confusion due to the nature of dental pain and confounding factors:
- Pulpal inflammation can cause intense pain; but at the same time teeth can develop total pulpal necrosis (death of pulp) without any symptoms.
- The type of pain can vary from sharp or shooting pain that is well localized; to dull and diffuse pain.
- Pain is not a reliable measure to evaluate the condition of the pulp. There is a wide variation in clinical presentation of pain in pulpal inflammation.
- Well understood pathophysiology: The underlying cause of dental caries and subsequent effects on the human body are well understood and documented.

Dental pain forms an ideal test domain for evaluation as the author has a background in dentistry. The main purpose of the dental version of the DS Model is to assist dentists in diagnosing dental pain.

6.3.2 Knowledge Acquisition

The main source of knowledge that was used to develop the dental pain DS Model was Cohen's Pathways of the Pulp (Hargreaves, Cohen and Berman, 2011). Other knowledge sources used for developing the diagnostic decision model and the clinical vignettes include:

- Endodontics: Part 2 Diagnosis and treatment planning (Carrotte, 2004).
- AAE Consensus Conference Recommended Diagnostic Terminology (AAE, 2009; Glickman, 2009)
- An Investigation Into Differential Diagnosis of Pulp and Periapical Pain: A Penn Endo Database Study (Iqbal, Kim and Yoon, 2007).
- Identify and Define All Diagnostic Terms for Pulpal Health and Disease States (Levin *et al.*, 2009).
- Electric pulp testing: a review (Lin and Chandler, 2008)
- Assessment of pulp vitality: a review (Gopikrishna, Pradeep and Venkateshbabu, 2009)
- Dental Diagnostic System (Kalenderian *et al.*, 2011a)

The knowledge sources described above was used to develop the dental version of the DS Model and the vignettes used in the study. The DS Model can be used as a basis for developing a diagnostic model in other domains such as Ophthalmology if needed.

6.3.3 Clinical vignette

Ideally in evaluation of CDSS models, researchers should use real-life patient data. However, because of strict regulations regarding the use of patient data it was decided against using actual patient data and rely on clinical vignettes for model evaluation. Clinical vignettes are a valid tool for evaluating how clinicians would perform under simulated conditions (Peabody *et al.*, 2004; Mettes *et al.*, 2010; Shah, Edgar and Evans, 2010; Kostopoulou *et al.*, 2015b). Since this study is comparing the performance of the DS Model against dentists it was decided to create clinical vignettes on a wide range of dental pain diagnoses and use them for evaluation.

The information model and its sections presented in the section above (section - 6.2) was used as the basis for developing the clinical vignettes. A sample vignette used for the evaluation has been presented below (Table 6-1). The vignettes were presented to the dentists in a tabular format, with a summary of the data on the top. The main sections of the information model have several sub-sections. Each sub-section

has several slots. Each slot has a value to denote that vignette. In a different vignette the sections and the slots will remain the same, however the values will change depending on the vignette. The vignettes were developed with the help of an Endodontist. The vignettes were developed first by the author and then independently validated by the Endodontist. The Endodontist has nearly 8 years' experience in primary care dentistry and has specialist training in the field of Endodontics.

Patient summary	Patient complains of pain in relation to tooth in lower right jaw. Patient had pain for the past week or so. Pain is a sharp and intense pain that lasts for few seconds. Pain is initiated by drinking cold water or when eating ice-cream. Pain subsides when stimulus is removed. Patient has history of diabetes. On examination, caries has been found in relation to #46 (Lower Right First Molar) Caries extends to inner dentine of the tooth. Pulp tests were conducted on the suspect tooth and neighbouring teeth. The tooth responds to cold stimulus with moderate pain. Pain subsides on removal of stimulus. Tooth responds to mild electric stimulus. Radiographic examination showed radiolucency in crown of #46 that extends to inner dentine. No radiographic changes observed periapically.

Chief Complaint		
	SLOT	**VALUE**
Patient Information		
	Gender	Male
	Birth Year	1975
	Occupation	Journalist
Medical History		
	General health status	No relevant history
	Current medical history	Diabetes Mellitus Type II
	Current medications	Metformin: 500 mg orally twice a day
	Allergies	None
Family History		
	Diabetes	Yes, mother has Diabetes Mellitus Type II
	Hypertension	Yes, father has Hypertension
	Cancer	None
	Infectious Disease	None
Patient Complaint	"Pain in relation to right lower jaw when eating or drinking cold food / drink"	
	Pain Present	Yes
	Location of pain	In relation to lower right area of jaw
	Pain initiated by	Cold food and drink
	Pain relieved by	Removal of stimulus
	Quality of pain	Sharp and intense pain
	Location of radiation of pain	N/A
	Intensity of Pain (On a scale of 0-10)	4 (Moderate Pain)
	When did you first notice symptoms?	When eating ice-cream
	History of pain	Pain has been present intermittently for the past week
	Locate tooth	Yes, patient can locate tooth
	Onset of pain	Stimulation required for onset
	Progression of pain	Pain present and pain character has not changed over time
	Does the pain keep you awake at night?	No
	Duration of pain	Pain lasts for seconds
Dental history		
	History of clenching teeth?	No
	History of tooth trauma?	No

	Do you wear Night guard?	No
	Previous RCT?	No
	Recent restoration?	No
Extra Oral Examination		
Lymph node exam		
	Lymph node firm	No
	Lymph node swelling	No
	Lymph node tenderness	No
Extra Oral swelling exam		
	Extra Oral swelling present	No
	Fluctuance of swelling	N/A
	Location of extra oral swelling	N/A
	Spread of extra oral swelling	N/A
	Type of extra oral swelling	N/A
	Extra Oral Sinus	No
Extra Oral tests		
	Pain present during functional evaluation of muscles?	No
	Pain present during palpation of muscles?	No
	Pain present during palpation over sinus?	No
	TMJ Exam	Within Normal Limits
Intra Oral Examination		
Soft Tissue Examination		
	Oral cancer exam	Within Normal Limits
	Hard palate exam	Within Normal Limits
	Soft palate exam	Within Normal Limits
	Floor of the mouth	Within Normal Limits
	Lips	Aphthous ulcer present on lower lip
	Pharynx and fauces	Within Normal Limits
	Tongue	Within Normal Limits
Hard tissue examination		
	Oral Hygiene	Fair
	Missing teeth	None
	Caries present?	Present
	Type of caries present	Primary active extensive dentin caries inner pulpal one-third of dentin
	Location of caries (in relation to)	Tooth Number: #46
Intra Oral swelling examination		
	Intra Oral swelling present	No
	Fluctuance of Intra Oral swelling	N/A
	Location of Intra Oral swelling	N/A
	Spread of Intra Oral swelling	Spread of Intra Oral swelling
	Intra Oral Sinus	No
Previous Restoration Examination		
	Existing restoration	Yes, Tooth number: #27
	Existing restoration type	Amalgam restoration
	Defective restoration	No
	Defective restoration type	N/A
Periodontal pockets examination		
	Periodontal pockets present	No
	Periodontal pockets depth	No
	Pseudo pockets present	No
	Pseudo pockets depth	No
	Furcation involvement?	No
	Tooth with furcation	No
	Degree of furcation involvement?	No
	Tooth Mobility?	No
	Mobile tooth	No

	Mobility grade	No
Clinical Tests		
Palpation test		
Apical palpation	Tooth tested	Tooth Number: #46
	Apical palpation response	Within Normal Limits
	Apical palpation response severity	N/A
Percussion test		
Axial percussion	Tooth tested	Tooth Number: #46
	Axial percussion response	Within Normal Limits
	Axial percussion response severity	N/A
Cold pulp test- (Application of ice to tooth of interest)		
	Tooth tested	Tooth Number: #46
	Nature of cold pulp test response	Positive response
	Positive cold pulp test severity	Moderate
	Positive cold pulp test response type	Pain subsides on removal of stimulus
Hot pulp test- (Application of heated Gutta Percha to tooth of interest)		
	Tooth tested	Tooth Number: #46
	Nature of hot pulp test response	Positive response
	Positive hot pulp test severity	Mild
	Positive hot pulp test response type	Within Normal Limits
Electric pulp test		
	Tooth tested	Tooth Number: #46
	Nature of electric pulp test response	Within Normal Limits
Bite test (using bite block)		
	Tooth tested	Tooth Number: #46
	Bite test result	Within Normal Limits
	Positive bite test result response	N/A
	Positive bite test result severity	N/A
Trans-illumination test		
	Trans-illumination test result	N/A
Test Cavity		
	Test Cavity result	N/A
Local anaesthetic test/Selective anaesthesia		
	Local anaesthetic test result	N/A
Radiographic Examination		
Bone level radiographic examination		
	Bone loss	Absent
	Bone level condition	Normal
Tooth radiographic examination		
	Radiolucencies	Present
	Radiolucencies type	Radiolucency reaching the inner one-third of dentin, clinically cavitated
	Radiolucency location (in relation to)	Tooth number: #46
	Radiopacities	Absent
Tooth Restoration radiographic examination		
	Defective margins in radiographs	N/A

Lamina dura Radiographic Examination		
	Condition of Lamina dura	Normal
Periodontal Tissue Radiographic Examination		
	Periapical radiopacities	Absent
	Periapical radiopacities tooth	N/A
	Periapical radiolucency	Absent
	Periapical radiolucency tooth	N/A
	Periradicular radiolucency	Absent
	Tooth resorption	Absent
	Widened periodontal ligament space	Absent

Table 6-1 - Sample patient vignette

Each vignette has 2 diagnoses- Pulpal and Apical diagnoses. Table 6-2 shows a breakdown of the type of clinical vignettes that were created with the help of the Endodontist and the pulpal and apical diagnoses of each vignette. All the vignettes can be accessed in tabular format in Appendix 11.

Vignette ID	Pulpal diagnosis	Apical diagnosis
1	Normal Pulpal Tissues	Normal Apical Tissues
2	Reversible Pulpitis	Normal Apical Tissues
3	Symptomatic Irreversible Pulpitis	Normal Apical Tissues
4	Asymptomatic Irreversible Pulpitis	Normal Apical Tissues
5	Pulp Necrosis	Asymptomatic Apical Periodontitis
6	Pain of Nonodontogenic Origin	Pain of Nonodontogenic Origin
7	Symptomatic Irreversible Pulpitis	Symptomatic Apical Periodontitis
8	Pulp Necrosis	Asymptomatic Apical Periodontitis
9	Pulp Necrosis	Acute Apical Abscess
10	Pulp Necrosis	Chronic Apical Abscess
11	Pulp Necrosis	Symptomatic Apical Periodontitis
12	Pulp Necrosis	Chronic Apical Abscess
13	Dentinal Hypersensitivity	Normal Apical Tissues
14	Symptomatic Irreversible Pulpitis	Normal Apical Tissues
15	Asymptomatic Irreversible Pulpitis	Normal Apical Tissues
16	Pulp Necrosis	Asymptomatic Apical Periodontitis
17	Pain of Nonodontogenic Origin	Pain of Nonodontogenic Origin
18	Symptomatic Irreversible Pulpitis	Symptomatic Apical Periodontitis
19	Pulp Necrosis	Asymptomatic Apical Periodontitis
20	Pulp Necrosis	Acute Apical Abscess
21	Pulp Necrosis	Chronic Apical Abscess
22	Asymptomatic Irreversible Pulpitis	Normal Apical Tissues
23	Symptomatic Irreversible Pulpitis	Normal Apical Tissues
24	Reversible Pulpitis	Normal Apical Tissues
25	Normal Pulpal Tissues	Normal Apical Tissues

Table 6-2 - Clinical vignettes diagnoses

The terminology used in Pulpal and Apical diagnosis has been derived from the AAE (American Academy of Endodontics) Consensus Conference Recommended Diagnostic Terminology (AAE, 2009). The diagnoses indicate the clinical state of the pulp. For example, Reversible Pulpitis indicate a reversible nature a pulpal inflammation (pulpitis). "itis" is a commonly used suffix in medicine to indicate an inflammation of a tissue (Wiktionary, 2016). The inflammation can be caused by several different causes including trauma, facture of the tooth and dental caries. However, in this study it was decided to create clinical vignettes that can be described without the use of images or photographs. The vignettes primarily included cases where the pathology has been caused by dental caries which can be described using existing terminology. The correct diagnoses for vignettes were determined by the Endodontist after creating the vignettes. The Pulpal and Apical diagnoses provided by the Endodontist comprised of the most likely diagnoses that the vignette would have. The Endodontist provided a next most likely diagnosis for each vignette. The most likely diagnosis was called the Rank 1 diagnosis, and the next most likely diagnosis was called Rank 2 diagnosis for this study.

6.3.4 Participants

A group of dentists (n=20) were recruited for the study. This study was undertaken by practising dentists who are either independent, part of a regional or national chain, or practising in an educational institution within India. The dentists were recruited via email. An email containing information about the trial and link to the trial web application were sent directly to the dentists (See Appendix 10 – section 10.1). The dentists were offered the equivalent of £10/- as compensation for completing 5 clinical vignettes (a set). The total duration taken for completing 5 vignettes was approximately 30 minutes.

6.3.4.1 Dentists background

Most the dentists recruited for the study fell in the 25-34 age groups (Figure 6-1). This meant that most the dentists who participated in the study had similar years of experience (Figure 6-2). Most of the participants had 6-8 years of clinical experience at the time of the study. This group of primary care dentists would therefore have enough experience to deal with the Endodontic cases presented in the vignettes. Only specialist Endodontists would have more knowledge dealing with these vignettes. This is good group to compare the performance of the DS Model as the dentists would have some experience managing cases like these in routine practices. Novice clinicians may not have the opportunity to manage different types of cases and therefore may not accurately diagnose the patients. Using novice dentists could therefore bias the results of the evaluation as it could artificially improve the performance of the CDSS as the DS Model was compared to the dentists. Half of the participants were Male (Figure 6-3).

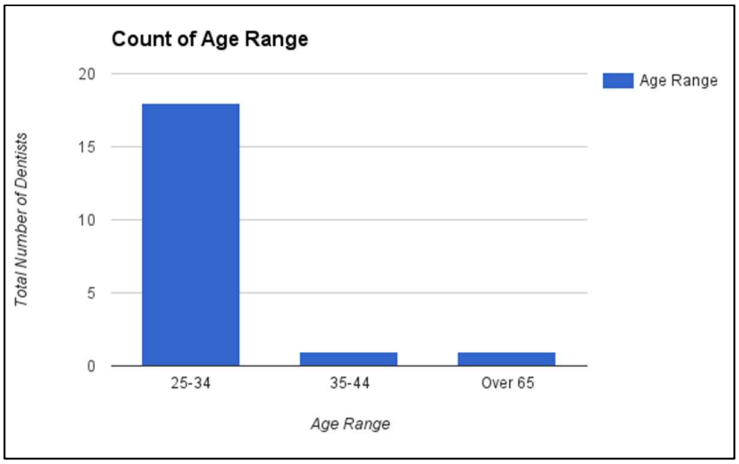


Figure 6-1 - Age range

Figure 6-2 - Year of qualification

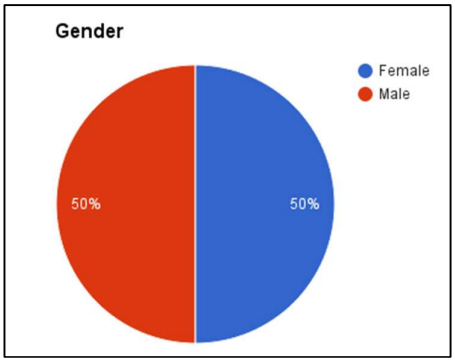


Figure 6-3 - Gender

More than half (60%) of the dentists recruited for the study had no post graduate qualifications (Figure 6-4). Dentists who had post-graduate qualifications such as 'Master of Dental Surgery' in subjects other than Endodontics were recruited for the study.

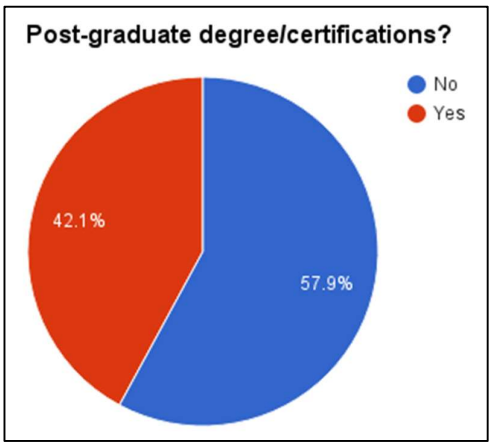


Figure 6-4 - Post-graduate qualifications

The dentists in the study came from a wide range of clinical practice backgrounds including independent practice, group practice etc. (Figure 6-5). Almost all dentists will come across cases of dental pain at some stage of their careers as dental caries is highly prevalent in India, especially among school children, and a significant proportion of patients (35%) experiencing different degrees of dental pain (Slade, 2001; Kumar, Acharya and Pentapati, 2014).

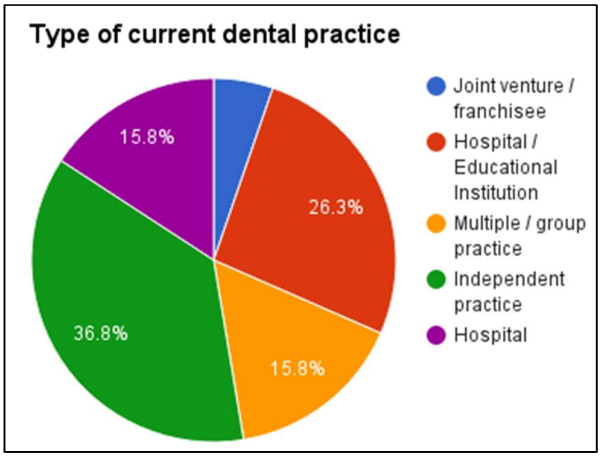


Figure 6-5 - Current dental practice

6.3.5 Procedure

6.3.5.1 Data collection from dentists

The study was approved by City University London Department of Computer Science Research Ethics Committee (See Appendix 10 – section 10.2). Participants were given the information about the study via email. The information contained details of the study purpose, duration, right to withdrawal from study, dentist data confidentiality, information about compensation, ethical committee clearances and study team contact details. If the participants had any questions they were encouraged to contact the study team.

This study is a controlled experiment in which each dentist was presented with 5 vignettes. The total duration taken to complete all 5 vignettes (including instructions) did not exceed 30 minutes. The vignettes were randomly assigned to the clinicians. The order in which the vignettes were presented was randomised to eliminate order effects. Some dentists agreed to look at additional sets of vignettes. Care was taken to provide these dentists with a different set of vignettes that were not included in the first round. Therefore, data equivalent of 24 attempts at a set (each containing 5) of vignettes or 120 individual attempts at the vignettes in total was collected.

Data collection in this study took place via the web. The dentists were sent an email with a link to an online survey form (Google Forms, 2016). Once the participants click on the link it will take them to the study information. Once they have read the information and agreed to its contents it will take them to the online consent form. Only when the participants have agreed to the terms and conditions specified in the consent form will they be taken to the data collection form.

In the data collection form, the dentists were provided with a link to the vignette that was presented in the tabular format shown in Table 6-1. The vignette was presented in an online spreadsheet software (Google Sheets, 2016) that contained a brief summary of the vignette. The data collected in the survey form is recorded in a different online spreadsheet of the same spreadsheet software (Google Sheets).

The dentists were provided with a set of signs, symptoms, past clinical history, family history and other data needed to make a diagnosis, including diagnostic test results and radiographic reports of vignettes. In addition, a summary of relevant findings was provided (Table 6-1). Below the link to the vignette and the summary of the vignette, the dentists were asked to select the most likely diagnosis from a drop-down menu (Rank 1) and provide a second most likely diagnosis (Rank 2) for the vignette. There will be a Rank 1 - Rank 2 Pulpal and Rank 1 - Rank 2 Apical diagnosis for each vignette (Figure 6-6).

Figure 6-6- Survey form - Pulpal and Apical diagnosis

The concept of a Rank 1 and Rank 2 diagnosis was explained to the dentist. If the dentist felt that the Rank 1 diagnosis for the vignette was Normal, then he/she can select "None of the above" from the options given for the Rank 2 diagnosis.

After Rank 1 and Rank 2 diagnoses were selected the dentists were asked to provide their opinion about the vignettes. The opinion survey form was below the form provided to obtain the Rank 1 and Rank 2 diagnoses. The dentists were asked to rank the difficulty of the vignette, their familiarity in diagnosing the vignette and how often the vignette appeared during their routine practice (Figure 6-7). The term "case" was used instead of vignette on the survey form because clinicians are more familiar with that term. The opinion about each vignette was obtained from the dentists (See section - 6.4).

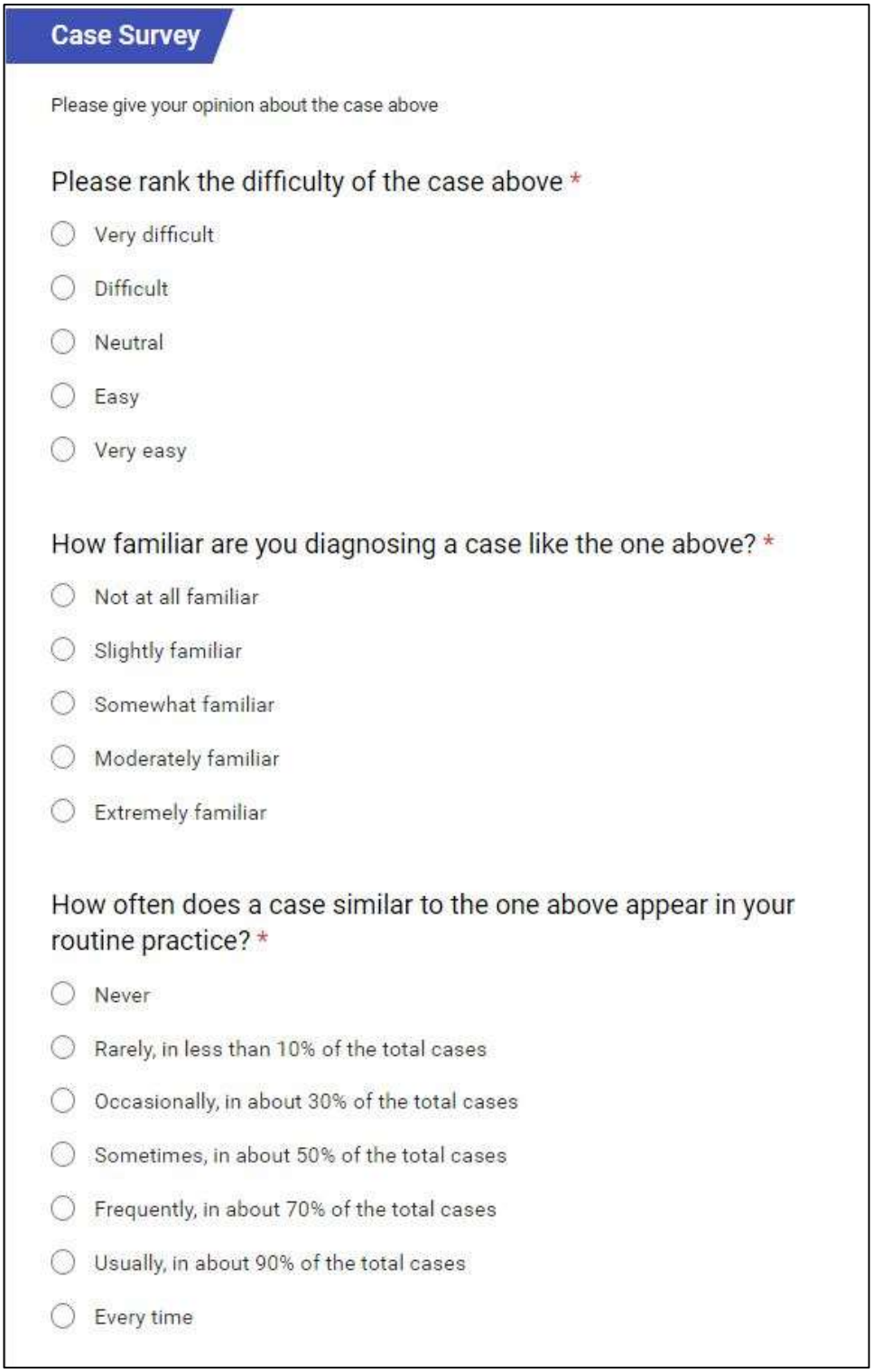


Figure 6-7 - Vignette survey form for dentists

Management recommendations for each vignette was collected from the dentists. However, since developing a CDSS based on the DS Model that could provide management recommendations was beyond the scope of this thesis the performance of the dentists could not be compared to the CDSS. The comparison was restricted to the diagnostic recommendations provided by the dentists.

After the 5 vignettes are completed the dentists were asked to provide demographic information (Age & Gender), education qualifications and clinical experience. The participating dentists did not have access to

the CDSS prototype or any additional decision support recommendations including guidelines while recording their clinical decision (See Appendix 10 – section 10.1 for a sample form sent to dentists)

After the web-based experiment was completed, the decisions made by each dentist for each vignette will be compared to the results of the evaluation of the CDSS using the dental DS Model and the ranking provided by the Endodontist. The Endodontist's Rank 1 and Rank 2 diagnosis recommendations were set as the gold-standard against which the comparison was made.

Only Rank 1 and Rank 2 were considered for the dentists as the dentists had difficulty in accurately providing a ranking beyond Rank 3. During pilot evaluation, the dentists could give a list of diagnoses that they thought was most probable for the vignettes, however when asked to rank the vignettes using their professional clinical judgement most of them found it difficult to identify diagnoses beyond Rank 3. The Endodontist mentioned the difficulty of obtaining a ranking beyond Rank. Therefore, it was decided to only record the Rank 1 and Rank 2 for the dentists.

6.3.5.2 CDSS evaluation data collection

The CDSS prototypes development was covered in the previous chapter (Chapter 5 – Case studies 1,2,3,4). The prototype developed in Case study 3 was used for evaluation.

Case Study 3 showed how the data in an EHR can be represented in a standardised manner using archetypes. The DIM was represented as archetypes and separated into different sections. The different sections represent the major steps in the diagnostic encounter. **APGD was used to generate patient data based on the archetype model in RDF data format. The APDG tool allows us to model different clinical vignettes in RDF format. Any of the case study prototypes developed in Chapter 5 can be used for evaluating the performance of the DS Model. However, the ease with which the different clinical vignettes can be created using DIM as the basis was the reason for selecting Case Study 3 as the basis for evaluating the dental pain DS Model.** Case Study 3 similarly demonstrates how patient data can be represented in different formats and using SPARQL rules the patient data can be mapped to inform/update the DS Model. To generate early diagnostic recommendations based on linked data on the web the CDSS must be able to use patient data in different formats and stored in disparate locations.

Once the data was collected and collated the CDSS (using the dental pain DS Model) was given the same 25 clinical vignettes in RDF format and the ranked diagnostic recommendations were recorded.

There are 2 different types of diagnostic recommendations generated by the CDSS in this study: Early and Late diagnosis. The diagnosis obtained after CC, EO and IO data was collected was considered as the early diagnosis. In a realistic clinical situation where the proposed CDSS using the DS Model provides early and late diagnostic recommendations the process of providing recommendations moves in stages following the sections of the information model. Once CC data is collected a diagnosis recommendation is provided by the CDSS (in the form of a list). Then the data from EO is combined (CC+EO) to provide another diagnosis recommendation. Then the data from IO is combined to generate another recommendation. Once CT and RE data are collected the final diagnosis can be obtained. Therefore, the evaluation of the CDSS will proceed in the following manner going through each section:

CC | CC+EO | CC+EO+IO | CC+EO+IO+CT | CC+EO+IO+CT+RE (Also called final diagnosis).

To simulate a clinical environment where the clinician captures data from the patient in an EHR database the CDSS evaluation study was conducted in the following manner.

1. **Generate patient data based on vignettes**: The APDG data generator tool (Huang *et al.*, 2013) was used to create patient RDF data based on the archetype model (see section - 5.4.3) used by the tool. The vignettes created with the help of the Endodontist were used as the basis to model the patient data. In total 25 RDF files was created using the APDG data generator with each file corresponding to a vignette. The table below (Table 6-3) shows an extract from a vignette expressed in tabular format. In this example, there is an archetype called Chief Complaint that corresponds to the CC section of the information model (Table 6-1) used to create the patient vignettes. Within the archetype there is a slot that can take 3 different values. If a vignette was created that had the value "Stimulation required for onset" for the slot "Onset of pain" then this information was mapped to an XML file (Figure 6-8) provided by the APDG tool. The probability for the slot to take a value can be set in the XML file. APDG can generate patient data based on a distribution range between 0 and 100. In this case, 100 was used since the data was based on an existing vignette and there was no need for variation. Using this process, XML files representing 25 vignettes was created. The APDG tool was run indicating the XML file to be used to generate the RDF data. For example, if the vignette ID was 10032, then a XML file corresponding to the vignette was created for that vignette and a RDF file called 10032.nt was created by APDG. NT is an abbreviation for N-Triples, a common serialization format for RDF alongside TURTLE and RDF/XML. (Also, see Appendix 9 for more information about APDG, XML files and RDF representation of vignettes).

Archetype	Slot	Value
Chief Complaint		
	Onset of pain	Gradual onset
		Stimulation required for onset
		Sudden onset
Intra Oral examination		

Table 6-3 - Extract from vignette

```
<Archetype concept="Chief Complaint">
      <Slot value="Onset of pain" type="enumeration">
            <DataRange>
              <enumeration value="Gradual onset">
              </enumeration>
              <enumeration value="Stimulation required for onset">
              </enumeration>
              <enumeration value="Sudden onset">
              </enumeration>
              <Distributions type="enumeration">
                        <Distribution item="Gradual onset" pfrom="0" pto="0"/>
                        <Distribution item="Stimulation required for onset" pfrom="0" pto="100"/>
                        <Distribution item="Sudden onset" pfrom="0" pto="0"/>
                </Distributions>
            </DataRange>
      </Slot>

</Archetype>
```

Figure 6-8 - XML extract based on vignette

2. **Data cleared from triple store**: As mentioned previously the architecture described in Chapter 5 – Case Study 3 (section - 5.4.3) was used for validating the CDSS using the dental version of the DS Model (Figure 6-9).

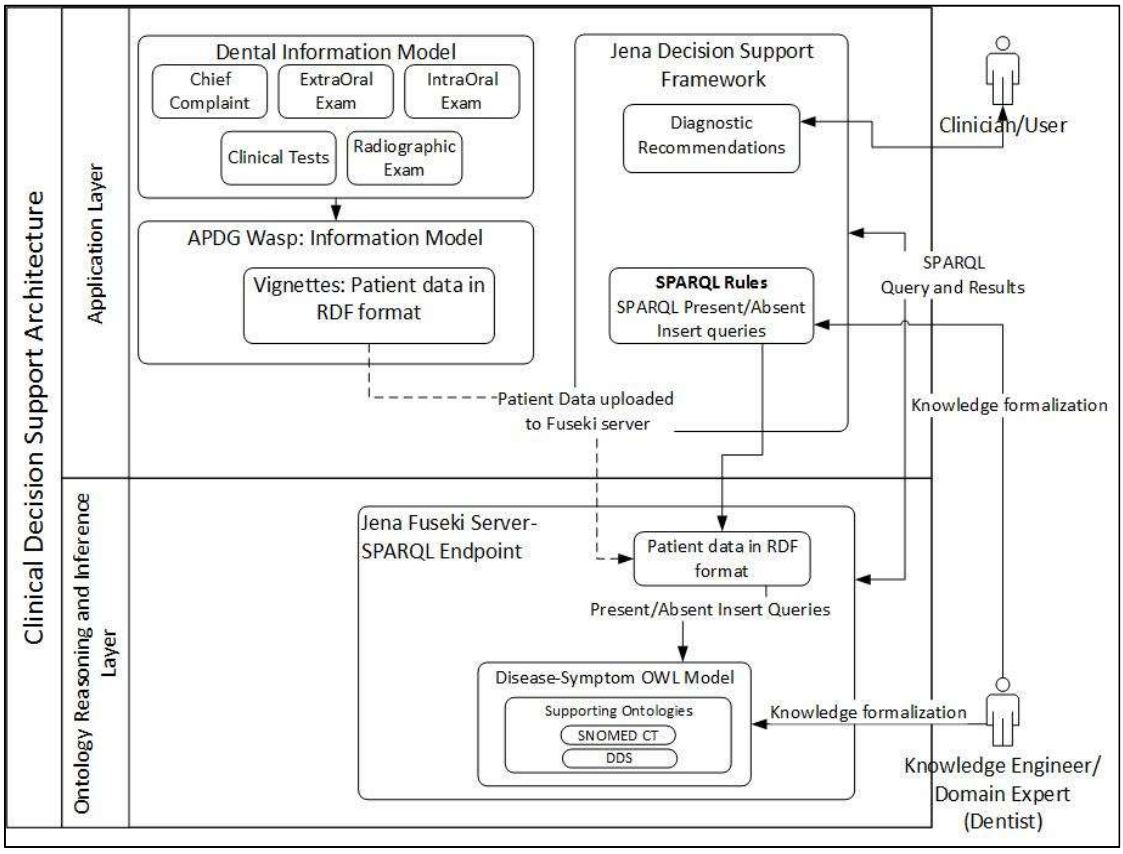


Figure 6-9 - Chapter 5 Case Study 3 architecture of CDSS using patient data in RDF format

For evaluation purposes, the Jena triple store accessed via the Jena Fuseki server was cleared of any existing patient data. This is done with the help of a SPARQL DELETE query.

3. **Data is added**: A patient data file (based on a vignette) in the form of RDF is added to the triple store's default empty graph model. All triple stores have an empty default graph into which the triples are inserted. If needed the triples can be inserted into a named graph. The DS Model OWL ontology file is added to the same graph.

4. **SPARQL present/absent insert query completed**: The SPARQL present/absent insert query for the data that comprises the CC section (archetype) of the information is executed. For example, the SPARQL insert query for the slot "Onset of pain" queries the corresponding value ("Stimulation required for onset") and inserts the following triple -> rp:Stimulation_required_for_onset *rp:hasObservationStatus* rp:9_ObservationPresent into the DS Model (Figure 6-10). This process is carried out in a loop until all the slots in the CC section has been queried.

```
INSERT  {?Observation1 rp:hasObservationStatus rp:9_ObservationPresent}
WHERE {
        SELECT ?value ?Observation1
        {?Observation1 a rp:Stimulation_required_for_onset.
        ?s wasp:hasSlot ?o.
        ?o rdfs:label ?slot.
        ?o wasp:value ?value.
                FILTER (str(?slot) = "Onset_of_pain")
                FILTER (str(?value) = "Stimulation required for onset")}
}
```

Figure 6-10 - SPARQL present/absent conditional insert query (Onset of pain)

5. **SPARQL insert query calculating weights**: The SPARQL insert query to calculate weights for all the observations that have *rp:hasObservationStatus* rp:9_ObservationPresent is completed. Now the different diagnoses in the dental version of the DS Model will have a corresponding weight attached.

6. **SPARQL query to obtain ranked list**: SPARQL SELECT query to obtain the ranked pulpal and apical diagnoses is completed. A ranked list of pulpal and apical diagnoses with the most likely diagnosis ranked on top is obtained.

Once all the steps (steps 2-6) mentioned above is completed a ranked diagnosis list is obtained for the CC section of that RDF patient data (based on the vignette that was given to dentists). This process simulates a real-life encounter where the dentist records the data in an EHR and the CDSS looks up the data to give a diagnosis at the CC section. The next step involves the repeating the entire process (steps 2-6 above) but with the addition of slots (data) from the EO section as well. This process completes until all the sections have been completed and a ranked list for each section (with data from the previous sections) is obtained:

CC

CC+EO

CC+EO+IO

CC+EO+IO+CT

CC+EO+IO+CT+RE (Final)

After completion of each section, the ranked diagnosis list is recorded in a spreadsheet for data analysis. The ranked diagnosis list obtained at the end of each section was compared to the dentists ranking (Rank 1 and Rank 2). This process is repeated for the next RDF patient data file (based on a vignette) and the sections are traversed in a step wise manner while recording the diagnostic ranking after each section is completed.

This section provides an overview of the methods used in this study. The next section provides more information about the vignette survey results.

6.4 Vignette survey results

All the dentists completed the survey form provided alongside the vignette to obtain their opinion about the vignette. The first question asked was "Please rank the difficulty of the case above". The results have been summarised below in (Table 6-4) and (Figure 6-11).

Question	Total count (N=120)
Very easy	1
Easy	22
Neutral	78
Difficult	19
Very Difficult	0

Table 6-4 - Survey question 1 - total count

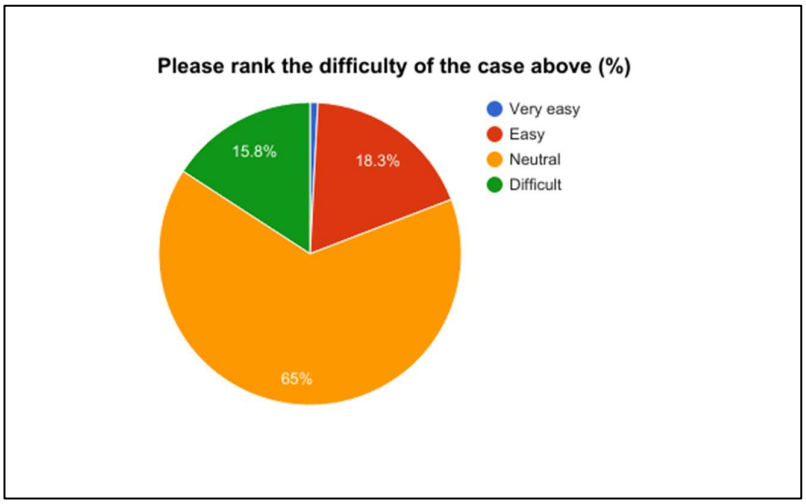


Figure 6-11 - Survey question 1 - percentage

The results show that most the dentists (N=78) felt the vignettes were Neutral as far as the difficulty was concerned. None of the dentists (N=0) felt the vignettes were "Very Difficult". So overall, the vignettes presented in this study were of the type that the dentists did not find too difficult to diagnose.

The second question asked was "How familiar are you diagnosing a case like the one above?". The results have been summarised below in (Table 6-5) and (Figure 6-12).

Question	Total count (N=120)
Extremely familiar	21
Moderately familiar	41
Somewhat familiar	34

Slightly familiar	23
Not at all familiar	1

Table 6-5 - Survey question 2 - total count

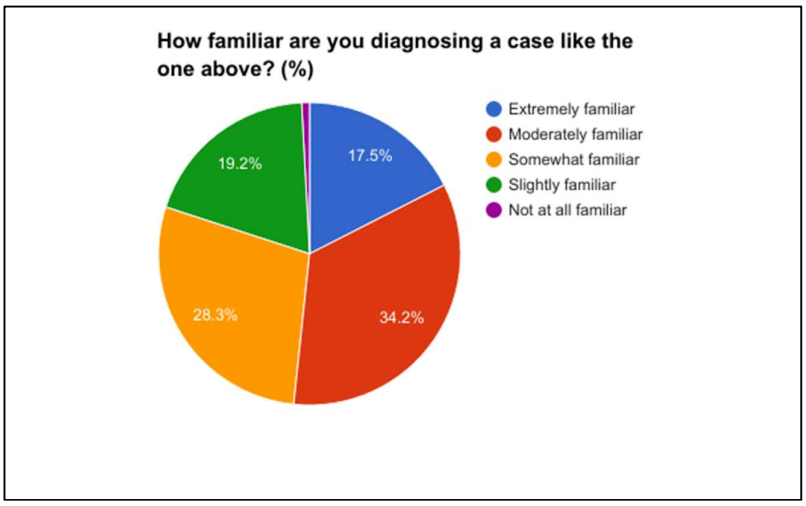


Figure 6-12 - Survey question 2 - percentage

The results show that overall the dentists were familiar with the vignettes and most the dentists were either Moderately (**N=41**) or Somewhat (**N=34**) familiar with the vignettes. So, the vignettes presented in this study mostly belonged to the type of cases that the dentists were familiar with and have seen before.

The third question asked was "How often does a case similar to the one above appears in your routine practice?". The results have been summarised below in (Table 6-6) and (Figure 6-13).

Question	Total count (N=120)
Usually, in about 90% of the total cases	4
Frequently, in about 70% of the total cases	26
Sometimes, in about 50% of the total cases	28
Occasionally, in about 30% of the total cases	43
Rarely, in less than 10% of the total cases	18
Never	1

Table 6-6 - Survey question 3 - total count

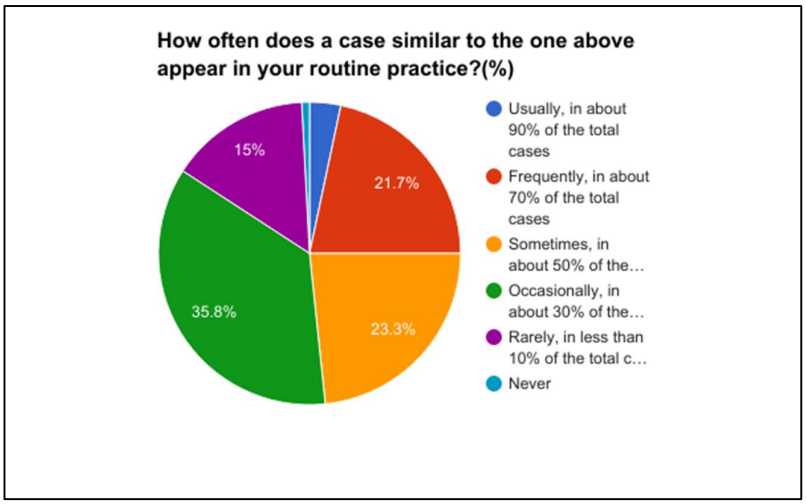


Figure 6-13 - Survey question 3 - percentage

The results show that a clinical case like the vignette have appeared in the dentist's routine clinical practice either Frequently (**N=26**), Sometimes (**N=28**) or Occasionally (**N=43**). Therefore, the vignettes are like the cases that the dentists diagnose and manage routinely and were not too rare in nature. Overall

the vignettes presented in this study to the dentists were of the type that the dentists did not find too difficult to diagnose, they are familiar with and routinely diagnose in their dental practice.

This section provides an overview of the vignette survey results. The next section provides an overview of the evaluation measures used in the study.

6.5 Evaluation Measures

The data computed from the CDSS was compared to the results collected from the dentists. The quality of the diagnostic ranking produced by the CDSS in each section of the diagnostic encounter when compared to the final diagnosis provided by the dentists was analysed. The quality of the ranking was measured using Normalized Discounted Cumulative Gain (NDCG) measure (See section - 6.6). The ability of the CDSS to get the Rank 1 and Rank 2 diagnosis right will be measured using Precision-Recall metrics (See section - 6.8). The precision-recall at all sections of the dental encounter represented in the information model was measured. These precision-recall metrics of the CDSS was compared to the dentists.

6.6 Normalized Discounted Cumulative Gain

6.6.1 Introduction to Normalized Discounted Cumulative Gain

The Normalized Discounted Cumulative Gain (Järvelin and Kekäläinen, 2002) metric was used to measure the ranking effectiveness of the DS Model when compared to the study participants and the Endodontist. It is a standard technique used in information retrieval algorithms in cases where the results are non-binary. It has been previously used to evaluate the performance of CDSS applications (Chen *et al.*, 2016). The CDSS using the DS Model provides a ranked list of 7 pulpal diagnoses and 6 apical diagnoses.

Each diagnosis is assigned a relevance score (utility). Diagnoses at the top of the ranking are more useful to the clinician than the diagnoses at the bottom of the ranking. Failure to consider a diagnosis is a major cause of diagnostic error (Schiff *et al.*, 2009) and therefore providing a ranked list helps the clinician in considering and prioritizing a diagnosis and order further diagnostic tests if needed. The analysis was conducted using NDCG under the assumption that clinicians are more likely to ignore diagnoses that are ranked lower down the list. The data collected from the Endodontist shows which diagnosis should be ranked in the top 2 ranks of the diagnosis list and was used as the gold standard against which the performance of the dentists and the CDSS was compared.

Using NDCG we can measure how well the CDSS performs in ranking the diagnoses. Precision is binary classification measure used in information retrieval science to identify the fraction of retrieved results that are relevant. Recall is the fraction of the relevant results that are retrieved. These metrics (Precision-Recall) can be used to ascertain the performance of the CDSS in getting Rank 1 and Rank 2 diagnoses correct (See section - 6.8).

However, at the early diagnostic recommendation section (CC, IO and EO) the correct diagnosis for that vignette may be ranked lower down the order by the CDSS due to incomplete information. The diagnosis may still be relevant and if ranked high enough the clinician may consider the diagnostic hypothesis. For example, 'Pain of nonodontogenic origin' is relatively rare when compared to pain caused by dental caries. When there is a deep dental caries the dentist may misdiagnose the cause of the pain as the tooth affected by the dental caries and may ignore or not consider the observations that indicate that the pain could be of non-odontogenic origin (i.e. non-dental origin). The CDSS should ideally rank the "Pain of Nonodontogenic Origin" diagnosis" high enough for the dentist to consider this diagnosis as a possibility, especially at early sections of the diagnosis.

If the CDSS consistently underperforms at the early section, then we can arrive at the conclusion that the CDSS using only information obtained from the patient directly during a diagnostic encounter cannot be

trusted to provide reliable diagnostic recommendations at the early section. We can then try to adapt the information model accordingly or rely on additional data sources using linked data technology to improve the diagnostic recommendations at early and the final diagnosis sections (if needed).

NDCG takes the utility of diagnosis at each rank and penalizes it by a logarithmic reduction factor. As the diagnosis goes lower the ranking, the utility of the diagnosis reduces. NDCG is used to measure the usefulness of a ranking on a scale from 0 to 1. A NDCG score of 1 represents a perfect match to the ideal ranking of the diagnoses. The ideal Rank 1 and Rank 2 were obtained from the Endodontist.

For example, for each vignette presented to the Endodontist, the expert was asked to rank the relevance of the diagnosis for the case as shown below. Each diagnosis was given a given a relevance score ranging from 0-3, with 3 being most relevant, 0 being least relevant and 1 and 2 somewhere in between (Table 6-7).

j	Diagnosis	rel_i	$\log_i 2$	$\frac{rel_i}{\log_2 i}$
1	Diagnosis-A	3	0	N/A
2	Diagnosis-B	1	1	1

Table 6-7 - Ideal DCG

Discounted Cumulative Gain (DCG) for this ranking is:

$$DCG = rel_1 + \sum_{i=2}^{2} \frac{rel_i}{\log_2 i}$$

Equation 1 - Discounted Cumulative Gain

$= 3 + 1$

$= 4$

Since this ranking and relevance score was provided by the expert it is called the Ideal DCG (IDCG). Similarly, the DCG for all dentists for each vignette was calculated (Equation 1). The dentists were given the same vignette and asked to select the Rank 1 and Rank 2 diagnoses that were most relevant for the case. The decision support provides a ranked list that consist of more diagnoses (n=6/7). To compare the DCG values of the decision support and the clinicians the DCG values need to be normalized.

Let's assume Dentist 1 selected the right diagnosis as Rank 1 & 2 and got a DCG= 4.

The normalized DCG (NDCG) of Dentist 1 for this case would be:

$$NDCG = \frac{DCG}{IDCG}$$

$= \frac{4}{4}$

$= 1$

The ranked list for the CDSS would be like the table below (Table 6-8).

i	Diagnosis	rel_i	$\log_i 2$	$\frac{rel_i}{\log_2 i}$
1	Diagnosis-A	3	0	N/A
2	Diagnosis-C	0	1	1
3	Diagnosis-B	1	1.585	0.6309148265
4	Diagnosis-E	0	2	0
5	Diagnosis-D	0	2.322	0
6	Diagnosis-G	0	2.584	0
7	Diagnosis-F	0	2.807	0

Table 6-8 - DCG for CDSS example

The DCG for the decision support will be: 3.6309

The NDCG for the decision support will be DCG / IDCG = 0.9077

From the above example the NDCG for the decision support is close to 1. However, the decision support was penalized for getting the position of Diagnosis B wrong. Since Dentist 1 got the ranking right, he got a NDCG score of 1, which is signifies a perfect match to the expert's ranking.

This section introduces the NDCG evaluation measure. The next section provides the results of the NDCG analysis.

6.7 Results of NDCG analysis

The results of the NDCG analysis was done in sections:

CC | CC+EO | CC+EO+IO | CC+EO+IO+CT | CC+EO+IO+CT+RE

Within each section the Pulpal and Apical average of the CDSS using the dental DS Model is compared with the Pulpal and Apical average of the dentists. The NDCG score of dentists was calculated based on the data provided by the dentists.

The NDCG scores for the dentists were the final diagnosis scores. The final diagnosis is made after taking into consideration all the data that is available (See Appendix 14 for raw data).

6.7.1 Significance tests

6.7.1.1 Independent 2-sample t-test (assuming unequal variances)

The independent 2-sample t-test (assuming unequal variances) was used to test the statistical difference between the means of both groups (CDSS and Dentists). The t-test is used to compare the means of the two independent groups and test if the population means are significantly different from each other. The independent 2-sample t-test was performed using statistical package R (version 3.3.1) (R Core Team, 2013). The function t.test in R was used to perform the test. The t-test works best when the data is obtained from a distribution that is normal or close to normal.

6.7.1.2 Wilcoxon rank-sum

In addition, the Wilcoxon rank-sum (also called the Mann-Whitney U test) was performed. Wilcoxon rank-sum test is a nonparametric statistical test.

When the Shapiro-Wilk test for normality (Shapiro and Wilk, 1965) (See Appendix 12 – section 12.1) was conducted on both the dentists data and the CDSS data it was found that the data did not come from a normal distributed population. Wilcoxon rank-sum test is the non-parametric equivalent of t-test and the results of the Wilcoxon rank sum test will be more valid when interpreting the results as the results of the Shapiro-Wilk test show the data does not have a normal distribution. (See Appendix 15 for raw data)

The Wilcoxon rank-sum test does not assume that the data has a normal distribution whereas the t-test assumes that the data has to come from a normally or near normally distributed population (Haynes, 2013). The Wilcoxon rank sum test was performed using statistical package R (version 3.3.1). The code used to perform the Wilcoxon rank sum is the function wilcox.test in R.

6.7.2 Dentists

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.71	0.86

Table 6-9 - NDCG average for dentists

Each vignette was analysed individually. The NDCG score for the dentists and the CDSS at different sections of the same vignette was then recorded and the data for all the vignettes collated and analysed. The NDCG score for the dentists were recorded for the Rank 1 and Rank 2 final diagnosis using the NDCG formula described in previous sections (section - 6.6.1). Table 6-9 shows the overall average than was obtained for the dentists across all 25 vignettes. As mentioned previously in the section about NDCG, a score of 1 is a perfect score and the ranking matches perfectly with that of the Endodontist. Clinicians in general practice (example dentists) are generally not expected to match the diagnostic performance of experts. Autopsy studies have shown that clinicians have a diagnostic error rate of 10-20% (Aalten, Samson and Jansen, 2006; Graber, 2013).

6.7.3 Chief Complaint (CC)

This section outlines the data obtained from the CDSS with only observations from the CC section of the information model.

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.81	0.64

Table 6-10 - NDCG average for CDSS – CC

6.7.3.1 Pulpal (CC)

In this section the Pulpal NDCG average of the CDSS (CC) (Table 6-10) is compared to the dentists' final Pulpal diagnosis (Table 6-9). The data shows that the mean NDCG value for dentist's final Pulpal diagnosis when compared to the CDSS Pulpal diagnosis using just the CC data is very close.

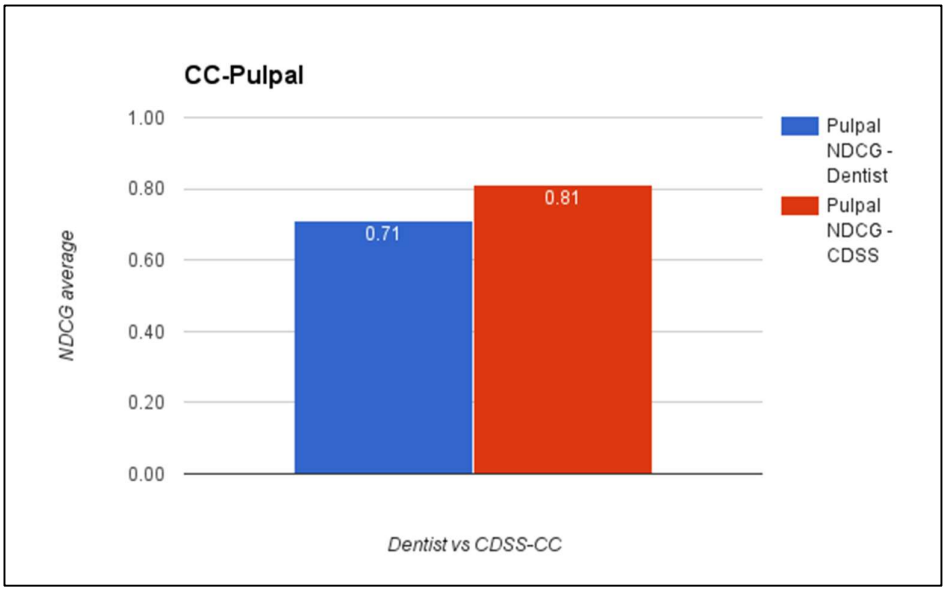


Figure 6-14 - Pulpal (CC)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences (p≤ 0.05)*
CC-Pulpal vs Dentists-Pulpal	0.01275046 0.18619496	t = 2.2786, df = 90.593, p-value = 0.02504	✓

Table 6-11 - CC-Pulpal vs Dentists-Pulpal - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences (p≤ 0.05)*
CC-Pulpal vs Dentists-Pulpal	-0.09221490 0.09666238	W = 1430, p-value = 0.7026	✗

Table 6-12 - CC-Pulpal vs Dentists-Pulpal -Wilcoxon

The Pulpal NDCG average comparison (Figure 6-14) shows that the difference between the dentists' final diagnosis and the CDSS at the CC section is not statistically significant according to the Wilcoxon rank sum test (Table 6-12) but statistically significant according to the t-test (Table 6-13).

6.7.3.2 Apical (CC)

In this section the Apical NDCG average of the CDSS (CC) (Table 6-10) is compared to the dentists' final Apical diagnosis (Table 6-9). The data shows that the mean NDCG value for dentists in final Apical diagnosis is much greater when compared to the Apical diagnosis of the CDSS using only CC data of the information model.

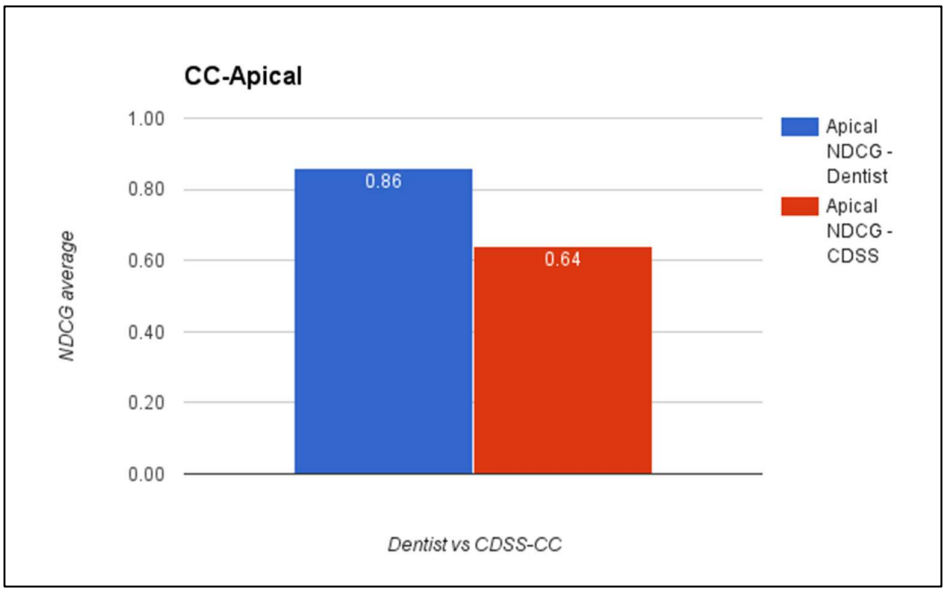


Figure 6-15 - Apical (CC)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
CC-Apical vs Dentists-Apical	-0.3102159 -0.1314513	t = -4.9453, df = 58.15, p-value = 6.816e-06	✓

Table 6-13 - CC-Apical vs Dentists-Apical - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
CC-Apical vs Dentists-Apical	-0.4596947 -0.2096987	W = 462, p-value = 4.566e-10	✓

Table 6-14 - CC-Apical vs Dentists-Apical - Wilcoxon

The Apical NDCG average comparison (Figure 6-15) shows that the difference between the dentists' final diagnosis and the CDSS at the CC section is statistically significant according to the Wilcoxon rank

sum test (Table 6-13) and the t-test (Table 6-14). The CDSS performs considerably poorer than the dentists in the CC section in Apical diagnosis. The dentist's final diagnosis NDCG average score was used as the cut-off for determining if the CDSS using the DS Model can perform adequately. If the CDSS performs poorer than the final NDCG average of a non-expert clinician, then we can identify this section of the information model as an area where improvement should be made. The ideal performance would closely match that of the Endodontist. However, the DS Model will not be able to achieve that level of performance at an early section since all the information has not collected yet. If the CDSS performs significantly better at ranking the diagnosis when compared to the non-expert clinician (dentists) then the CDSS could potentially support him/her at arriving at an early conclusion. Testing this effect on clinician's diagnostic performance is beyond the scope of this thesis. However we know from literature that providing early diagnostic recommendations can help the clinician in improving his/her final diagnosis as well (Kostopoulou *et al.*, 2015b). The results of the NDCG scores of final diagnosis of the dentists show that there is a need to improve their final diagnosis as well. The results of this section show that the DS Model needs improvement in Apical diagnosis in the CC section.

6.7.4 Chief Complaint (CC) + Extra Oral examination (EO)

This section outlines the data obtained from the CDSS using only observations from the CC + EO section of the information model.

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.77	0.80

Table 6-15 - NDCG average for CDSS - CC+EO

6.7.4.1 Pulpal (CC+EO)

In this section the Pulpal NDCG average of the CDSS (CC+EO) (Table 6-15) is compared to the dentists' final diagnosis (Pulpal) (Table 6-9).

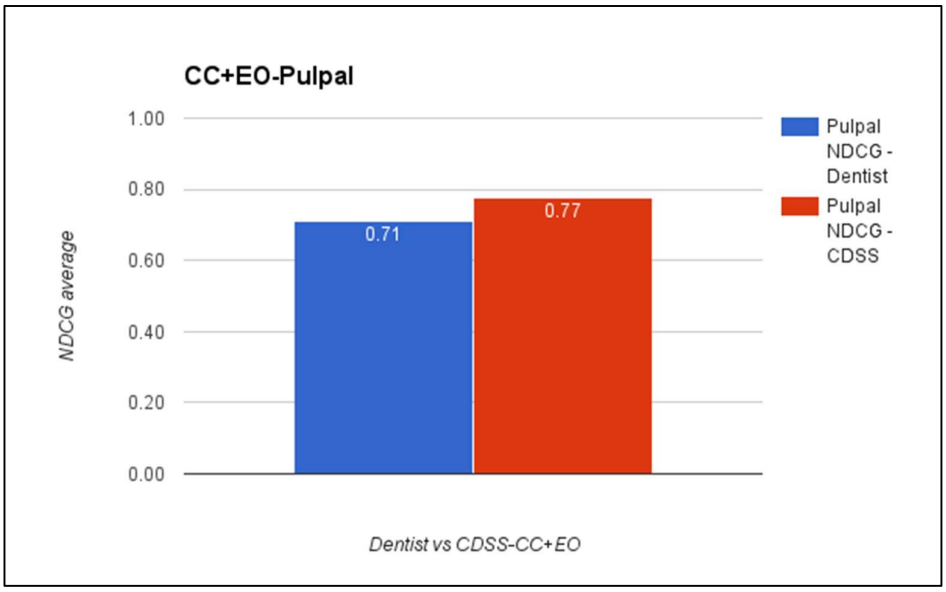


Figure 6-16 - Pulpal (CC+EO)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences (p≤ 0.05)*
CC+EO-Pulpal vs Dentists-Pulpal	-0.04133266 0.16212098	t = 1.1869, df = 61.72, p-value = 0.2398	X

Table 6-16 - CC+EO-Pulpal vs Dentists-Pulpal- T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences (p≤ 0.05)*
CC+EO-Pulpal vs Dentists-Pulpal	-0.1423007 0.1076113	W = 1466, p-value = 0.854	X

Table 6-17 - CC+EO-Pulpal vs Dentists-Pulpal - Wilcoxon

The Pulpal NDCG average comparison (Figure 6-16) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO section is not statistically significant according to the Wilcoxon rank sum test (Table 6-17) and the t-test (Table 6-16)

6.7.4.2 Apical (CC+EO)

In this section the Apical NDCG average of the CDSS (CC+EO) (Table 6-15) is compared to the dentists' final diagnosis (Apical) (Table 6-9).

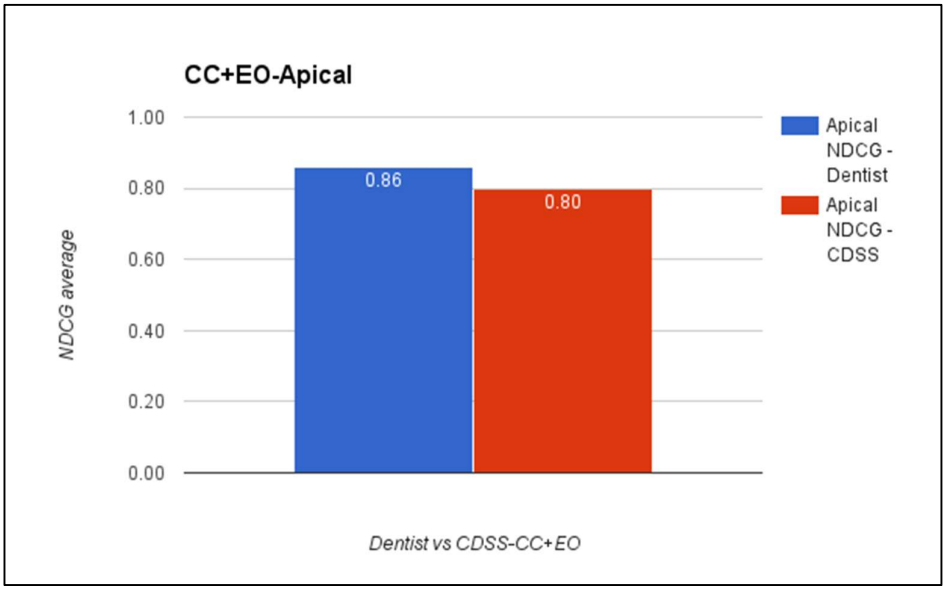


Figure 6-17 - Apical (CC+EO)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
CC+EO-Apical vs Dentists-Apical	-0.15525741 0.03679235	t = -1.2382, df = 51.205, p-value = 0.2213	✗

Table 6-18 - CC+EO-Apical vs Dentists-Apical - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
CC+EO-Apical vs Dentists-Apical	-1.423529e-01 -3.786639e-05	W = 966, p-value = 0.0007716	✓

Table 6-19 - CC+EO-Apical vs Dentists-Apical - Wilcoxon

The Apical NDCG average comparison (Figure 6-17) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO section is statistically significant according to the Wilcoxon rank sum test (Table 6-19) but not according to t-test (Table 6-18).

6.7.5 Chief Complaint (CC) + Extra Oral examination (EO) + Intra Oral examination (IO)

This section outlines the data obtained from the CDSS with only observations from the CC+EO+IO section of the information model.

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.81	0.80

Table 6-20 - NDCG average for CDSS – CC+EO+IO

6.7.5.1 Pulpal (CC+EO+IO)

In this section the Pulpal NDCG average of the CDSS (CC+EO+IO) (Table 6-20) is compared to the dentists' final diagnosis (Pulpal) (Table 6-9).

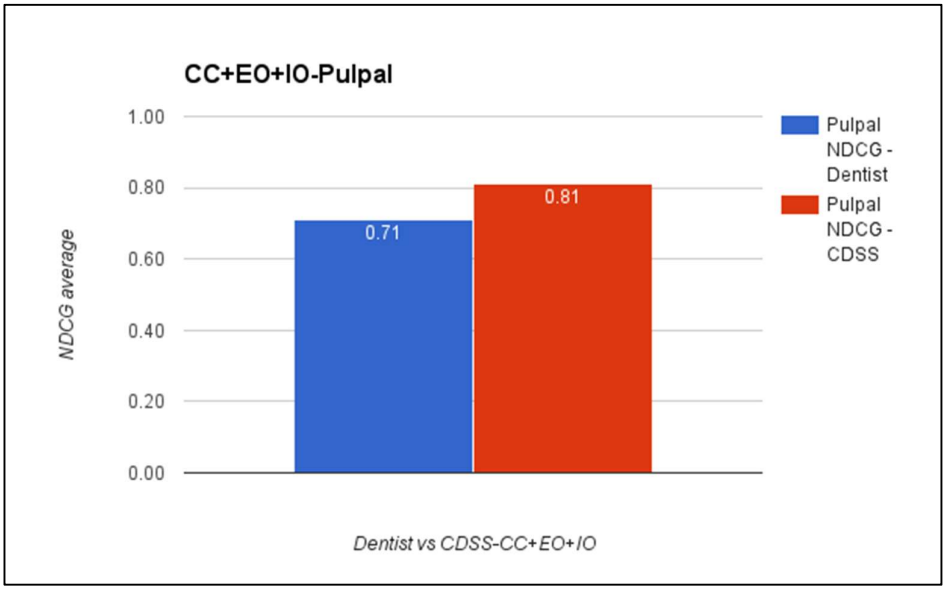


Figure 6-18 - Pulpal (CC+EO+IO)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences (p≤ 0.05)*
CC+EO+IO-Pulpal vs Dentists-Pulpal	-0.005935879 0.200554714	t = 1.8854, df = 59.902, p-value = 0.06423	✗

Table 6-21 - CC+EO+IO-Pulpal vs Dentists-Pulpal - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences (p≤ 0.05)*
CC+EO+IO-Pulpal vs Dentists-Pulpal	-0.09222195 0.15776828	W = 1548.5, p-value = 0.7908	✗

Table 6-22 - CC+EO+IO-Pulpal vs Dentists-Pulpal - Wilcoxon

The Pulpal NDCG average comparison (Figure 6-18) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO+IO section is not statistically significant according to the Wilcoxon rank sum test (Table 6-22) and the t-test (Table 6-21).

6.7.5.2 Apical (CC+EO+IO)

In this section the Apical NDCG average of the CDSS (CC+EO+IO) (Table 6-20) is compared to the dentists' final diagnosis (Apical) (Table 6-9).

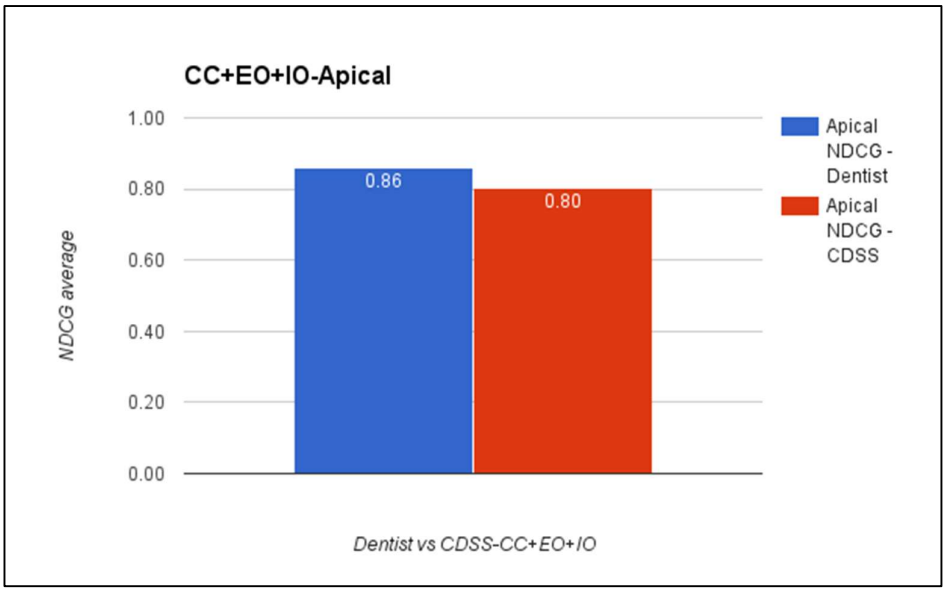


Figure 6-19 - Apical (CC+EO+IO)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO-Apical vs Dentists-Apical	-0.1514408 0.0381477	t = -1.1989, df = 52.312, p-value = 0.236	✗

Table 6-23 - CC+EO+IO-Apical vs Dentists-Apical - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO-Apical vs Dentists-Apical	-1.517864e-01 -3.694163e-05	W = 966, p-value = 0.0007716	✓

Table 6-24 - CC+EO+IO-Apical vs Dentists-Apical - Wilcoxon

The Apical NDCG average comparison (Figure 6-19) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO+IO section is statistically significant according to the Wilcoxon rank sum test (Table 6-24) but not according to t-test(Table 6-23).

6.7.6 Chief Complaint (CC) + Extra Oral examination + Intra Oral examination (IO) (EO) + Clinical Tests (CT)

This section outlines the data obtained from the CDSS with only observations from the CC+EO+IO +CT section of the information model.

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.94	0.86

Table 6-25 - NDCG average for CDSS - CC+EO+IO+CT

6.7.6.1 Pulpal (CC+EO+IO+CT)

In this section the Pulpal NDCG average of the CDSS (CC+EO+IO+CT) (Table 6-25) is compared to the dentists' final diagnosis (Pulpal) (Table 6-9).

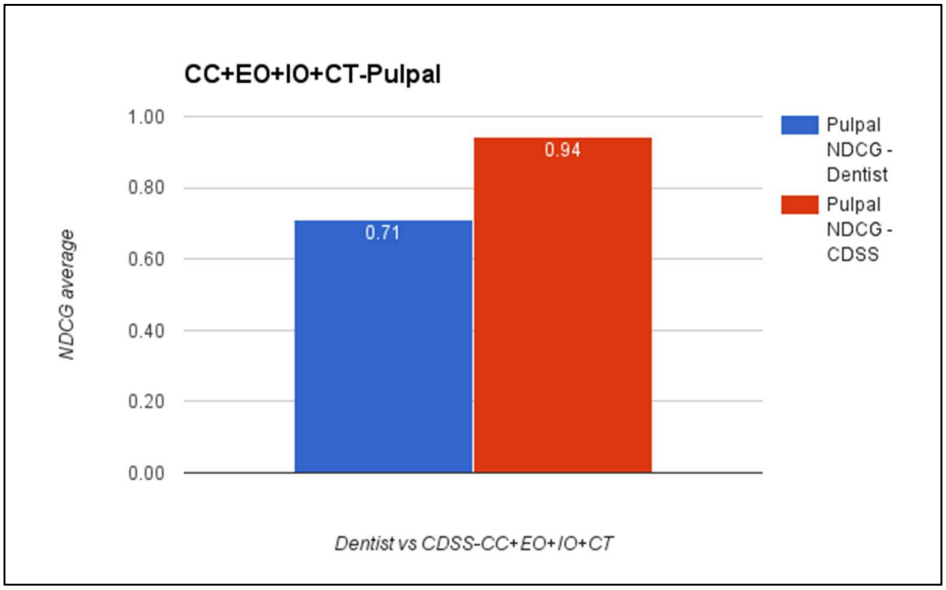


Figure 6-20 - Pulpal (CC+EO+IO+CT)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO+CT-Pulpal vs Dentists-Pulpal	0.1568956 0.2956173	t = 6.4481, df = 142.739, p-value = 1.636e-09	✓

Table 6-26 - CC+EO+IO+CT-Pulpal vs Dentists-Pulpal - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO+CT-Pulpal vs Dentists-Pulpal	3.514817e-05 2.499932e-01	W = 2010, p-value = 0.004355	✓

Table 6-27 - CC+EO+IO+CT-Pulpal vs Dentists-Pulpal - Wilcoxon

The Pulpal NDCG average comparison (Figure 6-20) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO+IO+CT section is statistically significant according to the Wilcoxon rank sum test (Table 6-27) and the t-test (Table 6-26).

6.7.6.2 Apical (CC+EO+IO+CT)

In this section the Apical NDCG average of the CDSS (CC+EO+IO+CT) (Table 6-25) is compared to the dentists' final diagnosis (Apical) (Table 6-9).

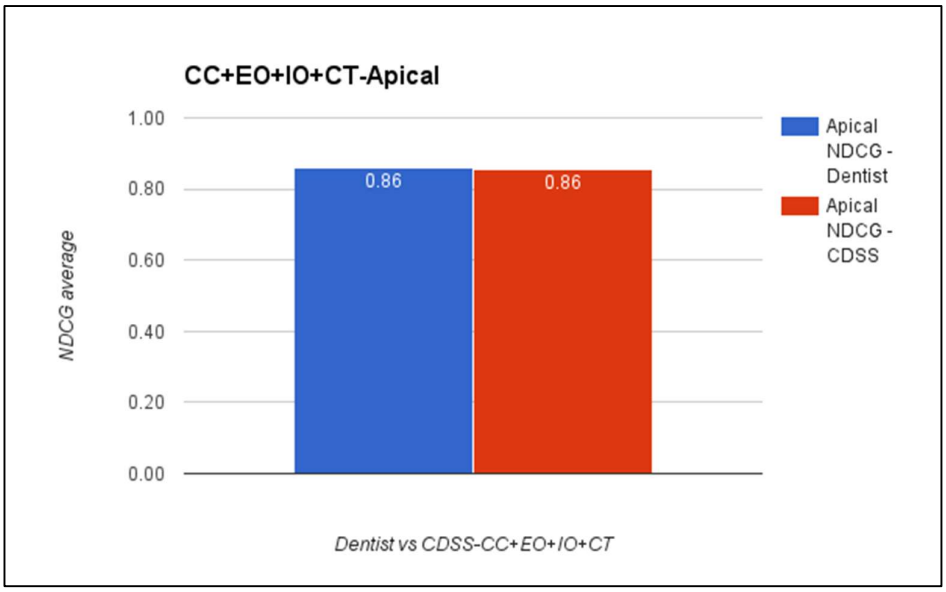


Figure 6-21 - Apical (CC+EO+IO+CT)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO+CT-Apical vs Dentists-Apical	-0.08795047 0.08276901	t = -0.0606, df = 63.819, p-value = 0.9518	✗

Table 6-28 - CC+EO+IO+CT-Apical vs Dentists-Apical - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
CC+EO+IO+CT-Apical vs Dentists-Apical	-0.09231064 -0.00003468	W = 1072, p-value = 0.006268	✓

Table 6-29 -CC+EO+IO+CT-Apical vs Dentists-Apical - Wilcoxon

The Apical NDCG average comparison (Figure 6-21) shows that the difference between the dentists' final diagnosis and the CDSS at the CC+EO+IO+CT section is statistically significant according to the Wilcoxon rank sum test (Table 6-29) but not according to t-test (Table 6-28).

6.7.7 Chief Complaint (CC) + Extra Oral examination + Intra Oral examination (IO) (EO) + Clinical Tests (CT) + Radiographic Examination (RE) (Final diagnosis)

This section outlines the data obtained from the CDSS with only observations from the CC+IO+EO+CT+RT (Final Diagnosis) section of the information model.

(n=25)	Pulpal NDCG	Apical NDCG
Mean	0.94	0.98

Table 6-30 - NDCG average for CDSS – Final diagnosis

6.7.7.1 Pulpal (final)

In this section the Pulpal NDCG average of the CDSS (final) (Table 6-30) is compared to the dentists' final diagnosis (Pulpal) (Table 6-9).

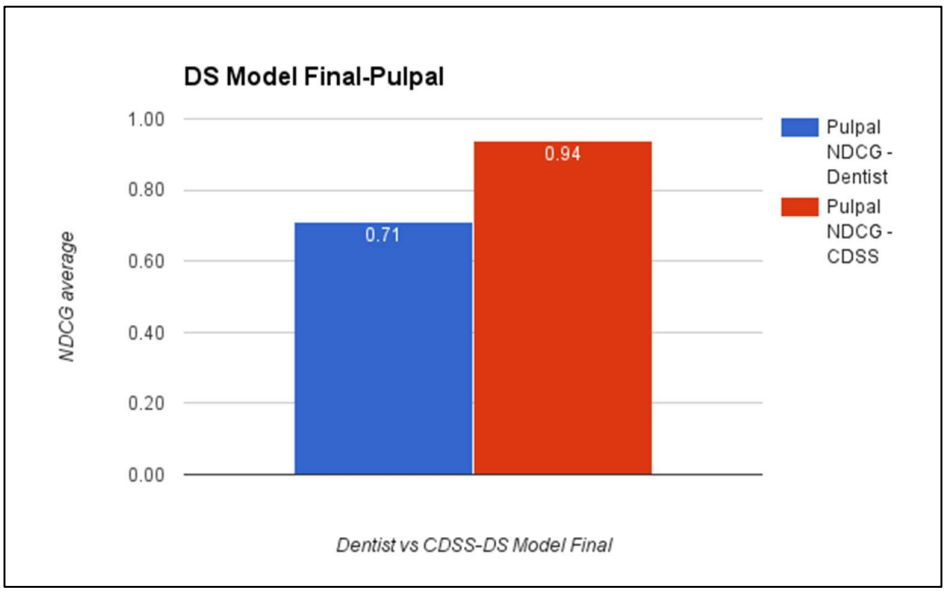


Figure 6-22 - Pulpal (final)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences (p≤ 0.05)*
DS Model (Final)-Pulpal vs Dentists-Pulpal	0.1537359 0.2922669	t = 6.3641, df = 142.657, p-value = 2.512e-09	✓

Table 6-31 - DS Model (Final)-Pulpal vs Dentists-Pulpal - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences (p≤ 0.05)*
DS Model(Final)-Pulpal vs Dentists-Pulpal	5.708955e-05 2.499524e-01	W = 1982.5, p-value = 0.007143	✓

Table 6-32 - DS Model (Final)-Pulpal vs Dentists-Pulpal - Wilcoxon

The Pulpal NDCG average comparison (Figure 6-22) shows that the difference between the dentists' final diagnosis and the CDSS at the final section is statistically significant according to the Wilcoxon rank sum test (Table 6-32) and the t-test (Table 6-31).

6.7.7.2 Apical (final)

In this section the Apical NDCG average of the CDSS (final) (Table 6-30) is compared to the dentists' final diagnosis (Apical) (Table 6-9).

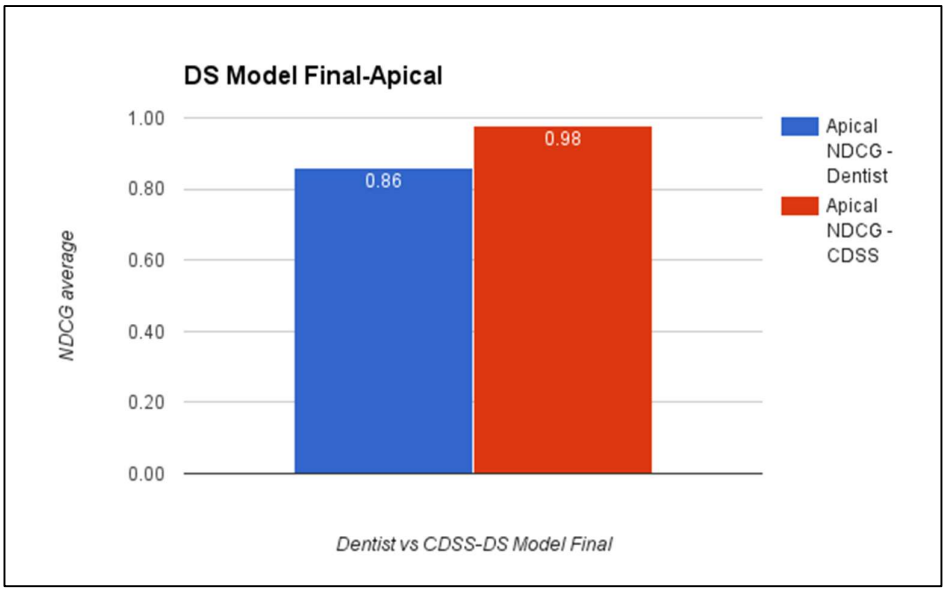


Figure 6-23 - Apical (final)

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences ($p \leq 0.05$)*
DS Model (Final)-Apical vs Dentists-Apical	0.06275468 0.17628843	t = 4.1639, df = 135.276, p-value = 5.539e-05	✓

Table 6-33 - DS Model (Final)-Apical vs Dentists-Apical - T test

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
DS Model (Final)-Apical vs Dentists-Apical	-8.687249e-05 3.065340e-06	W = 1605, p-value = 0.4701	✗

Table 6-34 - DS Model (Final)-Apical vs Dentists-Apical - Wilcoxon

The Apical NDCG average comparison (Figure 6-23) shows that the difference between the dentists' final diagnosis and the CDSS at the final section is not statistically significant according to the Wilcoxon rank sum test (Table 6-34) but significant according to t-test (Table 6-33).

6.7.8 T-Test: Two-Sample Assuming Unequal Variances-Summary of results

The following sections provide a summary of the results of the different sections of the information model.

6.7.8.1 Pulpal

According to the t-test (Table 6-35) the CDSS performed significantly better than the dentist's final diagnosis in the CC, CC+EO+IO+CT and the final sections in diagnosing Pulpal diagnosis (Figure 6-24). In the CC section the CDSS performs better than the dentists and there is drop in performance as more data is collected. The CDSS performs better than the dentists in the CC+EO+IO+CT and the final sections.

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences* *(p≤ 0.05)*
CC-Pulpal vs Dentists-Pulpal (CDSS better than dentists)	0.01275046 0.18619496	t = 2.2786, df = 90.593, p-value = 0.02504	✓
CC+EO-Pulpal vs Dentists-Pulpal	-0.04133266 0.16212098	t = 1.1869, df = 61.72, p-value = 0.2398	✗
CC+EO+IO-Pulpal vs Dentists-Pulpal	-0.005935879 0.200554714	t = 1.8854, df = 59.902, p-value = 0.06423	✗
CC+EO+IO+CT-Pulpal vs Dentists-Pulpal (CDSS better than dentists)	0.1568956 0.2956173	t = 6.4481, df = 142.739, p-value = 1.636e-09	✓
DS Model (Final)-Pulpal vs Dentists-Pulpal (CDSS better than dentists)	0.1537359 0.2922669	t = 6.3641, df = 142.657, p-value = 2.512e-09	✓

Table 6-35 - Overall NDCG Comparison-Pulpal - T test

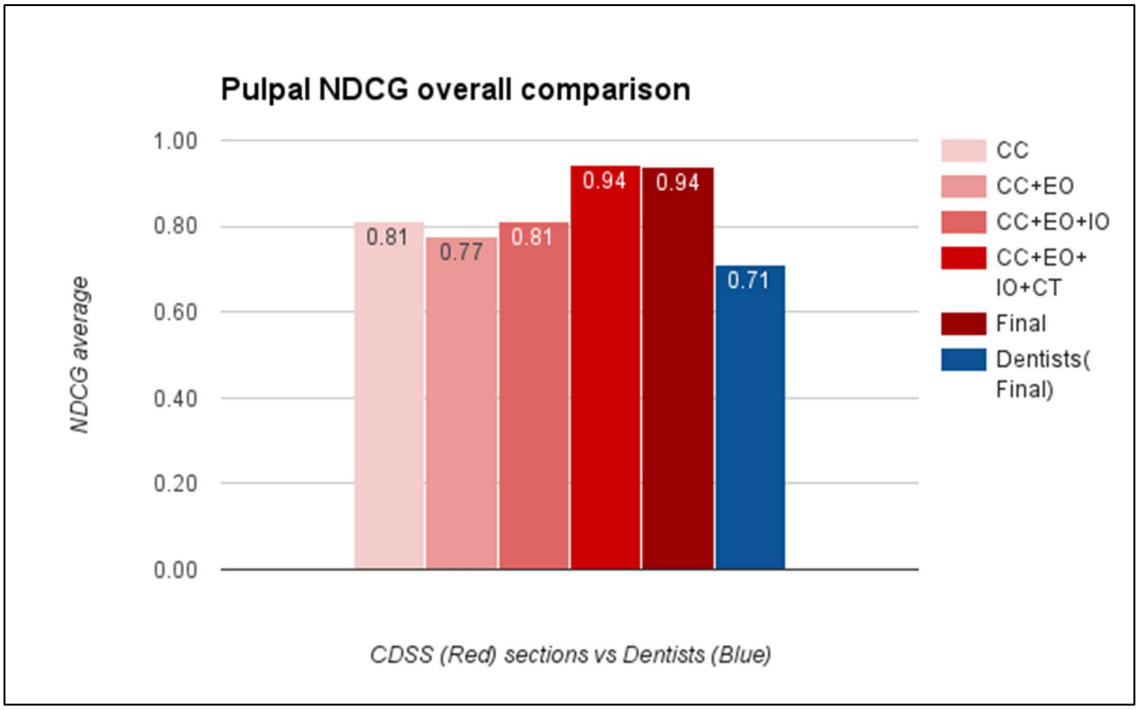


Figure 6-24 - Pulpal NDCG overall comparison

6.7.8.2 Apical

The t-test (Table 6-36) shows that the CDSS performed significantly poorer than the dentist's final diagnosis in the early section and performed better than the dentists in the Final section in diagnosing Apical diagnosis (Figure 6-25).

NDCG Comparison	*95 % Confidence Interval*	*Independent Two-Sample t Test Assuming Unequal Variances*	*Significant Differences* ($p \leq 0.05$)
CC-Apical vs Dentists-Apical (CDSS better than dentists)	-0.3102159 -0.1314513	t = -4.9453, df = 58.15, p-value = 6.816e-06	✓
CC+EO-Apical vs Dentists-Apical	-0.15525741 0.03679235	t = -1.2382, df = 51.205, p-value = 0.2213	✗
CC+EO+IO-Apical vs Dentists-Apical	-0.1514408 0.0381477	t = -1.1989, df = 52.312, p-value = 0.236	✗
CC+EO+IO+CT-Apical vs Dentists-Apical	-0.08795047 0.08276901	t = -0.0606, df = 63.819, p-value = 0.9518	✗
DS Model (Final)-Apical vs Dentists-Apical (CDSS better than dentists)	0.06275468 0.17628843	t = 4.1639, df = 135.276, p-value = 5.539e-05	✓

Table 6-36 - Overall NDCG Comparison-Apical - T test

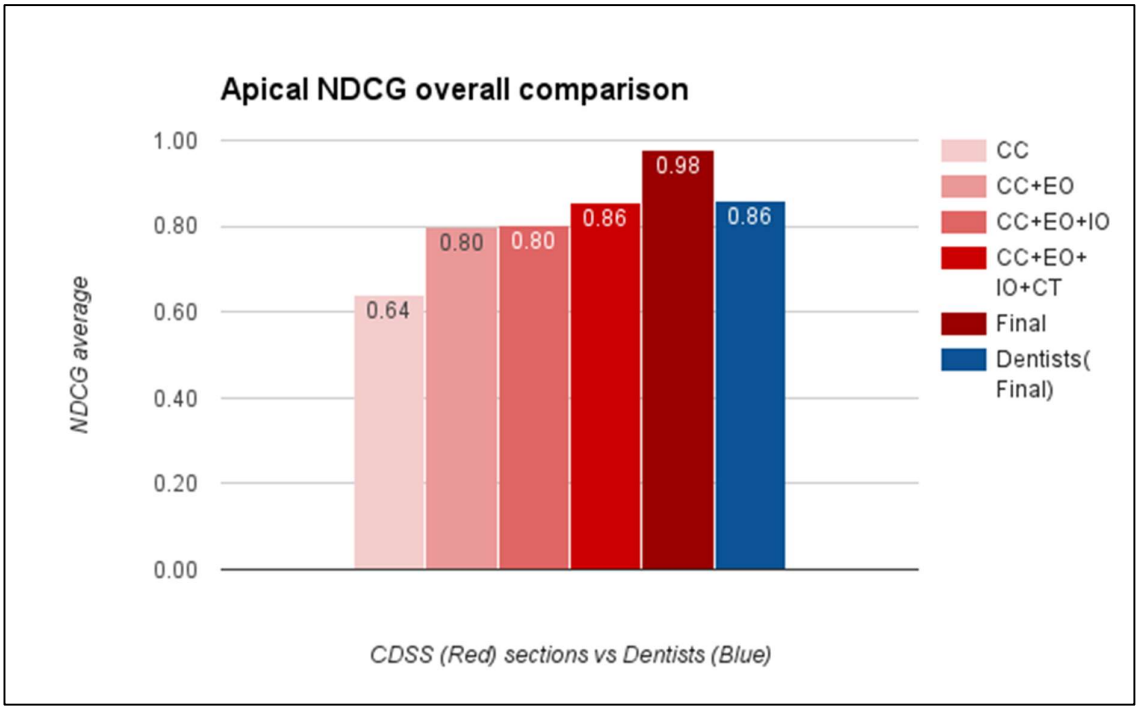


Figure 6-25 - Apical NDCG overall comparison

6.7.9 Wilcoxon rank sum test with continuity correction - Summary of results

6.7.9.1 Pulpal

The results of the Wilcoxon rank sum test (Table 6-37) show that the CDSS is better than the dentists only in the late sections (CC+EO+IO+CT and Final) in diagnosing Pulpal diagnosis (Figure 6-26). There is no significant difference when compared to the dentists in the other sections. Even with limited information the CDSS performs at least as good as the dentists in the early section, though there is scope for improvement.

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences* *(p≤ 0.05)*
CC-Pulpal vs Dentists-Pulpal	-0.09221490 0.09666238	W = 1430, p-value = 0.7026	✗
CC+EO-Pulpal vs Dentists-Pulpal	-0.1423007 0.1076113	W = 1466, p-value = 0.854	✗
CC+EO+IO-Pulpal vs Dentists-Pulpal	-0.09222195 0.15776828	W = 1548.5, p-value = 0.7908	✗
CC+EO+IO+CT-Pulpal vs Dentists-Pulpal (CDSS better than dentists)	3.514817e-05 2.499932e-01	W = 2010, p-value = 0.004355	✓
DS Model (Final)-Pulpal vs Dentists-Pulpal	5.708955e-05 2.499524e-01	W = 1982.5, p-value = 0.007143	✓

(CDSS better than dentists)			

Table 6-37 -Overall NDCG Comparison-Pulpal - Wilcoxon

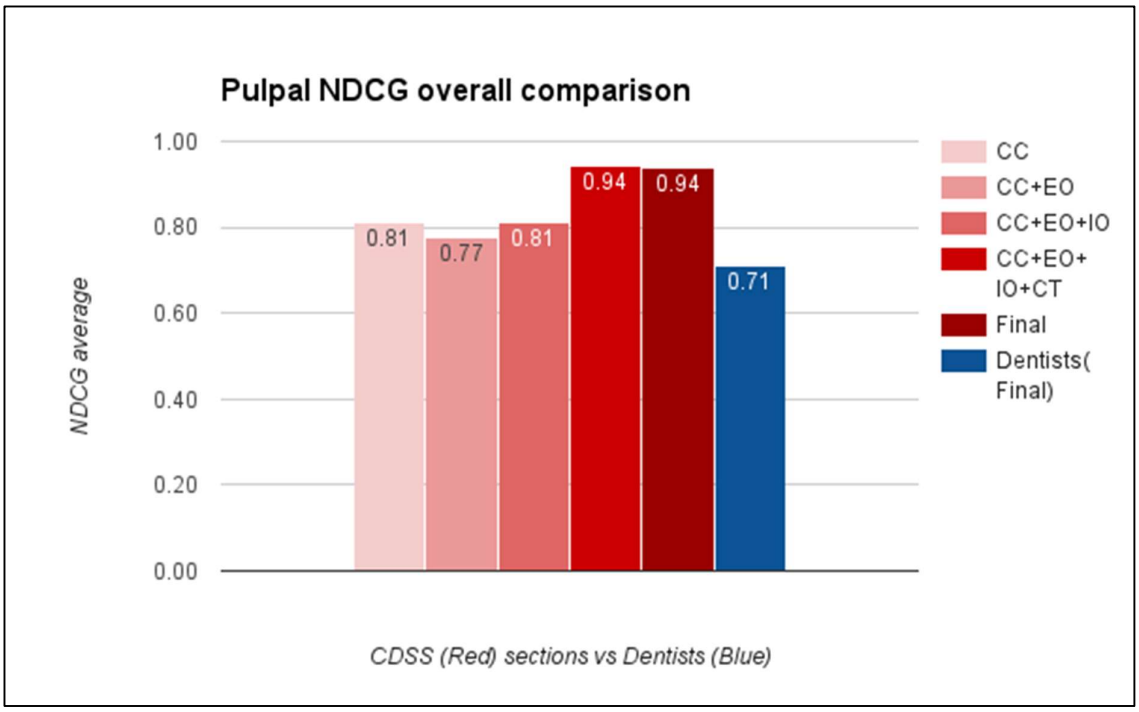


Figure 6-26 - Pulpal NDCG overall comparison

6.7.9.2 Apical

The results of the Wilcoxon rank sum test (Table 6-38) for Apical diagnosis show that the dentist's final diagnostic performance is better than the CDSS in all sections of the encounter (Figure 6-27). The CDSS did not perform significantly better than the dentists in the Final section. This result is markedly different from the t-test result (section - 6.7.8.2) where the CDSS performed significantly poorer in the CC section but showed that the CDSS performed significantly better than the dentist in the final section.

NDCG Comparison	*95 % Confidence Interval*	*Wilcoxon rank sum test with continuity correction*	*Significant Differences* ($p \leq 0.05$)
CC-Apical vs Dentists-Apical (CDSS better than dentists)	-0.4596947 -0.2096987	W = 462, p-value = 4.566e-10	✓
CC+EO-Apical vs Dentists-Apical (CDSS better than dentists)	-1.423529e-01 -3.786639e-05	W = 966, p-value = 0.0007716	✓
CC+EO+IO-Apical vs Dentists-Apical (CDSS better than dentists)	-1.517864e-01 -3.694163e-05	W = 966, p-value = 0.0007716	✓
CC+EO+IO+CT-Apical vs Dentists-Apical (CDSS better than dentists)	-0.09231064 -0.00003468	W = 1072, p-value = 0.006268	✓

DS Model (Final)-Apical vs Dentists-Apical	-8.687249e-05 3.065340e-06	W = 1605, p-value = 0.4701	✗

Table 6-38 - Overall NDCG Comparison-Apical - Wilcoxon

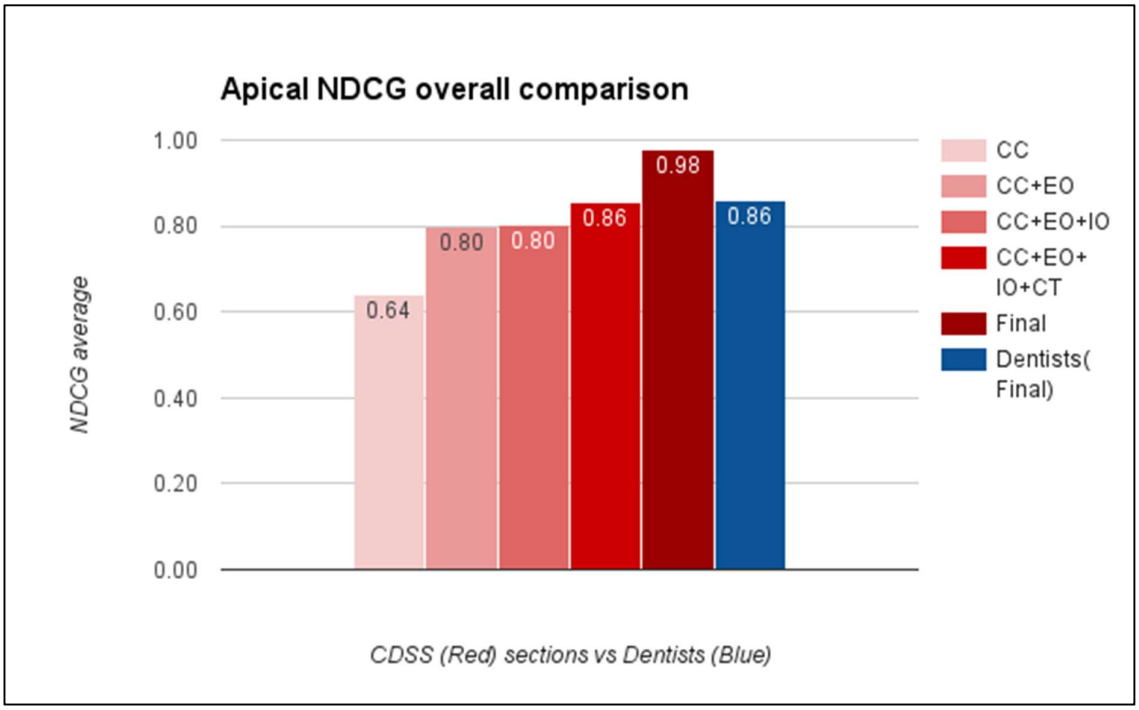


Figure 6-27 - Apical NDCG overall comparison

6.7.10 T-test vs Wilcoxon Rank Sum Test-Summary of results

6.7.10.1 Pulpal

A side-by-side comparison of the t-test and Wilcoxon rank sum test (Table 6-39) shows that there is general agreement in the t-test and Wilcoxon rank sum significance tests in the results as far as Pulpal diagnosis is concerned. The final diagnostic performance of the dentists in Pulpal diagnosis is better than the CDSS in the CC section of the diagnostic encounter and the CDSS performs better than the dentists in the CC+EO+IO+CT and final sections. In the CC+EO and CC+EO+IO the CDSS performed at least as good as the dentists. There is a need for improvement in diagnosis across all the early sections of the information model. In the CC, CC+EO and the CC+EO+IO sections the CDSS will require additional information to make a better diagnostic recommendation than the dentists.

NDCG Comparison	*Independent Two-Sample t Test* *Significant Differences* *($p \leq 0.05$)*	*Wilcoxon rank sum test* *Significant Differences* *($p \leq 0.05$)*
CC-Pulpal vs Dentists-Pulpal	✓	✗
CC+EO-Pulpal vs Dentists-Pulpal	✗	✗
CC+EO+IO-Pulpal vs Dentists-Pulpal	✗	✗

CC+EO+IO+CT-Pulpal vs Dentists-Pulpal	✓	✓
DS Model (Final)-Pulpal vs Dentists-Pulpal	✓	✓

Table 6-39 - Pulpal T-test vs Wilcoxon Rank Sum Test

6.7.10.2 Apical

A comparison of the t-test and Wilcoxon rank sum test shows that there is little consensus in the Apical diagnosis (Table 6-40). The final diagnostic performances of the dentists were significantly better than the CDSS in the CC section according to the t-test. According to the t-test the CDSS performs poorly in the CC section and performs at least as good as the dentists in the intermediate sections. The Wilcoxon test shows that the CDSS performs significantly poorer than the dentists in all the sections except the final section. Therefore, in the Apical diagnosis the CDSS requires additional information in all the sections except the final section to provide a diagnostic recommendation better than dentists. It is not known if the CDSS requires additional information in the final section for Apical diagnosis as the t-test and the Wilcoxon shows different results.

NDCG Comparison	*Independent Two-Sample t Test* *Significant Differences* *($p \leq 0.05$)*	*Wilcoxon rank sum test* *Significant Differences* *($p \leq 0.05$)*
CC-Apical vs Dentists-Apical	✓	✓
CC+EO-Apical vs Dentists-Apical	✗	✓
CC+EO+IO-Apical vs Dentists-Apical	✗	✓
CC+EO+IO+CT-Apical vs Dentists-Apical	✗	✓
DS Model (Final)-Apical vs Dentists-Apical	✓	✗

Table 6-40 - Apical T-test vs Wilcoxon Rank Sum Test

This section provides the results of the NDCG analysis. The next section provides an overview of the Precision-Recall analysis.

6.8 Precision-Recall Analysis

6.8.1 Introduction to Precision-Recall Analysis

The CDSS using the dental DS Model provides a Rank 1 and Rank 2 diagnosis for both the Pulpal and Apical diagnosis. The Pulpal Rank 1 can be considered as binary classification i.e. Correct or Incorrect. Consider the example of vignette no. 10 in Table 6-2. The Rank 1 Pulpal diagnosis for the vignette is Pulp Necrosis. If the CDSS accurately provides a diagnosis of Pulp Necrosis, then it is the correct answer. There can be 4 possible classification outcomes: True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

If the CDSS accurately classifies vignette no. 10 as Pulp Necrosis then it is a TP outcome, i.e. the CDSS has accurately classified the vignette as Pulp Necrosis when the diagnosis was Pulp Necrosis. If the CDSS accurately classifies the vignette as not Pulp Necrosis when the diagnosis is in fact not Pulp Necrosis, then it a TN outcome. If the CDSS falsely classifies a vignette as Pulp Necrosis when in fact the diagnosis is something else, then it is considered as a FP outcome. If the CDSS falsely classifies the vignette as not Pulp Necrosis when it is indeed Pulp Necrosis, then the outcome can be considered a FN outcome.

The results can be represented in a 2 x 2 contingency table. For example, the results of the Rank 1 Pulpal classification for the Pulp Necrosis diagnosis have been shown below (Table 6-41). All 25 vignettes were provided to the CDSS using the DS Model and the results were recorded in a contingency table for each diagnosis.

Pulp Necrosis	Correct (according to CDSS)	Incorrect (according to CDSS)
Correct (actual)	A (TP)	C (FN)
	10	0
Incorrect (actual)	B (FP)	D (TN)
	0	15

Table 6-41 - 2 x 2 contingency table – Pulp Necrosis Rank 1

In example result shown in the table obtained from the CDSS the CDSS has a TP outcome 10 times and TN outcome 15 times out of the total of 25 vignettes.

The same technique can be used to obtain a contingency table for all the Pulpal diagnoses (n=7) and Apical diagnoses (n=5). A Rank 1 and Rank 2 contingency table for each diagnosis can be obtained. The correct Rank 2 was provided by the Endodontist alongside the Rank 1 diagnosis. Therefore, for each individual diagnosis there will be 2 x 2 contingency table for Rank 1 and Rank 2 classifications. The same process was repeated for the dentists and a Rank 1 and Rank 2 contingency table was generated for each diagnosis. The total numbers of TP, TN, FP, and FN outcomes were recorded in these tables.

6.8.1.1 Precision

Once the contingency tables were generated for each diagnosis for the CDSS and the dentists the precision and recall metric was calculated. Precision is the percentage of diagnosed classifications that are correct. Precision calculates the proportion of correct responses of both the CDSS and dentists across all diagnoses and in both Rank 1 and Rank 2 positions. It is calculated using the formula below.

$$\text{Precision} = TP/(TP + FP) = A/(A + B)$$

Therefore, the precision for the CDSS in Pulp Necrosis Rank 1 is 1 or 100% (Table 6-41).

6.8.1.2 Recall

Recall can be defined as the percentage of vignettes that were correctly diagnosed in relation to the vignettes that were correct. It is calculated using the formula below.

$$\text{Recall} = TP/(TP + FN) = A/(A + C)$$

Therefore, the recall for the CDSS in Pulp Necrosis Rank 1 is 1 or 100% (Table 6-41).

The tables below (Table 6-42, Table 6-43, Table 6-44, Table 6-45) show the results of Pulpal (Rank 1 and Rank 2) and Apical (Rank 1 and Rank 2) precision and recall for CC section of the information model. The average precision and recall of the CDSS using the DS Model in Pulpal (Rank 1 and Rank 2) and Apical (Rank 1 and Rank 2) diagnoses for all sections of the information model (i.e. CC, CC+EO, CC+EO+IO, CC+EO+IO+CT, Final) was obtained. This was then compared to the Pulpal (Rank 1 and Rank 2) and Apical (Rank 1 and Rank 2) precision and recall of the dentists.

Pulpal Diagnosis (Rank 1) CC	Precision	Recall
Normal Pulpal Tissues	0.2	1
Reversible Pulpitis	0	0
Symptomatic Irreversible Pulpitis	0.5555555556	1
Asymptomatic Irreversible Pulpitis	0	0
Pulp Necrosis	1	0.1111111111
Pain of Nonodontogenic Origin	1	0.5
Dentinal Hypersensitivity	0.3333333333	1
Average	0.4412698413	0.5158730159

Table 6-42 - Pulpal Diagnosis (Rank 1) CC

Pulpal Diagnosis (Rank 2) CC	Precision	Recall
Normal Pulpal Tissues	0	0
Reversible Pulpitis	0.2	0.1
Symptomatic Irreversible Pulpitis	0	0
Asymptomatic Irreversible Pulpitis	0.5833333333	0.875
Pulp Necrosis	0	0
Pain of Nonodontogenic Origin	0.2	1
Dentinal Hypersensitivity	0	0
Average	0.1404761905	0.2821428571

Table 6-43 -Pulpal Diagnosis (Rank 2) CC

Apical Diagnosis (Rank 1) CC	Precision	Recall
Normal Apical Tissues	0	0
Asymptomatic Apical Periodontitis	0	0
Pain of Nonodontogenic Origin	0	0
Symptomatic Apical Periodontitis	0	0
Acute Apical Abscess	0.1538461538	1
Chronic Apical Abscess	0.25	1
Average	0.06730769231	0.3333333333

Table 6-44 - Apical Diagnosis (Rank 1) CC

Apical Diagnosis (Rank 2) CC	Precision	Recall
Normal Apical Tissues	0	0
Asymptomatic Apical Periodontitis	0	0
Pain of Nonodontogenic Origin	0	0
Symptomatic Apical Periodontitis	0	0
Acute Apical Abscess	0	0
Chronic Apical Abscess	0	0
Average	0	0

Table 6-45 - Apical Diagnosis (Rank 2) CC

The precision-recall data for all the sections of the information model (CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and Final) and dentists is presented in Appendix 13. Also, see Appendix 15 for raw data.

This section provides an overview of the Precision-Recall analysis. The next section provides the results of Precision-Recall analysis.

6.9 Results of Precision-Recall analysis

6.9.1 Wilcoxon rank sum test with continuity correction - Summary of results

The Wilcoxon rank sum test with continuity correction was performed for significance testing in the precision and recall analysis. Shapiro-Wilk test for normality was performed on the precision-recall data and the data was largely found to have a normal distribution (See Appendix 12 – section 12.2). Therefore, the Wilcoxon rank sum test was chosen was the test for significance. The test will be performed across several types of results. The Rank 1 and Rank 2 precision and recall of both Pulpal and Apical diagnosis (Figure 6-28) was compared to that of dentists.

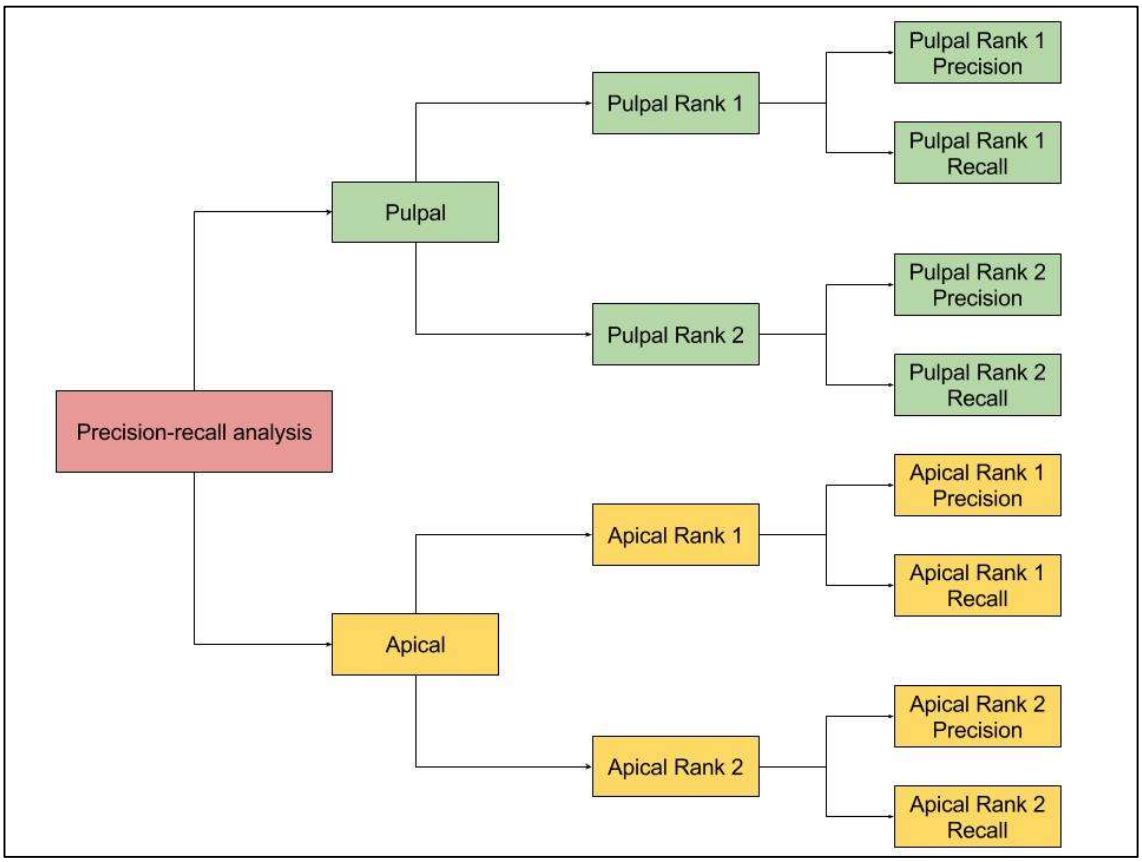


Figure 6-28 - Precision-recall analysis outline

6.9.1.1 Pulpal Rank 1

6.9.1.1.1 Pulpal Rank 1 Precision

In this section the Pulpal Rank 1 Precision of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Pulpal Rank 1 Precision (Table 6-46)(Figure 6-29).

Pulpal Rank 1	*Average Precision*
Rank 1-CC	0.44
Rank 1-CC+EO	0.15
Rank 1-CC+EO+IO	0.47
Rank 1-CC+EO+IO+CT	0.74
Rank 1- Final	1.00
Rank 1-Dentists (Final)	0.57

Table 6-46 - Pulpal Rank 1 Precision

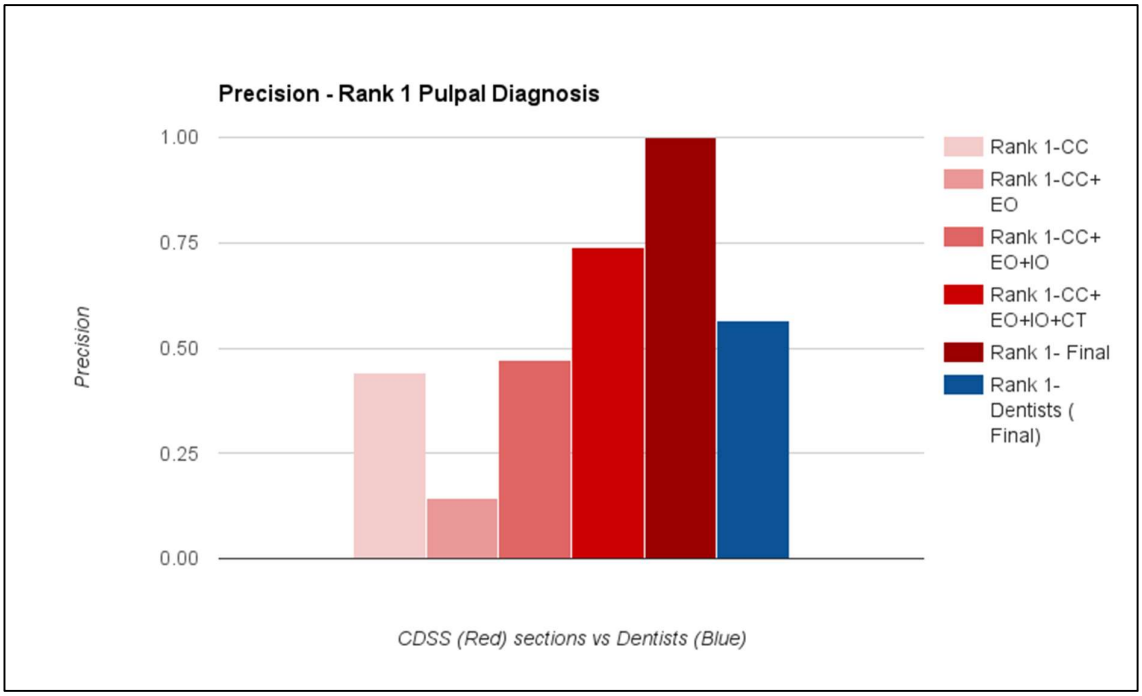


Figure 6-29 - Pulpal Rank 1 Precision

Precision Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences* ($p \leq 0.05$)
Pulpal Rank 1-CC-Precision vs Pulpal Rank 1-Dentists-Precision	-0.6250743 0.3600415	W = 18.5, p-value = 0.4808	✗
Pulpal Rank 1-CC+EO-Precision vs Pulpal Rank 1-Dentists-Precision (Dentists better than CDSS)	-0.7181599 -0.1400275	W = 4, p-value = 0.009764	✓
Pulpal Rank 1-CC+EO+IO-Precision vs Pulpal Rank 1-Dentists-Precision	-0.599953 0.3600594	W = 21, p-value = 0.7005	✗
Pulpal Rank 1-CC+EO+IO+CT-Precision vs Pulpal 1 Rank 1-Dentists-Precision	-0.3066343 0.6333031	W = 34, p-value = 0.2449	✗
Pulpal Rank 1-DS Model (Final)-Precision vs Pulpal Rank 1-Dentists-Precision (CDSS better than dentists)	0.1249891 0.6522447	W = 49, p-value = 0.001058	✓

Table 6-47 - Pulpal Rank 1 Precision Comparison

The average precision of the CDSS is different sections of the clinical encounter has been presented in Table 6-46. The same results have been presented in a bar chart above (Figure 6-29). The precision of the CDSS in the CC+EO section shows a reduction in performance. The dentist's final diagnosis is

significantly better than the CDSS at this section (Table 6-47) according to the Wilcoxon rank sum test. The Pulpal diagnostic performance improves as more information is collected about the patient and eventually surpasses that of the dentists in the final section.

6.9.1.1.2 Pulpal Rank 1 Recall

In this section the Pulpal Rank 1 Recall of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Pulpal Rank 1 Recall (Table 6-48) (Figure 6-30).

Pulpal Rank 1	*Average Recall*
Rank 1-CC	0.52
Rank 1-CC+EO	0.43
Rank 1-CC+EO+IO	0.64
Rank 1-CC+EO+IO+CT	0.77
Rank 1- Final	1.00
Rank 1-Dentists (Final)	0.66

Table 6-48 - Pulpal Rank 1 Recall

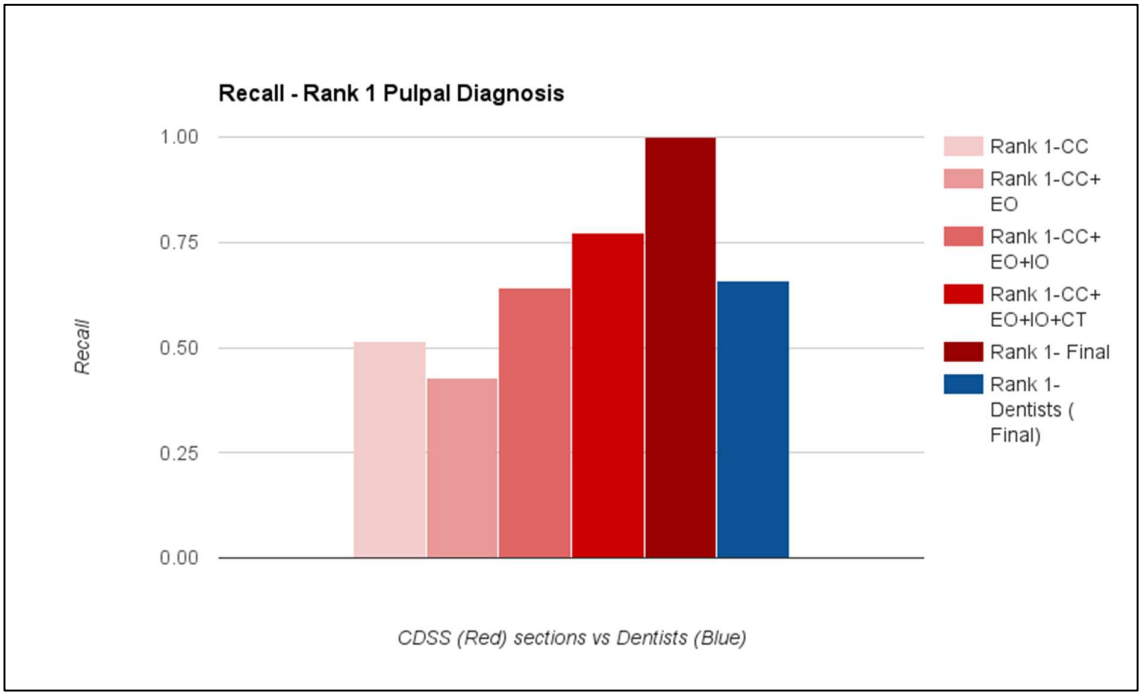


Figure 6-30 - Pulpal Rank 1 Recall

Recall Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences* ($p \leq 0.05$)
Pulpal Rank 1-CC-Recall vs Pulpal Rank 1-Dentists-Recall	-0.6666848 0.4999584	W = 20.5, p-value = 0.6466	✗
Pulpal Rank 1-CC+EO-Recall vs Pulpal Rank 1-Dentists-Recall	-0.727289 0.4999101	W = 18, p-value = 0.4276	✗
Pulpal Rank 1-CC+EO+IO-Recall vs	-0.5000814	W = 26.5, p-value = 0.8415	✗

Pulpal Rank 1-Dentists-Recall	0.5000537		
Pulpal Rank 1-CC+EO+IO+CT-Recall vs Pulpal 1 Rank 1-Dentists-Recall	-0.3333291 0.5666035	W = 31.5, p-value = 0.3868	✗
Pulpal Rank 1-DS Model (Final)-Recall vs Pulpal Rank 1-Dentists-Recall (CDSS better than dentists)	0.0000000 0.6250368	W = 42, p-value = 0.01136	✓

Table 6-49 - Pulpal Rank 1 Recall Comparison

The average Rank 1 recall of the CDSS across all sections showed similar reduction in recall performance of the CDSS however overall reduction in recall was not statistically significant (Table 6-49). The Rank 1 recall performance of the DS Model (Final) was significantly better than the dentists. The results show that the CDSS requires additional information in all sections except the final section to provide a Rank 1 diagnostic performance better than the dentists.

6.9.1.2 Pulpal Rank 2

6.9.1.2.1 Pulpal Rank 2 Precision

In this section the Pulpal Rank 2 Precision of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Pulpal Rank 2 Precision (Table 6-50)(Figure 6-31).

Rank 2	*Average Precision*
Rank 2-CC	0.14
Rank 2-CC+EO	0.15
Rank 2-CC+EO+IO	0.17
Rank 2-CC+EO+IO+CT	0.51
Rank 2-Final	0.41
Rank 2-Dentists (Final)	0.35

Table 6-50 - Pulpal Rank 2 Precision

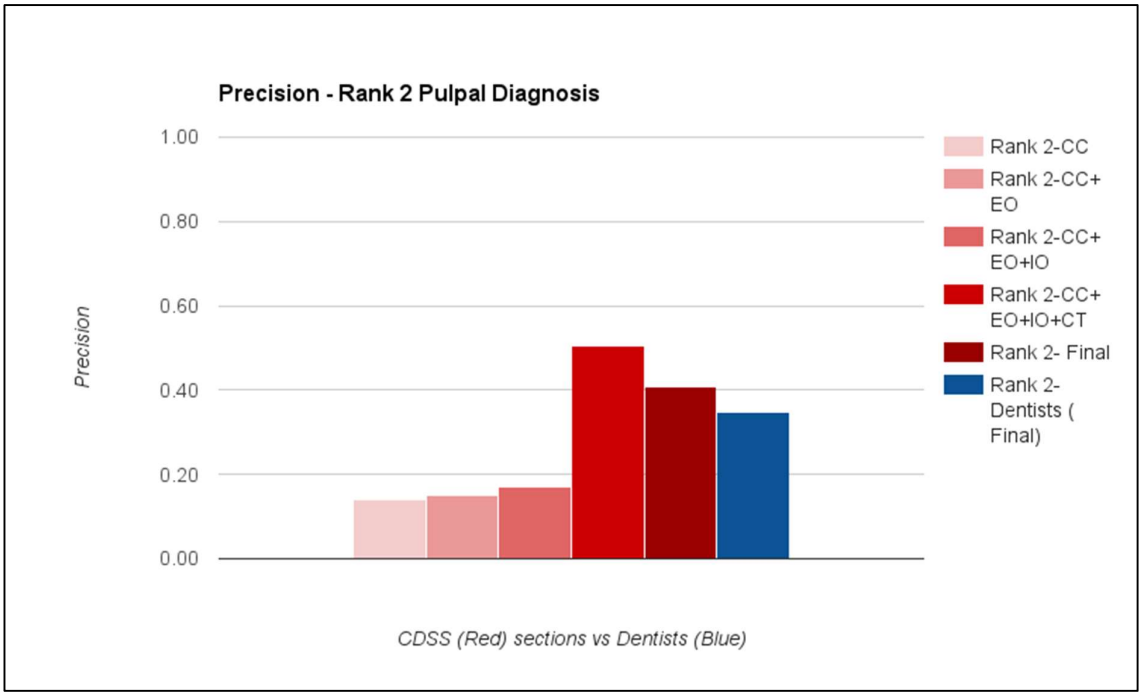


Figure 6-31 - Pulpal Rank 2 Precision

Precision Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
Pulpal Rank 2-CC-Precision vs Pulpal Rank 2-Dentists-Precision	-0.6824264 0.1999634	W = 18, p-value = 0.4124	✗
Pulpal Rank 2-CC+EO-Precision vs Pulpal Rank 2-Dentists-Precision	-0.7394827 0.1429098	W = 17, p-value = 0.3395	✗
Pulpal Rank 2-CC+EO+IO-Precision vs Pulpal Rank 2-Dentists-Precision	-0.3823533 0.5000075	W = 29, p-value = 0.6003	✗

Pulpal Rank 2-CC+EO+IO+CT-Precision vs Pulpal 1 Rank 2-Dentists-Precision	-0.6249563 0.9999360	W = 29.5, p-value = 0.5476	✗
Pulpal Rank 2-DS Model (Final)-Precision vs Pulpal Rank 2-Dentists-Precision	-0.6249411 0.6666719	W = 27.5, p-value = 0.7392	✗

Table 6-51 - Pulpal Rank 2 Precision Comparison

The Pulpal Rank 2 precision of the CDSS was poor across sections of the information model. The CDSS did not perform better or worse when compared to the dentists (Table 6-51). The results show that the CDSS requires additional information in all sections to improve its Pulpal Rank 2 precision.

6.9.1.2.2 Pulpal Rank 2 Recall

In this section the Pulpal Rank 2 Recall of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Pulpal Rank 2 Recall (Table 6-52)(Figure 6-32).

Rank 2	*Average Recall*
Rank 2-CC	0.28
Rank 2-CC+EO	0.28
Rank 2-CC+EO+IO	0.22
Rank 2-CC+EO+IO+CT	0.43
Rank 2-Final	0.43
Rank 2-Dentists (Final)	0.20

Table 6-52 - Pulpal Rank 2 Recall

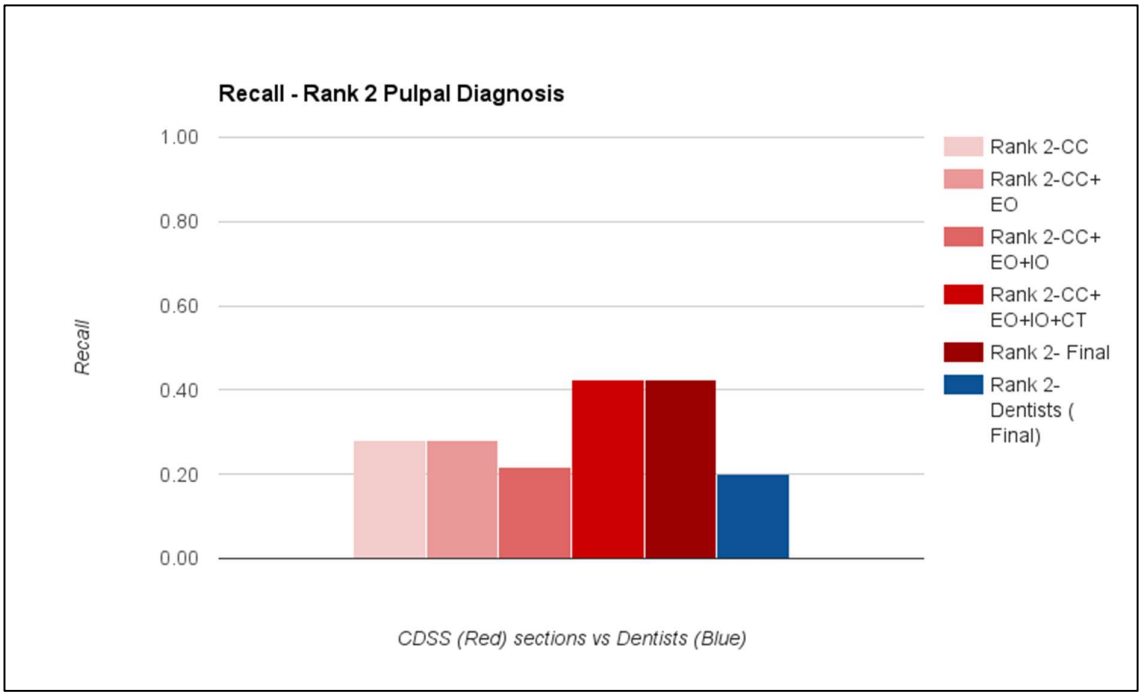


Figure 6-32 - Pulpal Rank 2 Recall

Recall Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences* *(p≤ 0.05)*

Pulpal Rank 2-CC-Recall vs Pulpal Rank 2-Dentists-Recall	-0.3749276 0.6554494	W = 23, p-value = 0.8915	✗
Pulpal Rank 2-CC+EO-Recall vs Pulpal Rank 2-Dentists-Recall	-0.3749276 0.6554494	W = 23, p-value = 0.8915	✗
Pulpal Rank 2-CC+EO+IO-Recall vs Pulpal Rank 2-Dentists-Recall	-0.05553853 0.99992293	W = 37, p-value = 0.1126	✗
Pulpal Rank 2-CC+EO+IO+CT-Recall vs Pulpal 1 Rank 2-Dentists-Recall	-0.2500661 0.8749952	W = 28.5, p-value = 0.6412	✗
Pulpal Rank 2-DS Model (Final)-Recall vs Pulpal Rank 2-Dentists-Recall	-0.2500661 0.8749952	W = 28.5, p-value = 0.6412	✗

Table 6-53 - Pulpal Rank 2 Recall Comparison

Like the Rank 2 precision results, the Pulpal Rank 2 Recall results were poor across all sections. The CDSS did not perform better or worse when compared to the dentists (Table 6-53). The results show that the CDSS requires additional information in all sections to improve its Pulpal Rank 2 recall.

6.9.1.3 Apical Rank 1

6.9.1.3.1 Apical Rank 1 Precision

In this section the Apical Rank 1 Precision of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Apical Rank 1 Precision (Table 6-54)(Figure 6-33).

Apical Rank 1	*Average Precision*
Rank 1-CC	0.07
Rank 1-CC+EO	0.27
Rank 1-CC+EO+IO	0.28
Rank 1-CC+EO+IO+CT	0.56
Rank 1- Final	1.00
Rank 1-Dentists (Final)	0.67

Table 6-54 - Apical Rank 1 Precision

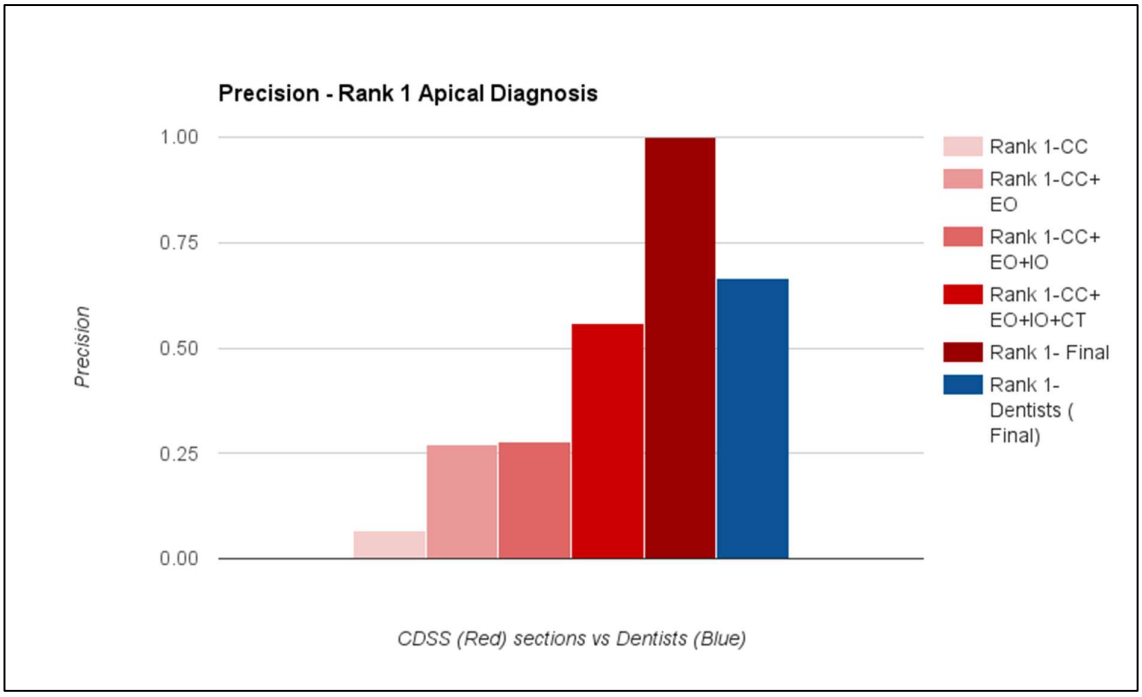


Figure 6-33 - Apical Rank 1 Precision

Precision Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
Apical Rank 1-CC-Precision vs Apical Rank 1-Dentists-Precision (Dentists better than CDSS)	-0.7619696 0.4614652	W = 0, p-value = 0.004337	✓
Apical Rank 1-CC+EO-Precision vs Apical Rank 1-Dentists-Precision	-0.7618905 0.2363782	W = 6, p-value = 0.06367	✗
Apical Rank 1-CC+EO+IO-Precision vs	-0.7618905 0.2363782	W = 6, p-value = 0.06367	✗

Apical Rank 1-Dentists-Precision			
Apical Rank 1-CC+EO+IO+CT-Precision vs Apical 1 Rank 1-Dentists-Precision	-0.5000110 0.3845922	W = 13.5, p-value = 0.5204	✗
Apical Rank 1-DS Model (Final)-Precision vs Apical Rank 1-Dentists-Precision (CDSS better than dentists)	0.2380685 0.4000415	W = 36, p-value = 0.002778	✓

Table 6-55 - Apical Rank 1 Precision Comparison

The Apical Rank 1 precision of the CDSS was poor in the early sections of the information model. It performed significantly worse than the dentists in the CC section. However, the precision improves as more information is collected and eventually when the data from radiology tests (DS Model-Final) section is obtained the CDSS performs better than the dentists (Table 6-55).

6.9.1.3.2 Apical Rank 1 Recall

In this section the Apical Rank 1 Recall of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Apical Rank 1 Recall (Table 6-56)(Figure 6-34).

Apical Rank 1	*Average Recall*
Rank 1-CC	0.33
Rank 1-CC+EO	0.50
Rank 1-CC+EO+IO	0.50
Rank 1-CC+EO+IO+CT	0.83
Rank 1- Final	1.00
Rank 1-Dentists (Final)	0.64

Table 6-56 - Apical Rank 1 Recall

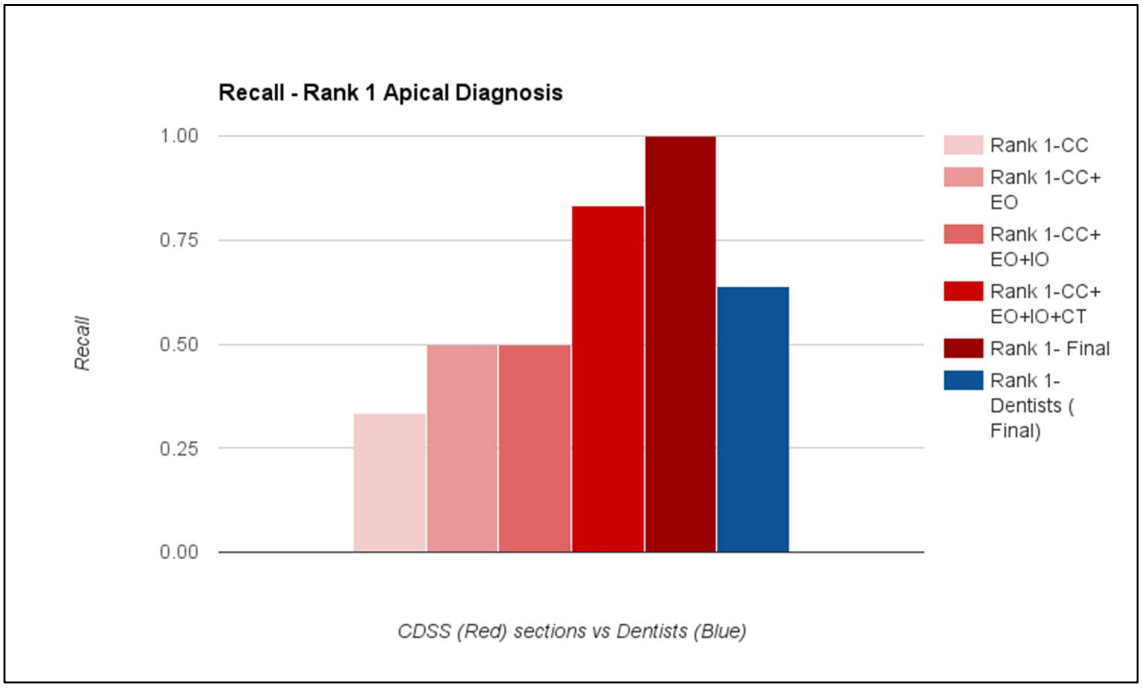


Figure 6-34 - Apical Rank 1 Recall

Recall Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences* *(p≤ 0.05)*
Apical Rank 1-CC-Recall vs Apical Rank 1-Dentists-Recall	-0.8421744 0.3749297	W = 12, p-value = 0.3691	✗
Apical Rank 1-CC+EO-Recall vs Apical Rank 1-Dentists-Recall	-0.8420758 0.3845539	W = 18, p-value = 1	✗
Apical Rank 1-CC+EO+IO-Recall vs Apical Rank 1-Dentists-Recall	-0.8420758 0.3845539	W = 18, p-value = 1	✗
Apical Rank 1-CC+EO+IO+CT-Recall vs Apical 1 Rank 1-Dentists-Recall	-0.2142694 0.3846181	W = 30, p-value = 0.0562	✗
Apical Rank 1-DS Model (Final)-Recall vs Apical Rank 1-Dentists-Recall (CDSS better than dentists)	0.1578436 0.3846157	W = 36, p-value = 0.002778	✓

Table 6-57 - Apical Rank 1 Recall Comparison

The Apical Rank 1 recall performance of the CDSS was poor in the early sections of the information model. The CDSS performed better as more information was collection and when the Clinical Tests section was complete the CDSS outperformed the dentists in Apical Rank 1 recall (Table 6-57).

6.9.1.4 Apical Rank 2

6.9.1.4.1 Apical Rank 2 Precision

In this section the Apical Rank 2 Precision of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Apical Rank 2 Precision (Table 6-58)(Figure 6-35).

Apical Rank 1	*Average Precision*
Rank 2-CC	0.00
Rank 2-CC+EO	0.00
Rank 2-CC+EO+IO	0.00
Rank 2-CC+EO+IO+CT	0.21
Rank 2-Final	0.38
Rank 2-Dentists (Final)	0.34

Table 6-58 - Apical Rank 2 Precision

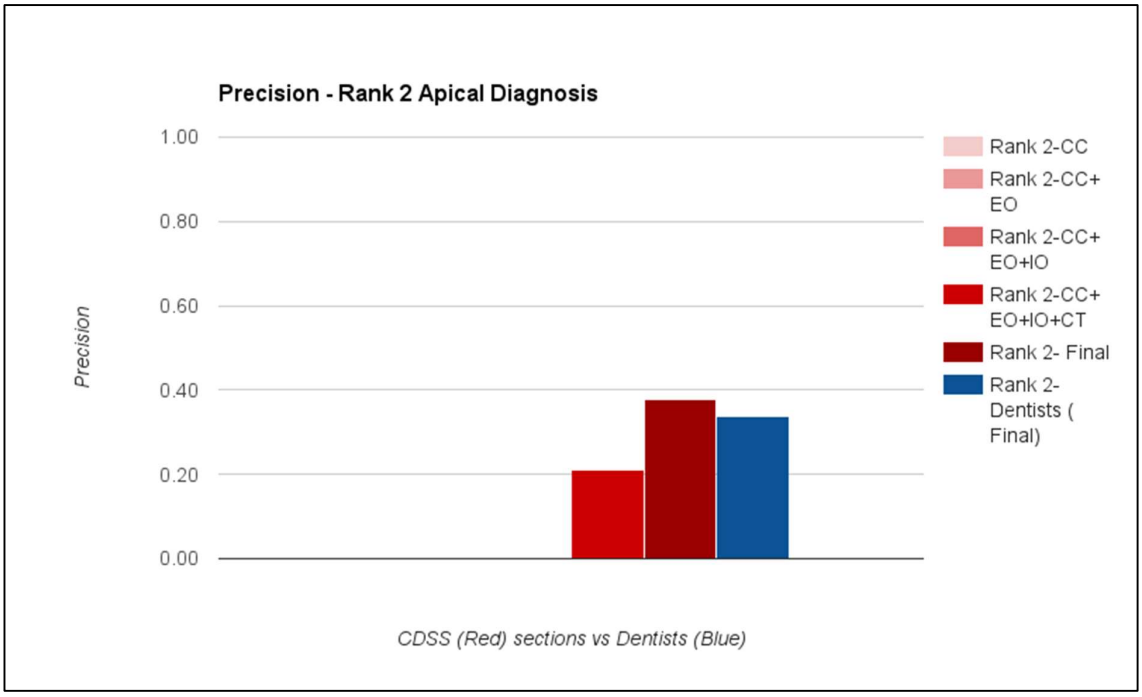


Figure 6-35 - Apical Rank 2 Precision

Precision Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences ($p \leq 0.05$)*
Apical Rank 2-CC-Precision vs Apical Rank 2-Dentists-Precision (Dentists better than CDSS)	-0.4737274 -0.2142115	W = 3, p-value = 0.009465	✓
Apical Rank 2-CC+EO-Precision vs Apical Rank 2-Dentists-Precision (Dentists better than CDSS)	-0.4737274 -0.2142115	W = 3, p-value = 0.009465	✓

Apical Rank 2-CC+EO+IO-Precision vs Apical Rank 2-Dentists-Precision (Dentists better than CDSS)	-0.4737274 -0.2142115	W = 3, p-value = 0.009465	✓
Apical Rank 2-CC+EO+IO+CT-Precision vs Apical 1 Rank 2-Dentists-Precision	-0.4737096 0.2666313	W = 14, p-value = 0.5604	✗
Apical Rank 2-DS Model (Final)-Precision vs Apical Rank 2-Dentists-Precision	-0.4736478 0.5999802	W = 19.5, p-value = 0.8703	✗

Table 6-59 - Apical Rank 2 Precision Comparison

The Apical Rank 2 precision of the CDSS was very poor in the early sections and the dentists performed better than the CDSS. These results are like the results obtained after analysis of Pulpal Rank 2 precision (Table 6-59). The results show that the CDSS requires additional information in all sections to improve its Apical Rank 2 precision.

6.9.1.4.2 Apical Rank 2 Recall

In this section the Apical Rank 2 Recall of the CDSS across all sections CC, CC+EO, CC+EO+IO, CC+EO+IO+CT and final is compared to the dentist's Apical Rank 2 Recall (Table 6-60)(Figure 6-36).

Apical Rank 1	*Average Recall*
Rank 2-CC	0.00
Rank 2-CC+EO	0.00
Rank 2-CC+EO+IO	0.00
Rank 2-CC+EO+IO+CT	0.33
Rank 2-Final	0.46
Rank 2-Dentists (Final)	0.34

Table 6-60 - Apical Rank 2 Recall

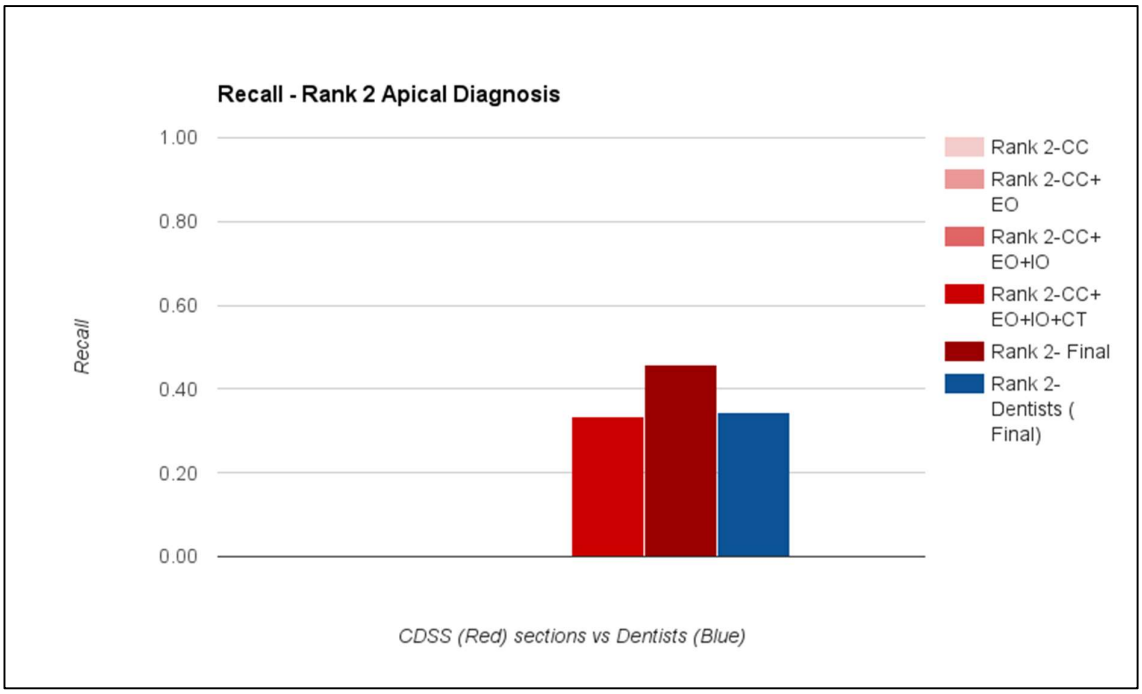


Figure 6-36 - Apical Rank 2 Recall

Recall Comparison	*95 % Confidence Interval*	*Wilcoxson rank sum test with continuity correction*	*Significant Differences* *(p≤ 0.05)*
Apical Rank 2-CC-Recall vs Apical Rank 2-Dentists-Recall (Dentists better than CDSS)	-0.5000438 -0.1999967	W = 3, p-value = 0.009465	✓
Apical Rank 2-CC+EO-Recall vs Apical Rank 2-Dentists-Recall (Dentists better than CDSS)	-0.5000438 -0.1999967	W = 3, p-value = 0.009465	✓
Apical Rank 2-CC+EO+IO-Recall vs Apical Rank 2-Dentists-Recall (Dentists better than CDSS)	-0.5000438 -0.1999967	W = 3, p-value = 0.009465	✓
Apical Rank 2-CC+EO+IO+CT-Recall vs Apical 1 Rank 2-Dentists-Recall	-0.5000407 0.7691873	W = 14, p-value = 0.5597	✗
Apical Rank 2-DS Model (Final)-Recall vs Apical Rank 2-Dentists-Recall	-0.5000660 0.7692014	W = 19.5, p-value = 0.87	✗

Table 6-61 - Apical Rank 2 Recall Comparison

The results of Apical Rank 2 recall show that the CDSS performed poorly in the early sections of the information model and improves after the Clinical Tests data is obtained. Even then the CDSS does not perform better than the dentists at recall of Apical Rank 2 diagnoses. The results show that the CDSS requires additional information in all sections to improve its Apical Rank 2 recall.

This section provides the results of the Precision-Recall analysis. The next section discusses the results of the NDCG and Precision-Recall analysis.

6.10 Discussion

This chapter details the evaluation of the DS Model and comparison of the results with primary care dentists. An implementation of the DS Model in dentistry was used in the evaluation. Two different evaluation metrics were used for evaluating the diagnostic performance of the model: NDCG and Precision-Recall. The dental encounter was split into different stages based on the different sections of the DIM dental information model. Each section of the information model represented various stages of the clinical encounter through which the dentist will traverse while collecting information from the patient. The DS Model was evaluated at each stage and compared to the final diagnostic performance of the dentists.

6.10.1 CDSS performance in the late stages

The results of Pulpal NCDG analysis show that for obtaining ranked recommendations the CDSS performs better than the dentists in the late stages of the clinical encounter (CC+EO+IO+CT and Final). The results of the Apical NDCG tests comparing the results with dentists were inconclusive. The t-test showed that the CDSS in the Final stage performed significantly better than the dentists (p-value = 5.539e-05). However, the Wilcoxon rank sum test does not show that the CDSS is better than the dentists.

As more information is obtained the CDSS performs better than the dentists. In the late stages of the clinical encounter the Clinical Tests and Radiographic Examinations are performed. These tests confirm the presence or absence of disease. In an ideal clinical setting where every clinical test and radiograph can be taken, and an accurate picture of the patient's condition can be obtained the CDSS using the DS Model could potentially help the dentists improve their final diagnosis.

Similarly, the DS Model is better than the dentists at the late stages and it is possible in certain clinical conditions it could help improve the final diagnosis of dentists and improve the diagnostic performance in the late stages especially in Pulpal diagnosis. However, further study is needed in this area to see if the ability of a CDSS to perform better than clinicians has an impact on their final diagnosis.

The results of the evaluation of the DS Model shows that the DS Model does not need more information in the late stages (CC+EO+IO+CT and Final) to improve the diagnostic performance. All the information collected in the previous sections when combined with the clinical tests should provide enough information for the CDSS to make an accurate ranked list for the patient. The results of the Precision-Recall evaluation show that the CDSS in the final stage is significantly better than the dentists in providing a Rank 1 diagnostic recommendation for both Pulpal and Apical diagnosis. The Rank 2 Pulpal and Apical diagnosis was poor in the late stages. However, it did perform as well as the dentists in the late stages.

The results of the Precision-Recall analysis support the findings of the NDCG analysis. The CDSS does not need more information in the late stages from external sources to provide a Rank 1 Pulpal and Apical diagnostic recommendation that is better than the dentists. Using semantic web and linked data technologies to link to other datasets to find more information to improve the diagnosis at this stage will not have a major impact in improving the diagnostic performance of the CDSS. Information around a concept can still be retrieved to support the clinician. For example, the dentist could be provided with the latest guideline specific information about a diagnostic test or the availability of new test. As the results in the Chapter 4 study with optometrists show, this could potentially help the dentist in choosing the appropriate test. There is potential savings in time and effort if clinicians perform the right test and patients are not subjected to unnecessary tests. The results of the evaluation of the DS Model shows that there is little scope additional improvement in diagnostic performance in the late stages by obtaining

external information about the patient. There is a need for integration and semantic interoperability. If some test results are missing from the patient's EHR then linked data and other technologies such as FHIR API's can be used to obtain a historical record about the patient. In this study all the information needed to decide was provided to both the dentists and the CDSS. In real-life situations in the presence of incomplete information a decision support tool that can integrate information can be valuable even in the late stages.

6.10.2 CDSS performance in the early stages

One of the key findings in the NDCG and Precision-Recall analysis in the early stages (CC, CC+EO, CC+EO+IO) is the presence of the dip in diagnostic performance especially in Pulpal diagnosis. This dip, which has been called the 'Intermediate dip' in this thesis, can be noticed in a side-by-side comparison of NDCG scores of each section (Figure 6-24). The t-test showed that the NDCG in the CC section of Pulpal diagnosis was statistically better than the dentists. However, the NDCG scores shows a dip in NDCG in the subsequent sections (CC+EO and CC+EO+IO). The scores rise when the Clinical Tests results are obtained and eventually becomes better than the dentists. The Wilcoxon significance test showed that the CC section NDCG score was not statistically better than the dentists (Table 6-37). Nevertheless, the intermediate dip can be seen in the chart (Figure 6-26).

Similar intermediate dip in diagnostic performance can be seen in the Precision-Recall analysis of Pulpal Rank 1 (Figure 6-29). The intermediate dip is more pronounced in the Precision-Recall graph for Pulpal Rank 1 when compared to the NDCG graph. In fact, the precision of Pulpal Rank 1 for the CC+EO is so marked that it poorer than the final Pulpal Rank 1 for dentists. There is a similar dip in recall of Pulpal Rank 1 diagnosis as well. However, the recall in the CC+EO section is not significantly worse than the dentists.

The intermediate dip is significant for several reasons. Conventionally the CDSS is expected to exhibit certain distinct behaviour. For example, it is expected that the diagnostic performance will be poor in the beginning stages and then gradually improve as more information is collected from the patient. If this pattern was constant across all CDSS inference models we can try various techniques to improve the performance. The CDSS can be linked to other data sources and the information model can be adapted accordingly to improve the performance until the results are satisfactory. However, the results of the study show that in the Pulpal diagnosis there is a greater need for improvement in the intermediate sections (CC+EO and CC+EO+IO). The intermediate dip can be partly explained by the weights adjusting accordingly as more information is obtained. The information collected in the CC section include signs and symptoms reported by the patient. Some signs and symptoms reported by the patient are indicative of certain diagnoses and therefore they will be ranked higher in the diagnostic list. However, as more information is collected in the subsequent sections the diagnoses that were ranked higher will go down the list, though not to the bottom of the list. This is corrected in the late stages (CC+EO+IO+CT and Final). This shallow dip can be observed in the NDCG graph for pulpal diagnosis (Figure 6-26). Since NDCG analysis looks at the ranking of the whole list and penalizes the score of the CDSS for incorrectly placing a correct diagnosis, the scores for Pulpal NDCG suggest that overall the relevant diagnosis is still ranked very high in the list. However, when we look at Rank 1 for Pulpal diagnosis in isolation we find that the fall in precision and recall scores are marked in the early stages and finally improves in the late stages.

The NDCG and precision-recall for Apical diagnosis did not show an intermediate dip in the early stages (Figure 6-27). Diagnostic performance of the CDSS in Apical diagnosis shows a gradual improvement across the stages as more information is collected. Though it is difficult to say if the CDSS performs better than the dentists in Apical diagnosis as the NDCG results were inconclusive. The NDCG scores in the CC section is significantly poorer than the dentists (according to both t-test and Wilcoxon test). Therefore, there is a definite need for more information in this section to improve diagnosis. These results are supported by the precision-recall analysis of Rank 1 Apical diagnosis as well. The precision in the CC section is significantly poorer than dentists (Figure 6-33 and Table 6-55). The recall performance of the CDSS in the early stages for Rank 1 Apical diagnosis was not significantly worse than dentists,

nevertheless it was low in the early stages and improved as the CDSS traversed the following sections (Table 6-56 and Figure 6-34). There is a need to provide more information to improve the diagnostic performance in the early stages as well.

6.10.3 Rank 2 CDSS performance

The Rank 2 precision of the CDSS was not good for both Pulpal (Figure 6-31) and Apical (Figure 6-35) diagnosis. The Rank 2 precision results were particularly bad in the early stages (CC, CC+EO, CC+EO+IO). Similar results were observed for the Rank 2 recall results for both Pulpal (Figure 6-32) and Apical (Figure 6-36) diagnosis. The CDSS cannot be recommended to make safe and reliable diagnostic recommendations for Rank 2 diagnosis. The CDSS did not perform better or worse than dentists in Rank 2 Pulpal diagnosis. Further study needs to be done in this area of ranking diagnostic recommendations. Clinicians are generally trained to produce a differential diagnosis list that includes several diagnoses that they should consider (Wears, 2009). Though little research has been done to compare the ranking produced by the clinicians with the ranking produced by CDSS applications. Some diagnostic CDSS applications such as DXPlain® produce a ranked list for differential diagnosis based on how strongly a finding suggests the presence of a disease and the prevalence of the disease (Bond *et al.*, 2012a). In the literature, evaluation of most diagnostic CDSS applications centred around evaluating the final Rank 1 diagnosis and most CDSS applications perform really well in this aspect (Bouhaddou, Lambert and Morgan, 1995; Graber and Mathew, 2008b; Umscheid and Hanson, 2012). These results have been confirmed by the results in this thesis.

6.10.4 Ranking of the diagnoses and its impact on cognitive biases

The DS Model uses a simple weighting scheme that was derived from the weighting scheme used in the Chapter 4 Optometry study. In this weighting scheme if the presence of an observation supports a diagnosis then a weight of +1 was given to the argument. If the presence of an observation strongly supports a diagnosis, then a weight of +2 was given to the argument. Similarly, negative weights were given for arguments that negate a diagnosis. The weighting scheme is simple and can be modelled very easily by the domain expert. A simple summation method was used to sum the weights of the arguments to obtain an overall weight for the diagnosis. This simple method of summation allows the CDSS using the DS Model to generate diagnostic recommendations without the need for specialized software and algorithms for inferencing.

However, the use of this simple weighting scheme and summation method means that there is increased risk of several diagnoses sharing the same rank when there is limited information. In the early stages of the clinical encounter there is limited information. For example, in CC section of the information model if the patient indicates the presence of pain then it supports several diagnoses. If the diagnoses are generated by the DS Model using 1 data point, then several diagnoses will share the same rank. However, in real-life situations the dentists must collect more information in the CC section. This includes nature of pain, location of pain, history of pain etc. Once all this information is collected in the CC section then number of diagnoses that share the same ranking will reduce.

Failure to consider a diagnosis is a major cause of diagnostic error and therefore providing a ranked list helps the clinician in considering and prioritizing a diagnosis and order further diagnostic tests if needed. The analysis was conducted using NDCG under the assumption that clinicians are more likely to ignore diagnoses that are ranked lower down the list. Using NDCG we can measure how well the CDSS performs in ranking the diagnoses. If the correct diagnosis is ranked low, then the NDCG score will be low. NDCG takes the utility of diagnosis at each rank and penalizes it by a logarithmic reduction factor. As the diagnosis goes lower the ranking, the utility of the diagnosis reduces. NDCG is used to measure the usefulness of a ranking on a scale from 0 to 1. A NDCG score of 1 represents a perfect match to the ideal ranking of the diagnoses. The ideal Rank 1 and Rank 2 was obtained from the Endodontist.

The results of the study show that in the CC section the DS Model performs significantly better than the dentists according to the t-test in the Pulpal diagnosis. A NDCG score of 0.81 was obtained (Figure 6-24). This shows that after collecting enough information in the CC section the DS Model can give a diagnostic recommendation that is very close to that of the expert's (Endodontist) final diagnosis. If both the correct Rank 1 and Rank 2 diagnoses were ranked low in the ranking, NDCG score would be very low. Therefore, in real life situations the dentists will be given a ranked diagnostic recommendation by the DS Model with the correct Rank 1 and Rank 2 diagnoses ranked high in the ranking. A diagnostic ranking in the CC stage that accurately matches the Endodontist's ranking is impossible in the early stages. However, a ranking that is very close to the Endodontist's ranking will be very useful to the dentist. Even with the simple weighting scheme and summation method the DS Model can provide a ranked diagnostic recommendation that is very good in the early stages.

The DS Model uses argumentation as the basis for generating the diagnostic recommendations. The ranked recommendations have explanations for making the recommendation. The arguments for and against a diagnosis can be presented alongside the diagnosis. For example, if Reversible Pulpitis is ranked high in the ranking then arguments such as 'Patient complains of pain' and 'Pain is sharp and localized' can be presented alongside the diagnosis. When the dentist collects information according to the hypothetico-deductive model he/she generates a diagnostic hypothesis. At the end of the CC section the DS Model generates the recommendation which the dentist can compare to his/her own mental model of the patient. The ranked diagnostic recommendations could potentially help the dentist consider alternative diagnoses. Argumentation helps organise the information and provide context for the observations that are present and help link the observations to the diagnosis. Researchers have suggested providing alternatives and organised presentation of information can help reduce diagnostic error (Graber, Franklin and Gordon, 2005b; Croskerry, Singhal and Mamede, 2013a). Premature closure bias is the most common cognitive cause for diagnostic error. When the DS Model provides a ranked diagnostic recommendation that closely matches the experts, the dentist can look at the top ranked recommendations and plan the next course of action. If the ranked recommendations were not very good, it could potentially lead the dentist to the wrong course of action. The dentist may spend more time collecting information via asking questions and performing tests that were not needed. Since the ranked diagnostic recommendations by the DS Model in the CC section are very good then the dentist can perform tests and asks questions that are more appropriate.

The Chapter 4 Optometry study demonstrated how guideline recommendations can help influence information gathering of optometrists. However, these guideline recommendations were not diagnosis specific as the diagnosis was generated at the end of the clinical encounter. The DS Model did not demonstrate how guideline recommendations can be generated. However, the DS Model can be adapted to provide guideline recommendations that are diagnosis specific. For example, if Reversible Pulpitis is ranked high then a plan for confirming this diagnosis can be presented early in the clinical encounter based on the limited data collected in the CC section. Guideline recommendations that explain the condition, provide anatomical/physiological information, provide tests that confirm or reject the diagnosis can be presented. Guideline recommendations could therefore potentially have a greater impact on information gathering of clinicians when combined with accurate early diagnostic recommendations. The impact of guideline recommendations in conjunction with early diagnostic recommendations need to be studied further.

Similarly, in the CC+EO section the overall Pulpal NDCG score is 0.77. The Pulpal NDCG score for CC+EO+IO is 0.81(Figure 6-24). The results show that the even though there is no significant different in the NDCG scores in these sections when compared to dentists in the study the NDCG scores are still very good. There is a small drop in the NDCG score in CC+EO section. This 'intermediate dip' is not significant as far as the NDCG scores are concerned. The Rank 1 and Rank 2 diagnoses are still high enough in the ranking to be clinical valuable. The drop in the score can be attributed to the CDSS obtaining more information in the CC+EO section. As more information is obtained the correct Rank 1

and Rank 2 diagnoses will move down the ranking generated by the CDSS. Therefore, the overall NDCG score will drop.

However, the results show that the overall NDCG score for Pulpal diagnosis in the CC+EO+IO section is 0.81 which means as more information is obtained the ranking adjust to provide a result that is close to the score in CC section (Figure 6-24). Confirmation bias can lead the clinician to look for evidence to support his/her diagnostic hypothesis rather than looking for evidence to reject. The changing ranking could potentially help in controlling confirmation bias. If the dentist has a mental model about the patient and if some other diagnosis is ranked higher then it forces the dentist to consider this alternate diagnosis. When there is no diagnostic CDSS recommendation especially in the early stages of the clinical encounter, there is no way of informing the dentist about the presence of another diagnosis that could potentially be of interest. The arguments inform the dentist why the alternate diagnosis must be explored.

25 vignettes were presented to the CDSS. Vignette number 10032 and 10038 will be used here to illustrate how the Rank 1 and Rank 2 diagnoses change with new information. In vignette number 10032 the Rank 1 Pulpal diagnosis provided by the Endodontist is Reversible Pulpitis and the Rank 2 Pulpal diagnosis is Dentinal Hypersensitivity. The Endodontist felt Reversible Pulpitis is the most likely diagnosis for this vignette and Dentinal Hypersensitivity is the next most likely diagnosis. The ranks for these 2 diagnoses change in the following manner as more data is collected. The corresponding NDCG scores obtained for this vignette are given in Table 6-62.

DIM Section	Reversible Pulpitis Rank	Dentinal Hypersensitivity Rank	NDCG score
CC	2	1	1
CC+EO	2	1	1
CC+EO+IO	1	2	1
CC+EO+IO+CT	1	2	1
DS Model Final	1	3	0.9

Table 6-62 - NDCG and ranking for 10032

The data for 10032 shows that ranking closely matches that of the Endodontist across all sections. In the Final section the Dentinal Hypersensitivity rank falls to Rank 3. However, the Rank 1 diagnosis is still Reversible Pulpitis. As more information is obtained the ranks adjust to reflect the correct diagnosis.

In vignette 10038 the Rank 1 Pulpal diagnosis is Pulp Necrosis and Rank 2 Pulpal diagnosis is Asymptomatic Irreversible Pulpitis. The NDCG score and ranking can be seen in Table 6-63.

DIM Section	Pulp Necrosis Rank	Asymptomatic Irreversible Pulpitis Rank	NDCG score
CC	1	2	1
CC+EO	7	2	0.7
CC+EO+IO	6	1	0.5
CC+EO+IO+CT	1	2	1
DS Model Final	1	2	1

Table 6-63 - NDCG and ranking for 10038

The data for 10038 shows how the ranking change as more information is collected. The Pulp Necrosis rank is 1 in the CC section. It falls drastically in the next 2 sections and then finally regains the top spot after the clinical test results are obtained. The intermediate dip is more pronounced in this vignette. The

reason for the intermediate dip can be explained with the type of questions asked and examination done in the early stages of the clinical encounter. The chief complaint and the nature of the complaint is highly indicative of certain diagnoses in certain cases. However, the observations that support Pulp Necrosis in extra-oral examination and intra-oral examination can be indicative of other diagnoses. Therefore, the rank falls drastically for Pulp Necrosis. However, the Rank 2 diagnosis is ranked high through all sections of the encounter. The arguments for a diagnosis can guide the dentist and explain why the diagnosis was ranked high. So, the chances of the CDSS negatively influencing the diagnostic performance are small and restricted to certain types of cases. In real life settings, the dentists will have to be informed that the early diagnostic recommendation is tentative and not to considered final and the diagnoses rankings are to be used in conjunction with the arguments and guideline recommendation (if available) as a guide to help in clinical test selection and interpretation of results. Overall, the NDCG scores for Pulpal diagnosis by the DS Model remains very high. However, the effect of changing ranking on premature closure and other cognitive biases of dentists and other primary care clinicians need to be studied further.

The Apical diagnosis overall NDCG scores show that it is 0.64 in the CC section. However, the NCDG scores adjust itself to 0.80 in CC+EO section as more information is obtained and remains high throughout (Figure 6-25). The reason for the low NDCG score in the CC section could be explained by the type of questions asked in that section. In the CC section the questions are oriented towards obtaining information about the nature of pain and location of pain. Apical diagnoses are characterised by the lack of pain in the early stages of pathology and in some cases, there is no pain at all. The Precision and Recall tests support the NDCG scores but show a more marked variation in the different stages (See section 6.9). The DS Model may not able to get the ranking perfect in the early stages but the NDCG scores show that the correct diagnoses are ranked high enough in the overall ranking.

The results show that even with the simple weighting scheme and simple summation method, the DS Model can provide effective diagnostic recommendations. When combined with arguments the DS Model can help inform the dentist about alternative diagnoses and the reasons for recommending the diagnoses. Compared to other CDSS reasoning methods argumentation provides an easy method for knowledge representation and presenting the results in a manner that is easy to understand by the clinician. The late diagnostic recommendations closely match the Endodontist's rankings. The performance of the DS Model varies depending on the type of case and the clinical presentation of the case. DS Model has been shown to provide effective early diagnostic recommendation in the early sections with an intermediate dip in performance. The DS Model was able to adjust the ranking appropriately to provide a final diagnostic recommendation that closely matches the Endodontist and exceeds the performance of the dentists recruited in the study.

7 Chapter 7 - General conclusions and discussion

7.1 Introduction

In this chapter the main findings are summarised with regards to research aims, general conclusions alongside contributions of study with optometrists, DS Model development and evaluation of the model is presented. The chapter outlines the strengths and limitations of the thesis and suggestions for further research are presented.

7.1.1 Aims

The aim of this thesis is a) to investigate the role of clinical guideline and argumentation-based diagnostic recommendations on primary care diagnosis and information gathering; b) to investigate and develop an argumentation-based model that can use data from disparate sources to assist in providing diagnostic recommendations across all stages of the clinical encounter; c) to develop CDSS prototypes that can use the model to deliver the diagnostic recommendations; d) to evaluate the model at different stages of the clinical encounter to see how a CDSS implementing the model will perform and identify areas of deficient performance of the model.

Objective 1: To review the literature on models of diagnostic reasoning and identify models used by diagnostic CDSS to represent knowledge and formulate diagnostic recommendations and their role in reducing diagnostic error. Review state of the art web technologies to identify how they address the problem of semantic interoperability and their application in CDSS.

Objective 2: Develop a diagnostic CDSS based on current argumentation-based CDSS technology and evaluate its effect on primary care clinician decision making and information gathering.

Objective 3: Develop a semantically interoperable diagnostic inference model that can integrate information from disparate sources and uses argumentation to generate diagnostic recommendations.

Objective 4: Implement a working CDSS prototype that implements the semantically interoperable diagnostic inference model and provides diagnostic recommendations.

Objective 5: Evaluate the CDSS prototype and the diagnostic inference model and its ability to support diagnostic decision making at all stages of the clinical encounter.

Objective 1 was addressed in Chapter 2 – Literature review. The literature review outlined some of the cognitive causes of diagnostic errors (section - 2.1.1). Some of the key cognitive causes for diagnostic error include biases such as premature closure bias and confirmation bias (Graber, Franklin and Gordon, 2005b). According to the hypothetico-deductive theory (Elstein, Schwartz and Schwarz, 2002) the clinician uses initial information obtained from the patient to form a list of working diagnostic hypotheses and then obtains additional information to confirm or reject diagnostic hypotheses. Once all the information is obtained the clinician combines and weighs information from all sources to confirm one diagnosis from the list of working hypotheses. Illness script theory provides a framework that describes how clinicians organize knowledge obtained from clinical literature and experience (Schmidt and Rikers, 2007a). Based on observations seen in the patient certain illness scripts are triggered in the mind of the clinician. The clinician then compares the data from the patient to the default values in the script. If the default value does not match, then another illness script is triggered until the right script that closely

matches the patient's condition is found. The clinician is constantly obtaining new information from the patient via questions, examination, tests and investigations. The illness script explains how the clinician moves from one diagnostic hypothesis to another until they arrive at the right diagnosis.

The Dual Process Theory explains how individuals make decision and judgements under uncertainty with limited information. The dual process theory consists of 2 types of processes, System 1 and System 2 (Evans and Stanovich, 2013). System 1 is the process that is associated with pattern recognition. It is intuitive, efficient and fast and is the system used to make the most common decisions. It is mostly activated in cases with routine clinical presentations and is dependent on the clinician's experience. It is used by busy clinicians to deal with patients with familiar clinical presentations. However diagnostic errors are highly associated with atypical presentations (Kostopoulou, Delaney and Munro, 2008b). When the clinical presentation of the case is atypical, and the clinician is not familiar with the case the clinical reasoning automatically switches to System 2 process of reasoning. System 2 is the process that is associated with slow, methodical, analytical reasoning. This process is deliberate and is most commonly activated when the patient presents with an atypical clinical presentation. An atypical presentation forces the clinician to slow down and consider all probable causes for the patient's current condition. Medical students and clinicians with less experience tend to use System 2 for reasoning and are therefore slower and more methodical when trying to diagnose a patient. System 1 and System 2 reasoning work in tandem and support each other (Saposnik, Redelmeier, Christian C. Ruff, *et al.*, 2016).

In premature closure bias the clinician prematurely halts the diagnostic decision-making process before all the data has been collected. The clinician confirms a working hypothesis early and do not collect all the information needed to make an accurate diagnosis. The clinician finds some key abnormal values or observation which triggers an illness script if the illness script closely matches other data obtained from the patient. The System 1 process of reasoning is activated, and the diagnostic decision is made fast without considering other alternatives. Researchers have identified other causes for diagnostic errors including problems coordinating information from multiple sources and the clinician having inadequate knowledge about the patient's condition (Graber, Gordon and Franklin, 2002a; Graber, Franklin and Gordon, 2005b).

Researchers have proposed several strategies to reduce diagnostic errors dues to cognitive factors (Croskerry, 2003a). This includes providing alternative diagnosis, providing access to clear and well-organized information, providing information in a structured manner etc. Providing early diagnostic recommendations is another proposed recommendation to help reduce diagnostic errors (Kostopoulou *et al.*, 2015b). Clinical guidelines are another solution that could impact diagnostic decision making by standardising the process of collecting, analysing and verifying clinical data. Clinical guidelines improve consistency of care and can potentially improve healthcare efficiency. They can alert clinicians to diagnostic tests and procedures that have greatest evidence and provide greatest benefit. Clinical guidelines can be used to streamline clinical decision making and provide a knowledge base for guideline-based CDSSs and clinical alert systems (Garg *et al.*, 2005a)

The literature indicated the role played by CDSSs in reducing diagnostic error and the review outlines the history of diagnostic CDSSs and the different historical methods used for capturing diagnostic criteria to support CDSS applications (section - 2.2). Diagnostic CDSSs identified in literature such as differential diagnosis generators provide alternative diagnosis to help the clinician consider alternative diagnosis. However, these systems provide the diagnostic recommendations at the end of the clinical encounter after all the information has been collected. Premature closure bias occurs when the clinician stops the information gathering process prematurely and arrives at a conclusion. Confirmation bias can lead the clinician to look for evidence to support his/her diagnostic hypothesis, rather than looking for evidence to reject. Therefore, alternative diagnoses provided to the clinician early in the diagnostic encounter is

potentially more useful to the clinician. When the alternative diagnoses are provided early in the diagnostic encounter (early diagnostic recommendations), this could help the clinician by encouraging the clinician to consider other alternate diagnoses and could prevent the clinician from halting the information gathering process. The reasons for making the recommendation must be given in a clear and organised manner as recommended by researchers. The recommendations must be easy to understand as well. When this information is provided in a structured manner to the clinician it helps the clinician understand the patient's condition better and help choose tests and investigations that are more appropriate for the patient (Croskerry, 2003a). The reasons for the diagnostic recommendations can potentially help in reducing confirmation bias by forcing the System 2 process of reasoning. System 2 is slow, analytical and methodical process of reasoning. The reasons(arguments) for a diagnosis can help force the clinician to slow down and think about alternatives and not restrict the enquiry to information that fits his/her mental model.

Guideline-based CDSSs and the role played by these systems have been identified in the literature review. Guideline-based CDSSs have been shown to improve adherence to guidelines and help provide patient specific guideline recommendations. Guideline-based CDSSs have been used in a wide variety of medical applications mainly in areas such as treatment planning. PROforma is a language that uses Argumentation Theory as the basis for generating decision recommendations and PROforma-based CDSSs have been successfully evaluated and deployed in several areas of medicine to support decision making.

Argumentation uses non-monotonic logic for inference. In non-monotonic logic existing conclusions can be withdrawn and new conclusions can be formulated based on current information. Argumentation-based CDSSs can therefore support defeasible reasoning where humans similarly retract existing conclusions and form new conclusions in the light of current information (Fox, Krause and Elvang-Gøransson, 1993). In argumentation the knowledge needed to support decision making is expressed in the form of arguments. For each decision a set of positive and negative arguments are generated. These arguments are used to generate decision recommendations by CDSSs. Argumentation-based CDSSs have another advantage when it comes to explaining the reasons for generating the recommendation. In an argumentation-based diagnostic CDSS the reasons for making the diagnostic recommendation can be provided in the form of arguments (Fox, Johns and Rahmanzadeh, 1998). These arguments can be presented in natural language and in a format that is easy to understand by the clinician (Fox and Parsons, 1998). Capturing the domain expert's knowledge is easier using argumentation as it uses a qualitative approach using subjective probabilities to model arguments (Fox *et al.*, 2007). It is not reliant on accurate numerical probabilities and complex statistical data to generate diagnostic recommendations (Longo, Kane and Hederman, 2012). This could potentially result in loss of accuracy when generating diagnostic recommendations, however in situations like early diagnosis the information is incomplete and therefore not every CDSS reasoning method can generate a useful diagnostic recommendation. Bayesian networks can generate diagnostic recommendation under uncertainty. However, translating the domain expert's knowledge base into a causal model and a Bayesian network is not easy, especially in domains that are not well understood. Argumentation uses a similar reasoning process used by clinicians and the clinicians have found the decision recommendations easy to understand when compared to other tools (Ouerdane, Maudet and Tsoukiàs, 2010). The ability to operate under uncertainty and incomplete information, ability to capture domain expertise in a simple and organised manner, ability to explain the decision recommendations in a manner that can be understood easily by clinicians and non-monotonicity are some of the reasons why argumentation was considered as a suitable candidate for generating early diagnostic recommendations (Fox, Patkar and Thomson, 2006a). Argumentation-based CDSSs have been developed in several areas such as diagnosis, treatment planning, risk assessment etc. However, the impact and feasibility of an argumentation-based diagnostic CDSSs on diagnostic decision making of primary care

clinicians has not been studied. The impact of guideline recommendations on the information gathering behaviour of clinicians during diagnostic decision making needs to be studied further.

Current generation of diagnostic CDSSs such as differential diagnostic generators and argumentation-based CDSSs do not generate early diagnostic recommendations due to limitation of data available during the clinical encounter. The clinician and CDSS is restricted to data available during a clinical encounter. Therefore, the CDSS must wait till end of an encounter to provide an accurate diagnostic recommendation. However, to influence cognitive biases such as premature closure and confirmation bias the recommendations must be ideally provided early in the encounter in a ranked manner. The web is an excellent source of data that includes data sources such as PHRs and EHRs that can provide information about the patient that can be used for decision making. The literature outlined some of the problems with current CDSS applications and how web technologies such as semantic web and linked data can help improve CDSS applications (section - 2.4). The main problem with data on the web is the diverse formats that the data is presented in and therefore semantic interoperability is a main issue that prevents CDSSs from using data from the web. RDF, OWL and other semantic web technologies help improve semantic interoperability. RDF has been identified as a candidate for a universal healthcare language (Booth, 2014b, 2015c). RDF enables smarter data use and automated data translations. Multiple data models and vocabularies can be easily combined and interrelated using RDF. RDF captures the information present in data and not syntax. RDF is easy to map from one data representation to another and RDF is self-describing. All these properties make RDF a suitable candidate for building a diagnostic inference model that can integrate information from disparate sources. Integrating information will allow the CDSS and the clinician to use the information to generate diagnostic recommendations. Using OWL/RDF enables us to reuse existing ontologies and terminologies to develop the ontology model.

The chapter outlined the different standards and technologies used in semantic web and how these standards and technologies can be used in diagnostic CDSSs. The limitations of using ontologies in diagnostic CDSSs were outlined (section - 2.5). OWL ontologies have been used to represent diagnostic criteria and used in ontology-driven diagnostic CDSSs. Existing ontology-based models for diagnostic CDSSs have used several techniques to deal with the uncertainty in diagnosis. Some methods have used an external probabilistic system to generate probabilities and generate the ranked diagnostic recommendations (García-Crespo *et al.*, 2010a). Other methods include representing the diagnostic criteria using OWL axioms (Ongenae *et al.*, 2010; Donfack Guefack *et al.*, 2012; Bau, Chen and Huang, 2014; Sherimon and Krishnan, 2016). However due to limits in expressiveness of DL-based axioms these methods are not well suited for representing uncertainty and cannot work with incomplete data. Most ontology-based CDSSs identified in literature do not provide a ranked diagnostic recommendation. A ranked diagnostic recommendation is important in early diagnostic CDSSs as the clinician can use the ranking to guide next steps for gaining more information. The ontology-based CDSSs that provide a ranked recommendation rely on complex ranking algorithms and specialized software to generate the ranking (Oberkampf *et al.*, 2012). The diagnostic CDSSs do not explain the reasons for making the recommendation. Using linked data repositories as a source of clinical information for CDSSs has not been explored in other diagnostic CDSSs.

The thesis is made of 3 parts. In the first study an existing off-the-shelf argumentation-based CDSS was used to develop a diagnostic CDSS for optometry. The feasibility and impact of this diagnostic CDSS on primary care diagnostic decision making was studied. In addition, the impact of guideline recommendations on information gathering behaviour of optometrists during a diagnostic encounter and management decision making was studied.

The next study involved developing an ontology-based CDSS model called DS Model that uses argumentation as the basis for generating early diagnostic recommendation. The DS Model demonstrated

how arguments for and against a diagnosis can be represented in OWL language. The CDSS prototypes using the DS Model can integrate information from disparate sources in a wide variety of data representations using semantic web and linked data technologies.

The DS Model was then subsequently evaluated and its performance across all stages of a clinical encounter was tested.

7.2 Feasibility and impact of a guideline-driven argumentation-based diagnostic CDSS on diagnostic performance and information gathering behaviour of optometrists

This section details how Objective 2 was achieved. Diagnostic CDSSs have been developed using a wide variety of methods including case-based reasoning, machine learning, Bayesian probabilities, fuzzy logic etc. These methods require the availability of well-structured clinical data to generate diagnostic recommendations. These methods need accurate data to create the inference models. Moreover, the diagnostic recommendations are presented in a manner that is not always intuitive to the clinicians. Argumentation allows domain experts to capture their domain expertise in an easy manner using qualitative methods. The recommendations generated by an argumentation-based CDSS can be presented in a manner that is easy to understand by the clinician. Argumentation-based CDSSs can support defeasible inference and reasoning and can retract existing conclusions and form new conclusions in the light of current information using non-monotonic logic. These properties make argumentation as suitable method for generating early diagnostic recommendations. However, the impact of argumentation-based CDSSs on the diagnostic performance of clinicians, especially primary care clinicians have not been studied before. Guideline recommendations from guideline-based CDSSs have shown to improve guideline adherence and improve standards of care. Additionally, the study evaluates the impact of guideline recommendations in influencing information gathering behaviour of optometrists. Chapter 4 presented the findings of the study involving community optometrists and a guideline driven argumentation-based diagnostic CDSS.

The study presented guideline recommendations at different stage of the clinical encounter simulated with the help of clinical vignettes. The guideline recommendations were presented with the help of a guideline-based CDSS and recommendations derived from current clinical evidence. Argumentation-based diagnostic recommendations for the vignettes were presented and the optometrist's diagnosis and management recommendations were recorded. The results of the study with intervention group (optometrists with support of CDSS) were compared to the results of an earlier study involving community optometrists who were asked to complete the same vignettes but did not have the support of any CDSS recommendations.

The results of the study showed that the guideline recommendations had some impact on influencing the information gathering of optometrists. The guideline recommendations were particularly useful in cases where there were abnormal values and the guideline recommendations were able to give condition specific guidance to the optometrist. The guidance is only effective in non-routine questions. If guidance asks the optometrists to ask certain questions that were part of routine clinical history, then the impact of the guidance was limited. However, when guidance alerted the optometrist about abnormal values not routinely seen and showed guideline recommendations about tests or investigations needed then the optometrists responded accordingly. The guideline recommendations were effective in improving management decision making especially when prescribing repeat tests. Guideline recommendations alone did not show any improvement in diagnostic decision making of optometrists.

The results showed the diagnostic recommendations of the CDSS was helpful in improving the diagnostic performance in certain conditions. The conditions that the CDSS helped improve the diagnosis was sight-threatening in nature and therefore there is a potential for an argumentation-based CDSS to help reduce diagnostic error. The management guideline recommendations helped improve prescribing supplementary tests especially in sight threatening cases. The guideline recommendations can flag up abnormal values and help those optometrists who may not have adequate knowledge/experience of the case. Guideline recommendations help clinicians by providing guidance on how to interpret images and clinical test results. Guideline recommendations can provide guidance from the latest clinical guidelines and guide the clinician towards appropriate tests. It helps reduce reliance on memory as the clinician does not have to remember bulky guidelines and protocols. The argumentation-based diagnostic CDSS was able to provide a ranked diagnostic recommendation and explain why the recommendation was made.

One of the key findings of the study was the CDSS itself was dependent on the data that the clinicians entered as part of the data collection process. In conditions where the observations are similar, erroneous interpretation of clinical data (especially eye images) resulted in the CDSS recommending diagnoses that were incorrect and the clinician accepting the results of the CDSS recommendations. This resulted in the CDSS accidently confirming the hypothesis of the clinician and thereby the CDSS was unable to correct the clinician. Providing diagnostic recommendations based on faulty data could potentially lead to more confirmation bias related diagnostic errors. This can be partially remedied by having additional checks on data quality and providing additional decision support explanations to avoid misinterpretation of data. When the weights of diagnoses and arguments are similar then additional alerts can be given to inform the optometrist that there is scope for confirmation bias related diagnostic error and therefore the optometrist needs to recheck data. The consequences of diagnostic error can be provided alongside the diagnosis to educate the optometrists. This study was a pilot study that evaluated the feasibility of an argumentation-based diagnostic CDSS on primary care. Therefore, the CDSS was a prototype application that used limited number of arguments, mainly clinical test results and investigations to generate the diagnostic recommendations. Faulty recommendations can be further reduced by improving the quality and quantity of arguments providing more detailed reasons for recommending a diagnosis. Increased number of arguments as modelled in the DS Model can help the CDSS make better recommendations as there is more information to differentiate between diagnoses.

The results of this study confirm some of the findings from previous studies on diagnostic CDSS applications and EHR systems (Grace *et al.*, 2013) where the quality of data entered affects the outcome. One of the recommended best practices for CDSS applications is to maintain high quality and complete data (Wright, Phansalkar, Bloomrosen, Robert A. Jenders, *et al.*, 2010). The data quality and correctness problems in EHR systems are well known (Wagner and Hogan, 1996; Hogan and Wagner, 1997; Curtis, Bollard and Dickson, 2002; Majeed, Car and Sheikh, 2008).

This study showed that the opinion of the optometrists towards CDSS was generally positive and even with minimal training and instructions the optometrists did not find the CDSS difficult or cumbersome to use (See section - 4.3.7). Almost half of the respondents are very likely to use the CDSS in their clinical practice. A majority felt the system was easy to use and felt that most people would find system easy to use. Most respondents felt that the CDSS was easy to learn and they didn't need the help of another individual for using the system. More than half the respondents felt that the system helped improve their decision making and they are likely to use the CDSS in practice. These results confirm the findings from other argumentation-based CDSSs that showed that clinicians generally found the recommendations useful and easy to understand (Fox, Patkar and Thomson, 2006a). The optometrists found potential for its application in supporting novice optometrists.

The CDSS prototype used in this study was developed using PROforma language and Tallis modelling tools and inference engine. The CDSS used a simple weighting scheme for the arguments that either support or negate a diagnosis. The weighting scheme used +1 for arguments that supported a diagnosis and +2 for arguments that are highly indicative of a diagnosis. Likewise, negative weights were given for arguments that negated a diagnosis. The presence of observations/absence of observations were used as arguments. There were no published guidelines that recommended the weights to be used as there are limited studies of argumentation-based diagnostic CDSSs. Therefore, this weighting scheme was developed and used for the study. Developing argumentation-based CDSSs using PROforma is a knowledge-based approach and requires contribution from domain experts. Developing arguments and adding weights for arguments requires tools such as Tallis. Executing PROforma models require the Tallis engine. The weights and arguments can be developed using a consensus approach. Several clinicians can come together to create diagnoses candidates, models the arguments and weights for the arguments. Web-based platforms such as OpenClinical.net (OpenClinical.net, 2015) have been developed to clinician and domain experts to collaborate on developing these models (Fox *et al.*, 2015).

The main problem with PROforma and other argumentation-based models is the use of proprietary format for modelling. The models require specialized proprietary software for modelling and execution. Integrating data from disparate data sources and EHRs require a lot of development effort and interface engines that can translate the data into formats that PROforma and similar argumentation-based CDSSs can understand. Integrating data from remote sources and data in a wide variety of formats is important for getting the data needed to generate effective early diagnostic recommendations. Otherwise the CDSS is restricted to data from the diagnostic encounter. When sharing PROforma models for collaboration the models can only be processed by tools like Tallis. This restricts the wide spread collaboration and development of argumentation-based diagnostic CDSS models and applications.

Diagnostic CDSS models that are machine-readable and machine-understandable are therefore needed. Diagnostic models that can be modelled using current generation semantic-web ontology modelling tools and executed using off-the-shelf tools are needed. Models that are represented using formats such as OWL/RDF which are inherently machine-readable and machine-understandable are needed. Literature review had identified several attempts at using OWL ontologies for modelling diagnostic criteria. However, these models have limits in expressing uncertainty or use external tools to generate the diagnostic rankings. The remaining parts of the thesis detailed the development of an argumentation-based OWL ontology-driven model that can capture diagnostic criteria and CDSS prototypes that can execute the model using off-the-shelf semantic web technology to generate diagnostic rankings.

This study demonstrated the feasibility of using the argumentation-based approach for creating diagnostic CDSSs and showed that argumentation-based diagnostic CDSSs could potentially help reduce diagnostic error. It showed that guideline recommendations can help optometrists in improving their information gathering.

7.3 DS Model and CDSS for early diagnostic recommendations

This section details how Objectives 3 and 4 were achieved. The results of Chapter 4 Optometry study showed that argumentation is an effective method for developing diagnostic CDSSs and could potentially reduce diagnostic error. Semantic interoperability is a key requirement for integrating data from remote sources on the web and EHRs.

Argumentation provides several advantages over other CDSS reasoning methods. Argumentation is inherently easier to model as it uses qualitative methods to capture domain expertise (Fox *et al.*, 2007). The DS Model developed in Chapter 5 uses argumentation as the basis for generating diagnostic recommendations. The domain experts can use natural language to describe the arguments for and

against a diagnosis. These statements can be translated easily into arguments for the CDSS. For example, if an observation 'Observation 1' supports a diagnosis 'Diagnosis 1' then this argument can be modelled in DS Model using OWL classes, properties and individuals. Weights for each argument can be modelled within the same model. DS Model provides a novel approach of using argumentation-based methods to model diagnostic criteria in OWL/RDF format. Existing argumentation-based CDSS models such as PROforma require specialized software for generating diagnostic recommendations. The CDSS prototypes demonstrated in Chapter 5 show how ranked diagnostic recommendations can be generated using off-the-shelf technologies. Any triple store capable of accepting RDF data and processing SPARQL queries could be used to generate diagnostic recommendations.

Chapter 5 outlined the development of the DS Model (See section - 5.2) and CDSS prototypes that can use the DS Model to generate diagnostic recommendations. Dentistry was used as the case study for demonstrating the model. The DS Model demonstrates the ability to represent arguments for and against a diagnosis and represent the degree to which each observation supported or negated a diagnosis using N-ary relations. SPARQL insert queries were used to map patient data in RDF format to the DS Model and update the DS Model about the status of the patient (See section - 5.3.4). SPARQL queries were used to generate a ranked list of diagnoses that had a total weighting based on the observations that are present/absent in the patient data.

The literature review outlined the development and evaluation of existing OWL ontology-based diagnostic inference models that can be used for generating diagnostic recommendations (See section - 2.5.1). Existing ontology-based diagnostic CDSS tools are limited by their ability to express uncertainty of diagnosis by using DL axioms. DL cannot represent negation as failure as DL uses Open World Assumption. The DS Model uses SPARQL rules as a method to close the world and infer which observations are present or absent in the data (See section - 5.6). Based on this information the SPARQL rules update the model about the present/absent observations following which the diagnostic recommendations are generated. Other ontology-based diagnostic CDSSs have used external probabilistic tools to generate the rankings. Rodriguez et al. (Rodriguez *et al.*, 2009) demonstrated a system where diagnosis and symptoms were modelled using OWL ontologies. Their model did not allow for weighting of symptoms to produce a ranked list of diagnosis. The model was subsequently improved upon by García-Crespo et al. (García-Crespo *et al.*, 2010a) and used probabilistic reasoning. However, the probabilities were not modelled within the OWL ontology and modelled separately. It was not capable of providing multi-level diagnosis, where the observation can support a finding, and the finding in turn support the diagnosis. This deficiency was later addressed by (Alejandro Rodríguez-González, Labra-Gayo, *et al.*, 2012), however it was still not possible to model the degree of support the presence of an observation provides a diagnosis within the model. The DS Model allows the modeller to capture the degree of support within the same model and then use SPARQL rules to generate a diagnostic recommendation. This simplifies the modelling process and makes the process of collaboration easier as all the information needed to model a diagnostic CDSS is captured in a single model. Others have used complex ranking algorithms to generate diagnostic rankings of diagnostic OWL ontology models (Oberkampf, Zillner and Bauer, 2012a). There is no need for another rule language or custom ranking algorithm when using SPARQL rules to map the patient data to the DS Model and provide the diagnostic ranking. SPARQL is a W3C standard (W3C®, 2013b) and is widely implemented in several commercial and open-source triple stores. Any triple store capable of accepting RDF data and processing SPARQL queries could be used to generate diagnostic recommendations.

The DS Model uses the same weighting scheme used in the Chapter 4 Optometry study for capturing the degree of support of each argument. The simple method for capturing weights makes it easier for domain experts to capture their knowledge in DS Model. SPARQL language has the additional advantage of being able to perform summation. DS Model used SPARQL rules to sum the weights for each argument to

provide an overall weight for each diagnosis. This overall weight is used as the basis for generating ranked diagnostic recommendations. SPARQL query is used for generating the ranked diagnosis in DS Model. **Using SPARQL rules to update the DS Model based on patient data, summing the weights and generating the ranked diagnostic recommendations is a novel method that has not yet been demonstrated in other ontology driven diagnostic CDSSs.**

The simple weighting scheme employed and using SPARQL inference to sum the weights means the diagnostic recommendations and associated weights can be generated easily when compared to other methods. SPARQL language has the additional advantage of being able to query remote endpoints for data. The data in remote places can be queried and DS Model can use this information to generate a diagnostic recommendation. The simple weighting scheme and summation method used ensures that any amount of data residing in any repository in any format can be used for generating diagnostic recommendations. Let us consider the example of a PHR that contains information about a patient's dental hygiene habits. Using information recorded daily by the patient the PHR can make an assessment about the patient's overall caries risk. For example, if a patient does not brush their teeth regularly the risk of caries is high. This high-risk assessment can be stored in the patient's PHR. If the patient visits the dentist for a check-up, the DS Model CDSS in the dentist's clinic can query the PHR's SPARQL endpoint and retrieve the caries risk assessment and use that information in diagnosis. If High Risk assessment is present, then it supports a diagnosis. Using Bayesian probabilities can result in a more accurate diagnosis, however the causal knowledge needed to create Bayesian networks does not exist in domains like dentistry that are not well understood. However, clinicians use their domain knowledge in decision making routinely. The domain expert can capture this knowledge in DS Model using arguments. The DS Model can be modelled using existing ontology modelling tools such as Protégé and Web-Protégé. **The simplicity of the modelling process using argumentation-based methods, the simple weighing scheme used to assign weights, the simple technique used to sum the weights and ability to use existing off-the-shelf semantic web tools to generate the diagnostic recommendations gives the DS Model several advantages over conventional methods for generating diagnostic recommendations.**

Chapter 5 outlined the development and the architecture of several prototype CDSS applications that can use the DS Model to generate diagnostic recommendations (See section - 5.4). The patient data in each of the prototype (presented as case study) was made available in different formats. In real-life clinical applications, the patient data will be represented in different formats. The case studies demonstrate how the same DS Model can be used with different representations of the patient data. In one case study (Case Study 1) the patient data was inserted directly into a triple store alongside the DS Model (See section - 5.4.1). This case study demonstrates how the existing triple-stores and semantic web technology can be used to generate the diagnostic recommendations without the use of any other specialized tools. Case Study 2 and 4 details how data in relational database can be transformed into RDF data and can be used for obtaining diagnostic recommendations. Case Study 2 demonstrates how patient data can be exposed via a SPARQL endpoint and the CDSS prototype can access the patient data (See section - 5.4.2). Linked data can be stored in triple stores or relational database and exposed via a SPARQL endpoint. The Case Study 2 prototype application can then access this data to provide diagnostic recommendations using the SERVICE keyword within the SPARQL query. The patient data can be stored in multiple SPARQL endpoints, and the CDSS application can access this data. Case Study 4 shows the DS Model can work with data from an existing open-source EHR (See section - 5.4.4). Case Study 3 shows how patient data represented in an archetype format can be used for diagnostic CDSS recommendations (See section - 5.4.3). This Case Study demonstrates how patient data can be represented in different formats and using SPARQL rules the patient data can be mapped to inform/update the DS Model.

All the case studies in this chapter demonstrate how the DS Model can work with iso-semantic representations of patient data (See section - 5.4.3.1). The data available on the web via linked data can be represented in a variety of formats and annotated using several different terminologies and vocabularies. The Yosemite Project describes how RDF can be used as a universal information representation format and the semantics of data can be captured in RDF format (Booth, 2015a). Once the meaning of data has been captured in RDF the data can be translated from one format to another without loss of meaning. Rules languages such as SPARQL rules can be used for the transformation. The SPARQL rules used for mapping the patient data to update the DS Model works in a comparable manner as the SPARQL rules of the Yosemite Project and the DS Model can therefore work in conjunction with the Yosemite Project (Yosemiteproject.org, 2016) (See section - 5.6).

This chapter outlined the development of an extension to the DS Model called the Multi-level DS Model (See section - 5.5). The Multi-level DS Model expands upon the DS Model and gives its more granularity when modelling intermediate states/sub-diagnostic states like Findings. It is a demonstration of how the sub-diagnostic states can be modelled by reusing an existing DS Model if the requirements dictate its development. The same DS Model can be expanded to include intermediate states via the Multi-level DS Model. The ability to model multiple levels has been a challenge faced by other researchers as well (Rodríguez *et al.*, 2009b). They had attempted to solve the problem of multi-level diagnosis by using an approach similar to the Multi-level DS Model where the diagnosis is supported by a finding, which in turn can be supported by another set of findings (Alejandro Rodríguez-González, Torres-Niño, Mayer, *et al.*, 2012). They have used the term 'finding' to denote all clinical observations. The Multi-level DS Model uses argumentation to generate recommendations for sub-diagnostic states and diagnosis.

The Multi-level DS Model is an extension to the DS Model that can capture additional detail in the form of intermediate sub-diagnostic states. Argumentation-based methods used in the DS Model was used to develop arguments that supported each Finding. The presence or absence of the Finding was used as arguments to support a diagnosis. The Finding is a sub-diagnostic state and there are areas where additional explanation needs to be provided to the clinician. Certain diagnostic tests are indicated when the dentist needs to confirm the current status of the pulp. The finding 'Inflammation of the pulp' can be triggered by the observations ("Pain present" and "Pain initiated by cold") during the clinical encounter and a ranked list of findings can be displayed to the clinician to help choose the appropriate test based on these findings or let the CDSS application choose the right test for the clinician. The findings can therefore be used as a trigger for generating guideline recommendations as well. However, Chapter 4 Optometry study showed how the guideline recommendations can be generated without the presence of findings, instead the Tallis CDSS uses abnormal values as triggers.

However, this additional level of explanation comes at the cost of increased complexity. The domain experts will need to model the relationships between the observations and the findings. In the DS Model the Diagnoses and the Observations can be reused from existing ontologies such as SNOMED-CT. Findings as a sub-diagnostic state are not captured in existing ontologies at present. Existing DS Models can be expanded to develop the Multi-level DS Model. The literature review did not provide examples of models like the Multi-level DS Model in other argumentation-based languages such as PROforma. Examples from literature that demonstrated modelling multiple levels of observations using ontologies did not model sub-diagnostic states as demonstrated in the Multi-level DS Model (Alejandro Rodríguez-González, Torres-Niño, Mayer, *et al.*, 2012). These can be adapted to create Multi-level DS Models in future, however, they have not found wide applicability to date. The key advantage of the DS Model is the simplicity of the modelling process including weights, and this is lost during development of the Multi-level DS Model. The Multi-level DS Model can be used in specific areas where the additional granularity is required. For routine diagnostic CDSS applications the DS Model is more than capable of

generating effective diagnostic recommendations as demonstrated in Chapter 5 and evaluated in Chapter 6.

The DS Model provides the clinician with a ranked diagnostic recommendation. Once data is available in the linked data format, intelligent agents and other software applications can search for the data and use them for generating the diagnostic recommendations. The intelligent agents can then search for all the data of the patient from data sources (including PHR, EHR) and any resource where the patient may store data on the web in a structured manner. If the data is stored in an unstructured format, then the data should be converted to a structured format and/or semantically annotated using established ontologies and terminologies. This data can be used in conjunction with the data reported by the patient to provide early diagnostic recommendations for the clinician. The use of argumentation-based methods will result in a simple explanation of the patient's condition. The clinician is presented with alternative diagnoses for the patient. The literature suggested that providing alternative diagnoses can help in System 1 reasoning process and help reduce diagnostic error. Similarly, Croskerry (Croskerry, 2003a) has identified several strategies to reduce diagnostic errors due to cognitive causes. This includes considering alternatives, reducing reliance on memory, providing access to clear and well-organized information etc. The working hypotheses for the clinical case are generated early in the diagnostic encounter based on a few clinical patterns that are observed in the patient. These working hypotheses may not be accurate always and are tested by acquiring additional information from the patient. Using the illness script model, the working hypotheses is triggered early in the diagnostic encounter. Generating the working hypotheses early in the encounter helps the clinician narrow the solution space. Using the working hypothesis, further information is gathered until one working hypothesis is confirmed. The argumentation-based diagnostic recommendations generated by the DS Model can help in this process by providing a list of diagnoses at various stages of the clinical encounter. It provides reasons for making these recommendations in the form of arguments. This helps in the process of generating working hypotheses and help in gathering additional information about the patient to help confirm or refute working hypotheses. The argumentation-based CDSS can help organizing the expert's knowledge into arguments and provide recommendations that are easy to understand. Graber et al. (Graber, Gordon and Franklin, 2002a; Graber, Franklin and Gordon, 2005b) has identified several reasons for diagnostic errors due to cognitive factors. The guideline recommendations in tandem with diagnostic recommendations in the form of arguments could potentially help address most of these reasons.

- Faulty knowledge or skills:
  - The clinician has inadequate knowledge about the patient's condition and this predisposes to cognitive errors: The guideline recommendations can help in flagging abnormal values and the argumentation-based early diagnostic recommendations can help the clinician relate the observations with a diagnosis.
  - The clinician does not have the required skill in a clinical condition: The guideline recommendations cannot directly help in improving skills but can indirectly help by interpreting clinical test results.
- Faulty data gathering:
  - Clinician has problems coordinating data from multiple sources: The argumentation-based recommendations can help in the process of explaining the patient's condition in an easy to understand manner. However current generation of CDSS tools including argumentation-based CDSS tools are restricted to data collected during the clinical encounter and cannot access information from other sources. The semantic web and linked data technologies can help in process of coordinating and integrating the information from multiple sources. The arguments in the diagnostic recommendations can help the clinician make sense of the data derived from these sources. Without the

help of such a CDSS the clinician will have to manually access the disparate sources for information and this is time-consuming or rely on the patient's memory for information which may not always be reliable.
        - Clinician fails to collect appropriate information during history and examination: The guideline recommendation can guide the clinicians towards gathering appropriate information. The early diagnostic recommendations can provide context for collecting this information and help link the information to a diagnosis.
        - Incorrect diagnostic test ordered: The guideline recommendations can help identify the appropriate tests to be taken to rule out certain diagnoses.
    - Faulty information processing
        - Clinician fails to generate appropriate context: The early diagnostic recommendations can help the clinician link the observations to a clinical diagnosis.
        - Clinician overestimates/underestimates the relevance of a finding: The weights of the diagnostic recommendation represent the relative support for the present/absent observations based on expert opinion. Detailed explanations for each argument can be given.
    - Faulty information synthesis- Most common cause of cognitive based errors
        - Premature closure bias: It is the tendency of the clinician to prematurely halt the diagnostic decision-making process before all the data has been collected and the diagnosis has been fully verified. The clinician does not consider all alternative diagnoses. Most diagnostic errors are due to premature closure biases (Ely et al. 2012b). The early diagnostic recommendations can help in informing the clinician about the presence of alternative diagnoses that he/she should consider and provides arguments for each diagnosis. The clinician could potentially be stopped from stopping the information collection process and the guideline recommendations can aid in the process of collecting additional information.
        - Clinician fails to follow up on an appropriate test: The CDSS can alert the clinician to follow up on previous tests.

The DS Model uses argumentation-based methods and expands upon research in the argumentation CDSSs community by providing a method to represent arguments in OWL/RDF. This combines the strengths of argumentation with the strengths of semantic web technology. DS Model demonstrated how the arguments can be modelled using existing ontology modelling tools and the diagnostic recommendations can be generated using existing off-the-shelf technology. The DS Model likewise expands upon existing ontology-based diagnostic CDSSs methods by using argumentation, a simple weighting scheme, summation method, SPARQL rules etc. to generate diagnostic recommendations without the need for specialized software. The DS Model combines argumentation and semantic web technologies to demonstrate a diagnostic CDSS that can integrate information to provide early diagnostic recommendations.

7.4 Evaluation of the DS Model across all stages of the clinical encounter

This section details how Objective 5 was achieved. The DS Model demonstrated how patient data can be used to generate early diagnostic recommendations. However, before linked data can be used in conjunction with the data reported by the patient and obtained via questioning and clinical tests, there is a need to know how the DS Model will perform without the support of any additional data. There is a need to know where the DS Model needs additional data (if needed) to improve the diagnostic performance. The DS Model uses a simple weighting scheme and summation method to generate the ranked diagnostic

recommendations. The study allowed us to evaluate the effectiveness of the simple weighting scheme and the summation method used to generate diagnostic ranking.

Chapter 6 evaluated the DS Model in both early and late diagnosis and compared the results of the evaluation with that of dentists in final diagnosis. Final diagnosis is the diagnostic ranking obtained with the support of all the data. In the early diagnosis scenario (Chief Complaint, Extra-Oral Examination, Intra-Oral Examination), all the information needed to make a diagnosis is not available to the CDSS application (See section - 6.2). Therefore, in theory the CDSS should not be able to provide a diagnostic recommendation that is at least as good as clinicians. The evaluation study described in Chapter 6 evaluates the DS Model and compares it to clinicians to see if this is indeed true and find out where the DS Model needs additional patient data via linked data or enhancement of the information model by collecting more information about the patient.

The evaluation was done with the help of clinical vignettes (n=25). The experiment was done with community dentists (n=20) (See section - 6.3.4). The clinical vignettes (n=5) were randomly assigned to each dentist who then provided Rank 1 and Rank 2 diagnostic recommendations for the vignettes via a web-based survey application. The dental version of the DS Model could provide diagnostic recommendations in the field of Endodontics in dentistry. The vignettes and the DS Model were developed with the support of an Endodontist and the author's own previous experience in dentistry (See section - 6.3.3). The Endodontist was given the same set of vignettes (n=25) and asked to provide the Rank 1 and Rank 2 diagnosis for each vignette. The Endodontist's opinion was considered the gold standard and used to compare the diagnostic performance of the community dentists and the CDSS prototype using the dental version of the DS Model.

A Pulpal and Apical Rank 1 and Rank 2 recommendation was obtained. The same vignettes were represented using an archetype format and one-by-one the ranked diagnostic recommendations for each vignette was obtained by a prototype CDSS application. The prototype was the same described in Case Study 3 of Chapter 5. A dental version of the DS Model for providing diagnostic recommendations was created as part of the model development in Chapter 5 and for evaluation of the DS Model with dentists in Chapter 6.

Normalized Discounted Cumulative Gain (NDCG) and Precision-Recall were used as the metrics for comparing the diagnostic performance of the dentists and the CDSS. NDCG measures the overall ranking of the dentists and the CDSS and compares it to the expert's ranking, while precision-recall measures the precision and recall of Rank 1 and Rank 2 of the dentists and the CDSS and compares it to the experts ranking (See sections - 6.5, 6.6 and 6.8).

The CDSS performed better than the dentists at overall ranking of diagnostic recommendation in the late stages. As more data is obtained in the late stages from clinical tests and radiographic examinations the CDSS could rank the diagnosis better than the dentists. The CDSS was better at overall ranking of Pulpal diagnosis than Apical in the late stages. This could partly be explained by signs, symptoms and interpretation of clinical test results. The nature of pain experienced by a patient with an Apical pathology and the radiographic results are more dependent on the subjective interpretation of the data by the dentist. The performance of the DS Model in diagnostic ranking of Apical diagnosis in the later stages can be improved by changing the weights of the DS Model and adapting the information to collect more information about the Apical pathology.

The results showed that the Rank 2 diagnostic performance of the CDSS in the late stages were comparable to that of dentists. The reason why the CDSS performed better at giving good Rank 2 recommendations could be partly explained by the script theory for medical diagnostic knowledge. The

signs and symptoms provided in the vignettes trigger certain illness scripts, which can then be used as the basis for generating a diagnostic hypothesis (Charlin, Tardif and Boshuizen, 2000; Schmidt and Rikers, 2007a). Depending on the vignette a Rank 1 diagnostic hypothesis may be triggered from the data in the vignette. However, a Rank 2 may not be clearly apparent as the data may strongly suggest only one diagnosis.

A Rank 2 diagnosis for CDSS may still be important for clinical decision making, especially in early diagnosis. If the CDSS can provide an accurate Rank 2 diagnosis early in the diagnostic encounter, this may trigger an illness script in the mental model of the clinician and may help him/her seek more information based on the triggered script. The results showed that the Rank 2 diagnosis was very poor for the CDSS, especially in early diagnosis. However, when the ranking was considered in its entirety (Rank 1-Rank 7 for Pulpal and Rank 1- Rank 5 for Apical) the NDCG scores showed that the overall ranking was comparable to the dentists in the early stages and became better than the dentists in the late stages (See sections 6.7.10.1 and 6.7.10.2). So, providing an accurate overall ranking could still be helpful to the dentists, as the presence of certain diagnoses in the ranking in the early stages of the clinical encounter could trigger the right illness scripts and help the clinician avoid ignoring relevant diagnoses. However, more research is needed to understand the impact of providing a ranked diagnostic recommendation to clinicians in early diagnostic scenario and the impact of the ranking and weights on diagnostic hypotheses generation.

The performance of the CDSS in the early stages (CC, CC+EO, CC+EO+IO) (CC=Chief Complaint, EO=Extra Oral Examination, IO=Intra Oral Examination) for Pulpal diagnosis was characterised by the presence of an intermediate dip in diagnostic performance. The diagnostic performance in NDCG results and precision-recall showed a comparable performance to the dentists in the CC stage. However, as more information is collected the CDSS shows a dip in performance in the next stage (CC+EO). The performance subsequently improves and eventually becomes better than the dentists. The CDSS is expected to perform the worst in the earliest stage of the clinical encounter and gradually improve as more data is collected during the encounter. However, the results show that this may not be the case in all situations. Depending on the domain and the information model used, the intermediate dip can cause the CDSS to be less reliable as the clinical encounter proceeds through the different stages.

The Apical diagnostic performance (NDCG and precision-recall) is poor in the early stages and gradually improves as more data is collected. Therefore, the clinician should be informed about the possibility of the Pulpal diagnosis performance of the CDSS reducing and subsequently improving as more data is collected. The intermediate dip is more pronounced in the precision-recall of Rank 1 Pulpal diagnosis (See section - 6.9.1.1) than the NDCG scores of the Pulpal diagnosis (See sections - 6.7.8.1 and 6.7.9.1).

The DS Model uses a simple weighting scheme and summation method to generate the ranked diagnostic recommendations. This simple method of summation allows the CDSS using the DS Model to generate diagnostic recommendations without the need for specialized software and algorithms for inferencing. However, the use of this simple weighting scheme and summation method means that there is increased risk of several diagnoses sharing the same rank when there is limited information. The results show that even with simple weighting scheme the overall ranking is close to the Endodontist in the early stages. In some vignettes, the overall ranking falls in the CC+EO and CC+EO+IO sections. However, the ranking quickly adjust as more information is obtained and becomes very good in the later stages. Even if the correct diagnosis falls in the ranking the arguments for each diagnosis is presented alongside the diagnosis and this can explain the reasons for making the recommendation. The dentist can then look at the ranking and use that information to guide the next steps of information collection.

Overall the results show that in the early stages of a diagnostic encounter there is a need for more information for the CDSS to perform better than the average dentist. This information can come from linked data sources and the DS Model is equipped to incorporate this additional information to provide diagnostic recommendations. The simple weighting scheme and summation method provides an overall ranking of diagnosis that is comparable to dentists. The clinicians using the DS Model for diagnostic recommendations in the early stages of the clinical encounter will have to be informed about the potential for rankings to change as more information is collected. The ranking in the early diagnosis must be considered tentative and the clinician must use the arguments and other recommendations alongside the clinical data to form diagnostic hypotheses. The ranking in the early stages of the clinical encounter can be improved by integrating information from diverse sources and improving the quality and quantity of data available to the CDSS.

The results of this study can be used to focus the efforts of finding more information via linked data in the early stage of the clinical encounter and combine this information with other information collected during the encounter and the results of clinical tests and radiographic examination. The results of this study can be used by other researchers to develop information models that are better at providing early diagnosis and not just good late diagnosis.

7.5 Limitations of the research

7.5.1 Chapter 4 Optometry study

In the Chapter 4 study with optometrists the sample size for the intervention group that had the CDSS support was relatively small (n=36) when compared to the control group study. The Chapter 4 Optometry study was a pilot study with optometrists that evaluated a prototype argumentation-based diagnostic CDSS for the feasibility and impact of the system on diagnostic performance. The aim of the study was to find out if argumentation is a suitable approach for developing diagnostic CDSSs. Argumentation-based diagnostic CDSSs in literature have not evaluated the impact of argumentation on diagnostic error. Since this was a pilot study it was decided to use a smaller number of optometrists.

The study was conducted in a simulated environment and was limited by the small number of vignettes used in the study and smaller number of optometrists who were recruited for the study (intervention group). Since this was a pilot study that preceded the development of the ontology-based diagnostic CDSS, we decided to use a smaller group of optometrists for evaluation. A more detailed study using randomised controlled study design is needed to fully understand the impact of guidelines and argumentation-based diagnostic recommendations on the diagnostic performance on clinicians such as optometrists. The small numbers involved in the intervention group was a key limitation of the study and prevented deeper analysis of the results. More studies are needed to see to what extent the argumentation-based diagnostic recommendations can help impact specific cognitive causes of diagnostic error. However, it was able to sufficiently demonstrate how the argumentation-based diagnostic recommendations could potentially improve diagnostic decision making in primary care optometrists.

Another limitation of the Chapter 4 Optometry study was the historical comparison with the control group study and the way the vignettes were presented. Little demographic data was available for comparison with the intervention group. The control group study presented the same vignette data in a web-page that had buttons simulating a clinical encounter. This was a web-based application that represented paper-based vignettes/paper patient records but captured the data in an electronic manner. The intervention group study used a CDSS that simulated an EHR and the real-life way the guideline and diagnostic recommendations are presented to the optometrists. CDSS recommendations cannot be given in an interface that is simulates a paper-based patient record system. The CDSS should be able to guide the optometrist towards the right clinical tests or investigations which cannot be done in an interface that

displays all the data in a single page. Therefore, a decision was made to alter the way the guideline and diagnostic recommendations are provided to the optometrists in the intervention group study. This may have impacted the way the data is collected by the optometrists.

In the control group study all the data collected was presented in the same page. The optometrists are not penalised for asking inappropriate questions or performing unnecessary tests. This would not be ideal as in real-life situation optometrists can repeat tests and ask any questions according to their clinical judgement. However, they are encouraged to complete clinical examinations in 60 minutes. Therefore, a counter that displayed the total time (in real clinical time) elapsed was provided to the optometrists. This counter was not provided in the intervention group study as analysis of the control group study data showed that the optometrists asked most of the questions anyway. The optometrists in the intervention group were not penalised for not asking certain questions or performing investigations. Analysis of the intervention group study data showed that there was a difference in the total number of questions asked or investigations performed between the two groups in only 18-29% of the questions.

This pilot study is a proof-of-concept and is therefore insufficiently powered for a formal comparative study. A historical comparison study is much weaker than a fully-powered randomised controlled trial and this is another weakness of this study. More randomised controlled studies using vignettes or simulated patients are therefore needed to fully understand how guideline and diagnostic recommendations impact optometrist performance before implementation in clinical practices.

The intervention group optometrists were only provided CET points as compensation for completing a vignette. The control group optometrists were provided with monetary compensation to complete each vignette. Each vignette takes around 20-30 minutes to complete. Therefore, the decision was taken to ask each intervention group optometrist to only complete one vignette. Asking the busy intervention group optometrists to do more would have been unfair and unethical without enough monetary compensation. The study did not have the funds to adequately compensate the intervention group optometrists beyond providing them with CET points.

The decision support recommendations were simulated to reflect real life settings. The cost and effort of developing a full decision support for the Chapter 4 Optometry study hindered the efforts in testing and evaluating the decision support in real life setting with real patients. However, the study does indicate where CDSS may have an impact and areas where the CDSS recommendations can be improved. The data provided enough results to justify the development of a full decision support and details clinical trials with real patients.

Though the CDSS application was found to be effective in improving the diagnosis in certain vignettes, the application was limited in many ways. The data used for making the diagnosis came directly from the patient from clinical history, examination and clinical tests and investigations. The CDSS developed for this study did not demonstrate defeasible reasoning and integrating information from multiple sources. However, it was able to successfully demonstrate how an argumentation-based diagnostic CDSS can provide guideline and diagnostic recommendations and the usability survey found that the optometrists found the recommendations useful.

Another limitation faced in this thesis was development of arguments for each candidate diagnosis. Ideally there should be several diagnosis candidates for each diagnosis, and the CDSS should recommend some from the list. We were forced to restrict the total number of candidates as developing arguments for each candidate for an extensive list of candidates would be a time-consuming task and would involve a lot of resources, especially clinician time.

7.5.2 DS Model development and evaluation

7.5.2.1 DS Model

The DS Model was developed in the dentistry domain. Dentistry is mainly a primary care domain. Diagnosis of a patient's condition can happen in hospital and the patient's bedside. The DS Model will need to be developed and adapted where necessary to be able to function equally well at the patient's bedside in a hospital. Clinicians in primary care are increasing facing reduced time to examine the patient. A comparable situation is being faced by clinician in secondary and tertiary care as well. The DS Model though the use of semantic web and linked data technologies can potentially reduce the time taken to find all information about the patient. The extent to which the DS Model can help make this process easier needs to be studied further as this was not covered in this thesis.

A domain expert can use any currently available ontology modelling tools to develop a DS Model for their domain. The diagnostic recommendations can be generated without the use of specialized ranking algorithms or software for inference. This greatly reduces barriers to developing and executing the DS Model when compared to existing argumentation-based tools such as PROforma or currently available ontology-based diagnostic CDSS tools. However, the domain experts will still need to be trained in the basics of OWL language if they are to use currently available ontology modelling tools. Tools such as Protégé are being widely used for developing ontologies and terminologies such as SNOMED-CT and ICD-11. Therefore, there is a pool of trained clinicians already available that can potentially create DS Models in their domains. However, this pool of clinicians is limited in number when compared to overall number of domain experts in medicine. Developing graphical user interfaces like the one provided by Tallis can potentially help ordinary clinicians in ontology modelling. Nevertheless, there is still a learning curve associated with the modelling tools that the domain experts must overcome.

The other problem encountered during this research was the unavailability of linked patient data. Due to concerns regarding privacy and patient confidentiality patient data is not available in linked data format. However, there are web-based PHRs and other sources for patient data currently available and the data can be transformed to RDF format and exposed via SPARQL endpoints to make them available for CDSS applications to access.

The DS Model is limited in the type of relations it can represent. The DS Model can capture the link between clinical observations and diagnoses. However, to represent sub-diagnostic states and other intermediate state such as Findings the DS Model will need to be expanded. Expanding the DS Model results in a more complex Multi-level DS Model and this sacrifices the simple and effective modelling approach used in the DS Model.

7.5.2.2 Use of clinical vignettes

The Chapter 6 study investigated the diagnostic performance of a CDSS prototype using the DS Model as its inference model and compared the performance at different stages (early and late) with the final diagnostic performance of dentists. Vignettes simulating commonly seen clinical scenarios in dentistry were used to evaluate the dental version of the DS Model and the performance of the dentists. Even though an attempt was made to accommodate a wide range of clinical scenarios, the vignettes represent only a small proportion of the clinical cases a dentist may encounter.

Vignettes have some advantages when compared to using actual patients or real case records. Using actual patients is not feasible when comparing the diagnostic performance of a large group of dentists. Using real case records has its challenges when it comes to patient data confidentiality, obtaining a variety of clinical scenarios and overcoming the problem of missing data. However, vignettes used in this study did

not factor in situations that can impact the diagnostic decision making of the dentist. The participants in the study were not given time limits to complete the study. In real life situations, the dentists will be under time constraints to diagnose and manage the patient. The participating dentists were given all the information needed to manage the patient including the test results. In real-life clinical settings, clinicians may pick and choose the clinical tests and may only perform additional tests if the previous tests were inconclusive. The dentists were not given the option of choosing clinical tests in this study, however, the optometrists in the optometry study was given the option of choosing the appropriate clinical tests. Dentists must factor in the patient's ability to pay for clinical tests or radiographs when prescribing them. This study did not include the financial implications of diagnosis and treatment when developing the vignettes.

Both the studies involving optometrists and dentists used clinical vignettes for data collection. Clinical vignettes have advantages however they cannot completely simulate factors such as time pressures, personalities of patients, team-work, financial constraints etc. which can impact their decision making in real-life settings. Therefore, future evaluation of performance of clinicians will need more realistic vignettes. Use of actors, simulations etc. are other alternatives, however these tools have limitations such as increased cost and complexity of modelling.

7.5.2.3 Clinicians and ranking of diagnoses

One the main challenges faced in this study was the ranking of diagnosis. Ideally, the Endodontist and the participant dentists should be asked to provide a complete ranked list for both NDCG analysis. The CDSS provides a ranked diagnostic list for both Pulpal and Apical diagnosis. However, during pilot tests of vignettes it was found that the Endodontist and pilot study dentists (n=2) found it difficult to provide an appropriate ranked list of diagnoses beyond 2nd rank. Therefore, ranking of vignettes was restricted to Rank 1 and Rank 2 diagnoses.

The hypothetico-deductive model for clinical diagnosis suggests that clinicians generate multiple competing hypotheses from patient data and collect more data to confirm or refute the hypothesis (Elstein, Shulman and Sprafka, 1978). However, studies have shown that clinicians structure knowledge that they have gained from textbooks and experience in different forms. The 'Illness script' is a widely accepted concept that tries to provide a framework for understanding how clinicians organize the knowledge that they gain from literature and clinical experience and use it to diagnose and treat patients (Charlin, Tardif and Boshuizen, 2000). A script can be described as a set of attributes each with a set of values. The script has a structure like the archetype model used in this study (Table 6-3), however it is cognitive model and each attribute has a default value for each situation. For example, the dentist will have an illness script for Reversible Pulpitis along with scripts for other diagnoses. Scripts are formed for non-diagnostic tasks as well. Each attribute within the script will have different probabilities of occurring. The value with the greatest chance of occurring is the default value. The extract of an illness script for Reversible pulpitis is shown below in Table 7-1 with the default values in red. Similar illness scripts are present for other diagnoses with different default values.

Attributes	Value
Onset of pain	Gradual onset
	Stimulation required for onset
	Sudden onset

Attributes	Value
Pain location	No pain
	On tooth
	Dull sensation of pressure over the maxilla
	Infraorbital pain

Table 7-1 - Illness script example (default value in red)

Based on observations present in the patient, the illness script(s) for that condition is triggered and the clinicians looks for the default value in the patient (Schmidt and Rikers, 2007a). If the default value is not present or unacceptable values are found for the attributes, another illness script is triggered, and clinicians looks for the attributes and values in that script. The clinician will traverse the network of knowledge in their memory, gained from experience and biomedical knowledge, to find the script that best suits/explains the clinical situation in front of him/her. In this manner, the clinician will generate several different hypotheses and reject the hypothesis that does not fit his/her mental model and arrive at a working diagnosis.

Illness scripts are an efficient way of organizing knowledge and allows the clinician to rapidly seek information, search for the appropriate illness script, find the right script and verify the values (Charlin *et al.*, 2007). However, because the illness script model does not use a system of weights similar to the mechanism used by DS Model, it is difficult for the clinician to provide a ranked set of hypotheses beyond the first few diagnoses. It is possible for the dentists to identify the most likely diagnosis (Rank 1) and depending on the vignette the dentists can identify the Rank 2 diagnosis as well. In fact, the more well defined the vignette/scenario the more difficult it is to identify the Rank 2 diagnosis as there will be one script that best explains the condition of the patient and there is usually no script that can be considered that next most likely hypothesis.

7.6 Future directions

The arguments supporting and negating a diagnosis can be obtained from domain experts in the form of natural language statements. For example, dentists can be asked what arguments they use when diagnosing a case of Reversible Pulpitis. The dentist can write down the arguments in natural language. Currently translating these statements into the DS Model requires the help of an ontologist or knowledge engineer. However, using natural language processing the text can be broken down into components and meaning of the statement can be captured in RDF format. Once captured in RDF format, using SPARQL rules the statements can be automatically translated into DS Model arguments. Natural language processing can be used to translate existing PROforma diagnostic CDSS models into DS Models. Both PROforma-based diagnostic CDSS models and DS Model used argumentation as the basis. If semantics of a PROforma model is captured in RDF format, then using SPARQL rules the RDF PROforma model can be translated to DS Model so that a DS Model CDSS can reuse existing PROforma models. Similarly, existing DS Models can be translated into a RDF PROforma format. This RDF format can be translated back into proprietary PROforma model currently used so that existing Tallis engine can execute the model. Future research will investigate this round tripping methodology.

DS Models in each domain can be produced and published on the semantic web. DS Models that have been annotated with other ontologies and terminologies like SNOMED-CT, LOINC, ICD etc. are more easily discoverable by intelligent agents as the model is both machine-readable and understandable. Web-based platforms that can enable modelling and sharing of DS Models need to be developed that can help domain experts collaborate with others in creating DS Models. Platforms like OpenClinical.net are already doing similar work with PROforma-based models. They can be expanded to support DS Models as well.

The DS Model produces ranked diagnostic recommendations with arguments for and against each diagnosis. The CDSS prototypes have demonstrated how remote sources can be queried for information and data in a wide variety of formats can be integrated to generate diagnostic recommendations. There are linked data repositories such as DBpedia and Wikidata that hold information about different types of concepts. For example, the DBpedia page for Toothache contains structured information about the concept (http://dbpedia.org/page/Toothache). This information can be accessed via the DBpedia SPARQL endpoint. The structured data within the concept are in turn linked to other concept via HTTP links. When the patient complains of toothache and that information is recorded in the EHR, then additional information about the concept can be queried using SPARQL and then displayed to the clinician if needed. If guideline recommendations are similarly published in the form of structured RDF data and exposed via a SPARQL endpoint, then guideline recommendations can be displayed in the same manner as the DBpedia information.

DS Model has demonstrated how an ontology-based diagnostic CDSS using argumentation can use linked data resources to generate diagnostic recommendations. If patient data is available as linked data resources CDSSs or intelligent agents can query the data to generate early diagnostic recommendations or trigger guideline recommendations. However, the architecture that allows linked data to be open and easily accessible makes it less suitable for exposing patient data without compromising patient confidentiality (Corsar, Edwards and Nelson, 2013). There is a need for developing linked data architectures for securely hosting linked data without compromising the privacy of the patient. Technologies such as blockchain can potentially be used to solve some of these problems (Mettler, 2016). Blockchain is a distributed system which records and stores transaction records. All the data is available publicly but only to those with the right credentials and permissions. In a blockchain there is no central authority such as governments or healthcare organizations that hold the data. The data is stored and distributed amongst all participants in the distributed network. Blockchain uses established cryptographic techniques to allow each participant in the distributed system to freely participate in the exchange of information without predefined contracts in a secure manner. Patients can freely share their data using blockchain in an encrypted format and only CDSSs and clinicians with the right credentials will be able to access this information. Future research will investigate exposing patient data using linked data and securing the information using blockchain technology.

The Learning Healthcare System is a proposed system where data is collected routinely in clinical practice using EHRs and is used to ask specific questions about the environment (Friedman, Wong and Blumenthal, 2010). The results of this analysis are fed back into the system where the outcomes are routinely monitored and the data in analysed further. This results in a continuously learning healthcare system that can in theory lead to better quality, patient outcomes, effectiveness and efficiency in the healthcare system and more patient-centred care. CDSSs are a central part of this learning healthcare ecosystem (Foley and Vale, 2017). CDSSs help deliver up-to-date patient-centred content and recommendations directly to the clinician. The results of analysis of patient data in a learning healthcare system can be used to update the logic and content of the CDSSs and the system can monitor patient data to see how the change has affected the clinical outcomes and understand where the CDSSs models can be improved. Platforms like OpenClinical.net (Fox *et al.*, 2015) are collaborating with organisations like the Learning Healthcare Project (The Learning Healthcare Project, 2017) to design and develop machine

readable content for deploying in a learning healthcare system (Fox, 2015). This content can then be implemented in clinical environments and the outcomes analysed using machine learning and other data analytics methods and the changes needed are studied and pushed to the clinical environment. Future research will investigate how impact of the DS Model in a real-life clinical environment can be analysed using machine learning techniques and the results used to produce better arguments and weights for the DS Model. The updated DS Model can then be pushed to the deployed environments and the learning healthcare cycle can be repeated until an appropriate and mature DS Model is developed.

Using techniques currently being developed as part of the Yosemite Project, rules can be developed for transforming data from one format to another to support semantic interoperability. The same techniques can be used to develop mapping rules from patient data to the DS Model. At present the mapping from patient data to DS Model to update the DS Model (present/absent SPARQL insert queries) must be created for each data format. Future research involves working with the Yosemite Project to standardise the present/absent insert queries (mapping rules), crowd-source the mapping rules and host them on a shared web-based platform. Further research is needed to align the two techniques and develop methods to better crowd-source the rules needed for mapping.

Developing and evaluating DS Model in other domains (optometry) are needed to test if the intermediate dip in performance of the CDSS is a phenomenon observable in other domains as well, or if it's something unique to dentistry. Further research is needed to further understand the cause of this intermediate dip and how linked data can potentially help alleviate this problem. Likewise, further research is needed to see if the intermediate dip can be seen with inference models that do not use weights to rank the diagnosis. For example, with CDSS applications that uses Bayesian inference models.

Evaluating the impact of the DS Model and the Multi-level DS Model on clinician diagnosis and management decision making is needed before adoption of the DS Model in clinical practice. Diagnostic error is a problem in primary care and both studies involving clinicians (optometrists and dentists) confirms the presence of this problem. Even with CDSS help the optometrists were still getting diagnostic errors in some non-sight threatening cases. The role played by CDSS in reducing diagnostic error needs to be studied in further depth with greater emphasis in the type of diagnostic recommendations that can best help reduce diagnostic error, for example, ranking with weights or probability measures.

Incorporating diagnostic recommendations with guideline recommendations is another area that needs further study. At present the guideline recommendations are generated separately from the diagnostic recommendations. The diagnostic recommendations can itself trigger guideline recommendations. The interplay between guideline and diagnostic recommendations and its exact role in improving diagnostic performance needs further evaluation. The role of early diagnostic recommendations and the impact of early recommendations on cognitive biases will need to be studied further to better tailor the recommendations for each case.

7.7 Summary of contributions

In addition to opening avenues for future research into ontology driven CDSS and modelling of diagnostic criteria in OWL/RDF this research has made three major contributions to the literature on the use of argumentation-based CDSS in primary care diagnosis and argumentation-based ontology driven CDSS for diagnostic recommendations.

Firstly, the thesis evaluated the feasibility and impact of argumentation-based CDSS on the diagnostic performance of optometrists. An off-the-shelf argumentation-based modelling tool and execution engine was used to develop a diagnostic CDSS that can generate diagnostic recommendations. The thesis found

that argumentation is a viable method for generating diagnostic recommendations that can potentially help reduce diagnostic error especially in sight-threatening cases in optometry. The thesis found that guideline recommendations do have a positive impact on information gathering of optometrists and can potentially help optometrists in asking questions and performing tests as per current standards of care. The CDSS is dependent on the quality of data that is entered into the system. The faulty interpretation of medical images can lead the clinician to enter the wrong data and causing the CDSS to provide wrong clinical recommendations and supporting the wrong diagnostic hypothesis of the clinician. This problem can be remedied by improving the quality and quantity of arguments used for the diagnosis. The guideline recommendations were similarly found to have a positive impact on management decision making of optometrists. The optometrists found the CDSS easy to use and found its recommendations useful. This is one of the first studies evaluating guideline based CDSSs amongst community optometrists. The results of this study can be used to better design and develop guideline-based and diagnostic CDSSs for optometry.

Secondly, this thesis continued from the study with optometrists and developed an argumentation-based ontology driven diagnostic ontology model and CDSS that can generate early diagnostic recommendations. The ontology model called DS Model can capture domain expertise using arguments in favour and against each diagnosis. This novel approach to developing ontology driven diagnostic CDSSs combines argumentation and semantic web technologies. The DS Model can be modelled using any off-the-shelf the ontology modelling tools. The thesis expands upon research in areas of argumentation-based CDSSs and ontology based CDSSs. DS Model developed for this thesis used the strength of argumentation methods where the modelling process is a lot easier and the diagnostic recommendations can be explained to the clinician and combined it with the strength of semantic web technologies. The current generation argumentation-based CDSSs tools use proprietary software and formats and have problems with semantic interoperability. Using semantic web technologies, the thesis was able to demonstrate the ability of CDSSs using the ontology model to integrate data in a wide variety of formats to generate diagnostic recommendations. The ontology model used a simple weighting method developed for the optometry study CDSS prototype to model the degree of support of each argument favouring and negating a diagnosis. The ontology model demonstrated a novel summation method using SPARQL that can generate diagnostic recommendations without the use of specialized software and algorithms and overcomes some of the problems facing current generation ontology driven diagnostic CDSSs. The thesis explored the development of an extension to this ontology-driven diagnostic model that can provide recommendations for sub-diagnostic states where necessary. The dental diagnostic ontology model is the first ontology-driven diagnostic inference model that has been developed in the field of Endodontics.

Lastly, an evaluation of the DS Model in all stages of the clinical encounter was performed. The evaluation showed that the DS Model can provide a ranked diagnostic recommendation that is comparable to dentists in the early stages of the clinical encounter. The diagnostic performance can be improved using linked data technology to incorporate more information into the decision making. As more information is collected depending on the type of case the performance of the DS Model decreases slightly, and then improves as more information is collected to provide a ranked recommendation that is better than dentists. The evaluation showed that even with a simple weighting method and summation method used in the DS Model the diagnostic recommendation was comparable to dentists. With limited information in the early stages of the clinical encounter the DS Model was able to provide a good ranked diagnostic recommendation validating the approach used in this thesis.

8 References

AAE (2009) 'AAE Consensus Conference Recommended Diagnostic Terminology', *Journal of Endodontics*. doi: 10.1016/j.joen.2009.09.035.

Aalten, C. M., Samson, M. M. and Jansen, P. A. F. (2006) 'Diagnostic errors; the need to have autopsies.', *The Netherlands journal of medicine*, 64(6), pp. 186–90.

Abbott, P. (2000) 'Are dental radiographs safe?', *Australian dental journal*, 45(3), pp. 208–13.

Acharya, A., Mital, D. P. and Schleyer, T. K. (2009a) 'Electronic dental record information model', *International Journal of Medical Engineering and Informatics*, 1(4), p. 418.

Acharya, A., Mital, D. P. and Schleyer, T. K. (2009b) 'Electronic dental record information model', *International Journal of Medical Engineering and Informatics*, 1(4), p. 418. doi: 10.1504/IJMEI.2009.026810.

Adams, P. *et al.* (2012) 'Clinical decision support systems to improve utilization of thromboprophylaxis: a review of the literature and experience with implementation of a computerized physician order entry program.', *Hospital practice (1995)*, 40(3), pp. 27–39. doi: 10.3810/hp.2012.08.987.

Aikins, J. S. *et al.* (1983) 'PUFF: an expert system for interpretation of pulmonary function data.', *Computers and biomedical research, an international journal*, 16(3), pp. 199–208.

Allemang, D. and Hendler, J. (2011) *Semantic web for the working ontologist: effective modeling in RDFS and OWL*. Second Edi. Elsevier.

Andreassen, S. *et al.* (1987) 'MUNIN: a causal probabilistic network for interpretation of electromyographic findings', in *IJCAI'87 Proceedings of the 10th international joint conference on Artificial intelligence - Volume 1*. Morgan Kaufmann Publishers Inc., pp. 366–372.

Apache Jena (2016) *Apache Jena Reasoners and rule engines : Jena inference support*, *jena.apache.org*. Available at: https://jena.apache.org/documentation/inference/.

Apache Software Foundation (2015) *Apache Jena - Fuseki: serving RDF data over HTTP*, *jena.apache.org*. Available at: https://jena.apache.org/documentation/serving_data/ (Accessed: 30 January 2016).

Archer, P. (2017) *Dive into the Semantic Data Lake | W3C Blog*, *www.w3.org/blog*. Available at: https://www.w3.org/blog/2017/04/dive-into-the-semantic-data-lake/ (Accessed: 4 October 2017).

Aron, Y. (2013) *Information privacy for linked data*. Massachusetts Institute of Technology.

Atkinson, K., Bench-Capon, T. and Modgil, S. (2006) 'Argumentation for Decision Support', in *Database and Expert Systems Applications*. Springer, Berlin, Heidelberg, pp. 822–831. doi: 10.1007/11827405_80.

Baader, F. *et al.* (2003) *The description logic handbook: theory, implementation, and applications*. Cambridge University Press.

Barnett, G. O. *et al.* (1987a) 'DXplain. An evolving diagnostic decision-support system.', *JAMA : the journal of the American Medical Association*, 258(1), pp. 67–74.

Barnett, G. O. *et al.* (1987b) 'DXplain. An evolving diagnostic decision-support system.', *JAMA : the journal of the American Medical Association*, 258(1), pp. 67–74.

Barraclough, K. (2013) *Avoiding errors in general practice*. Wiley-Blackwell.

Bau, C.-T., Chen, R.-C. and Huang, C.-Y. (2014) 'Construction of a clinical decision support system for undergoing surgery based on domain ontology and rules reasoning.', *Telemedicine journal and e-health : the official journal of the American Telemedicine Association*, 20(5), pp. 460–72.

Beckett, D. *et al.* (2014) *RDF 1.1 Turtle - Terse RDF Triple Language*, *w3.org*. World Wide Web Consortium (W3C). Available at: http://www.w3.org/TR/turtle/ (Accessed: 8 November 2016).

Begum, S. *et al.* (2011) 'Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments', *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 41(4), pp. 421–434. doi: 10.1109/TSMCC.2010.2071862.

Belleau, F. *et al.* (2008) 'Bio2RDF: Towards a mashup to build bioinformatics knowledge systems', *Journal of Biomedical Informatics*, 41(5), pp. 706–716. doi: 10.1016/j.jbi.2008.03.004.

Bender, I. B. (2000) 'Pulpal pain diagnosis--a review.', *Journal of endodontics*, 26(3), pp. 175–9.

Berner, E. S. *et al.* (1994a) 'Performance of four computer-based diagnostic systems.', *The New England journal of medicine*, 330(25), pp. 1792–6. doi: 10.1056/NEJM199406233302506.

Berner, E. S. *et al.* (1994b) 'Performance of four computer-based diagnostic systems.', *The New England journal of medicine*, 330(25), pp. 1792–6.

Berner, E. S. (2009) 'Diagnostic error in medicine: introduction.', *Advances in health sciences education : theory and practice*, 14 Suppl 1, pp. 1–5. doi: 10.1007/s10459-009-9187-x.

Berners-Lee, T. (2006) *Linked Data - Design Issues*, *www.w3.org/DesignIssues/LinkedData.html*. Available at:

https://www.w3.org/DesignIssues/LinkedData.html (Accessed: 1 January 2017).

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) 'The Semantic Web', *Scientific American*, 284(5), pp. 34–43.

Bertaud-Gounot, V., Duvauferrier, R. and Burgun, A. (2012a) 'Ontology and medical diagnosis.', *Informatics for health & social care*, 37(2), pp. 51–61.

Bertaud-Gounot, V., Duvauferrier, R. and Burgun, A. (2012b) 'Ontology and medical diagnosis.', *Informatics for health & social care*, 37(2), pp. 51–61.

Bertaud Gounot, V. *et al.* (2011a) 'Creating an ontology driven rules base for an expert system for medical diagnosis.', *Studies in health technology and informatics*, 169, pp. 714–8.

Bertaud Gounot, V. *et al.* (2011b) 'Creating an ontology driven rules base for an expert system for medical diagnosis.', *Studies in health technology and informatics*, 169, pp. 714–8.

Bettin, P. and Di Matteo, F. (2013) 'Glaucoma: present challenges and future trends.', *Ophthalmic research*, 50(4), pp. 197–208.

Bichindaritz, I. (2006) 'Mémoire: A framework for semantic interoperability of case-based reasoning systems in biology and medicine.', *Artificial intelligence in medicine*, 36(2), pp. 177–92.

Bizer, C. (2009) 'The Emerging Web of Linked Data', *IEEE Intelligent Systems*, 24(5), pp. 87–92. doi: 10.1109/MIS.2009.102.

Bizer, C. *et al.* (2012) 'The meaningful use of big data', *ACM SIGMOD Record*. ACM, 40(4), p. 56. doi: 10.1145/2094114.2094129.

Bizer, C. and Cyganiak, R. (2007) *D2RQ — Lessons Learned*, *W3C Workshop on RDF Access to Relational Databases*. Cambridge, MA, USA.

Bizer, C., Heath, T. and Berners-Lee, T. (2009) 'Linked Data - The Story So Far', *International Journal on Semantic Web and Information Systems*, 5(3), pp. 1–22. doi: 10.4018/jswis.2009081901.

Bizer, C. and Seaborne, A. (2004) 'D2RQ - treating non-RDF databases as virtual RDF graphs', in *Proceedings of the 3rd international semantic web conference (ISWC2004)*.

Bizer, C. and Seaborne, A. (2004) 'D2RQ - treating non-RDF databases as virtual RDF graphs', in *ISWC 2004*. Hiroshima.

Blomqvist, E. (2014) 'The use of Semantic Web technologies for decision support – a survey', *Semantic Web*. IOS Press, pp. 177–201. doi: 10.3233/SW-2012-0084.

Boley, H. (2006) 'The RuleML Family of Web Rule Languages', in. Springer Berlin Heidelberg, pp. 1–17. doi: 10.1007/11853107_1.

Bond, W. F. *et al.* (2012a) 'Differential diagnosis generators: an evaluation of currently available computer programs.', *Journal of general internal medicine*, 27(2), pp. 213–9. doi: 10.1007/s11606-011-1804-8.

Bond, W. F. *et al.* (2012b) 'Differential diagnosis generators: an evaluation of currently available computer programs.', *Journal of general internal medicine*. Springer, 27(2), pp. 213–9. doi: 10.1007/s11606-011-1804-8.

Booth, D. (2014a) *The Yosemite Project An RDF Roadmap for Healthcare Information Interoperability - Part 1 - YouTube*, *YouTube.com*. Available at: https://www.youtube.com/watch?v=krsPw2Oc63g (Accessed: 19 February 2017).

Booth, D. (2014b) *Why RDF as a Universal Healthcare Exchange Language?*, *http://dbooth.org/*. Available at: http://dbooth.org/2014/why-rdf/ (Accessed: 9 May 2016).

Booth, D. (2015a) *Different source formats, same RDF*, *yosemiteproject.org*. Available at: http://yosemiteproject.org/wp-content/uploads/2016/09/Different-source-formats-same-RDF.jpg.

Booth, D. (2015b) *Translating patient data*, *yosemiteproject.org*. yosemiteproject.org. Available at: http://yosemiteproject.org/interoperability-roadmap/.

Booth, D. (2015c) *Yosemite Project – Part 5 Introduction and RDF Representation of FHIR for Clinical Data*, *YouTube.com*. Available at: https://www.youtube.com/watch?v=G2YI21AXoDU (Accessed: 9 May 2016).

Booth, D. (2015d) *Yosemite Project | Interoperability Roadmap*, *yosemiteproject.org*. Available at: http://yosemiteproject.org/interoperability-roadmap/ (Accessed: 24 February 2017).

Bordage, G. (1999) 'Why did I miss the diagnosis? Some cognitive explanations and educational implications.', *Academic medicine : journal of the Association of American Medical Colleges*, 74(10 Suppl), pp. S138-43.

Borok, L. S. (1995) 'The use of relational databases in health care information systems.', *Journal of health care finance*, 21(4), pp. 6–12.

Boshuizen, H. P. A. and Schmidt, H. G. (1992) 'On the Role of Biomedical Knowledge in Clinical Reasoning by Experts, Intermediates and Novices', *Cognitive Science*, 16(2), pp. 153–184. doi: 10.1207/s15516709cog1602_1.

Bouamrane, M.-M., Rector, A. and Hurrell, M. (2008) 'Gathering Precise Patient Medical History with an Ontology-Driven Adaptive Questionnaire', in *2008 21st IEEE International Symposium on Computer-Based Medical Systems*. IEEE, pp. 539–541. doi: 10.1109/CBMS.2008.24.

Boudoulas, K. D. *et al.* (2015) 'The shortcomings of clinical practice guidelines.', *Cardiology*. Karger Publishers, 130(3), pp. 187–200. doi: 10.1159/000371572.

Bouhaddou, O., Lambert, J. G. and Morgan, G. E. (1995) 'Iliad and Medical HouseCall: evaluating the impact of common sense knowledge on the diagnostic accuracy of a medical expert system.', *Proceedings / the ... Annual Symposium on Computer Application [sic] in Medical Care. Symposium on Computer Applications in Medical Care*, pp. 742–6.

Boxwala, A. A. *et al.* (2004) 'GLIF3: a representation format for sharable computer-interpretable clinical practice guidelines', *Journal of Biomedical Informatics*, 37(3), pp. 147–161. doi: 10.1016/j.jbi.2004.04.002.

Boyd, K. (2017) *What Is Ocular Hypertension? - American Academy of Ophthalmology*, *www.aao.org*. Available at: https://www.aao.org/eye-health/diseases/what-is-ocular-hypertension (Accessed: 27 March 2017).

Brachman, R. J. and Schmolze, J. G. (1985) 'An Overview of the KL-ONE Knowledge Representation System', *Cognitive science*, 9(2), pp. 171–216.

Bradford, A. *et al.* (2009) 'Missed and delayed diagnosis of dementia in primary care: prevalence and contributing factors.', *Alzheimer disease and associated disorders*. NIH Public Access, 23(4), pp. 306–14. doi: 10.1097/WAD.0b013e3181a6bebc.

Brickley, D. and Guha, R. V. (2014) *RDF Schema 1.1*, *w3.org*. Available at: https://www.w3.org/TR/rdf-schema/ (Accessed: 17 December 2016).

Brickley, D. and Miller, L. (2014) *FOAF Vocabulary Specification*, *xmlns.com/foaf/spec/*. Available at: http://xmlns.com/foaf/spec/ (Accessed: 18 February 2017).

Brooke, J. (1996) 'SUS-A quick and dirty usability scale', *Usability evaluation in industry*. London, 189(194), pp. 4–7.

BRUNSWIK, E. (1955) 'Representative design and probabilistic theory in a functional psychology.', *Psychological review*, 62(3), pp. 193–217.

Bryan, C. and Boren, S. A. (2008) 'The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature.', *Informatics in primary care*, 16(2), pp. 79–91.

Bucci, G., Sandrucci, V. and Vicario, E. (2011) 'Ontologies and Bayesian Networks in Medical Diagnosis', in *2011 44th Hawaii International Conference on System Sciences*. IEEE, pp. 1–8. doi: 10.1109/HICSS.2011.333.

Buil-Aranda, C. *et al.* (2013) 'SPARQL Web-Querying Infrastructure: Ready for Action?', in Alani, H. et al. (eds) *The Semantic Web – ISWC 2013*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science).

Bury, J. *et al.* (2004) 'A quantitative and qualitative evaluation of LISA, a decision support system for chemotherapy dosing in childhood acute lymphoblastic leukaemia.', *Studies in health technology and informatics*, 107(Pt 1), pp. 197–201.

Bury, J. *et al.* (2005) 'LISA: a web-based decision-support system for trial management of childhood acute lymphoblastic leukaemia', *British Journal of Haematology*, 129(6), pp. 746–754. doi: 10.1111/j.1365-2141.2005.05541.x.

Cahan, A. *et al.* (2003) 'Probabilistic reasoning and clinical decision-making: do doctors overestimate diagnostic probabilities?', *QJM : monthly journal of the Association of Physicians*, 96(10), pp. 763–9.

Campbell, K. E., Oliver, D. E. and Shortliffe, E. H. (1998) 'The Unified Medical Language System: toward a collaborative approach for solving terminologic problems.', *Journal of the American Medical Informatics Association : JAMIA*, 5(1), pp. 12–6.

Carroll, J. J. *et al.* (2004) 'Jena: implementing the semantic web recommendations', in *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters - WWW Alt. '04*. New York, New York, USA: ACM Press, pp. 74–83.

Carrotte, P. (2004) 'Endodontics: Part 2 Diagnosis and treatment planning.', *British dental journal*. Nature Publishing Group, 197(5), pp. 231–8.

Ceusters, W. and Smith, B. (2010) 'A unified framework for biomedical terminologies and ontologies.', *Studies in health technology and informatics*, 160(Pt 2), pp. 1050–4.

Chang-Shing Lee, Mei-Hui Wang and Hagras, H. (2010) 'A Type-2 Fuzzy Ontology and its Application to Personal Diabetic Diet Recommendation', *IEEE Transactions on Fuzzy Systems*, 18(2), pp. 374–395. doi: 10.1109/TFUZZ.2010.2042454.

Chang, Y.-S., Hung, W.-C. and Juang, T.-Y. (2013) 'Depression Diagnosis Based on Ontologies and Bayesian Networks', in *2013 IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, pp. 3452–3457.

Chankai, I., Prestrud, A. A. and Brooks, A. D. (2005) 'Converting Semi-structured Clinical Medical Records into Information and Knowledge', in *21st International Conference on Data Engineering Workshops (ICDEW'05)*. IEEE, pp. 1162–1162.

Charlin, B. *et al.* (2007) 'Scripts and clinical reasoning', *Medical Education*, 41(12), pp. 1178–1184. doi: 10.1111/j.1365-2923.2007.02924.x.

Charlin, B., Tardif, J. and Boshuizen, H. P. (2000) 'Scripts and medical diagnostic knowledge: theory and applications for clinical reasoning instruction and research.', *Academic medicine : journal of the Association of American Medical Colleges*, 75(2), pp. 182–90.

Chen, Q. *et al.* (2016) 'An ontology-driven, case-based clinical decision support model for removable partial denture design', *Scientific Reports*, 6(1), p. 27855. doi: 10.1038/srep27855.

Cimino, J. J. and Zhu, X. (2006a) 'The practical impact of ontologies on biomedical informatics.', *Yearbook of medical informatics*, pp. 124–35.

Cimino, J. J. and Zhu, X. (2006b) 'The practical impact of ontologies on biomedical informatics.', *Yearbook of medical informatics*, pp. 124–35.

de Clercq, P. A. *et al.* (2004) 'Approaches for creating computer-interpretable guidelines that facilitate decision support.', *Artificial intelligence in medicine*, 31(1), pp. 1–27.

College of Optometrists (2016) *Clinical Management Guidelines*, *http://www.college-optometrists.org/*. Available at: http://www.college-optometrists.org/en/professional-standards/clinical_management_guidelines/guidelines_by_condition.cfm.

College of Optometrists and Royal College of Ophthalmologists (2010) *Joint Supplementary College Guidance on Supervision in relation to Glaucoma-related Care by Optometrists*, *https://www.rcophth.ac.uk*. Available at: https://www.rcophth.ac.uk/professional-resources/glaucoma/.

Corcoglioniti, F. *et al.* (2015) 'Processing billions of RDF triples on a single machine using streaming and sorting', in *Proceedings of the 30th Annual ACM Symposium on Applied Computing - SAC '15*. New York, New York, USA: ACM Press, pp. 368–375. doi: 10.1145/2695664.2695720.

Corrigan, D. (2015) 'An ontology driven clinical evidence service providing diagnostic decision support in family practice.', *AMIA Joint Summits on Translational Science proceedings. AMIA Joint Summits on Translational Science*. American Medical Informatics Association, 2015, pp. 440–4.

Corsar, D., Edwards, P. and Nelson, J. (2013) 'Personal privacy and the web of linked data', in *Proceedings of the 2013th International Conference on Society, Privacy and the Semantic Web - Policy and Technology - Volume 1121*. Sydney, Australia: CEUR-WS.org, pp. 11–21.

COSSAC (2011) *The Tallis Toolset*, *archive.cossac.org/tallis/*. Available at: http://archive.cossac.org/tallis/.

Coulson, A. S. *et al.* (2001) 'RAGs: A novel approach to computerized genetic risk assessment and decision support from pedigrees.', *Methods of information in medicine*, 40(4), pp. 315–22.

Crespo, K. E., Torres, J. E. and Recio, M. E. (2004) 'Reasoning process characteristics in the diagnostic skills of beginner, competent, and expert dentists.', *Journal of dental education*, 68(12), pp. 1235–44.

Croskerry, P. (2002) 'Achieving quality in clinical decision making: cognitive strategies and detection of bias.', *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, 9(11), pp. 1184–204.

Croskerry, P. (2003a) 'Cognitive forcing strategies in clinical decisionmaking.', *Annals of emergency medicine*, 41(1), pp. 110–20. doi: 10.1067/mem.2003.22.

Croskerry, P. (2003b) 'Cognitive forcing strategies in clinical decisionmaking.', *Annals of emergency medicine*, 41(1), pp. 110–20.

Croskerry, P. (2003c) 'The importance of cognitive errors in diagnosis and strategies to minimize them.', *Academic medicine : journal of the Association of American Medical Colleges*, 78(8), pp. 775–80.

Croskerry, P., Singhal, G. and Mamede, S. (2013a) 'Cognitive debiasing 1: origins of bias and theory of debiasing.', *BMJ quality & safety*, 22 Suppl 2, pp. ii58-ii64. doi: 10.1136/bmjqs-2012-001712.

Croskerry, P., Singhal, G. and Mamede, S. (2013b) 'Cognitive debiasing 1: origins of bias and theory of debiasing.', *BMJ quality & safety*, 22 Suppl 2, pp. ii58-ii64.

Croskerry, P., Singhal, G. and Mamede, S. (2013c) 'Cognitive debiasing 2: impediments to and strategies for change.', *BMJ quality & safety*, 22 Suppl 2, pp. ii65-ii72. doi: 10.1136/bmjqs-2012-001713.

Croskerry, P., Singhal, G. and Mamede, S. (2013d) 'Cognitive debiasing 2: impediments to and strategies for change.', *BMJ quality & safety*, 22 Suppl 2, pp. ii65-ii72.

Curtis, K., Bollard, L. and Dickson, C. (2002) 'Coding errors and the trauma patient--is nursing case management the solution?', *Australian health review : a publication of the Australian Hospital Association*, 25(2000), pp. 73–80.

Custers, E. J., Boshuizen, H. P. and Schmidt, H. G. (1996) 'The influence of medical expertise, case typicality, and illness script component on case processing and disease probability estimates.', *Memory & cognition*, 24(3), pp. 384–99.

Cyganiak, R. *et al.* (2012) *The D2RQ Mapping Language | The D2RQ Platform*, *d2rq.org*. Available at: http://d2rq.org/d2rq-language.

Cyganiak, R. (2014) *The Linking Open Data cloud diagram*, *http://lod-cloud.net/*. Available at: http://lod-cloud.net/.

D'Aquin, M. *et al.* (2008) 'Toward a New Generation of Semantic Web Applications', *IEEE Intelligent Systems*, 23(3), pp. 20–28. doi: 10.1109/MIS.2008.54.

Dailey, C. A. (1952) 'The Effects of Premature Conclusion Upon the Acquisition of Understanding of a Person', *The Journal of Psychology*. Taylor & Francis Group , 33(1), pp. 133–152. doi: 10.1080/00223980.1952.9712825.

Dawes, R. M., Faust, D. and Meehl, P. E. (1989) 'Clinical versus actuarial judgment.', *Science (New York, N.Y.)*, 243(4899), pp. 1668–74.

Van de, A. H. and Delbecq, A. L. (1974) 'The Effectiveness of Nominal, Delphi, and Interacting Group Decision Making Processes.', *Academy of Management Journal*. Academy of Management, 17(4), pp. 605–621. doi: 10.2307/255641.

Diagnosis Pro® (2014) *DiagnosisPro - The Ultimate Medical & Differential Diagnosis Reminder Tool*, *en.diagnosispro.com/*. Available at: http://en.diagnosispro.com/ (Accessed: 1 August 2014).

DoH (2013) *The UK Strategy for Rare Diseases*.

de Dombal, F. T. *et al.* (1972) 'Computer-aided Diagnosis of Acute Abdominal Pain', *BMJ*, 2(5804), pp. 9–13.

Donfack Guefack, V. *et al.* (2012) 'Ontology driven decision support systems for medical diagnosis - an interactive form for consultation in patients with plasma cell disease.', *Studies in health technology and informatics*, 180, pp. 108–12.

Douali, N., De Roo, J. and Jaulent, M.-C. (2012) 'Decision support system based semantic web for personalized patient care.', *Studies in health technology and informatics*, 180, pp. 1203–5.

Dugdale, D. C., Epstein, R. and Pantilat, S. Z. (1999) 'Time and the patient-physician relationship.', *Journal of general internal medicine*. Springer, 14 Suppl 1(Suppl 1), pp. S34-40. doi: 10.1046/J.1525-1497.1999.00263.X.

Edgar, D. *et al.* (2010) 'Referral behaviour among optometrists: increase in the number of referrals from optometrists following the publication of the April 2009 NICE guidelines for the diagnosis and management of COAG and OHT in England and Wales and its implications', *Optometry in Practice*, 11(1), pp. 33–38.

El-Kareh, R., Hasan, O. and Schiff, G. D. (2013) 'Use of health information technology to reduce diagnostic errors.', *BMJ quality & safety*, 22 Suppl 2, pp. ii40-ii51. doi: 10.1136/bmjqs-2013-001884.

Elstein, A. S., Schwartz, A. and Schwarz, A. (2002) 'Clinical problem solving and diagnostic decision making: selective review of the cognitive literature.', *BMJ (Clinical research ed.)*. BMJ Group, 324(7339), pp. 729–32.

Elstein, A. S., Shulman, L. S. and Sprafka, S. A. (1978) *Medical problem solving : an analysis of clinical reasoning*.

Ely, J. W., Graber, M. L. and Croskerry, P. (2011) 'Checklists to reduce diagnostic errors.', *Academic medicine : journal of the Association of American Medical Colleges*, 86(3), pp. 307–13. doi: 10.1097/ACM.0b013e31820824cd.

Ely, J. W., Kaldjian, L. C. and D'Alessandro, D. M. (2012a) 'Diagnostic errors in primary care: lessons learned.', *Journal of the American Board of Family Medicine : JABFM*, 25(1), pp. 87–97. doi: 10.3122/jabfm.2012.01.110174.

Ely, J. W., Kaldjian, L. C. and D'Alessandro, D. M. (2012b) 'Diagnostic errors in primary care: lessons learned.', *Journal of the American Board of Family Medicine : JABFM*, 25(1), pp. 87–97. doi: 10.3122/jabfm.2012.01.110174.

Emery, J. *et al.* (1999) 'Computer support for recording and interpreting family histories of breast and ovarian cancer in primary care (RAGs): qualitative evaluation with simulated patients.', *BMJ (Clinical research ed.)*, 319(7201), pp. 32–6.

Emery, J. *et al.* (2000) 'Computer support for interpreting family histories of breast and ovarian cancer in primary care: comparative study with simulated cases.', *BMJ (Clinical research ed.)*. British Medical Journal Publishing Group, 321(7252), pp. 28–32. doi: 10.1136/BMJ.321.7252.28.

EN 13606 Association (2016) *Semantic interoperability of health information*, *en13606.org*. Available at: http://www.en13606.org/the-ceniso-en13606-standard/semantic-interoperability (Accessed: 9 May 2016).

Epner, P. L., Gans, J. E. and Graber, M. L. (2013) 'When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine.', *BMJ quality & safety*. BMJ Publishing Group Ltd, 22 Suppl 2(Suppl 2), pp. ii6-ii10. doi: 10.1136/bmjqs-2012-001621.

Espinosa, J. A. and Nolan, T. W. (2000) 'Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study.', *BMJ (Clinical research ed.)*, 320(7237), pp. 737–40.

European Glaucoma Society (2008) *Terminology and guidelines for glaucoma*. PubliComm.

Evans, J. S. B. T. and Stanovich, K. E. (2013) 'Dual-Process Theories of Higher Cognition', *Perspectives on Psychological Science*, 8(3), pp. 223–241. doi: 10.1177/1745691612460685.

Evans, R. S., Lloyd, J. F. and Pierce, L. A. (2012) 'Clinical use of an enterprise data warehouse.', *AMIA ... Annual Symposium proceedings. AMIA Symposium*. American Medical Informatics Association, 2012, pp. 189–98.

Farooq, K. *et al.* (2012) 'An Ontology Driven and Bayesian Network Based Cardiovascular Decision Support Framework', in *Advances in Brain Inspired Cognitive Systems*. Springer Berlin Heidelberg, pp. 31–41. doi: 10.1007/978-3-642-31561-9_4.

Fellbaum, C. (2010) 'WordNet', in Poli, R., Healy, M., and Kameas, A. (eds) *Theory and Applications of Ontology: Computer Applications SE - 10*. Springer Netherlands, pp. 231–243. doi: 10.1007/978-90-481-8847-5_10.

Fernández, S. *et al.* (2016) *SPARQL Endpoint interface to Python*, *rdflib.github.io/sparqlwrapper/*. Available at: https://github.com/RDFLib/sparqlwrapper.

flickr.com (2016) *Flickr Services*, *www.flickr.com/services/api/*. Available at: https://www.flickr.com/services/api/ (Accessed: 1 January 2017).

Foley, T. J. and Vale, L. (2017) 'What role for learning health systems in quality improvement within healthcare providers?', *Learning Health Systems*. doi: 10.1002/lrh2.10025.

Forrest, C. B. *et al.* (2013) 'Improving adherence to otitis media guidelines with clinical decision support and physician feedback.', *Pediatrics*, 131(4), pp. e1071-81. doi: 10.1542/peds.2012-1988.

Fossum, M. *et al.* (2011) 'An evaluation of the usability of a computerized decision support system for nursing homes.', *Applied clinical informatics*. Schattauer Publishers, 2(4), pp. 420–36. doi: 10.4338/ACI-2011-07-RA-0043.

Fox, J. *et al.* (2007) 'Argumentation-Based Inference and Decision Making--A Medical Perspective', *IEEE Intelligent Systems*, 22(6), pp. 34–41. doi: 10.1109/MIS.2007.102.

Fox, J. *et al.* (2015) 'OpenClinical.net: A platform for creating and sharing knowledge and promoting best practice in healthcare', *Computers in Industry*, 66, pp. 63–72. doi: 10.1016/j.compind.2014.10.001.

Fox, J. (2015) *Prof John Fox OpenClinical.net - YouTube*, *www.youtube.com*. Available at: https://www.youtube.com/watch?v=opc77Iqlk8I (Accessed: 4 October 2017).

Fox, J., Johns, N. and Rahmanzadeh, A. (1998) 'Disseminating medical knowledge: the PROforma approach.', *Artificial intelligence in medicine*, 14(1–2), pp. 157–81.

Fox, J., Krause, P. J. and Elvang-Gøransson, M. (1993) 'Argumentation as a General Framework for Uncertain Reasoning', in *UAI'93 Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*. Washington, DC, pp. 428–434.

Fox, J. and Parsons, S. (1998) 'Arguing about beliefs and actions', in *Applications of Uncertainty Formalisms*. Springer, Berlin, Heidelberg, pp. 266–302. doi: 10.1007/3-540-49426-X_13.

Fox, J., Patkar, V. and Thomson, R. (2006a) 'Decision support for health care: the PROforma evidence base.', *Informatics in primary care*, 14(1), pp. 49–54.

Fox, J., Patkar, V. and Thomson, R. (2006b) 'Decision support for health care: the PROforma evidence base.', *Informatics in primary care*, 14(1), pp. 49–54.

Friedman, C. P., Wong, A. K. and Blumenthal, D. (2010) 'Achieving a nationwide learning health system.', *Science translational medicine*. American Association for the Advancement of Science, 2(57), p. 57cm29. doi: 10.1126/scitranslmed.3001456.

Frith, C. D. (2007) *Making up the mind : how the brain creates our mental world*. Blackwell Pub.

Fürber, C. and Hepp, M. (2010) 'Using SPARQL and SPIN for Data Quality Management on the Semantic Web', in *Business Information Systems*. Springer Berlin Heidelberg, pp. 35–46. doi: 10.1007/978-3-642-12814-1_4.

Gao, Y. *et al.* (2006) 'SWAN: A distributed knowledge infrastructure for Alzheimer disease research', *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(3), pp. 222–228. doi: 10.1016/j.websem.2006.05.006.

García-Crespo, Á. *et al.* (2010a) 'ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements', *Expert Systems with Applications*. Pergamon Press, Inc., 37(3), pp. 2621–2628. doi: 10.1016/j.eswa.2009.08.016.

García-Crespo, Á. *et al.* (2010b) 'ODDIN: Ontology-driven differential diagnosis based on logical inference and probabilistic refinements', *Expert Systems with Applications*. Pergamon Press, Inc., 37(3), pp. 2621–2628.

Garg, A. X. *et al.* (2005a) 'Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review.', *JAMA : the journal of the American Medical Association*. American Medical Association, 293(10), pp. 1223–38. doi: 10.1001/jama.293.10.1223.

Garg, A. X. *et al.* (2005b) 'Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review.', *JAMA : the journal of the American Medical Association*, 293(10), pp. 1223–38.

Gearon, P., Passant, A. and Polleres, A. (2013) *SPARQL 1.1 Update*, *w3.org/TR/sparql11-update*. Available at: https://www.w3.org/TR/sparql11-update/ (Accessed: 28 December 2016).

Gennari, J. H. *et al.* (2003) 'The evolution of Protégé: an environment for knowledge-based systems development', *International Journal of Human-Computer Studies*, 58(1), pp. 89–123. doi: 10.1016/S1071-5819(02)00127-1.

Glasspool, D. W. *et al.* (2003) 'Interactive Decision Support for Medical Planning', in *Artificial Intelligence in Medicine*. Springer, Berlin, Heidelberg, pp. 335–339. doi: 10.1007/978-3-540-39907-0_45.

Glickman, G. N. (2009) 'AAE Consensus Conference on Diagnostic Terminology: background and perspectives.', *Journal of endodontics*, 35(12), pp. 1619–20.

Goldberg, L. J. *et al.* (2005) 'The Significance of SNODENT.', *Studies in health technology and informatics*, 116, pp. 737–42.

Google Forms (2016) *Google Forms - create and analyze surveys, for free.*, *Google*. Available at: https://www.google.co.uk/forms/about/.

Google Sheets (2016) *Google Sheets - create and edit spreadsheets online, for free.*, *Google*. Available at: https://www.google.co.uk/sheets/about/.

Gopikrishna, V., Pradeep, G. and Venkateshbabu, N. (2009) 'Assessment of pulp vitality: a review.', *International journal of paediatric dentistry / the British Paedodontic Society [and] the International Association of Dentistry for Children*, 19(1), pp. 3–15.

Graber, M. (2005) 'Diagnostic errors in medicine: a case of neglect.', *Joint Commission journal on quality and patient safety / Joint Commission Resources*, 31(2), pp. 106–13.

Graber, M. A. and VanScoy, D. (2003) 'How well does decision support software perform in the emergency department?', *Emergency medicine journal : EMJ*, 20(5), pp. 426–8.

Graber, M., Gordon, R. and Franklin, N. (2002a) 'Reducing diagnostic errors in medicine: what's the goal?', *Academic medicine : journal of the Association of American Medical Colleges*, 77(10), pp. 981–92.

Graber, M., Gordon, R. and Franklin, N. (2002b) 'Reducing diagnostic errors in medicine: what's the goal?', *Academic medicine : journal of the Association of American Medical Colleges*, 77(10), pp. 981–92.

Graber, M. L. *et al.* (2012a) 'Cognitive interventions to reduce diagnostic error: a narrative review.', *BMJ quality & safety*. BMJ Publishing Group Ltd, 21(7), pp. 535–57. doi: 10.1136/bmjqs-2011-000149.

Graber, M. L. *et al.* (2012b) 'Cognitive interventions to reduce diagnostic error: a narrative review.', *BMJ quality & safety*, 21(7), pp. 535–57.

Graber, M. L. (2013) 'The incidence of diagnostic error in medicine.', *BMJ quality & safety*, 22 Suppl 2, pp. ii21-ii27. doi: 10.1136/bmjqs-2012-001615.

Graber, M. L., Franklin, N. and Gordon, R. (2005a) 'Diagnostic error in internal medicine.', *Archives of internal medicine*, 165(13), pp. 1493–9.

Graber, M. L., Franklin, N. and Gordon, R. (2005b) 'Diagnostic error in internal medicine.', *Archives of internal medicine*, 165(13), pp. 1493–9. doi: 10.1001/archinte.165.13.1493.

Graber, M. L. and Mathew, A. (2008a) 'Performance of a web-based clinical diagnosis support system for internists.', *Journal of general internal medicine*, 23 Suppl 1, pp. 37–40.

Graber, M. L. and Mathew, A. (2008b) 'Performance of a web-based clinical diagnosis support system for internists.', *Journal of general internal medicine*, 23 Suppl 1, pp. 37–40. doi: 10.1007/s11606-007-0271-8.

Grace, A. *et al.* (2013) 'Evaluating the effectiveness of clinical decision support systems: the case of multimorbidity care', *Journal of Decision Systems*, 22(2), pp. 97–108. doi: 10.1080/12460125.2013.780320.

Grando, M. A. *et al.* (2013) 'Argumentation-logic for creating and explaining medical hypotheses', *Artificial Intelligence in Medicine*. Elsevier, 58(1), pp. 1–13. doi: 10.1016/J.ARTMED.2013.02.003.

Greenough, A. (2002) 'Help from ISABEL for paediatric diagnoses.', *Lancet*, 360(9341), p. 1259.

Grenon, P., Smith, B. and Goldberg, L. (2004) 'Biodynamic ontology: applying BFO in the biomedical domain.', *Studies in health technology and informatics*, 102, pp. 20–38.

Griggs, R. C. *et al.* (2009) 'Clinical research for rare disease: Opportunities, challenges, and solutions', *Molecular Genetics and Metabolism*, 96(1), pp. 20–26. doi: 10.1016/j.ymgme.2008.10.003.

Grosan, C. and Abraham, A. (2011) 'Rule-Based Expert Systems', in *Intelligent Systems*. Springer Berlin Heidelberg, pp. 149–185. doi: 10.1007/978-3-642-21004-4_7.

Groves, M., O'Rourke, P. and Alexander, H. (2003) 'The clinical reasoning characteristics of diagnostic experts.', *Medical teacher*, 25(3), pp. 308–13. doi: 10.1080/0142159031000100427.

Gruber, T. R. (1993) 'A translation approach to portable ontology specifications', *Knowledge Acquisition*, 5(2), pp. 199–220. doi: 10.1006/knac.1993.1008.

Guarino, N. (1997) 'Semantic matching: Formal ontological distinctions for information organization, extraction, and integration', in Pazienza, M. (ed.) *Information Extraction A Multidisciplinary Approach to an Emerging Information Technology SE - 8*. Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 139–170. doi: 10.1007/3-540-63438-X_8.

Gunn, A. A. (1991) 'The acute abdomen: the role of computer-assisted diagnosis.', *Baillière's clinical gastroenterology*, 5(3 Pt 1), pp. 639–65.

Gutteridge, C. (2012a) *Graphite PHP Linked Data Library*, *graphite.ecs.soton.ac.uk/*. Available at: http://graphite.ecs.soton.ac.uk/.

Gutteridge, C. (2012b) *SPARQL RDF Library for PHP*, *graphite.ecs.soton.ac.uk/*. Available at: http://graphite.ecs.soton.ac.uk/sparqllib/.

Hadzic, M. and Chang, E. (2005) 'Ontology-based Multi-agent Systems Support Human Disease Study and Control', in Czap, H. et al. (eds) *Proceedings of the 2005 conference on Self-Organization and Autonomic Informatics (I)*. Amsterdam: IOS Press, pp. 129–141.

Harbison, J. (1991) 'Clinical decision making in nursing', *Journal of Advanced Nursing*. Blackwell Publishing Ltd, 16(4), pp. 404–407. doi: 10.1111/j.1365-2648.1991.tb03429.x.

Hargreaves, K. M., Cohen, S. and Berman, L. H. (2011) *Cohen's Pathways of the Pulp*. Mosby Elsevier.

Harman, L. B., Flite, C. A. and Bond, K. (2012) 'Electronic Health Records: Privacy, Confidentiality, and Security', *Virtual Mentor*. Amer Med Assoc, 14(9), pp. 712–719. doi: 10.1001/virtualmentor.2012.14.9.stas1-1209.

Harmelen, F. Van and McGuinness, D. (2004) *OWL Web Ontology Language Overview*, *w3.org*. Available at: https://www.w3.org/TR/owl-features/ (Accessed: 31 January 2016).

Hartig, O., Bizer, C. and Freytag, J.-C. (2009) 'Executing SPARQL Queries over the Web of Linked Data', in *The Semantic Web - ISWC 2009*. Chantilly, VA, USA: Springer Berlin Heidelberg, pp. 293–309. doi: 10.1007/978-3-642-04930-9_19.

Hasan, S. A. *et al.* (2015) 'Ontology-Driven Semantic Search for Brazilian Portuguese Clinical Notes.', *Studies in health technology and informatics*, 216, p. 1022.

Hassanzadeh, O. and Miller, R. J. (2015) 'Automatic Curation of Clinical Trials Data in LinkedCT', in *The Semantic Web - ISWC 2015*. Springer International Publishing, pp. 270–278. doi: 10.1007/978-3-319-25010-6_16.

Haug, P. J. *et al.* (2013) 'An ontology-driven, diagnostic modeling system.', *Journal of the American Medical Informatics Association : JAMIA*, 20(e1), pp. e102-10.

Hayes-Roth, F., Waterman, D. A. and Lenat, D. B. (1983) 'Building expert systems'. Addison-Wesley Longman Publishing Co., Inc.

Hayes, P. J. and Patel-Schneider, P. F. (2014) *RDF Schema 1.1- rdfs:label*, *www.w3.org*. Available at: https://www.w3.org/TR/rdf-schema/#ch_label.

Haynes, W. (2013) 'Wilcoxon Rank Sum Test', in *Encyclopedia of Systems Biology*. New York, NY: Springer New York, pp. 2354–2355. doi: 10.1007/978-1-4419-9863-7_1185.

Heath, T. and Bizer, C. (2011) 'Linked Data: Evolving the Web into a Global Data Space', *Synthesis Lectures on the Semantic Web: Theory and Technology*, 1(1), pp. 1–136. doi: 10.2200/S00334ED1V01Y201102WBE001.

van Heijst, G., Schreiber, A. T. and Wielinga, B. J. (1997) 'Using explicit ontologies in KBS development', *International Journal of Human-Computer Studies*, 46(2–3), pp. 183–292.

Hendler, J. (2001) 'Agents and the Semantic Web', *IEEE Intelligent Systems*. IEEE, 16(2), pp. 30–37.

Heneghan, C. *et al.* (2009) 'Diagnostic strategies used in primary care', *BMJ*, 338.

HL7 (2015) *Rdf - FHIR v1.0.2*, *www.hl7.org*. Available at: https://www.hl7.org/fhir/rdf.html.

HL7 (2016) *Index - FHIR v1.0.2*, *www.hl7.org*. Available at: https://www.hl7.org/fhir/.

Hmelo, C. E. (1998) 'Cognitive Consequences of Problem-Based Learning for the Early Development of Medical Expertise', *Teaching and Learning in Medicine*. Routledge, 10(2), pp. 92–100. doi: 10.1207/S15328015TLM1002_7.

Hoekstra, R. *et al.* (2015) 'Hubble: Linked Data Hub for Clinical Decision Support', in *The Semantic Web: ESWC 2012 Satellite Events*. Springer Berlin Heidelberg, pp. 458–462. doi: 10.1007/978-3-662-46641-4_45.

Hogan, W. R. and Wagner, M. M. (1997) 'Accuracy of Data in Computer-based Patient Records', *Journal of the American Medical Informatics Association*, 4, pp. 342–355. doi: 10.1136/jamia.1997.0040342.

Hogeveen, S. E. *et al.* (2012) 'Comparison of international breast cancer guidelines: are we globally consistent? cancer guideline AGREEment.', *Current oncology (Toronto, Ont.)*. Multimed Inc., 19(3), pp. e184-90. doi: 10.3747/co.19.930.

HOLT, A. *et al.* (2005) 'Medical applications in case-based reasoning', *The Knowledge Engineering Review*. Cambridge University Press, 20(03), p. 289. doi: 10.1017/S0269888906000622.

Horrocks, I. *et al.* (2004) *SWRL: A Semantic Web Rule Language Combining OWL and RuleML*, *w3.org/Submission*. Available at: https://www.w3.org/Submission/SWRL/.

Horst, O. V *et al.* (2015) 'Prevalence of pain in the orofacial regions in patients visiting general dentists in the Northwest Practice-based REsearch Collaborative in Evidence-based DENTistry research network.', *Journal of the American Dental Association (1939)*. Elsevier, 146(10), p. 721–8.e3. doi: 10.1016/j.adaj.2015.04.001.

Horton, P., Kiely, P. M. and Chakman, J. (2006) 'The Australian optometric workforce 2005', *Clinical and Experimental Optometry*. Blackwell Publishing Asia, 89(4), pp. 229–240. doi: 10.1111/j.1444-0938.2006.00048.x.

Horvitz, E. J., Breese, J. S. and Henrion, M. (1988) 'Decision theory in expert systems and artificial intelligence', *International Journal of Approximate Reasoning*. Elsevier, 2(3), pp. 247–302. doi: 10.1016/0888-613X(88)90120-X.

Howard, J. *et al.* (2006) 'Reducing missed injuries at a level II trauma center.', *Journal of trauma nursing : the official journal of the Society of Trauma Nurses*, 13(3), pp. 89–95.

Hripcsak, G. (1994) 'Writing Arden Syntax Medical Logic Modules.', *Computers in biology and medicine*, 24(5), pp. 331–63.

Hu, B. *et al.* (2011) 'The HealthAgents ontology: knowledge representation in a distributed decision support system for brain tumours', *The Knowledge Engineering Review*, 26(03), pp. 303–328.

Huang, Z. *et al.* (2013) 'Knowledge-Based Patient Data Generation', in Riaño, D. et al. (eds) *Process Support and Knowledge Representation in Health Care*. Springer International Publishing, pp. 83–96. doi: 10.1007/978-3-319-03916-9_7.

Huff, S. M. *et al.* (2014) '17.7 Iso-semantic models', in Greenes, R. A. (ed.) *Clinical decision support : the road to broad adoption*. 2nd Editio. Elsevier/AP, pp. 476–480.

Iheozor-Ejiofor, Z. *et al.* (2015) 'Water fluoridation for the prevention of dental caries.', *The Cochrane database of systematic reviews*, (6), p. CD010856. doi: 10.1002/14651858.CD010856.pub2.

Institute of Medicine (2011) 'Trustworthy Clinical Practice Guidelines: Challenges and Potential', in Graham, R. et al. (eds) *Clinical Practice Guidelines We Can Trust*. Washington (DC): National Academies Press (US).

Ioannidis, J. P. and Lau, J. (2001) 'Evidence on interventions to reduce medical errors: an overview and recommendations for future research.', *Journal of general internal medicine*, 16(5), pp. 325–34.

Iqbal, M., Kim, S. and Yoon, F. (2007) 'An investigation into differential diagnosis of pulp and periapical pain: a PennEndo database study.', *Journal of endodontics*, 33(5), pp. 548–51.

Järvelin, K. and Kekäläinen, J. (2002) 'Cumulated gain-based evaluation of IR techniques', *ACM Transactions on Information Systems*. ACM, 20(4), pp. 422–446.

jena.apache.org (2016) *Apache Jena - Reasoners and rule engines: Jena inference support*, *jena.apache.org/documentation/inference/*. Available at: https://jena.apache.org/documentation/inference/ (Accessed: 27 December 2016).

Jones, R. C. M. *et al.* (2014) 'Opportunities to diagnose chronic obstructive pulmonary disease in routine care in the UK: a retrospective study of a clinical cohort.', *The lancet. Respiratory medicine*, 2(4), pp. 267–76. doi: 10.1016/S2213-2600(14)70008-6.

Kachalia, A. *et al.* (2007) 'Missed and delayed diagnoses in the emergency department: a study of closed malpractice claims from 4 liability insurers.', *Annals of emergency medicine*, 49(2), pp. 196–205.

Kahneman, D. and Egan, P. (2011) *Thinking, fast and slow*. Vol 1. New York, NY: Farrar, Straus and Giroux.

Kalenderian, E. *et al.* (2011a) 'The development of a dental diagnostic terminology.', *Journal of dental education*, 75(1), pp. 68–76.

Kalenderian, E. *et al.* (2011b) 'The development of a dental diagnostic terminology.', *Journal of dental education*, 75(1), pp. 68–76.

Karopka, T., Schmuhl, H. and Demski, H. (2014) 'Free/Libre open source software in health care: a review.', *Healthcare informatics research*. Korean Society of Medical Informatics, 20(1), pp. 11–22. doi: 10.4258/hir.2014.20.1.11.

Kawamoto, K. *et al.* (2005) 'Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success.', *BMJ (Clinical research ed.)*, 330(7494), p. 765.

Kayyali, B., Knott, D. and Van Kuiken, S. (2013) 'The big-data revolution in US health care: Accelerating value and innovation. McKinsey & Company'.

Khan, S., Clarke, J. and Kotecha, A. (2012) 'Comparison of optometrist glaucoma referrals against published guidelines.', *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 32(6), pp. 472–7. doi: 10.1111/j.1475-1313.2012.00943.x.

Khawaja, A. *et al.* (2016) *Primary Open-Angle Glaucoma - EyeWiki*, *eyewiki.aao.org*. Available at: http://eyewiki.aao.org/Primary_Open-Angle_Glaucoma (Accessed: 27 March 2017).

Kifer, M. and Boley, H. (2013) *RIF Overview (Second Edition)*, *www.w3.org/TR/rif-overview/*. Available at: https://www.w3.org/TR/rif-overview/ (Accessed: 1 January 2017).

Klaczynski, P. A. *et al.* (2012) 'Ontology and Semantic Web', in *Encyclopedia of the Sciences of Learning*. Boston, MA: Springer US, pp. 2507–2510. doi: 10.1007/978-1-4419-1428-6_527.

Klein, G. (2008a) 'Naturalistic decision making.', *Human factors*, 50(3), pp. 456–60.

Klein, G. (2008b) 'Naturalistic decision making.', *Human factors*, 50(3), pp. 456–60.

Knublauch, H., Hendler, J. A. and Idehen, K. (2011) *SPIN - Overview and Motivation*, *www.w3.org/Submission*. Available at: https://www.w3.org/Submission/spin-overview/.

Kohn, L. T., Corrigan, J. M. and Donaldson, M. S. (1999) *To Err is Human: Building A Safer Health System*, *National Academies Press*. Washington, DC.

Kostopoulou, O. (2008) 'Do GPs report diagnostic errors?', *Family practice*, 25(1), pp. 1–2.

Kostopoulou, O. *et al.* (2015a) 'Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients.', *The British journal of general practice : the journal of the Royal College of General Practitioners*. British Journal of General Practice, 65(630), pp. e49-54.

Kostopoulou, O. *et al.* (2015b) 'Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients.', *The British journal of general practice : the journal of the Royal College of General Practitioners*. British Journal of General Practice, 65(630), pp. e49-54. doi: 10.3399/bjgp15X683161.

Kostopoulou, O., Delaney, B. C. and Munro, C. W. (2008a) 'Diagnostic difficulty and error in primary care--a systematic review.', *Family practice*, 25(6), pp. 400–13. doi: 10.1093/fampra/cmn071.

Kostopoulou, O., Delaney, B. C. and Munro, C. W. (2008b) 'Diagnostic difficulty and error in primary care--a systematic review.', *Family practice*, 25(6), pp. 400–13.

Kronz, J. D., Westra, W. H. and Epstein, J. I. (1999) 'Mandatory second opinion surgical pathology at a large referral hospital.', *Cancer*, 86(11), pp. 2426–35.

Krötzsch, M., Simancik, F. and Horrocks, I. (2012) 'A Description Logic Primer'.

Kumar, Y. S., Acharya, S. and Pentapati, K. C. (2014) 'Prevalence of dental pain and its relationship to caries experience in school children of Udupi district', *European Archives of Paediatric Dentistry*, 15(6), pp. 371–375. doi: 10.1007/s40368-014-0124-1.

Kuperman, G. J. *et al.* (2007) 'Medication-related clinical decision support in computerized provider order entry systems: a review.', *Journal of the American Medical Informatics Association : JAMIA*, 14(1), pp. 29–40.

Kwiatkowska, M. *et al.* (2007) 'Knowledge-based data analysis: first step toward the creation of clinical prediction rules using a new typicality measure.', *IEEE transactions on information technology in biomedicine : a publication of the IEEE Engineering in Medicine and Biology Society*, 11(6), pp. 651–60.

Labrie, N. and Schulz, P. J. (2014) 'Does Argumentation Matter? A Systematic Literature Review on the Role of Argumentation in Doctor–Patient Communication', *Health Communication*, 29(10), pp. 996–1008. doi: 10.1080/10410236.2013.829018.

Laney, D. (2001) '3D data management: Controlling data volume, velocity and variety', *META Group Research Note*, 6, p. 70.

Langarizadeh, M. and Moghbeli, F. (2016) 'Applying Naive Bayesian Networks to Disease Prediction: a Systematic Review.', *Acta informatica medica : AIM : journal of the Society for Medical Informatics of Bosnia & Herzegovina : casopis Drustva za medicinsku informatiku BiH*. The Academy of Medical Sciences of Bosnia and Herzegovina, 24(5), pp. 364–369. doi: 10.5455/aim.2016.24.364-369.

Larntz, K. (1978) 'Small-Sample Comparisons of Exact Levels for Chi-Squared Goodness-of-Fit Statistics', *Journal of the American Statistical Association*, 73(362), p. 253. doi: 10.2307/2286650.

Lassila, O. and McGuinness, D. (2001) 'The role of frame-based representation on the semantic web', *Linköping Electronic Articles in Computer and Information Science*, 6(5), p. 2001.

Leape, L. L. and Berwick, D. M. (2005) 'Five years after To Err Is Human: what have we learned?', *JAMA : the journal of the American Medical Association*, 293(19), pp. 2384–90. doi: 10.1001/jama.293.19.2384.

Leavitt, N. (2010) 'Will NoSQL Databases Live Up to Their Promise?', *Computer*, 43(2), pp. 12–14. doi: 10.1109/MC.2010.58.

Ledley, R. S. and Lusted, L. B. (1959) 'Reasoning Foundations of Medical Diagnosis: Symbolic logic, probability, and value theory aid our understanding of how physicians reason', *Science*, 130(3366), pp. 9–21. doi: 10.1126/science.130.3366.9.

Lee, C.-S. and Wang, M.-H. (2011) 'A fuzzy expert system for diabetes decision support application.', *IEEE transactions on systems, man, and cybernetics. Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society*, 41(1), pp. 139–53. doi: 10.1109/TSMCB.2010.2048899.

Lee, D. *et al.* (2014) 'Literature review of SNOMED CT use.', *Journal of the American Medical Informatics Association : JAMIA*, 21(e1), pp. e11-9.

Lenzer, J. (2013a) 'Why we can't trust clinical guidelines.', *BMJ (Clinical research ed.)*. British Medical Journal Publishing Group, 346, p. f3830.

Lenzer, J. (2013b) 'Why we can't trust clinical guidelines.', *BMJ (Clinical research ed.)*, 346(jun14_2), p. f3830. doi: 10.1136/bmj.f3830.

Leslie, H. (2012) *Introduction to Archetypes and Archetype classes - Health Information Models - openEHR wiki*, *openehr.atlassian.net*. Available at: https://openehr.atlassian.net/wiki/display/healthmod/Introduction+to+Archetypes+and+Archetype+classes.

Levin, L. G. *et al.* (2009) 'Identify and define all diagnostic terms for pulpal health and disease states.', *Journal of endodontics*, 35(12), pp. 1645–57.

Liang, Y., Zhai, J. and Li, J. (2013) 'Fuzzy System Modeling and Representation Based on Linguistic Variable Ontology and SWRL', *International Journal of Database Theory & Application*, 6(2).

Lin, J. and Chandler, N. P. (2008) 'Electric pulp testing: a review.', *International endodontic journal*, 41(5), pp. 365–74.

Lindberg, C. (1990) 'The Unified Medical Language System (UMLS) of the National Library of Medicine.', *Journal (American Medical Record Association)*, 61(5), pp. 40–2.

Lindgren, H. (2011) 'Towards personalized decision support in the dementia domain based on clinical practice guidelines', *User Modeling and User-Adapted Interaction*. Springer Netherlands, 21(4–5), pp. 377–406. doi: 10.1007/s11257-010-9090-4.

Liu, K. *et al.* (2013) 'Using electronic dental record data for research: a data-mapping study.', *Journal of dental research*, 92(7 Suppl), p. 90S–6S.

lod2.eu (2016) *LODStats - 9960 datasets*, *stats.lod2.eu*. Available at: http://stats.lod2.eu/stats (Accessed: 27 December 2016).

Long, W. (1989) 'MEDICAL DIAGNOSIS USING A PROBABILISTIC CAUSAL NETWORK', *Applied Artificial Intelligence*. Taylor & Francis, Inc., 3(2–3), pp. 367–383.

Longo, L. and Hederman, L. (2013) 'Argumentation Theory for Decision Support in Health-Care: A Comparison with Machine Learning', in *International Conference on Brain and Health Informatics*. Springer, Cham, pp. 168–180. doi: 10.1007/978-3-319-02753-1_17.

Longo, L., Kane, B. and Hederman, L. (2012) 'Argumentation theory in health care', in *2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, pp. 1–6. doi: 10.1109/CBMS.2012.6266323.

Louie, B. *et al.* (2007) 'Data integration and genomic medicine', *Journal of Biomedical Informatics*, 40(1), pp. 5–16. doi: 10.1016/j.jbi.2006.02.007.

Lubarsky, S. *et al.* (2015) 'Using script theory to cultivate illness script formation and clinical reasoning in health professions education.', *Canadian medical education journal*. University of Saskatchewan, 6(2), pp. e61-70.

Majeed, A., Car, J. and Sheikh, A. (2008) 'Accuracy and completeness of electronic patient records in primary care', *Family Practice*, 25, pp. 213–214. doi: 10.1093/fampra/cmn047.

Mandl, K. D. and Kohane, I. S. (2015) 'Federalist principles for healthcare data networks', *Nature Biotechnology*. Nature Research, 33(4), pp. 360–363. doi: 10.1038/nbt.3180.

Marco-Ruiz, L. *et al.* (2016) 'Publication, discovery and interoperability of Clinical Decision Support Systems: A Linked Data approach', *Journal of Biomedical Informatics*, 62, pp. 243–264. doi: 10.1016/j.jbi.2016.07.011.

Marco-Ruiz, L. and Bellika, J. G. (2015) 'Semantic Interoperability in Clinical Decision Support Systems: A Systematic Review.', *Studies in health technology and informatics*, 216, p. 958.

Maria Yaneli, A.-A. *et al.* (2013) 'Assessment of Bayesian Network Classifiers as Tools for Discriminating Breast Cancer Pre-diagnosis Based on Three Diagnostic Methods', in. Springer, Berlin, Heidelberg, pp. 419–431. doi: 10.1007/978-3-642-37807-2_36.

McCarthy and John (1977) 'Epistemological problems of artificial intelligence', *Proceedings of the 5th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann Publishers Inc., pp. 1038–1044.

McDermott, D. and Doyle, J. (1980) 'Non-monotonic logic I', *Artificial Intelligence*. Elsevier, 13(1–2), pp. 41–72. doi: 10.1016/0004-3702(80)90012-0.

McHugh, M. L. (2013) 'The Chi-square test of independence', *Biochemia Medica*, pp. 143–149. doi: 10.11613/BM.2013.018.

van Melle, W. (1979) 'A domain-independent production-rule system for consultation programs', in *IJCAI'79 Proceedings of the 6th international joint conference on Artificial intelligence - Volume 2*. Morgan Kaufmann Publishers Inc., pp. 923–925.

Merriam-Webster (2016) *Differential Diagnosis | Definition of Differential Diagnosis by Merriam-Webster*, *www.merriam-webster.com*. Available at: https://www.merriam-webster.com/dictionary/differential diagnosis.

Mettes, T. G. *et al.* (2010) 'Routine oral examination: Clinical vignettes, a promising tool for continuing professional development?', *Journal of Dentistry*, 38(5), pp. 377–386. doi: 10.1016/j.jdent.2010.01.004.

Mettler, M. (2016) 'Blockchain technology in healthcare: The revolution starts here', in *2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, pp. 1–3. doi: 10.1109/HealthCom.2016.7749510.

Miller, G. A. (1956) 'The magical number seven plus or minus two: some limits on our capacity for processing information.', *Psychological review*, 63(2), pp. 81–97.

Miller, R. A. (1994) 'Medical diagnostic decision support systems--past, present, and future: a threaded bibliography and brief commentary.', *Journal of the American Medical Informatics Association : JAMIA*, 1(1), pp. 8–27.

Miller, R. A., Pople, H. E. and Myers, J. D. (1982) 'Internist-1, an experimental computer-based diagnostic consultant for general internal medicine.', *The New England journal of medicine*, 307(8), pp. 468–76.

Miller, R., Masarie, F. E. and Myers, J. D. (1986) 'Quick medical reference (QMR) for diagnostic assistance.', *M.D. computing : computers in medical practice*, 3(5), pp. 34–48.

Minsky, M. (1975) 'Minsky's frame system theory', in *Proceedings of the 1975 workshop on Theoretical issues in natural language processing - TINLAP '75*. Morristown, NJ, USA: Association for Computational Linguistics, p. 104.

Mirhaji, P. (2016) *Yosemite Project | Semantic Data Lake in Healthcare*, *http://yosemiteproject.org*. Available at: http://yosemiteproject.org/recorded-webinars/webinars/semantic-data-lake-in-healthcare/ (Accessed: 4 October 2017).

Mohammed, O., Benlamri, R. and Fong, S. (2012) 'Building a diseases symptoms ontology for medical diagnosis: An integrative approach', in *The First International Conference on Future Generation Communication Technologies*. IEEE, pp. 104–108. doi: 10.1109/FGCT.2012.6476567.

Molino, G. *et al.* (1986) 'CHECK: a diagnostic expert system Combining HEuristic and Causal Knowledge', *Int. J. of Biomedical Measurement, Informatics and Control*, 1(4), pp. 182–193.

Moore, R. C. (1985) 'Semantical considerations on nonmonotonic logic', *Artificial Intelligence*. Elsevier, 25(1), pp. 75–94. doi: 10.1016/0004-3702(85)90042-6.

Morley, A. M. S. and Murdoch, I. (2006) 'The future of glaucoma clinics.', *The British journal of ophthalmology*, 90(5), pp. 640–5. doi: 10.1136/bjo.2005.085522.

Mowatt, G. *et al.* (2008) 'Screening tests for detecting open-angle glaucoma: systematic review and meta-analysis.', *Investigative ophthalmology & visual*

science, 49(12), pp. 5373–85. doi: 10.1167/iovs.07-1501.

Murdoch, I. E., Lawrenson, J. G. and Myint, J. (2015) 'Personal communication'.

Murdoch, T. B. and Detsky, A. S. (2013) 'The Inevitable Application of Big Data to Health Care', *JAMA*. American Medical Association, 309(13), p. 1351. doi: 10.1001/jama.2013.393.

Murdock, B. (2016) *How Dental Problems Can Cause Swollen Lymph Nodes - Murdock and Searly Family Dentistry*, *americanforkdentists.com*. Available at: http://americanforkdentists.com/dental-problems-cause-lymph-nodes/.

Musen, M., Shahar, Y. and Shortliffe, E. (2006) 'Clinical Decision-Support Systems', in Shortliffe, E. and Cimino, J. (eds) *Biomedical Informatics SE - 20*. Springer New York (Health Informatics), pp. 698–736. doi: 10.1007/0-387-36278-9_20.

Myers, J. D., Pople, H. E. and Miller, R. A. (1982) 'CADUCEUS: A Computerized Diagnostic Consultation System in Internal Medicine', *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, p. 44.

NCBI (2016) *Home - PubMed - NCBI*, *www.ncbi.nlm.nih.gov/pubmed/*. Available at: https://www.ncbi.nlm.nih.gov/pubmed/ (Accessed: 1 January 2017).

Neches, R. *et al.* (1991) 'Enabling technology for knowledge sharing', *AI Magazine*. American Association for Artificial Intelligence, 12(3), pp. 36–56.

NHS Choices (2015) *Macular degeneration - NHS Choices*, *www.nhs.uk*. Department of Health. Available at: http://www.nhs.uk/conditions/macular-degeneration/Pages/Introduction.aspx (Accessed: 27 March 2017).

NICE (2009) *Glaucoma: diagnosis and management | Guidance and guidelines | NICE*, *nice.org.uk*. NICE. Available at: https://www.nice.org.uk/guidance/cg85 (Accessed: 18 June 2016).

Noll, J., Beecham, S. and Seichter, D. (2011) 'A Qualitative Study of Open Source Software Development: The Open EMR Project', in *2011 International Symposium on Empirical Software Engineering and Measurement*. IEEE, pp. 30–39.

NORD (2016) *Home - NORD (National Organization for Rare Disorders)*, *https://rarediseases.org/*. Available at: https://rarediseases.org/ (Accessed: 17 December 2016).

Norman, G. (2009) 'Dual processing and diagnostic errors.', *Advances in health sciences education : theory and practice*, 14 Suppl 1, pp. 37–49.

Noy, N. *et al.* (2006) *Defining N-ary Relations on the Semantic Web*, *w3.org/TR*. Available at: https://www.w3.org/TR/swbp-n-aryRelations/#example1.

Noy, N. F. and Mcguinness, D. L. (2001) *Ontology Development 101: A Guide to Creating Your First Ontology*.

Oberkampf, H. *et al.* (2012) 'Towards a ranking of likely diseases in terms of precision and recall', in *Proceedings of the 1st International Workshop on Artificial Intelligence and NetMedicine*. Montpellier, France, pp. 11–20.

Oberkampf, H., Zillner, S. and Bauer, B. (2012a) 'Interpreting Patient Data using Medical Background Knowledge.', in *ICBO*.

Oberkampf, H., Zillner, S. and Bauer, B. (2012b) 'Interpreting Patient Data using Medical Background Knowledge', in Cornet, R. and Stevens, R. (eds) *Proceedings of the 3rd International Conference on Biomedical Ontology (ICBO 2012)*. Graz, Austria: CEUR-WS.org.

Ongenae, F. *et al.* (2010) 'Towards computerizing intensive care sedation guidelines: design of a rule-based architecture for automated execution of clinical guidelines', *BMC Medical Informatics and Decision Making*, 10(1), p. 3. doi: 10.1186/1472-6947-10-3.

Onisko, A., Druzdzel, M. J. and Wasyluk, H. (1998) 'A probabilistic causal model for diagnosis of liver disorders', in *Proceedings of the Seventh International Symposium on Intelligent Information Systems (IIS—98)*, p. 379.

Ontobee.org (2016) *Ontobee: OHD*, *http://www.ontobee.org/*. Available at: http://www.ontobee.org/ontology/OHD (Accessed: 30 January 2016).

OpenClinical.net (2015) *OpenClinical.net*, *openclinical.net*. Available at: http://openclinical.net/.

openEHR Foundation (2016) *What is openEHR?*, *openehr.org*. Available at: http://www.openehr.org/what_is_openehr (Accessed: 31 January 2016).

OpenEMR (2016) *OpenEMR Project*, *www.open-emr.org*. Available at: http://www.open-emr.org/.

Osheroff, J. A. *et al.* (2007) 'A Roadmap for National Action on Clinical Decision Support', *Journal of the American Medical Informatics Association*, 14(2).

Ouerdane, W., Maudet, N. and Tsoukiàs, A. (2010) 'Argumentation Theory and Decision Aiding', in *Trends in Multiple Criteria Decision Analysis*. Springer, Boston, MA, pp. 177–208. doi: 10.1007/978-1-4419-5904-1_7.

Patel, V. L. *et al.* (1998) 'Representing clinical guidelines in GLIF: individual and collaborative expertise.', *Journal of the American Medical Informatics Association : JAMIA*. American Medical Informatics Association, 5(5), pp. 467–83.

Patel, V. L., Arocha, J. F. and Kaufman, D. R. (1994) 'Diagnostic Reasoning and Medical Expertise', *Psychology of Learning and Motivation*, 31, pp.

187–252.

Patel, V. L., Evans, D. A. and Groen, G. J. (1989) 'Biomedical knowledge and clinical reasoning', *Cognitive science in medicine: Biomedical modeling*, pp. 53–112.

Patel, V. L., Kaufman, D. R. and Arocha, J. F. (2002) 'Emerging paradigms of cognition in medical decision-making', *Journal of Biomedical Informatics*, 35(1), pp. 52–75.

Patil, R. S. and Senyk, O. (1987) 'Efficient Structuring of Composite Causal Hypotheses in Medical Diagnosis', in *Proceedings of the Annual Symposium on Computer Application in Medical Care*. American Medical Informatics Association, p. 23.

Patil, R. S., Szolovits, P. and Schwartz, W. B. (1981) 'Causal Understanding of Patient Illness in Medical Diagnosis', *IJCAI*, 81, pp. 893–899.

Patkar, V. *et al.* (2006) 'Evidence-based guidelines and decision support services: A discussion and evaluation in triple assessment of suspected breast cancer.', *British journal of cancer*. Nature Publishing Group, 95(11), pp. 1490–6. doi: 10.1038/sj.bjc.6603470.

Patkar, V. *et al.* (2012) 'Using computerised decision support to improve compliance of cancer multidisciplinary meetings with evidence-based guidance.', *BMJ open*. British Medical Journal Publishing Group, 2(3), p. e000439. doi: 10.1136/bmjopen-2011-000439.

Pauer, F. *et al.* (2016) 'Adopting Quality Criteria for Websites Providing Medical Information About Rare Diseases', *Interactive Journal of Medical Research*, 5(3), p. e24. doi: 10.2196/ijmr.5822.

Pauker, S. G. *et al.* (1976a) 'Towards the simulation of clinical cognition. Taking a present illness by computer.', *The American journal of medicine*, 60(7), pp. 981–96.

Pauker, S. G. *et al.* (1976b) 'Towards the simulation of clinical cognition. Taking a present illness by computer.', *The American journal of medicine*, 60(7), pp. 981–96.

Peabody, J. W. *et al.* (2004) 'Measuring the quality of physician practice by using clinical vignettes: a prospective validation study.', *Annals of internal medicine*, 141(10), pp. 771–80.

Peleg, M. *et al.* (2003) 'Comparing computer-interpretable guideline models: a case-study approach.', *Journal of the American Medical Informatics Association : JAMIA*. American Medical Informatics Association, 10(1), pp. 52–68. doi: 10.1197/JAMIA.M1135.

Peleg, M. (2013) 'Computer-interpretable clinical guidelines: a methodological review.', *Journal of biomedical informatics*, 46(4), pp. 744–63. doi: 10.1016/j.jbi.2013.06.009.

PEPID™ (2014) *DDX | Medical Symptom Checker and Differential Diagnosis Generator*, *www.pepid.com/*. Available at: http://www.pepid.com/products/ddx/ (Accessed: 1 August 2014).

Perno, J. F. *et al.* (2005) 'Significant reduction in delayed diagnosis of injury with implementation of a pediatric trauma service.', *Pediatric emergency care*, 21(6), pp. 367–71.

Phua, D. H. and Tan, N. C. K. (2013) 'Cognitive aspect of diagnostic errors.', *Annals of the Academy of Medicine, Singapore*, 42(1), pp. 33–41.

Pople, H. E. (1982) 'Heuristic methods for imposing structure on ill-structured problems: The structuring of medical diagnostics', *Artificial intelligence in medicine*, 51, pp. 119–190.

Powell, J. A., Darvell, M. and Gray, J. A. M. (2003) 'The doctor, the patient and the world-wide web: how the internet is changing healthcare.', *Journal of the Royal Society of Medicine*. Royal Society of Medicine Press, 96(2), pp. 74–6.

Prakken, H. and Vreeswijk, G. (2001) 'Logics for Defeasible Argumentation', in *Handbook of Philosophical Logic*. Dordrecht: Springer Netherlands, pp. 219–318. doi: 10.1007/978-94-017-0456-4_3.

Prcela, M., Gamberger, D. and Jovic, A. (2008) 'Semantic web ontology utilization for heart failure expert system design.', *Studies in health technology and informatics*, 136, pp. 851–6.

Purcell, G. P., Wilson, P. and Delamothe, T. (2002) 'The quality of health information on the internet', *BMJ*, 324(7337).

Quilitz, B. and Leser, U. (2008) 'Querying Distributed RDF Data Sources with SPARQL', in *The Semantic Web: Research and Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 524–538. doi: 10.1007/978-3-540-68234-9_39.

R Core Team (2013) 'R: A language and environment for statistical computing'. Vienna, Austria.

Raghupathi, W. *et al.* (2014) 'Big data analytics in healthcare: promise and potential', *Health Information Science and Systems*. BioMed Central, 2(1), p. 3. doi: 10.1186/2047-2501-2-3.

Rajendran, V. V. and Swamynathan, S. (2015) 'MOSS-IR: Multi-Ontology Based Search System for Information Retrieval in E-health Domain', *Procedia Computer Science*. Elsevier, 47, pp. 179–187. doi: 10.1016/j.procs.2015.03.196.

Rakhmawati, N. A. *et al.* (2013) 'A Comparison of Federation over SPARQL Endpoints Frameworks', in *Knowledge Engineering and the Semantic Web*. Springer Berlin Heidelberg, pp. 132–146. doi: 10.1007/978-3-642-41360-5_11.

Ramnarayan, P. *et al.* (2003) 'ISABEL: a web-based differential diagnostic aid for paediatrics: results from an initial performance evaluation.',

Archives of disease in childhood. BMJ Publishing Group Ltd and Royal College of Paediatrics and Child Health, 88(5), pp. 408–13. doi: 10.1136/ADC.88.5.408.

Raredisease.org (2013) *Rare Disease Resources*, *raredisease.org*. Available at: http://raredisease.org/index.html (Accessed: 17 December 2016).

Ratnarajan, G. *et al.* (2013) 'The effect of changes in referral behaviour following NICE guideline publication on agreement of examination findings between professionals in an established glaucoma referral refinement pathway: the Health Innovation & Education Cluster (HIEC) Glaucoma P', *The British journal of ophthalmology*, 97(2), pp. 210–4. doi: 10.1136/bjophthalmol-2012-302352.

Ray, P. and Diehl, A. D. (2013) 'The Ocular Disease Ontology.', in *ICBO*, p. 119.

RDFLib Team (2013) *rdflib 4.2.2-dev — rdflib 4.2.2-dev documentation*, *https://rdflib.readthedocs.io*. Available at: https://rdflib.readthedocs.io/en/stable/.

Rector, A. L. *et al.* (1997) 'The GRAIL concept modelling language for medical terminology.', *Artificial intelligence in medicine*, 9(2), pp. 139–71.

Rector, A. L. *et al.* (2003) 'OpenGALEN: open source medical terminology and tools.', in *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, p. 982.

Rector, A. L. and Nowlan, W. A. (1994) 'The GALEN project.', *Computer methods and programs in biomedicine*, 45(1–2), pp. 75–8.

Reiter, R. (1980) 'A logic for default reasoning', *Artificial Intelligence*. Elsevier, 13(1–2), pp. 81–132. doi: 10.1016/0004-3702(80)90014-4.

Reiter, R. (1987) 'Nonmonotonic Reasoning', *Annual Review of Computer Science*. Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA , 2(1), pp. 147–186. doi: 10.1146/annurev.cs.02.060187.001051.

Reyes-Ortiz, J. A. *et al.* (2013) 'Ontology-based Knowledge Representation for Supporting Medical Decisions', *Advances in Soft Computing Techniques*, pp. 127–136.

Riaño, D. *et al.* (2012) 'An ontology-based personalization of health-care knowledge to support clinical decisions for chronically ill patients.', *Journal of biomedical informatics*, 45(3), pp. 429–46. doi: 10.1016/j.jbi.2011.12.008.

Rob Zazueta (2013) *What is a RESTful API? | Mashery*, *www.mashery.com/blog*. Available at: https://www.mashery.com/blog/what-restful-api.

Rodolfo, I. *et al.* (2016) 'How Far in the Future will We Start From?', in *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems - CHI EA '16*. New York, New York, USA: ACM Press, pp. 899–909. doi: 10.1145/2851581.2851593.

Rodríguez-González, A. *et al.* (2011) 'Designing an ontology to support the creation of diagnostic decision support system', in *Proceedings of the 2nd ACM Conference on Bioinformatics, Computational Biology and Biomedicine - BCB '11*. New York, New York, USA: ACM Press, p. 612.

Rodríguez-González, A., Torres-Niño, J., Mayer, M. A., *et al.* (2012) 'Analysis of a Multilevel Diagnosis Decision Support System and Its Implications: A Case Study', *Computational and Mathematical Methods in Medicine*, 2012.

Rodríguez-González, A. *et al.* (2012) 'SeDeLo: using semantics and description logics to support aided clinical diagnosis.', *Journal of medical systems*, 36(4), pp. 2471–81.

Rodríguez-González, A., Labra-Gayo, J. E., *et al.* (2012) 'SeDeLo: using semantics and description logics to support aided clinical diagnosis.', *Journal of medical systems*, 36(4), pp. 2471–81. doi: 10.1007/s10916-011-9714-1.

Rodríguez-González, A., Torres-Niño, J., Hernández-Chan, G., *et al.* (2012) 'Using agents to parallelize a medical reasoning system based on ontologies and description logics as an application case', *Expert Systems with Applications*, 39(18), pp. 13085–13092.

Rodríguez-González, A. and Alor-Hernández, G. (2013) 'An approach for solving multi-level diagnosis in high sensitivity medical diagnosis systems through the application of semantic technologies.', *Computers in biology and medicine*, 43(1), pp. 51–62.

Rodriguez, A. *et al.* (2009) 'ADONIS: Automated diagnosis system based on sound and precise logical descriptions', in *2009 22nd IEEE International Symposium on Computer-Based Medical Systems*. IEEE, pp. 1–8. doi: 10.1109/CBMS.2009.5255449.

Rodríguez, A. *et al.* (2009a) 'MEDBOLI: Medical Diagnosis Based on Ontologies and Logical Inference', in *2009 International Conference on eHealth, Telemedicine, and Social Medicine*. IEEE, pp. 233–238. doi: 10.1109/eTELEMED.2009.43.

Rodríguez, A. *et al.* (2009b) 'MEDBOLI: Medical Diagnosis Based on Ontologies and Logical Inference', in *2009 International Conference on eHealth, Telemedicine, and Social Medicine*. IEEE, pp. 233–238.

Roitman, H. *et al.* (2011) 'Exploratory search over social-medical data', in *Proceedings of the 20th ACM international conference on Information and knowledge management - CIKM '11*. New York, New York, USA: ACM Press, p. 2513. doi: 10.1145/2063576.2064006.

Romero-Tris, C., Riaño, D. and Real, F. (2010) 'Ontology-based retrospective and prospective diagnosis and medical knowledge personalization', in *KR4HC'10 Proceedings of the ECAI 2010 conference on Knowledge representation for health-care*. Springer-Verlag, pp. 1–15.

Routledge, R., Routledge and Rick (2005) 'Fisher's Exact Test', in *Encyclopedia of Biostatistics*. Chichester, UK: John Wiley & Sons, Ltd. doi: 10.1002/0470011815.b2a10020.

Saber Tehrani, A. S. *et al.* (2013) '25-Year summary of US malpractice claims for diagnostic errors 1986-2010: an analysis from the National

Practitioner Data Bank.', *BMJ quality & safety*, 22(8), pp. 672–80. doi: 10.1136/bmjqs-2012-001550.

Sachdeva, S., Mchome, S. and Bhalla, S. (2010) 'Web Services Security Issues in Healthcare Applications', in *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*. IEEE, pp. 91–96. doi: 10.1109/ICIS.2010.134.

Sahoo, S. S., Sheth, A. and Zhang, G.-Q. (2012) 'Role of semantic web in health informatics', in *Proceedings of the 2nd ACM SIGHIT symposium on International health informatics - IHI '12*. New York, New York, USA: ACM Press, p. 881. doi: 10.1145/2110363.2110480.

Samwald, M. *et al.* (2011) 'Linked open drug data for pharmaceutical research and development.', *Journal of cheminformatics*. Springer, 3(1), p. 19. doi: 10.1186/1758-2946-3-19.

Samwald, M. *et al.* (2012) 'The Arden Syntax standard for clinical decision support: Experiences and directions', *Journal of Biomedical Informatics*. Academic Press, 45(4), pp. 711–718. doi: 10.1016/J.JBI.2012.02.001.

Samwald, M. *et al.* (2013) 'An RDF/OWL knowledge base for query answering and decision support in clinical pharmacogenetics.', *Studies in health technology and informatics*, 192, pp. 539–42.

Sanchez, E. *et al.* (2011) 'An Architecture for the Semantic Enhancement of Clinical Decision Support Systems', in König, A. et al. (eds) *Knowlege-Based and Intelligent Information and Engineering Systems SE - 62*. Springer Berlin Heidelberg (Lecture Notes in Computer Science), pp. 611–620. doi: 10.1007/978-3-642-23863-5_62.

Saposnik, G., Redelmeier, D., Ruff, C. C., *et al.* (2016) 'Cognitive biases associated with medical decisions: a systematic review', *BMC Medical Informatics and Decision Making*. BioMed Central, 16(1), p. 138.

Saposnik, G., Redelmeier, D., Ruff, C. C., *et al.* (2016) 'Cognitive biases associated with medical decisions: a systematic review', *BMC Medical Informatics and Decision Making*. BioMed Central, 16(1), p. 138. doi: 10.1186/s12911-016-0377-1.

Sarntivijai, S. *et al.* (2016) 'Linking rare and common disease: mapping clinical disease-phenotypes to ontologies in therapeutic target validation.', *Journal of biomedical semantics*. BioMed Central, 7, p. 8. doi: 10.1186/s13326-016-0051-7.

Schank, R. C. and Abelson, R. P. (1977) *Scripts, plans, goals, and understanding : an inquiry into human knowledge structures*. L. Erlbaum Associates.

Scheuermann, R. H., Ceusters, W. and Smith, B. (2009) 'Toward an ontological treatment of disease and diagnosis.', *Summit on translational bioinformatics*, 2009, pp. 116–20.

Schiff, G. D. *et al.* (2009) 'Diagnostic error in medicine: analysis of 583 physician-reported errors.', *Archives of internal medicine*. American Medical Association, 169(20), pp. 1881–7.

Schiff, G. D. and Bates, D. W. (2010) 'Can electronic clinical documentation help prevent diagnostic errors?', *The New England journal of medicine*, 362(12), pp. 1066–9. doi: 10.1056/NEJMp0911734.

Schleyer, T. K. *et al.* (2013a) 'An ontology-based method for secondary use of electronic dental record data.', in *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science*, pp. 234–8.

Schleyer, T. K. *et al.* (2013b) 'An ontology-based method for secondary use of electronic dental record data.', *AMIA Joint Summits on Translational Science proceedings AMIA Summit on Translational Science*, 2013, pp. 234–8.

Schmidt, H. G. and Rikers, R. M. J. P. (2007a) 'How expertise develops in medicine: knowledge encapsulation and illness script formation.', *Medical education*, 41(12), pp. 1133–9.

Schmidt, H. G. and Rikers, R. M. J. P. (2007b) 'How expertise develops in medicine: knowledge encapsulation and illness script formation.', *Medical education*, 41(12), pp. 1133–9. doi: 10.1111/j.1365-2923.2007.02915.x.

Schriml, L. M. *et al.* (2012) 'Disease Ontology: a backbone for disease semantic integration.', *Nucleic acids research*, 40(Database issue), pp. D940-6. doi: 10.1093/nar/gkr972.

Schröder, M. and Wagner, G. (eds) (2003) *Rules and Rule Markup Languages for the Semantic Web*. Berlin, Heidelberg: Springer Berlin Heidelberg (Lecture Notes in Computer Science). doi: 10.1007/b94129.

Schulz, S. and Martínez-Costa, C. (2013) 'How Ontologies Can Improve Semantic Interoperability in Health Care', in Riaño, D. et al. (eds) *Process Support and Knowledge Representation in Health Care SE - 1*. Springer International Publishing (Lecture Notes in Computer Science), pp. 1–10. doi: 10.1007/978-3-319-03916-9_1.

Searls, D. B. (2005) 'Data integration: challenges for drug discovery', *Nature Reviews Drug Discovery*. Nature Publishing Group, 4(1), pp. 45–58. doi: 10.1038/nrd1608.

Sendhilkumar, S., Mahalakshmi, G. S. and Rajasekar, S. N. (2009) 'Ontology-based automatic query refinement', *International Journal of Artificial Intelligence and Soft Computing*. Inderscience Publishers, 1(2/3/4), p. 316. doi: 10.1504/IJAISC.2009.027298.

Sesen, M. B. *et al.* (2013) 'Bayesian Networks for Clinical Decision Support in Lung Cancer Care', *PLoS ONE*. Edited by R. A. Calogero. Public Library of Science, 8(12), p. e82349. doi: 10.1371/journal.pone.0082349.

Sevdalis, N. *et al.* (2010) 'Diagnostic error in a national incident reporting system in the UK.', *Journal of evaluation in clinical practice*, 16(6), pp. 1276–81. doi: 10.1111/j.1365-2753.2009.01328.x.

Seyfang, A., Miksch, S. and Marcos, M. (2002) 'Combining diagnosis and treatment using ASBRU.', *International journal of medical informatics*, 68(1–3), pp. 49–57.

Shadbolt, N., Berners-Lee, T. and Hall, W. (2006) 'The Semantic Web Revisited', *IEEE Intelligent Systems*, 21(3), pp. 96–101.

Shah, R., Edgar, D. F. and Evans, B. J. W. (2010) 'A comparison of standardised patients, record abstraction and clinical vignettes for the purpose of measuring clinical practice.', *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 30(3), pp. 209–24. doi: 10.1111/j.1475-1313.2010.00713.x.

Shah, S. and Murdoch, I. E. (2011) 'NICE - impact on glaucoma case detection.', *Ophthalmic & physiological optics : the journal of the British College of Ophthalmic Opticians (Optometrists)*, 31(4), pp. 339–42. doi: 10.1111/j.1475-1313.2011.00843.x.

Shah, T., Rabhi, F. and Ray, P. (2014) 'Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions', *Cluster Computing*, 18(1), pp. 351–367.

Shalom, E. *et al.* (2015) 'A multiple-scenario assessment of the effect of a continuous-care, guideline-based decision support system on clinicians' compliance to clinical guidelines.', *International journal of medical informatics*, 84(4), pp. 248–62. doi: 10.1016/j.ijmedinf.2015.01.004.

Shapiro, S. S. and Wilk, M. B. (1965) 'An analysis of variance test for normality (complete samples)', *Biometrika*. Oxford University Press, 52(3–4), pp. 591–611. doi: 10.1093/biomet/52.3-4.591.

Sherimon, P. C. and Krishnan, R. (2016) 'OntoDiabetic: An Ontology-Based Clinical Decision Support System for Diabetic Patients', *Arabian Journal for Science and Engineering*. Springer Berlin Heidelberg, 41(3), pp. 1145–1160. doi: 10.1007/s13369-015-1959-4.

Shortliffe, E. H. *et al.* (1975) 'Computer-based consultations in clinical therapeutics: Explanation and rule acquisition capabilities of the MYCIN system', *Computers and Biomedical Research*, 8(4), pp. 303–320.

Shu-Hsien Liao (2005a) 'Expert system methodologies and applications—a decade review from 1995 to 2004', *Expert Systems with Applications*, 28(1), pp. 93–103. doi: 10.1016/j.eswa.2004.08.003.

Shu-Hsien Liao (2005b) 'Expert system methodologies and applications—a decade review from 1995 to 2004', *Expert Systems with Applications*, 28(1), pp. 93–103.

Sibbald, M., de Bruin, A. B. H. and van Merrienboer, J. J. G. (2013) 'Checklists improve experts' diagnostic decisions.', *Medical education*, 47(3), pp. 301–8. doi: 10.1111/medu.12080.

Sibbald, M., Panisko, D. and Cavalcanti, R. B. (2011) 'Role of clinical context in residents' physical examination diagnostic accuracy.', *Medical education*, 45(4), pp. 415–21. doi: 10.1111/j.1365-2923.2010.03896.x.

Singh, H. *et al.* (2012a) 'System-related interventions to reduce diagnostic errors: a narrative review.', *BMJ quality & safety*. BMJ Publishing Group Ltd, 21(2), pp. 160–70. doi: 10.1136/bmjqs-2011-000150.

Singh, H. *et al.* (2012b) 'System-related interventions to reduce diagnostic errors: a narrative review.', *BMJ quality & safety*. BMJ Publishing Group Ltd, 21(2), pp. 160–70.

Singh, H. *et al.* (2013a) 'Types and origins of diagnostic errors in primary care settings.', *JAMA internal medicine*, 173(6), pp. 418–25. doi: 10.1001/jamainternmed.2013.2777.

Singh, H. *et al.* (2013b) 'Types and origins of diagnostic errors in primary care settings.', *JAMA internal medicine*, 173(6), pp. 418–25.

Singh, H., Meyer, A. N. D. and Thomas, E. J. (2014) 'The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving US adult populations', *BMJ Quality & Safety*, p. bmjqs-2013-002627-. doi: 10.1136/bmjqs-2013-002627.

Singh, H. and Weingart, S. N. (2009) 'Diagnostic errors in ambulatory care: dimensions and preventive strategies.', *Advances in health sciences education : theory and practice*. NIH Public Access, 14 Suppl 1(0 1), pp. 57–61. doi: 10.1007/s10459-009-9177-z.

Sirin, E. *et al.* (2007a) 'Pellet: A practical OWL-DL reasoner', *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), pp. 51–53. doi: 10.1016/j.websem.2007.03.004.

Sirin, E. *et al.* (2007b) 'Pellet: A practical OWL-DL reasoner', *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2), pp. 51–53.

Sittig, D. F. *et al.* (2006) 'A survey of factors affecting clinician acceptance of clinical decision support.', *BMC medical informatics and decision making*, 6(1), p. 6. doi: 10.1186/1472-6947-6-6.

Sittig, D. F. *et al.* (2008) 'Grand challenges in clinical decision support', *Journal of Biomedical Informatics*, 41(2), pp. 387–392. doi: 10.1016/j.jbi.2007.09.003.

Slade, G. D. (2001) 'Epidemiology of dental pain and dental caries among children and adolescents.', *Community dental health*, 18(4), pp. 219–27.

Smith, A. K. *et al.* (2007) 'LinkHub: a Semantic Web system that facilitates cross-database queries and information retrieval in proteomics', *BMC Bioinformatics*, 8(Suppl 3), p. S5. doi: 10.1186/1471-2105-8-S3-S5.

Smith, E. E., Shoben, E. J. and Rips, L. J. (1974) 'Structure and process in semantic memory: A featural model for semantic decisions.', *Psychological Review*, 81(3), pp. 214–241. doi: 10.1037/h0036351.

SNOMED-CT (2016) *Toothache-Code System Concept*, *http://phinvads.cdc.gov/*. Available at: http://phinvads.cdc.gov/vads/http:/phinvads.cdc.gov/vads/ViewCodeSystemConcept.action?oid=2.16.840.1.113883.6.96&code=27355003.

Sonderegger-Iseli, K. *et al.* (2000) 'Diagnostic errors in three medical eras: a necropsy study.', *Lancet*, 355(9220), pp. 2027–31.

Spackman, K. A., Campbell, K. E. and Côté, R. A. (1997) 'SNOMED RT: a reference terminology for health care.', *Proceedings : a conference of the American Medical Informatics Association. AMIA Fall Symposium*, pp. 640–4.

Stein, J. D. and Challa, P. (2007) *Diagnosis and Treatment of Normal-Tension Glaucoma - American Academy of Ophthalmology*, *www.aao.org*. Available at: https://www.aao.org/eyenet/article/diagnosis-treatment-of-normal-tension-glaucoma (Accessed: 27 March 2017).

Sullivan, F., Gardner, M. and van Rijsbergen, K. (1999) 'An information retrieval service to support clinical decision-making at the point of care.', *The British journal of general practice : the journal of the Royal College of General Practitioners*, 49(449), pp. 1003–7.

Sutton, D. R. and Fox, J. (2003) 'The Syntax and Semantics of the PROforma Guideline Modeling Language', *Journal of the American Medical Informatics Association : JAMIA*. American Medical Informatics Association, 10(5), p. 433. doi: 10.1197/jamia.m1264.

Sutton, D. R., Taylor, P. and Earle, K. (2006) 'Evaluation of PROforma as a language for implementing medical guidelines in a practical context', *BMC Medical Informatics and Decision Making*, 6(1), p. 20. doi: 10.1186/1472-6947-6-20.

Tang, P. C. *et al.* (2006) 'Personal Health Records: Definitions, Benefits, and Strategies for Overcoming Barriers to Adoption', *Journal of the American Medical Informatics Association*, 13(2).

Taylor, P., Fox, J. and Pokropek, A. T. (1999) 'The development and evaluation of CADMIUM: a prototype system to assist in the interpretation of mammograms', *Medical Image Analysis*. Elsevier, 3(4), pp. 321–337. doi: 10.1016/S1361-8415(99)80027-9.

Teunissen, P. W. *et al.* (2009) 'The influence of context on residents' evaluations: effects of priming on clinical judgment and affect.', *Advances in health sciences education : theory and practice*, 14(1), pp. 23–41. doi: 10.1007/s10459-007-9082-2.

Thammasitboon, S. and Cutrer, W. B. (2013) 'Diagnostic decision-making and strategies to improve diagnosis.', *Current problems in pediatric and adolescent health care*, 43(9), pp. 232–41. doi: 10.1016/j.cppeds.2013.07.003.

The Learning Healthcare Project (2017) *The Learning Health Care Project*, *www.learninghealthcareproject.org*. Available at: http://www.learninghealthcareproject.org/ (Accessed: 4 October 2017).

Thornton, J. G., Lilford, R. J. and Johnson, N. (1992) 'Decision analysis in medicine.', *BMJ (Clinical research ed.)*. BMJ Publishing Group, 304(6834), pp. 1099–103.

Torres-Urquidy, M. H. and Schleyer, T. (2006) 'Evaluation of the Systematized Nomenclature of Dentistry using case reports: preliminary results.', *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, p. 1124.

Toulmin, S. E. (1958) *The uses of argument*. Cambridge University Press.

Trowbridge, R. L., Dhaliwal, G. and Cosby, K. S. (2013) 'Educational agenda for diagnostic error reduction.', *BMJ quality & safety*, 22 Suppl 2, pp. ii28-ii32. doi: 10.1136/bmjqs-2012-001622.

Tu, S. W. and Musen, M. A. (2001) 'Modeling data and knowledge in the EON guideline architecture.', *Studies in health technology and informatics*, 84(Pt 1), pp. 280–4.

Tudorache, T. *et al.* (2013) 'WebProtégé: A Collaborative Ontology Editor and Knowledge Acquisition Tool for the Web.', *Semantic web*. NIH Public Access, 4(1), pp. 89–99. doi: 10.3233/SW-2012-0057.

Tural, C. *et al.* (2002) 'Clinical utility of HIV-1 genotyping and expert advice: the Havana trial.', *AIDS (London, England)*, 16(2), pp. 209–18.

Umscheid, C. A. and Hanson, C. W. (2012) 'A follow-up report card on computer-assisted diagnosis--the grade: C+.', *Journal of general internal medicine*, 27(2), pp. 142–4. doi: 10.1007/s11606-011-1944-x.

Uschold, M. and Gruninger, M. (2009) 'Ontologies: principles, methods and applications', *The Knowledge Engineering Review*. Cambridge University Press, 11(02), p. 93.

Vandervalk, B., McCarthy, L. and Wilkinson, M. (2008) 'CardioSHARE: Web Services for the Semantic Web', *Semantic Web Challenge*, 2008.

Vardell, E. and Bou-Crick, C. (2012) 'VisualDx: a visual diagnostic decision support tool.', *Medical reference services quarterly*, 31(4), pp. 414–24.

Vardell, E. and Moore, M. (2011) 'Isabel, a clinical decision support system.', *Medical reference services quarterly*, 30(2), pp. 158–66.

Verborgh, R. (2015) *600,000 queryable datasets—and counting | Ruben Verborgh*, *ruben.verborgh.org/blog*. Available at: https://ruben.verborgh.org/blog/2015/01/30/600000-queryable-datasets-and-counting/.

Vernon, S. A. *et al.* (2011) 'Community optometrist referral of those aged 65 and over for raised IOP post-NICE: AOP guidance versus joint college guidance--an epidemiological model using BEAP.', *The British journal of ophthalmology*, 95(11), pp. 1534–6. doi: 10.1136/bjo.2010.194092.

Vernon, S. A. and Ghosh, G. (2001) 'Do locally agreed guidelines for optometrists concerning the referral of glaucoma suspects influence referral practice?', *Eye (London, England)*, 15(Pt 4), pp. 458–63. doi: 10.1038/eye.2001.155.

w3.org (2005) *List of Rule Systems - W3C RIF-WG Wiki*, *w3.org/2005/rules*. Available at: https://www.w3.org/2005/rules/wg/wiki/List_of_Rule_Systems (Accessed: 27 December 2016).

W3C® (2007) *Semantic Web Layer Cake*, *www.w3.org*. Available at: https://www.w3.org/2007/03/layerCake.png (Accessed: 30 January 2016).

W3C® (2013a) *RDF 1.1 JSON Alternate Serialization (RDF/JSON)*, *www.w3.org*. Available at: https://www.w3.org/TR/rdf-json/ (Accessed: 30 January 2016).

W3C® (2013b) *SPARQL Query Language for RDF*, *www.w3.org*. Available at: https://www.w3.org/TR/rdf-sparql-query/ (Accessed: 30 January 2016).

W3C® (2014) *RDF - Semantic Web Standards*, *www.w3.org*. Available at: https://www.w3.org/RDF/ (Accessed: 30 January 2016).

W3C OWL Working Group (2012) *OWL Web Ontology Language Overview*, *w3.org*. Available at: https://www.w3.org/TR/owl-features/ (Accessed: 31 January 2016).

Wachter, R. M. (2010) 'Why diagnostic errors don't get any respect--and what can be done about them.', *Health affairs (Project Hope)*, 29(9), pp. 1605–10. doi: 10.1377/hlthaff.2009.0513.

Wagner, M. M. and Hogan, W. R. (1996) 'The accuracy of medication data in an outpatient electronic medical record.', *Journal of the American Medical Informatics Association : JAMIA*, 3(3), pp. 234–44.

Walton, R. T. *et al.* (1997) 'Evaluation of computer support for prescribing (CAPSULE) using simulated cases.', *BMJ (Clinical research ed.)*. British Medical Journal Publishing Group, 315(7111), pp. 791–5. doi: 10.1136/BMJ.315.7111.791.

Wang, A. Y. *et al.* (2001) 'Mapping between SNOMED RT and Clinical terms version 3: a key component of the SNOMED CT development process.', in *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pp. 741–5.

Wang, H.-T. and Tansel, A. U. (2013) 'Composite Ontology-Based Medical Diagnosis Decision Support System Framework', *Communications of the IIMA*, 13(2).

Warner, H. R. (1989) 'Iliad: moving medical decision-making into new frontiers.', *Methods of information in medicine*, 28(4), pp. 370–2.

WARNER, H. R. *et al.* (1961) 'A mathematical approach to medical diagnosis. Application to congenital heart disease.', *JAMA*, 177, pp. 177–83.

Warren, J., Beliakov, G. and van der Zwaag, B. (2000) 'Fuzzy logic in clinical practice decision support systems', in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences*. IEEE Comput. Soc, p. 10. doi: 10.1109/HICSS.2000.926789.

Wasyluk, H., Oniśko, A. and Druzdzel, M. J. (2001) 'Support of diagnosis of liver disorders based on a causal Bayesian network model.', *Medical science monitor : international medical journal of experimental and clinical research*, 7 Suppl 1, pp. 327–32.

Waxman, H. S. and Worley, W. E. (1990) 'Computer-assisted adult medical diagnosis: subject review and evaluation of a new microcomputer-based system.', *Medicine*, *69*(3), pp. 125–36.

Wears, R. L. (2009) 'What makes diagnosis hard?', *Advances in health sciences education : theory and practice*, 14 Suppl 1, pp. 19–25. doi: 10.1007/s10459-009-9181-3.

Weber, G. M. *et al.* (2009) 'The Shared Health Research Information Network (SHRINE): A Prototype Federated Query Tool for Clinical Data Repositories', *Journal of the American Medical Informatics Association*, 16(5), pp. 624–630. doi: 10.1197/jamia.M3191.

Weiss, S. M. *et al.* (1978) 'A model-based method for computer-aided medical decision-making', *Artificial intelligence*, 11, pp. 145–172.

Wellbery, C. (2011) 'Flaws in clinical reasoning: a common cause of diagnostic error.', *American family physician*, 84(9), pp. 1042–8.

West Health Institute (2016) *HCI-DC 2014 White Paper Event Summary: Igniting an Interoperable Healthcare System | West Health*, *http://www.westhealth.org/*. Available at: http://www.westhealth.org/projects/igniting-an-interoperable-healthcare-system-white-paper-summarizes-hci-dc-2014-event-with-guidance-on-driving-interoperability/ (Accessed: 9 May 2016).

Wicks, P. *et al.* (2010) 'Sharing health data for better outcomes on PatientsLikeMe.', *Journal of medical Internet research*. Journal of Medical Internet Research, 12(2), p. e19. doi: 10.2196/jmir.1549.

Wiktionary (2016) *-itis - Wiktionary*, *en.wiktionary.org*. Available at: https://en.wiktionary.org/wiki/-itis.

Woolf, S. H. *et al.* (1999) 'Clinical guidelines: potential benefits, limitations, and harms of clinical guidelines.', *BMJ (Clinical research ed.)*, 318(7182), pp. 527–30.

Wright, A., Phansalkar, S., Bloomrosen, M., Jenders, R. A., *et al.* (2010) 'Best Practices in Clinical Decision Support: the Case of Preventive Care Reminders.', *Applied clinical informatics*. Schattauer Publishers, 1(3), pp. 331–345. doi: 10.4338/ACI-2010-05-RA-0031.

Wright, A., Phansalkar, S., Bloomrosen, M., Jenders, R. A., *et al.* (2010) 'Best Practices in Clinical Decision Support: The Case of Preventive Care Reminders', *Applied Clinical Informatics*. Schattauer Publishers, 1(3), p. 331. doi: 10.4338/aci-2010-05-ra-0031.

Wu Zhao, He Yanxiang and Jin Hui (2005) 'A model of intelligent distributed medical diagnosis and therapy system based on mobile agent and ontology', in *Eighth International Conference on High-Performance Computing in Asia-Pacific Region (HPCASIA'05)*. IEEE, p. 6 pp.-587. doi:

10.1109/HPCASIA.2005.6.

Yatani, H. *et al.* (2014) 'Systematic review and recommendations for nonodontogenic toothache.', *Journal of oral rehabilitation*, 41(11), pp. 843–52. doi: 10.1111/joor.12208.

Yosemiteproject.org (2016) *Introduction - Interoperability Roadmap - The Yosemite Project*, *http://yosemiteproject.org/*. Available at: http://yosemiteproject.org/interoperability-roadmap/ (Accessed: 9 May 2016).

Zaveri, A. *et al.* (2013) 'Publishing and interlinking the Global Health Observatory dataset', *Semantic Web*. IOS Press, 4(3), pp. 315–322. doi: 10.3233/SW-130102.

Zenuni, X. *et al.* (2015) 'State of the Art of Semantic Web for Healthcare', *Procedia - Social and Behavioral Sciences*. Elsevier, 195, pp. 1990–1998. doi: 10.1016/j.sbspro.2015.06.213.

Zhang, X. *et al.* (2014) 'Ontology driven decision support for the diagnosis of mild cognitive impairment.', *Computer methods and programs in biomedicine*, 113(3), pp. 781–91.

Zhu, M. *et al.* (2014) 'Scalable Numerical SPARQL Queries over Relational Databases', in *EDBT/ICDT Workshops*, pp. 257–262.

Zwaan, L. *et al.* (2010) 'Patient record review of the incidence, consequences, and causes of diagnostic adverse events.', *Archives of internal medicine*, 170(12), pp. 1015–21. doi: 10.1001/archinternmed.2010.146.

Appendix

See Volume II of this thesis