CITY UNIVERSITY LONDON

Systemic Analysis and Modelling of Diagnostic Errors in Medicine

---

by

Shijing Guo

A thesis submitted for the degree of Doctor of Philosophy

Centre for Health Informatics

School of Mathematics, Computer Science & Engineering

City University London

May 2016

Acknowledgments

I would like to express my greatest appreciation to my supervisor Prof. Abdul Roudsari. I am always motivated by his full support and his knowledge of the field during the whole period of my doctoral studies. My thesis work would not have been accomplished without his encouragement, aspiring advice and expert guidance. Also, I would like to express my special gratitude to Dr. Artur d'Avila Garcez. He is always patient, and his encouragement and advice are valued greatly.

I would like to extend my appreciation to all the people from Health Informatics centre and Computer Science department for insightful advice and help. Thanks also go to Dr. Xiang Li for sharing his knowledge in data analysis.

I also wish to thank all clinicians who participated the studies in this thesis. Their feedback is valuable.

A special thank goes to my sister Lijing Guo for her greatest support and help for data collection. I would like to express my sincere appreciation to my parents and my husband for their endless love.

ABSTRACT

Diagnostic accuracy is an important index of the quality of health care service. Missed, wrong or delayed diagnosis has a direct effect on patient safety. Diagnostic errors have been discussed at length; however it still lacks a systemic research approach.

This thesis takes the diagnostic process as a system and develops a systemic model of diagnostic errors by implementing system dynamics modelling combined with regression analysis. It aims to propose a better way of studying diagnostic errors as well as a deeper understanding of how factors affect the number of possible errors at each step of the diagnostic process and how factors contribute to patient outcomes in the end.

It is executed following two parts:

In the first part, a qualitative model is developed to demonstrate how errors can happen during the diagnostic process; in other words, the model illustrates the connections among key factors and dependent variables. It starts from discovering key factors of diagnostic errors, producing a hierarchical list of factors, and then illustrates interrelation loops that show how relevant factors are linked with errors. The qualitative model is based on the findings of a systematic literature review and further refined by experts' reviews.

In the second part, a quantitative model is developed to provide system behaviour simulations, which demonstrates the quantitative relations among factors and errors during the diagnostic process. Regression modelling analysis is used to estimate the quantitative relationships among multi factors and their dependent variables during the diagnostic phase of *history taking and physical examinations*. The regression models are further applied into quantitative system dynamics modelling 'stock and flow diagrams'. The quantitative model traces error flows during the diagnostic process, and simulates how the change of one or more variables affects the diagnostic errors and patient outcomes over time. The change of the variables may reflect a change in demand from policy or a proposed external intervention.

The results suggest the systemic model has the potential to help understand diagnostic errors, observe model behaviours, and provide risk-free simulation experiments for possible strategies.

Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background and motivation

Diagnostic errors are the incorrect diagnoses after clinical examination or technical diagnostic procedures[1], which are widely recognised as missed, wrong or unintentionally delayed diagnosis[2]. Diagnostic errors are taken as frequent errors with negative outcomes in medical care. The rate of diagnostic errors occurring in medicine can reach up to 10% ~ 15%, which is determined in the relevant autopsy studies [3] [4] [5]. At the same time, some researchers[6] [7] claim that about 40% ~ 60% errors may be still undetected by autopsies. Diagnostic errors not only induce severe consequences to patient safety, which sometimes is irreparable, but also cause large extra finance payouts. Relevant tort shows it was nearly twice as common as claims for other medication errors and resulted in the largest payouts[8]. It is also claimed that diagnostic errors take the largest part in the ambulatory malpractice claims and can cost up to $300 000 per claim approximately on average in the US.[9] According to the research done by CRICO[10], it was discovered that diagnosis-related claims were the highest in frequency and severity of all malpractice cases[11]. In terms of these diagnosis-related claims, cancer was the most common missed or misdiagnosis- related disease, while myocardial infarctions lay in the second position.[11] Meanwhile, diagnostic errors are usually undiscovered and easily unreported. The medical incident reporting system is still the essential way of reporting diagnostic errors [6], although it has some limitations and unsatisfied outcomes in reporting them. The limitations of reporting them, such as insufficient records and poor quality of records, make the further analysing or detecting work even harder.

Many researchers believe diagnostic errors could be potentially reduced.[12] An increasing number of researches about diagnostic errors were witnessed during the last decade. Most of these methods focus on one of the causes or a specific disease, which analyse diagnostic errors at length[13] [14], and the results are diverse and limited [15] [16] [17] [18]. A systemic view of diagnostic errors and diagnostic process errors has not known and the possibility of analysing diagnostic errors from a system level has received little attention, although system-level solution is suggested by the Institute of Medicine as the most powerful way to reduce medicine errors[19].

Moreover, diagnosis is a complex process and is commonly multifactorial[2]. The diagnostic process involves a series of phases including history taking and physical examinations, diagnostic tests, referrals and follow-up phase. These phases are connected together, and

affect each other. Errors that can happen at any phase of the diagnostic process may lead to a diagnostic error directly or indirectly [20], and these errors are usually referred to as diagnostic process errors. The causes of the errors at each phase may also affect the final decision-making, and can be the causes of diagnostic errors. Causes of diagnostic errors bring about both positive cause-effect relations and negative cause-effect relations. A positive cause-effect relation means the cause is positive to the effect. In other words, if the cause increases (decreases), then the effect, which is the diagnostic errors or diagnostic process errors, increases (decreases). A positive cause-effect relation means that if the cause increases (decreases), then the effect decreases (increases).

Therefore, diagnostic errors need to systemically analyse the causes, and illustrate the whole picture of causes and errors in the diagnostic process, and provide a way to seek system-level solutions.

1.2 Aims

This thesis aims to take the diagnostic process as a system, and systemically model diagnostic errors. The qualitative model is to provide a structured and hierarchical picture of the causes of diagnostic errors. The quantitative model aims to study errors in a system of the diagnostic process, and to understand interrelations of model variables: the factors, which are the causes having either positive effect or negative effect on diagnostic errors; the errors, which are diagnostic errors, diagnostic process errors; and patient outcomes. It is to provide a way of observing model response to the changes of model variables using simulation experiments. The changes of model variables could indicate the effect of projected changes in demand or proposed interventions; thus, it can present the guidance of possible strategies in terms of diagnostic error reduction.

Specifically, the following aims are to be achieved:

- To identify the key factors or the leading causes of diagnostic errors and to provide a structured and hierarchical picture of the causes/factors of diagnostic errors
- To represent errors in a system of diagnostic process from where errors initially occur, how errors are delivered out of the model, to error effect on patient outcomes, and to identify quantitative interrelations between model variables.
- To understand the model behaviours and analyse diagnostic errors through the whole picture of the diagnostic process.
- To evaluate the constructed model.

1.3 Objectives

To achieve the above aims, this thesis seeks to accomplish the following specific objectives:

- To conduct a systematic review of recent researches and methods to reduce diagnostic errors in order to identify the key factors of diagnostic errors.
- To design a qualitative model that reflects the interrelations between factors, as well as the interrelations between factors and errors, in order to provide a structured and hierarchical picture of the factors of diagnostic errors.
- To collect feedback from clinicians to refine the multifactor model in order to increase the reliability of the qualitative model.
- To further represent the qualitative model into the structure of a quantitative model showing interrelations between model variables in order to quantitatively represent errors in the entire diagnostic process.
- To conduct regression analysis and to apply the algorithms in the system dynamics modelling to determine quantitative interrelations between model variables.
- To conduct risk-free simulation experiments under different scenarios and to observe the changes of model outputs when changing one or more model variables in order to observe model behaviours and analyse diagnostic errors through the whole picture of the diagnostic process.
- To adopt a comprehensive evaluation assessment for the quantitative model in order to evaluate the model and determine model applicability.

Table 1.1 shows the links between each aim and its objectives of this thesis.

Aims	Objectives
To identify factors and to provide a structured picture of the factors.	• To conduct a systematic review • To design a qualitative model • To refine the qualitative model
To represent errors in the diagnostic process and to identify quantitative interrelations.	• To represent the qualitative model into a quantitative model • To conduct regression analysis and to apply the algorithms
To understand the model behaviours and analyse diagnostic errors.	• To conduct simulation experiments
To evaluate the model.	• To adopt an evaluation assessment

Table 1.1 Links between each aim and its objectives

1.4 Organisation of the thesis

The rest of the thesis is organised as below:

Chapter 2 contains the primary literature review about the background of diagnostic errors.

Chapter 3 introduces the relevant methods, and summarises all methods used in the thesis.

Chapter 4 starts with a systematic literature review of diagnostic errors, and then enters in the discovery of key factors and relevant methods reducing diagnostic errors.

Chapter 5 uses the findings from the previous chapter to illustrate a qualitative multi-factor model by causal loop diagrams. Relevant factors and cause-effect relations are further revised based on clinicians' opinions.

Chapter 6 shows the development of the quantitative model for diagnostic errors. It discusses individual phases of the diagnostic process covering the phases of the diagnostic process as well as the phase of after-diagnosis, and maps the key factors from the qualitative model and errors into different phases.

Chapter 7 is devoted to the regression analysis of the interrelations of factors and the number of errors in the diagnostic hypotheses during the first diagnostic phase: *history taking and physical examinations*. It first introduces how to quantify the non-numerical variables, and then collects the relevant data for the analysis. Regression modelling is carried out in this chapter, and the results of the regression models are described and evaluated using different metrics. The equations of regression models are further applied to the system dynamics modelling to accomplish the quantitative relations between system elements, considering that system dynamics modelling has limitations in determining the quantitative relational equations when it is implemented alone.

Chapter 8 presents the model simulation. The first part is to collect data for model simulation, and the later part is to perform simulation experiments. Simulation results of error flows and patient outcomes are discussed, and model behaviours are observed under different scenarios.

Chapter 9 focuses on model evaluation. It summaries the evaluation methods for regression models, and further implements a list of evaluation assessments for system dynamics models. The results of the assessments are provided and discussed.

Chapter 10 discusses the contributions of the thesis and the future work.

Chapter 11 concludes the work of the thesis.

The original work presented in this thesis is based on the following peer-reviewed publications:

- Full paper: Guo, S., Roudsari, A. and Garcez, A. (May 2015) A System Dynamics Approach to Analyze Laboratory Test Errors. *Studies in health technology and informatics, 26th European Medical Informatics Conference (MIE2015), 210, 266-270*
- Full paper: Guo, S., Roudsari, A. and Garcez, A. (Jan 2015). Modelling clinical diagnostic errors: a system dynamics approach. *Studies in health technology and informatics, 208, 160-164.*
- Full paper: Guo, S., Roudsari, A. and Garcez, A. (2014). A causal loop approach to the study of diagnostic errors. *Studies in health technology and informatics, 25th European Medical Informatics Conference (MIE2014), 205, 73-77.*

Chapter 2 Preliminary Literature Review

2.1 Introduction

This chapter starts by introducing general medical errors to help understand the role of diagnostic errors from the aspect of patient safety in healthcare. Then, it further describes the background and characteristics of diagnostic errors in detail, which provides the information about the current situation and specific challenges for researchers.

2.2 Medical errors and diagnostic errors

A medical error is "the failure to complete a planned action as intended or the use of a wrong plan to achieve an aim", defined by the US Institute of Medicine[21]. In other words, all actions which are committed by health professionals and cause harm to patients are taken as medical errors, including "diagnostic errors, medication errors, errors in the performance of surgical procedures, in the use of other types of therapy, in the use of equipment, and in the interpretation of laboratory findings"[22]. The frequency and magnitude of medical errors were not well known until the 1990s, and after recognising that medical errors impact on one in ten patients in the world; the World Health Organisation calls patient safety an endemic concern.[23] Nevertheless, diagnostic errors, as a part of medical errors, started to come to public attention in the last decade. Most relevant researches have been conducted since the beginning of the 21st century, and they aimed to understand the causes of diagnostic errors and the ways of reducing diagnostic errors. However, most studies are in-depth partial analysis of the diagnostic error problem, such as improving doctor education to avoid diagnostic errors and finding the most efficient educational methods in terms of diagnostic errors, and few studies observe the entire diagnostic process or patient pathway in terms of diagnostic errors. The particular characteristics of diagnostic errors make the problem analysis and solution seeking even harder.

Diagnostic errors are the missed, wrong or delayed diagnosis, based on a classification used by the Australian Patient Safety Foundation. It is the errors happening after clinical examination or technical diagnostic procedures. Graber et al in 2002[13] showed that diagnostic errors can be divided into three categories according to the causes of the errors: ''No-fault errors'', ''System errors'' and ''Cognitive errors'', presented in Table 2.1.

No-fault errors	System errors	Cognitive errors
•Unusual presentation of disease •Uncertainty regarding the state of the world •Lack of patient cooperation •Limitations of medical knowledge •Failure of noremative precesses	•Technical failure •Organisational failure	•Inadequate knowledge •Faulty data gathering •Faulty information processing •Faulty metacognition

Table 2.1 Categories of diagnostic errors[13]

"No-fault errors" are the cases where the illness is silent, or masked, or presented in an atypical presentation such that the correct diagnosis, with the current state of medical knowledge, would not be expected.[24] ''Cognitive errors'' are the errors cause by adequate knowledge or faulty data gathering, or inaccurate clinical reasoning, or faulty verification.[13] [24] "No-fault errors" and ''Cognitive errors'' are usually considered as more harmful errors, whilst little has been known to reduce them[79]. ''System errors'' are related to the system, including faults of technologies or flaws of relevant systems, such as faulty tests or patients with abnormal test results being neglected.

2.3 Diagnostic error characteristics

Compared to other medical errors, diagnostic errors have their special characteristics, which are discussed as below.

- Diagnostic errors indicate severe harm in relation to patient safety. Relevant research[25] in the UK shows 25% of diagnostic incidents resulted in death, 23% in disability and 15% in cognitive impairment and/or disability[6]. One research result[26] of hospitalised patient samples from 21 hospitals in the Netherlands showed that 23% of adverse events related to diagnostic process contributed to death. But still it is believed that diagnostic errors are potentially preventable.[12]
- Diagnostic errors are usually undetected, recognised late and easily unreported to the current incident reporting system.

  In the primary care service, general practitioners (GPs) are not only the important executor during the diagnosis, but also have an essential role of preventing and reporting diagnostic errors [27].

Incident reporting systems, such as the National Reporting and Learning System (NRLS) in the UK, are the essential tools currently used for reporting and detecting diagnostic errors, and also provide the main source of diagnostic errors for further analysis[6]. Nevertheless, incident reporting systems have obvious limitations and unsatisfied outcomes in reporting diagnostic errors, mainly because the systems are challenged in reporting every diagnostic error, and recording high-quality records. For the diagnostic error cases that have been successfully reported, the report quality is variable. Reports can hardly avoid unintended false information, and the accuracy and clarity of the reports should be improved.

The other ways of finding diagnostic errors are usually through clinical follow-ups or necropsies[28]. Specially, autopsy is considered by many researchers to play an important role in providing reliable information for misdiagnosed cases.[3 4]

Thus, poorly reported diagnostic errors and lack of high-quality records make the relevant studies more challenging.

- Diagnosis sometimes can be under time constraint especially in emergency departments, and is a complex process involving many phases and factors. These factors have an impact on the actions during the diagnostic process and then affect the diagnostic decisions directly or indirectly, while, at the same time, the factors may be related to each other as well. The entire diagnostic process can be reflected in the diagnostic process, which includes initial access to the patient, history taking, physical exams, tests, doctor assessment, consultation and follow-up procedures. Schiff et al in 2005 illustrated where and what errors may occur during the diagnostic process[20], as shown in Figure 2.1.

  As we can see, for a better study of diagnostic errors and finding ways of error reduction, it is important to interpret the diagnostic process, show associated factors at each step, and then provide a systemic view of relevant factors as well as interrelations of variables.

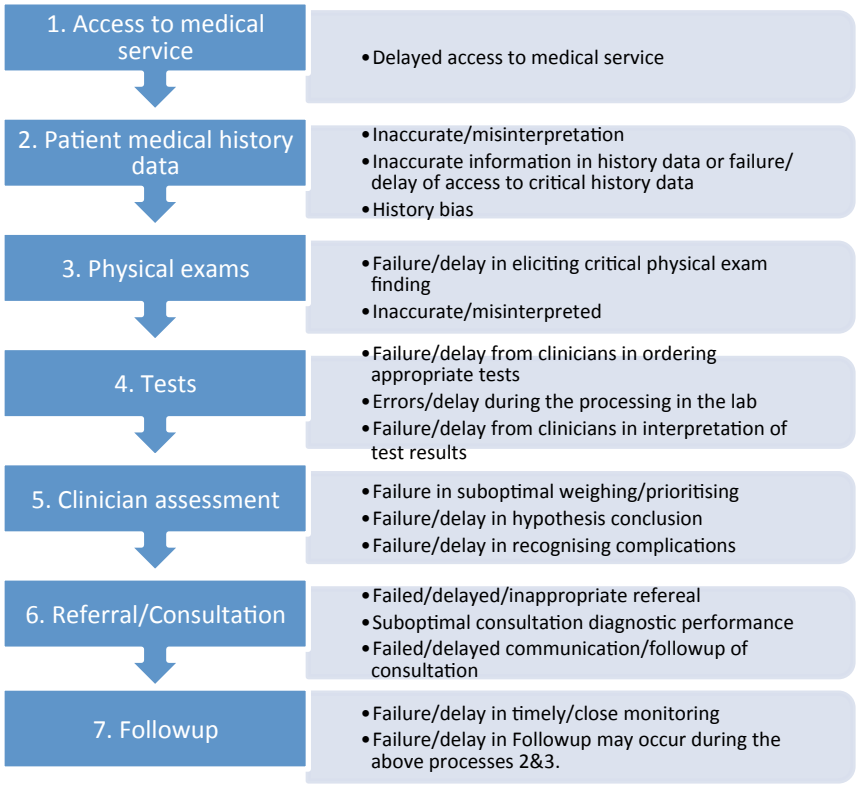


Figure 2.1 Where and what errors may occur[20]

- In spite of many relevant researches about diagnostic errors, more effort still needs to be made to find effective ways of reducing diagnostic errors and improving patient safety, especially in terms of systemic analysis. Many studies focus on the individual health provider or single factor for error-improvement strategies. However, diagnostic errors should be observed from the entire diagnostic process as a system, and methods for changes from a system view, which reflect the impact of all key health providers and key factors, are believed to be a more effective way of seeking solutions.[29 30] Relevant works from Reason[31 32] reflect that systemic analysis has its advantage of reducing errors, using the "Swiss Cheese model"[31], shown as Figure 2.2. The model illustrates that although many layers lie between hazards and losses, errors can happen if flaws in each layer are aligned.

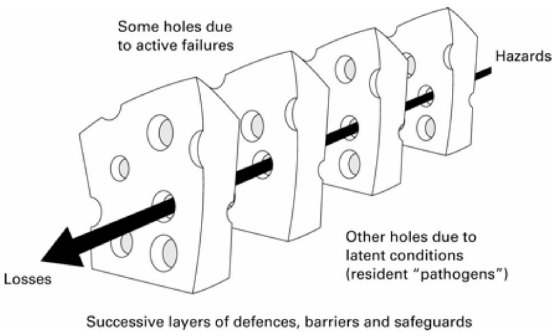


Figure 2.2 Swiss cheese model[31]

Additionally, one research[33] in 2002 suggests that methods of reducing diagnostic errors from a view of the entire diagnostic process, such as either avoiding diagnostic errors or correcting diagnostic errors in the early stage, represent a possible and effective way of preventing harm to patients, and fail-safe methods will fail to improve patient outcomes.

Thus, the diagnostic error problem requires the analysis of the entire diagnostic process as a system.

- The way of measuring diagnostic errors remains unknown. A more scientific and precise method or magnitude to measure diagnostic errors is needed[34].

2.4 Summary

Five characteristics of diagnostic error are discussed in this chapter. It has been shown that diagnostic errors have negative impact on patient safety and trigger the demand to reduce its occurrence and impact, and diagnosis is a complex diagnostic process covering different factors at each phase. Although many researches have been carried on seeking ways of reducing diagnostic errors, most are either disease-focused or individual provider-focused. According to the primary literature review, few studies provide systemic analysis of factors and errors during each phase of diagnostic process. Current ways of reducing diagnostic error need to be further reviewed and summarised. What is more, no study was found in terms of quantifying relevant factors and modelling diagnostic errors. Thus, the diagnostic process should be taken as a system, and diagnostic errors need to be studied and modelled from the view of the entire system, so that models of diagnostic error can represent the factors of diagnostic errors and the interrelations of key variables in the system, and also provide ways of seeking possible methods from a systemic understanding of the system.

The next chapter includes the methods used in this thesis to address the problems above and the reasons for their choice.

Chapter 3 Methods

3.1 Introduction

This chapter introduces the methods used in the thesis. It begins by comparing models of errors in healthcare. Then, it introduces system dynamics modelling and regression modelling, including their application hypotheses of how they might help with the diagnostic error problem. Data collection methods are also presented in this chapter.

3.2 System modelling

System modelling is an essential tool of analysing the elements or components in the system. It has significant benefits in managing the interactions of the elements, and understanding the functionality of the system, and it especially enables the analyst to predict the effect of changes to the system. [35]

System modelling has been widely applied, and can represent a system in various ways with different aims. It is often classified into the following types: data processing model which shows how the data are processed at different phases; composition model showing how entities are composed of other entities; architectural model showing principal sub-systems; classification model showing how entities have common characteristics; and stimulus/response model showing the system's reaction to events. [36]

The downside of modelling may be considered as some types of modelling may produce too much documentation, and a model with too many details may be difficult for users to understand sometimes. [36]

A typical modelling and simulation process mainly involves developing a simulation model, designing a simulation experiment, and performing simulation analysis. Specifically, the process covers the possible steps as below[35]:

- Identify the problem.
- Formulate the problem.
- Collect and process real system data.
- Formulate and develop a model
- Validate the model, and iterate between model refinement and validation.
- Document the model for future use.
- Select an appropriate experimental design.

- Establish experimental conditions for runs.
- Perform simulation runs.
- Interpret and present results.
- Recommend further course of action.

Diagnosis is a complex process involving multiple phases and many factors. Studying diagnostic errors requires a way of showing the correlations of the factors, analysing the problem from the entire diagnostic process. Modelling errors in the diagnostic process can help analysing root causes by constructing the structure of cause and effect variables. At the same time, it is a method of describing the entire system and analysing the problem from a system viewpoint. It is a systemic approach to discover how errors happen, as well as to propose better management policies and organisational structures[37]. Models in healthcare provide a constructive way of discovering real causes and offer theoretical underpinnings for both researchers and clinicians.

3.3 Comparing models of errors in healthcare

Based on the searching of models in error management or error control in healthcare and complex systems, several models were found in the studies of modelling medical errors; these models are structural-equation models (SEM), Bayesian hierarchical models, and system dynamics models. This section discusses these three types of modelling methods separately.

a. Structural-equation modelling

SEM is "a comprehensive statistical approach to testing hypotheses about relations"[38] among observed (measured) variables and unobserved (latent) variables. It usually includes two parts: a "measurement model" and a "structural regression model"[39 40]. A "measurement model" defines latent variables, which are not directly observed but rather inferred from one or more observed variables. A "structural regression model" links latent variables together, via statistical methods, to observe the quantitative impact of inputs on the outputs and to estimate relations between inputs and outputs.

SEM is very similar to traditional statistical methods, which are based on linear statistical models. At the same time, it is an "advantaged" version of traditional methods. It can solve the relational problems for both observed and unobserved variables, while traditional methods analyse observed variables only. It applies multiple tests, such as chi-square or

Root Mean Squared Error of Approximation (RMSEA), to the determination of the model fit.[41]

It was used to explore the factors of medical errors as well as the estimation of factor impact in a prospective cohort study in 2012[42]. The study was designed to find the cause-and-effect relationship between potential predictors from nurses, such as age and degree of depression, and medical errors. The outputs suggested SEM could be a countermeasure for the factors of medical errors.

Generally speaking, SEM provides a reliable way of determining cause-and-effect relationships, especially for unobserved variables. However, it requires a good quantity of data to perform and it is not applicable to reflecting the component structure of a system. In addition, the potential causes have to be predicted first.

b. Bayesian hierarchical modelling

Bayesian hierarchical modelling, a particular type of Bayesian network, is also a type of statistical model that represents the probabilistic relationships. One study[43] in 2010 using a Bayesian hierarchical model discovered causes of a type of medical error, called "near misses" which had the potential to cause serious harm but did not. This model, using a mathematical approach, analysed clinical evidence or data to find out the causes or the contributing factors of the error.

Compared to the SEM model, the relationships in the Bayesian hierarchical model are not limited to the cause-effect relationship, but it requires a much larger amount of evidence as the input data to determine the relationships in a quantitative way.

c. System dynamics modelling

System dynamics modelling is different from statistically based SEM and Bayesian hierarchical modelling methods. It uses simulations to provide the relevant data flow information in the system components over time. It not only illustrates complex internal relationships that affect system behaviours, but also reflects the component structure of a system. Thus, it is an approach widely employed for complex systems.[44] Eric Wolstenholme explained system dynamics as "*problem solving and analysis of complex real world systems by methodological means, where the emphasis is on promoting holistic understanding rather than piecemeal solutions*"[63]. By simulating an over-time look of the output flow under both positive and negative input factors, system dynamics modelling provides the guidance for potential interventions or the policy for the problem.

System dynamic modelling has a significant advantage in representing known relations and reflecting components in complex systems. However, as with SEM, it also requires the causes to be determined first. Moreover, it has limitations in identifying the relationship functions between multi-variables and their effect variables. Moreover, it requires real-time data to plot real-time simulation outputs.

3.4 System dynamics modelling

This section introduces the system dynamics modelling method in further detail, explains the reason to implement system dynamics modelling in diagnostic error analysis, and shows how system dynamics modelling can perform systemic analysis of diagnostic errors.

3.4.1 Introduction of system dynamics modelling and its applications

Jay W. Forrester of the Massachusetts Institute of Technology firstly proposed the system dynamic modelling during mid-1950s, and published his work in the book *Industrial Dynamics*[37], where system dynamics modelling was described in length in the application of helping General Electric (GE) manager better understand industrial processes. Until the late 1960s, system dynamics modelling had been applied almost exclusively to corporate or managerial problems.[45]

Since the late 1960s, Jay W. Forrester had been working with John Collins and they published a book titled *Urban Dynamics*, which served as the sign that system dynamics started to broaden its application into non-corporate areas.[52] It was the key that led to the later two well-known projects in the early 1970s in system dynamics area*: World Dynamics* and the *Limits to Growth*. In 1970, Jay Forrester was invited to a meeting in Bern by a group called the Club of Rome, and the world problems discussed at the meeting became the basis for the model in *World Dynamics.*[52] *World Dynamics* was published in 1971, which extended system dynamics application into modelling important interrelationships between world population, industrial production, pollution, resources, and food.[45] The model also predicted a collapse of the world socioeconomic system sometime during the twenty-first century. The Club of Rome further funded Meadows and her associates to conduct the *Limits to Growth* where system dynamics modeling was applied in explaining world population growth and economic growth. It was published in 1972, and discussed the growth, overshoot, and collapse of the world economy using system dynamics models.[46]

The Fifth Discipline by Peter Senge in 1990[47], which described systems thinking in helping to convert companies into learning organizations, popularized the qualitative systems dynamics in organizational learning and managerial application. At about the same time, Eric Wolstenholme developed "System Enquiry"[63] and gave its definition by combining its original definitions and its later broadening of applications: " A rigorous method for qualitative description, exploration and analysis of complex systems in terms of their processes, information, organisational boundaries and strategies; which facilitates quantitative simulation modelling and analysis for the design of system structure and control"[63]. Since 1990, a number of researchers of system dynamics, including Richardson[48], Richmond[49], and Vennix[50], started to recognize the issue of client involvement and develop protocols for group model building that is a method for analyzing data with a group of people. John Sterman further promoted system thinking in the analysis of policy and strategy in 2000.[51]

System dynamics modelling can be applied to any dynamic system with any time and spatial scale[51]. System dynamics modelling initially arose in relation to corporate or managerial problems[37] and then has broadened its application into non-corporate areas since the late 1960s[52]. Nowadays it has been used widely in real life, including healthcare. It was implemented in healthcare applications since the 1970s, and has been witnessed in a large number of applications of healthcare researches in recent years. These researches involves several aspects of health and social care[53], such as health reform[54], capacity planning[55], older people's services[56], disease management[57] and mental health[58]. The health interest group in International System Dynamics Society established in 1983, was organized in 2003.[62] Its applications in healthcare and social care system aim to illustrate the structure of system resources and give suggestions on significant resource that can be saved, without influencing performance[59]. Furthermore, its applications in disease researches help to interpret experimental results and understand the dynamics of results, for example research into HIV and human immune system[60]. System dynamics modelling is currently adopted as the major tool in the modelling kit for the Operational Research(OR) Group in the Department Health, England and has been used in a wide range of health policy and programme development and implementations[61], including: assessing public health risks; screening for disease, such as screening for cervical cancer and for chlamydia; managing waiting lists for hospital treatment; planning the healthcare workforce; and developing emergency health and social care.

Homer and Hirsch in 2006 summarised its current application in healthcare into five aspects[62]:

1) *Disease epidemiology*
2) *Substance abuse epidemiology*
3) *Patient flows in emergency and extended care*
4) *Health care capacity and delivery*
5) *Interactions between health care or public health capacity and disease epidemiology.*

The system dynamics modelling can be both "solution oriented" and "learning oriented". It is commonly recognised as a tool for discovering problems and suggesting solutions. At the same time, it is also a tool for learning about the system[61]. In particular, when there is no single optimal solution, it helps to learn system behaviours and to propose possible strategies.

3.4.2 Two phases of system dynamics

System dynamics involves two phases: "qualitative system dynamics" and "quantitative system dynamics".[51]

Qualitative system dynamics:

The qualitative system dynamics phase is the model construction and analysis. It defines how individual variables are working in the system, via visualising the interrelations among variables, and makes clear how interrelated variables affect each other using arrow links. It can be implemented in order to "quickly capture hypotheses about the causes of dynamics; elicit and capture the mental models of individuals or teams; and ccommunicate the important feedbacks which you believe are responsible for a problem."[51]

A qualitative system dynamics model is often known as a "*Causal Loop Diagram*"(CLD), as it uses a "cause and effect diagram" to illustrate the factors or causes of the problem, and to represent the cause-effect relationships among model variables. It uses nodes and arrows to graphically represent the variables and interrelations in a system. The variables of the system are represented as nodes, and arrows link the elements together to represent cause-effect relations.

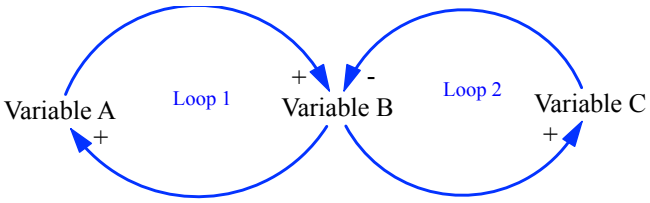


Figure 3.1 A causal loop diagram example

Figure 3.1 is a CLD example. Three variables A, B, C are represented in nodes in the figure, and arrows show cause-effect links or relationships between the two linked variables. The arrow polarities indicate link polarities. A points to B with a positive polarity at the end of the arrow, which means A is the cause of B and A has positive effect to B. In other words, if A increases, B would increase. On the other hand, C has an arrow with a negative polarity to B, which indicates C has negative effect to B. Following the arrows, B also gives feedback to A &C and feedback loops are formed. By tracing the effect of a change around the loop, the loop polarity can be determined. Loops, such as loop 1, that can reinforce change are named as positive loops or reinforcing loops, while loops that are self-correcting are negative loops or balancing loops, such as loop 2.

Quantitative system dynamics:

Quantitative system dynamics represents the system using "Stock and Flow Diagrams". A simple diagram is shown as Figure 3.2.

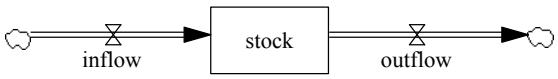


Figure 3.2 A simple stock and flow diagram

As we can see, the above diagram is made from a rectangle with an input arrow and an output arrow. The rectangle, referring to as "Stock", indicates a quantitative stock. The input arrow indicates the inflow of the stock, and it increases the stock level. On the other hand, output arrow indicates the outflow of the stock, and it decreases the stock level. [63]

Mathematical representations, shown as an integral equation in Eq. (3.1) and a differential equation in Eq. (3.2), can be exploited to explain level changes of the stock over a period of time from initial time t0 to current time t. Stocks are known as integrals or state variables, and flows are known as rates or derivatives.

$$stock(t) = \int_{t_0}^{t} [inflow(s) - outflow(s)]ds + stock(t_0) \quad (3.1)$$

where s represents any time between the initial time t_0 and the current time t.

$$\frac{d(stock)}{dt} = inflow(t) - outflow(t) \quad (3.2)$$

3.4.3 Its application hypothesis

A basic structure of decision making for real world problems follows a circular loop, and includes three components: 1, recognising real world problems; 2, collecting or retrieving for decision makers reliable qualitative and quantitative information from the real world; 3, decision makers making decisions, and decisions implemented back into the real world system.

How system dynamics models are applied to real world problems is illustrated in Figure 3.3, introduced by Sterman in 2000 [51]. System dynamics modelling represents the dynamics of a system in the second component "Information Feedback" by discovering and representing the feedback process. System dynamics modelling simulation is the stage of proposing possible strategy, structure and decision rules, which helps decision makers with making decisions.

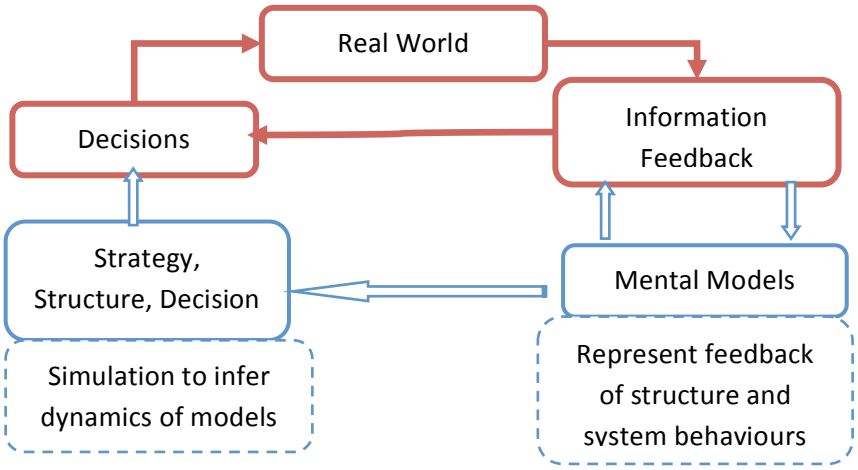


Figure 3.3 How system dynamics models implement to real world problems[51]

Similarly with other applications in complex systems, system dynamics modelling can contribute to a systemic analysis of diagnostic errors in many ways from both "solution oriented" and "learning oriented" aspects. The following hypotheses can be used in implementing system dynamics modelling in the context of diagnostic error analysis.

- It helps with learning the relations between factors and effect variables, as well as understanding factors, especially for factors not readily quantified.

  Diagnostic errors involve many factors during the whole diagnostic process. After discovering all key factors of diagnostic errors, CLDs can help illustrating the links between the factors and relevant effect variables. Also, further quantitative analysis helps to understand how the individual factor contributes to the system.
- It provides a systemic analysis of the diagnostic error problem.

  The diagnostic process can be used as a guide in mapping the key phases, and links phases together. System dynamics modelling can follow the process and provide a whole picture for tracing errors in the process and how it is linked to decision-making errors as well as patient outcomes.
- System dynamics modelling, linking different phases of the system together, performs simulation of the number of errors in all phases and then suggests potential strategies for reducing diagnostic errors.

  During each phase, new errors may occur when doctors collect diagnostic information or diagnostic clues. At the same time, errors may be realized or corrected in the following phases. As shown in Figure 3.4, stock and flow diagrams can be used to interpret the number of errors at each phase and error flows in the system. The simulation of the quantitative model helps to analyse how the level will be changed by the changes of relevant variables over a period of time.

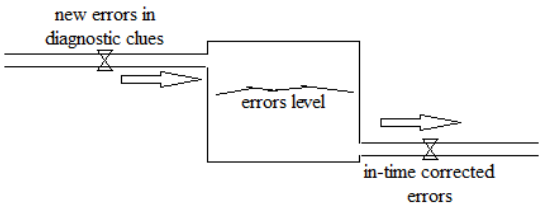


Figure 3.4 Hypothesis of applying stock and flow diagrams for diagnostic errors

Chapter 5 and Chapter 6 will give further explanations on the application of system dynamics modelling including both the qualitative model and the quantitative model.

3.4.4 Why use system dynamics modelling

The reasons to choose the system dynamics modelling method is mainly because of the following considerations:

1. Its adaptation to complex real world systems.

Diagnosis is a complex process. Many factors involved in the diagnostic process can affect the final decision making directly or indirectly, and many phases in the diagnostic process are connected and sometimes worked together. System dynamics modelling is adapted to complex systems. It could help to analyse the causes, identify or structure the internal relationships of model elements, and link all phases together in order to conduct the analysis and to observe the whole system behaviours.

2. Its coverage of both positive and negative factors.

System dynamics models can present both positive and negative factors. Specifically, The factors are called "*causes*" in the CLDs, where the "*causes*" can have both positive effect and negative effect.

3. Its both "solution oriented" and "learning oriented" applications.

By performing simulation experiments, it helps to discover system behaviours. Especially, when one or more factors of the system are changed, it provides a clear view of the changes in the relevant key variables. For the system with variables that can be readily quantified, it is known as a way of predicting an optimal solution to a problem[64]. If the system is closely associated with variables that cannot be readily quantified and there will be no single optimal solution, system dynamics modelling becomes a tool for learning.[61]

4. Its current application in healthcare.

System dynamics modelling has been widely used and known in healthcare. Extending its application to diagnostic errors helps with learning about diagnostic errors, and provides an opportunity of linking or combining the model of diagnostic errors with other existing models together.

5. An economic approach.

System dynamics modelling can implement computer simulations that are very low cost[62], which is another important reason why it has been largely applied in both engineering work as well as in healthcare. It provides risk-free experiments, and then encourages creative thinking of possible solutions to the problem.

3.5 Regression analysis

This section discusses the hypotheses involved in applying regression analysis to the systemic model of diagnostic error, where the implementation of regression analysis mainly aims to overcome the limitation of the system dynamics modelling.

3.5.1 Introduction of regression analysis and its applications

Although system dynamic modelling provides a method for looking into the entire system instead of partial analysis, it has its limitations at the same time. It is limited to determining parameters as well as relationship functions between multi-variables and relevant dependent variables. The initial diagnosis phase, *history taking and physical examinations*, is an essential phase for decision making in the diagnostic process, and it involves many key factors that affect the error rate in the initial diagnostic hypotheses. The quantitative relationship functions between the factors and the number of errors in the diagnostic hypotheses are unknown. Thus, a method is required to determine the interrelationship functions in this part and to overcome the limitation of system dynamics modelling.

Regression analysis is a statistical method and is widely used to estimate the relationship functions among variables based on observed data, and regression analysis is used in the thesis regarding the issue above. It is applicable for the relationship between a dependent variable and one or more independent variables and the relationship can be either linear or non-linear.

Generally, regression modelling consists of four parts:

1) Cohort construction. This is to define aims, targets and, sometimes, time window.
2) Feature/factor engineering. It includes data cleansing, data imputation, feature construction and feature selection. The common methods used for feature selection are using filters, a wrapper or embedded optimisation. There are three reasons to conduct feature selection: simplification of models for an easy interpretation; shorter training times; and enhancing generalisation and avoiding overfitting.
3) Regression modelling. Firstly, regression analysis requires identifying the possible types of function. For example, logistic regression is often chosen when the dependent variable is dichotomous. Secondly, it identifies the parameters of the function using training data. For example, a linear relationship can be represented as a function in a form of $Y= p + qX$, where p & q are the parameters. The linear

relationship can be determined by identifying the parameters p & q using training data.

4) Model evaluation. The evaluation approaches cover using cross validation or external validation to compare the model outputs to the real data, or using metrics to analyse model outputs, such as variance or R^2.

3.5.2 Its application hypothesis

Regression modelling can be implemented in estimating the quantitative relationships between multi factors and their dependent variables, such as the number of errors, during the phase of *history taking and physical examinations*. It complements system dynamics modelling by providing modelling relationship functions between multi-variables and their dependent variable. These functions obtained by regression analysis are further applied in the system dynamics model.

In order to accomplish determining the relationship functions, following the same recipe for the SEM method introduced in section 3.3, there are two tasks: defining variables and determining regression models. For the first task, the results of the CLDs and the structure of the stock and flow diagrams in system dynamics modelling can be used to define the variables as well as the qualitative variable relations. The second task is to conduct regression analysis and determine the relationship functions.

For unobserved or non-numerical variables involved in the system, there are two ways of scaling the variables. One method is used in the SEM, and it measures the unobserved variables by one or more observed variables as indicators. Take a variable “depression” as an example. It can be measured by several observed variables such as “loss of interest”, “sleep problem”. The downside of the method is that it requires more data for the indicators. The other method is using a *Likert scale*, which can be used to scale these variables into measurable ordinal variables. For example, “sleep problem” can be scaled into three groups: “light”, “moderate” and “severe”. Both methods can be applied to measuring the factors of diagnostic errors.

Regression models generally use the goodness of fit to test the model fit. It can either use new observed data to test the model-predicted results, or summarise the discrepancy between original observed values and the values expected under the model[65]. The general measures could be applied to testing the model fit, such as significance testing, chi-squared testing or the analysis of variance.

3.5.3 Why use regression analysis?

The main reasons why regression analysis is applied to modelling diagnostic errors are listed as follows:

- Regression analysis is widely used in determining the relationship functions in different areas, and sufficient software and methods support conducting regression analysis.
- It fits the process of system dynamics modelling well. The CLD in system dynamics modelling provides the information for its feature-engineering step, and its modelling results help system dynamics modelling provide relationship functions and thus overcome the limitations of system dynamics modelling.
- A Bayesian network is not applicable because it requires even larger amounts of data to be collected, and the relations in the problem of diagnostic errors are limited to cause-and-effect relations.

3.6 Software

The research described in this thesis is conducted mainly using two pieces of softwares: Vensim and SPSS. The software used for the system dynamics modelling is Vensim[66]. Vensim is a free software package. It can illustrate the CLD, build hieratical cause trees, demonstrate and simulate quantitative models. The software used for the regression analysis is SPSS[67], which can conduct correlation analysis perform, different curve estimations, regression analysis as well as testing model fit.

3.7 Data collection methods

Data used in the thesis mainly come from two sources: public data from literature and data from experts. For the variables that have sufficient data in the literature, literature data are chosen as a prior data source. For the variables that lack appropriate literature sources, data are collected from experts.

There are many methods used to gather data from experts, and they include interviews or semi-structured interviews, questionnaires, participant observation, and expert elicitation. Questionnaires and expert elicitation are chosen to collect the data from participants. A questionnaire is a convenient approach to receive a large quantities of feedbacks from a wide range of participants. It is chosen to collect the data for regression analysis because it

provides a number of data sets for individual variables. Expert elicitation is a method used to obtain estimations of variables from expert opinions, and specifically works for the assessment with insufficient, incomplete, controversial or inconsistent information.[68] It is chosen in the thesis to provide estimated values with confidence bands for simulation tests.

3.8 Summary of methods used in the thesis

This thesis adopts a systemic model for diagnostic errors based on the system dynamics modelling approach, and also combines computer science regression methods into the system dynamics modelling.

The first part of this thesis focuses on the qualitative system dynamics modelling, which uses CLDs to qualitatively present the factors affecting diagnosis and their interrelations. A systematic literature review and discussions with experts are implemented as the main methods in this part. The systematic review provides the source and evidence to identify the key factors of diagnostic errors. An initial CLD is designed based on the translation of the knowledge from the systematic review. Then, discussions with clinicians about the initial CLD are followed to further refine the qualitative model and improve its reliability. The second part is about quantitative system dynamics modelling which implements the stock and flow diagrams to quantitatively simulate the factors and diagnostic errors. The quantitative system dynamics modelling takes the diagnostic process as a system, and maps the diagnostic phases following the diagnostic process, which covers all the phases from "history taking and physical examination" to the "after-diagnosis" phase. Based on the previous CLD, it presents the error flows in different phases of diagnosis and relations with relevant factors. Regression analysis is used in this part to identify the relationship functions of factors and their dependent variables in the phase of history taking and physical examinations. After applying the regression results to the quantitative model, simulations are conducted to demonstrate how changing one or more variables affects diagnostic outputs, where the variable changing may indicate the effect of an external intervention or a policy. The evaluation of the model is described in the last part of the thesis.

The modelling process of this thesis is summarized and illustrated in Figure 3.5. This thesis firstly identifies the problem of diagnostic error studies, and finds out that the study of diagnostic errors requires a systemic approach of representing the diagnostic process as well as the interrelations of the variables during the diagnosis, which indicates that the

perception of the problem of diagnostic errors leads to a modelling purpose. Then, the model formulation is further developed in order to understand the diagnosis process and the diagnostic errors. During the modeling process, model identification is a key ingredient. It includes two aspects: model structure development and model parameter identification. Model identification starts with identifying key factors of diagnostic errors. A qualitative model, as the essential model structure, is constructed and refined after this process. Based on the qualitative model, a quantitative model is developed, where model parameters are identified. Furthermore, simulation experiments are conducted and the quantitative model is evaluated. Meanwhile, feedback from model simulation and evaluation can help to refine model structure and model parameters.[69] The process of between model developing and model evaluation may be iterated in order to do refine the model based on the feedback of simulation and evaluation results. In the end, the model is documented and test results are presented for future use.

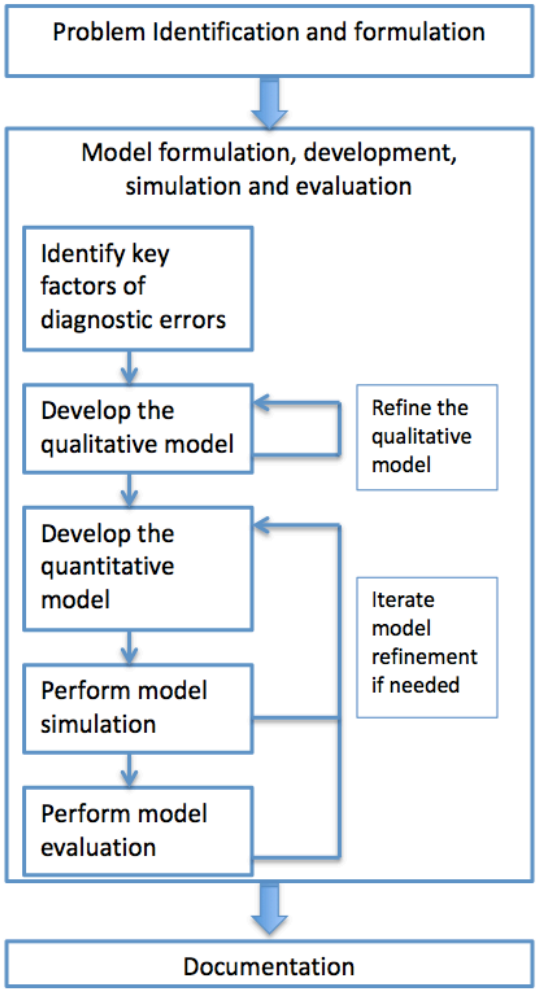


Figure 3.5 Modelling process of this thesis

More specifically, the methods used in the research are listed as below, following an order of research process:

1) Identify key factors of diagnostic errors and understand relevant researches on diagnostic errors:

Purpose: To identify the key inputs of the model and take a broad look at current ways of reducing diagnostic errors.

Methods: A systematic literature review of relevant studies is carried out in this step.

Outputs: Key factors of diagnostic errors are discovered. A summary of current methods used on reducing diagnostic errors is produced.

2) Develop and refine the qualitative model:

Purpose: To illustrate a qualitative model reflecting the relationships among the key factors and diagnostic errors, and to modify and refine the qualitative model in order to increase the model's reliability.

Methods:

- Causal loop diagrams are constructed as the qualitative systems dynamic models to analyse the diagnostic process and the key factors affecting the final diagnosis.
- Discussions with clinicians about the initial qualitative model are conducted. Feedback and suggestions of the clinicians are collected from the discussions.

Outputs: A causal loop diagram of diagnostic errors with its factors is developed, and the qualitative model is further refined based on the feedback and suggestions.

3) Develop the quantitative model:

Purpose: To transfer the qualitative model into a quantitative model structure, and to modify the model in order to present relevant factors as well as error flows during the diagnostic process.

Methods: Stock and flow diagrams are used for constructing the quantitative model.

Outputs: A quantitative model structure that shows factors, possible patient case flows and patient outcomes is constructed

4) Identify quantitative relations:

Purpose: To determine the unknown relational equations in the quantitative model.

Methods:

- A Likert scale is used to quantify the non-numerical factors.
- Questionnaires are conducted at this step to provide the relevant data source for regression analysis.
- Regression analysis is used to model the relationship function between multi-factors and dependent variables.
- Significance testing and the analysis of variance are implemented to test the model fit

Outputs: Non-numerical factors are quantified. Data from the questionnaires are collected and analysed using regression modelling methods. The relationship functions between multi-factors and dependent variables are identified, evaluated and further applied into the stock and flow diagram. And finally, the quantitative model is developed after this step.

5) Perform model simulation:

Purpose: To visualise how model elements work in the system and observe model behaviours, and at the same time, to simulate the changes inside the model when changing one or more factors.

Methods:

- A literature review is performed to collect relevant data from literature for the simulation experience. At the same time, expert elicitation is performed to collect data for the variables that are not available from literature source.
- Cook's classic model[181] is used for the data gathered from the expert elicitation.
- Scenario analysis is used for the model simulation. The quantitative model is simulated under different assumptions or scenarios.

Outputs: Relevant data are retrieved from the literature, and data from the experts are prepared for the simulation. Then, the qualitative model is simulated under different scenarios.

6) Perform model evaluation:

Purpose: To test and evaluate the model outputs

Methods: A list of assessments for dynamic model evaluation from Sterman (2000)[51] is adopted, which covers: boundary adequacy, structure assessment, dimensional consistency, parameter assessment, extreme conditions, integration error, behaviour reproduction, behaviour anomaly, family member, supervised behaviours, sensitivity analysis, and system improvement

Outputs: Evaluation results of the model are documented.

3.9 Conclusion

This thesis implements system dynamics modelling as a problem-solving method to analyse root causes and interrelations among variables of the system. It proposes solutions via the holistic understanding of the entire complex real world systems, rather than piecemeal solutions.[63]

Additionally, since system dynamics modelling has its limitations in identifying variable correlation functions, computer-based regression modelling is also conducted to model the correlations between the factors and dependent variables, which then overcomes the deficiency of implementation of system dynamics modelling.

The thesis follows a typical approach to model development. It initially identifies the purpose of the proposed model. Next, it identifies organisational variables, and develops a qualitative model reflecting the root causes of diagnostic errors. Afterwards, it develops the system into a quantitative model that describes the behaviours of the model. Finally, it simulates the system showing the information flows or data flows.

In the next chapter, a systematic review of the literature is conducted, which aims to identify the relevant factors and interrelations as model variables, and further look at the current methods used to reduce diagnostic errors.

Chapter 4 A Systematic Review of Literature

4.1 Introduction

A systematic review of literature is performed in this chapter. The systematic review helps further understand diagnostic errors by identifying the key factors associated with diagnostic errors, understanding how these factors are linked through diagnostic process, and analysing current methods used to reduce diagnostic errors. The findings of key factors are used for identifying model variables, and can also be found in the peer-reviewed paper by GUO et al (2014)[70].

The searching method of the systematic review is introduced at the beginning of this chapter together; the two main findings, key factors and current methods of diagnostic error reduction, are individually discussed in the results part. Relevant papers about key factors of diagnostic errors and current methods of reducing diagnostic errors are retrieved from relevant databases. After further detailed review, these papers reflect the key factors of diagnostic errors into six categories according to different aspect focuses. Also, both electronic and non-electronic methods of reducing diagnostic errors are summarised in results.

4.2 Searching method

Papers published in English between 2002 and 2012, which tried to address diagnostic errors, were retrieved from PubMed and relevant databases.

4.2.1 Searching questions

The review is to answer the following questions:

- What are the key factors associated with diagnostic errors?
- What are the current methods of reducing diagnostic errors?

4.2.2 Sources of literature

The following databases were covered in the search: PubMed, CINAHL with Full Text, EMbase, PsycINFO, Web of Science and IEEE Xplore. The timeframe limit for the published literature is from 2002 to 2012. The language filter was set to English language only.

The following search phrases were used during the search to cover all research-levels of key factors that affect the number of diagnostic errors. The effect can be discussed in two ways:

causing errors (positive effect on the error number) or reducing errors (negative effect on the error number).

Phrases used in the search:
Diagnostic Error[MeSH terms];
Delayed Diagnosis[MeSH terms];
Misdiagnosis [MeSH Terms];
Reduce diagnostic errors;
Prevent diagnostic errors;
Manage diagnostic errors;
Cause

Table 4.1 Search phrases in the systematic review

4.2.3 Literature selection criteria

Papers were excluded if: (i) the paper was a commentary or general review paper; (ii) the paper was used for a particular medical disease study only. In other words, the selected papers should focus on consequences of and contributors to misdiagnosis instead of discussing clinical features.

4.2.4 Process of retrieving articles

Name of databases	**Number of articles**
PubMed	261
CINAHL with Full Text	157
EMbase	55
PsycINFO	27
Web of Science	79
IEEE	17
	Total: 596

Table 4.2 Number of papers initially retrieved from databases

A total number of 596 papers are initially retrieved from databases, and details are shown in Table 4.2. Then, the papers are further selected using duplication removal, screening titles and abstracts, and applying a number of criteria. A total of 65 papers are selected in the end, and full papers are reviewed. The flow diagram for the selection of the relevant articles is shown in Figure 4.1.

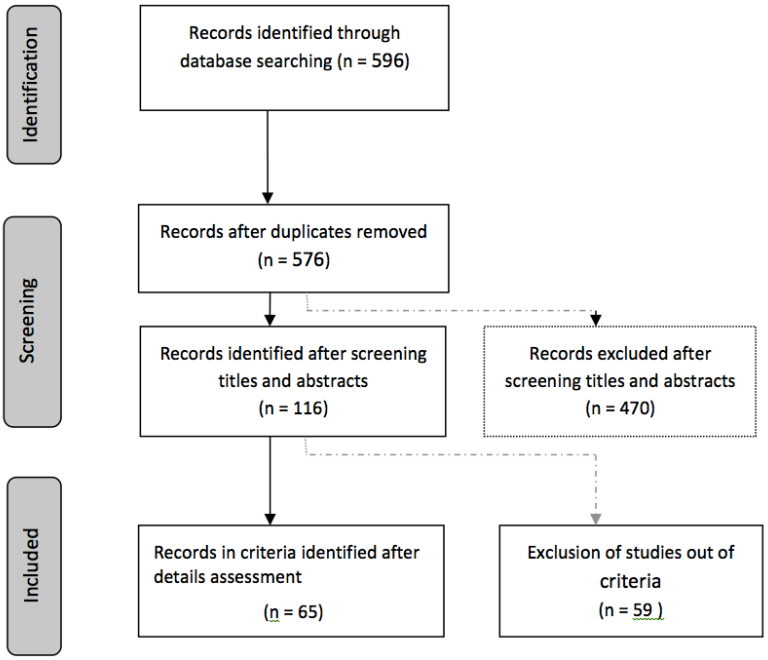


Figure 4.1 Flow diagram of study selection

4.3 Findings

A total of 65 papers are retrieved and viewed for the study, among which 35 papers focus on studies of factors only, 17 papers focus on solutions only, and 13 papers discuss factors and provide solutions or strategies as well. At the same time, six papers discuss multi-aspect factors. A limited number of articles focus on the evaluation of possible solutions. Though disease feature studies were not included, some 17 papers used in the study of diagnostic errors are clinical disease related, which shows relevant researches are closely connected with clinical disease study. Figure 4.2 illustrates the distribution of papers in terms of focus or usage.

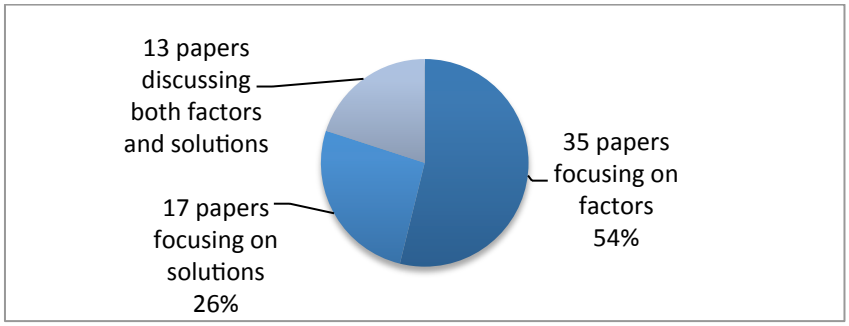


Figure 4.2 Paper distribution

Findings show diagnostic errors are closely connected with other medical errors occurring during the diagnostic process. In particular, biochemical test errors, radiological test errors

and improper referrals may lead to a final diagnostic error. It is also found that there are more quantitative researches regarding biochemical and radiological test errors than other relevant researches, such as history taking, physical exams and clinical reasoning.

Factors involved in diagnostic errors cover many aspects. Factors are classified according to the relevant aspect in subsection 4.3.1, and relevant studies were divided into groups for supporting detailed discussions of the factors. Relevant methods of error reductions are described in subsection 4.3.2.

4.3.1 Factors of diagnostic errors

The factors that affect a diagnosis are diverse, and sometimes are linked to each other. These factors are summarised into six categories according to different foci, shown below.

- Clinical disease features: It covers two aspects: whether a disease is well researched and the clarity of disease symptoms or presentation. When the clinical findings are untypical or unclear, the disease is more likely to be misdiagnosed.
- Educational background of patients or doctors: Patient educational background is highly related to patient health awareness; that is about whether proper actions could be carried out after a symptom was presented. The educational background of doctors indicates the experience and knowledge of doctors that are gained from the past education, including relevant training specifically designed for avoiding diagnostic errors. Experience and knowledge covers the abilities of doctors to observe clinical signs, understand collected clinical information, reason with clinical information and organise treatment plans.
- Collecting and reasoning with diagnostic clues: Diagnostic clues are the evidential information used to make diagnostic decisions. Successfully collecting this clinical diagnostic information involves gathering correct information from patient medical history, patient-doctor communication, initial physical examinations as well as further diagnostic information collected from tests or consulting from other healthcare providers. Also, clinical reasoning plays an essential role. This involves using doctors' knowledge to retrieve the right key clues from a series of collected diagnostic information. In other words, clinical reasoning involves properly weighting collected diagnostic information to retrieve key information as to diagnostic clues. Doctors are more likely to make an error when they receive ambiguous or conflicting diagnostic information.

- Psychological factors: This mainly involves biases and doctor awareness of an error or high-risk cases.
- Follow-up after a diagnosis: Close follow-up helps to discover and correct existing diagnostic errors before they can have a severe impact on patients.
- Other factors

a. Clinical disease features: understanding signs and symptoms

Different diseases can have similar or diverse clinical manifestations, and features of diseases here mean the signs or clinical information, which help doctors make a determination of one disease or distinguish one disease from another.[71] Thus, whether the features of diseases are well known by doctors has a direct effect on the accuracy of clinical diagnosis.

Disease features that may lead to a misdiagnosis are mainly focused on two aspects. Table 4.3 lists the keys factors with relevant examples in terms of clinical disease features.

- **Whether a disease is well researched**

  When the clinical findings are rare, the disease is more likely to be misdiagnosed. Unsuspected rare causes, or patient groups that have a lower index of suspicion can increase the risk of a delayed or missed diagnosis. When the presentations of the disease are "too similar or too rare", the disease with atypical presentations is more likely to be misdiagnosed.

  Also, a particular disease, though it can be caused by different reasons, usually has similar clinical test or image findings. A good example is that hereditary angioedema(HAE) has the similar clinical picture as do other forms of angioedema.[75]
- **The clarity of disease**

  The clarity of disease is related to the clinical manifestation or presentations in individual patients. In other words, it is the clarity of disease symptoms or signs presented by patients. In particular, it may involve the early stage of a disease. Relevant study shows that diseases can be poorly diagnosed in their early stages due to the low clarity of disease features at the very early stage.

Main categories	Factors	Examples	Papers
Clinical disease features	Whether a disease is well researched	Unsuspected rare causes or unsuspected age group; Atypical disease symptoms or signs;	Paper 72 Paper 73
	Clarity of a disease	Early stage of a disease; Different diseases may have similar clinical presentation in patients, or a disease may have a high variability of clinical presentation, such as anaphylaxis.	Paper 74 Paper 75 Paper 76

Table 4.3 Key factors in terms of features of diseases

b. Educational background

The education of both patients and doctors plays an important role in the diagnostic process. Keys factors with relevant papers are described in Table 4.4.

- **Patient educational background**

  A patient's educational background concerns the relevant healthcare education that is available to the public in order to increase their awareness regarding a disease and the proper actions to be taken when feeling unwell. Research shows that a better educational background is linked with a higher awareness and better actions. Thus, public education in relation to relevant medical information, especially for epidemic diseases, is encouraged so that patients can be aware of the first signs and symptoms. It directly affects patient access to healthcare in a timely manner.

- **Doctor's knowledge and experience**

  A doctor's knowledge and experience can also be interpreted as the skills required to deliver correct diagnosis. It refers to the relevant knowledge or experience from professional medical education, training or work experience. Also, training, which is specifically designed to avoid diagnostic errors, may help deliver the knowledge or awareness of diagnostic errors, such as negative case studies. The doctor's abilities obtained from knowledge and experience include observing clinical signs, understanding collected clinical information, reasoning with clinical information and organising treatment plans.

Main categories	Factors	Examples	Papers
Educational background	Health awareness		Paper 77 Paper 78
	Doctor's experience and knowledge background	Medical knowledge and experience; Specifically designed training to avoid diagnostic errors	Paper 79 Paper 80 Paper 81

Table 4.4 Key factors in terms of educational background

c. Ambiguous or conflicting information in diagnostic clues:

Diagnostic clues are the evidential information used to make diagnostic decisions. They are affected by the quality of the original information that is collected during diagnosis. Also, they are impacted on by the clinical reasoning that is used to filter and retrieve key clues from original collected information.

The original diagnostic information covers the information that is collected from each phase of the diagnostic process, such as patient medical history, patient-doctor communications, initial physical examinations, as well as further information collected from diagnostic tests or consultation information from other healthcare providers. More specifically, ambiguous information can result from poor communication between healthcare providers and patients, which is more frequently witnessed with patients who have language difficulty or verbal commutation difficulty, such as infants or mental health patients. Ambiguous information in test results is mainly due to the errors in test management, for example missed tests, delayed responses to abnormal findings, and the interpretation of test results. The most common errors in tests are the interpretations of "faint test bands" and false-negative or false-positive test outputs.[87] In addition to collecting information, reasoning with diagnostic information is also crucial, and it indicates weighing suboptimal diagnostic information and choosing leading clues. Table 4.5 summaries the keys factors as well as relevant examples in this category.

Main categories	Factors	Examples	Papers
Collecting and retrieving/reasoning diagnostic clues	Collecting diagnostic information from patient medical history	Incomplete medical history	Paper 82 Paper 83
	Collecting diagnostic information from communications between patients and doctors	Fetal disease; Diagnostic overshadowing: misinterpreting some of the physical symptoms as symptoms of the mental illness	Paper 84 Paper 85
	Collecting diagnostic information from tests	Test data management; False-negative/positive results; Interpretation of "faint" test bands	Paper 86 Paper 87 Paper 88
	Collecting diagnostic information from referral; Retrieving and reasoning regarding diagnostic clues		Paper 89 Paper 90 Paper 91 Paper 92 Paper 93 Paper 94 Paper 95

Table 4.5 Key factors in terms of diagnostic clues

d. Psychological factors or bias

Psychological factors closely affect clinical reasoning in the process when clinicians seek to understand diagnostic information and assess the probabilities of diseases. The factors and related papers are listed in Table 4.6.

Psychological factors are mainly of two types: bias and awareness of an error. Different biases for the most part relate to psychological factors.

Biases, often called as cognitive bias or cognitive factors, refer to the cognitive functions of performance and learning, such as attention, memory, and reasoning.[96] They are found to be an important cause of diagnostic errors.[97 98] At least 40 types of bias that may affect clinical reasoning have been found[99], such as conformation bias which is about physicians'

desire to confirm a preliminary diagnosis while failing to seek contradictory evidence, history bias which is related to misleading information from patient medical history, or selection bias which is witnessed when patients are referred to other healthcare providers[100].

Main categories	Factors	Examples	Papers
Psychological factors	Bias	Over confidence Conformation bias; Selection bias; History bias	Paper 101 Paper 102 Paper 103 Paper 104 Paper 105 Paper 106 Paper 107 Paper 108
	Awareness of an error		Paper 109

Table 4.6 Psychological factors

e. Follow-up

Follow-up after a diagnosis is mainly used for discovering and corresponding errors in time so that further harms resulted from errors on patient outcomes can be prevented. Close follow-up helps clinicians receive feedback of the initial diagnosis and revise treatment plans in time when needed. Relevant papers are listed in Table 4.7.

Main categories	Factors	Papers
Follow-up	Follow-up and feedback of previous diagnosis	Paper 110 Paper 111 Paper 112 Paper 113

Table 4.7 Follow-up factors

f. Other factors:

Two other factors were found during the systematic review, but these factors were not taken as the key factors since they rarely happen or have only a minor effect.

These two factors are delayed responses to abnormal clinical findings and the work environment. It has been stated in one study in 2009[114] that delayed responses to abnormal findings may delay the diagnosis. The study also suggests that clinical data management, especially following up abnormal findings, should be improved. Another research in 2007[115] argues that the work environment of diagnostic tests, such as ambient light conditions for viewing radiological images, may affect the test results.

In summary, relevant factors of diagnostic errors can be divided into the six categories described above, and Table 4.8 summaries the categories and lists the related reference papers.

Key factors	Support papers
Clinical disease features	Paper72 to Paper76
Educational background	Paper77 to Paper81
Ambiguous or conflicting information in diagnostic clues	Paper82 to Paper95
Psychological factors	Paper101 to Paper109
Follow-up	Paper110 to Paper113
Other factors	Paper114, Paper115
Papers which cover more than one factor	Paper 20 Paper90 Paper91 Paper 116 Paper117 Paper 118

Table 4.8 Six categories of key factors and relevant papers

4.3.2 Methods for diagnostic error reduction

This subsection discusses the methods that were employed in the selected papers to help reduce diagnostic errors. Methods are divided into two large groups, "non-electronic methods" and "electronic methods", based on whether they use computer technology to solve the problem. Furthermore, they are also separated into small groups depending on the factors they are working on. Most of the methods tried to assist with one factor only.

a. Non-electronic methods

Non-electronic methods are varied in terms of subjects. "Improving education" and "improving clinical guidelines" cover a large percentage of relevant research papers. Creative methods from other high-risk professions and other recommendations are also found to solve the problem.

- **Improving education for patients and doctors:**

Education for patients mainly aims to increase public awareness. It helps shorten the time between "having symptoms" to "accessing a medical service", and thus avoiding late

diagnosis in the first step. Education for doctors is about improving doctors' study both regarding medical knowledge and clinical guidelines. Also, some researches focus on training physicians in relation to decision thinking and decision making.

- **Improving clinical guidelines:**

Several papers investigate in depth the gaps in clinical guidelines so as to uncover links with diagnostic errors. According to these papers, the following processes are suggested for improvement: investigation period service, access to the patient history records, analysing clinical data, and clinical follow-up.

- **Other methods:**

Other methods include creative ideas, improving the work environment and increasing evidence-based recommendations. In particular, creative ideas can be adopted from other high-risk, high-reliability professions, such as aviation.

Table 4.9 outlines relevant papers.

Non-electronic methods	**Support papers with outlines**
Education for patients	Papers 77: education for public aimed at increasing high awareness
Education for doctors	Paper 74: medical education and public information about leprosy's signs and symptoms; Paper 81: calling for better medical education in relation to cluster headache
Improving clinical guidelines	Paper 119: improving guidelines regarding symptoms of extradigital glomus tumour Paper 120: a treatment algorithm to avoid missing similar injuries is proposed.
	Refine clinical rules: Paper 121: improving clinical rules to identify patients who have intracerebral haemorrhage.
	Improve the investigation period service: Paper 122: better integrating services during the investigation period, before final diagnosis.

	Encourage access to patient medical history records: Paper 123: encouraging the access to patient medical history records, including access to previous mammograms, but should avoid "misleading diagnostic information from clinical history" at the same time.
	Improve how to analysis the collected data: Paper 110: a method to analyse the collected data
	Close clinical follow-up: Paper 114: improving clinical data management, focusing on following up abnormal findings
Creative ideas, which come from other application fields, especially high-risk, high-reliability professions	Paper 124: creative ideas from applications relating to airline pilots and nuclear plant operators. Paper 125: situational awareness is a model that is primarily used in aviation human factors research that can encompass both the cognitive and the systems roots of such errors.
Work environment	Paper 115: different light conditions for viewing radiological images
Increasing evidence-based recommendations	Paper 126: increasing evidence-based recommendations for clinical decisions. Paper 127: encouraging evidence-based medicine

Table 4.9 Non-electronic methods and relevant papers

b. Electronic methods

Electronic methods used in solving this problem are mainly concerned with the following two aspects:

- **Solutions focusing on laboratory tests**

These solutions focus on the mistakes or delays during laboratory tests, and are divided into two parts. First, advanced electronic laboratory equipment or intervention can help prevent errors from laboratory test results, such as false positive or false negative test results, which will directly affect the process of making a diagnosis. Secondly, electronic laboratory systems or interventions for test workflow can counteract leaks for potential workflow errors. For example, missed tests or wrongly labelled patient names on test samples can cause delays or mistakes, which can be prevented by electronic systems or interventions.

- **Solutions focusing on decision making**

These solutions implement Internet technologies or computing knowledge into clinical decision support systems or interventions that help clinicians make decisions. Specifically, they contribute to reducing diagnostic errors in three ways: predicting high-risk patient cases, providing diagnostic information for doctors, and detecting errors in time after a diagnosis.

Table 4.10 summaries the electronic methods described above and relevant papers.

Electronic methods	**Support papers with outlines**
Advanced equipment for laboratory test results	Paper 128: Advanced imaging technology
Electronic laboratory system or interventions for laboratory test workflow	Paper 129: encouraging interventions for the workflow management of tests. The workflow of tests includes: presentation and management test data, as well as insuring oppropriate follow-up of tests. Paper 130: e-Chasqui laboratory information system
	Paper 108: blinded review may reduce errors in the interpretation of skin biopsies
Clinical decision support systems or interventions in helping make a diagnosis	**Using computing knowledge to work as a reminder by identifying high risk diagnosis:** Paper 131: using "affinity set by topology concept" to find, obtain and classify key attributes. Paper 94: using “a knowledge base” from a diagnostic decision support system to identify “high-information clinical findings” of a certain disease which may be related to a high-risk diagnosis, such as colon or breast carcinomas
	Computer-assisted diagnostic database (extending clinical criteria from an expert system): Paper 132: a computer-assisted diagnosis database for headache
	Web-based search engine helping diagnosis: Paper 133: discussing *Google* in helping diagnosis

	Using "ontology" to classify diagnostic criteria: Paper 134: updating existing description logic ontology; new ontology will classify a patient's characteristics or diagnostic criteria under a particular disease.
Clinical decision support systems or system interventions in detecting diagnostic errors	**Electronic health records:** Paper 14: detecting diagnostic errors by finding information patterns in electronic health records. **System-related interventions:** Paper 135: encouraging system-related interventions **Systematic re-examination:** Paper 95: suggesting systematic re-examination of leading diagnostic clues. Paper 109: querying an initial diagnostic hypothesis may help to reduce diagnostic errors. Paper 111: correcting an initial error before it affects patients Paper 112: modifying the initial diagnosis during the follow-up of a patient is important in diagnosing epilepsy.

Table 4.10 Electronic methods and relevant papers

Overall, most methods aim to promote Internet Technology (IT) support, clinical guidelines and evidence-based medicine, and acknowledge the fact that a single method may have multiple effects on outcomes. Taking electronic health records (EHR) as an example, EHR provides an electronic source of medical evidence, which benefits to both clinicians and researchers. Clinicians can benefit from easy access to the individual patient medical history. At the same time, clinicians are able to search records horizontally, such as checking previous patient records with similar symptoms. Similarly, researchers are also able to conveniently retrieve relevant data from a large quantity of data based on clinical evidence.

Figure 4.3 further develops Figure 2.1 in terms of where and what errors may occur, and then maps the methods above into the process.

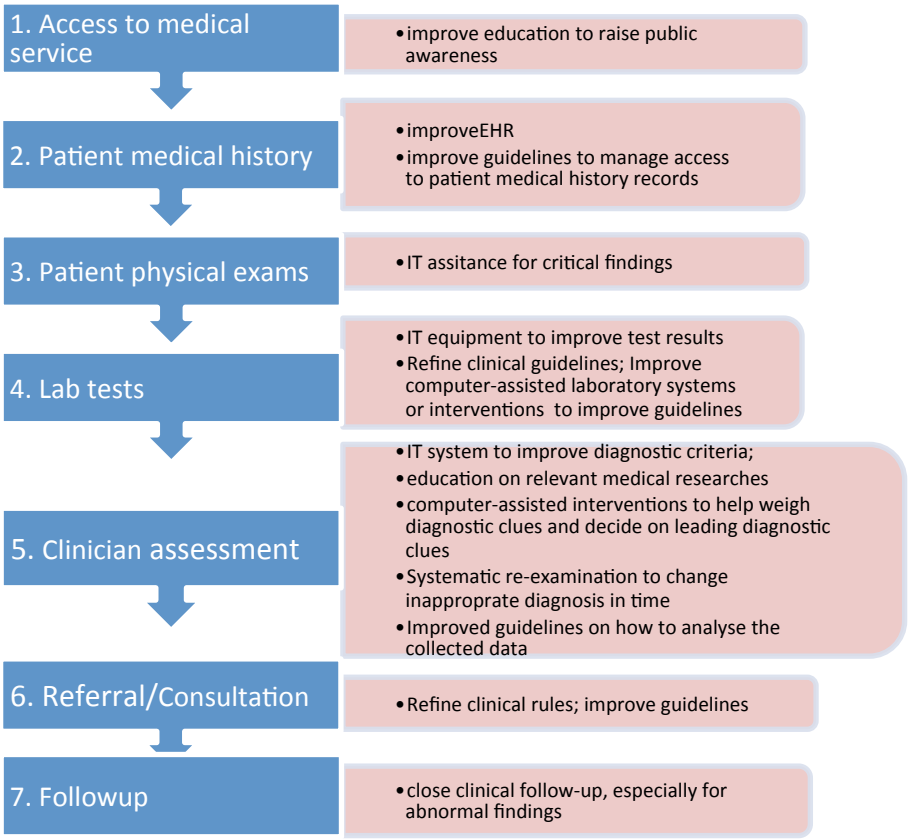


Figure 4.3 What can be done during diagnostic process

4.4 Summary

The factors that affect a diagnosis cover several different aspects, and can be divided into six main categories: clinical disease features, educational background, ambiguous or conflicting information in diagnostic clues, psychological factors or bias, follow-up, and other factors. Specifically, clinical disease features explain diagnosis from a standpoint of disease related factors, including "whether a disease is well researched" and the "clarity of a disease". The factor of educational background indicates both "public health awareness" and "doctor's experience and knowledge". In addition, diagnosis also depends on gathering sufficient correct diagnostic information and weighing suboptimal diagnostic clues. Moreover, psychological factors, such as bias, and close follow-up also have an influence on patient outcomes.

Current methods of reducing diagnostic errors are also reviewed. Depending on whether internet technology is applied, the methods can be divided into non-electronic methods and electronic methods. Non-electronic methods cover education in relation to patients and doctors, improving clinical guidelines, a proper work environment and encouraging

evidence-based medicine. Electronic methods include advanced equipment and IT interventions which help to reduce diagnostic errors in three ways: predicting high-risk patient cases, providing diagnostic information for doctors, or detecting errors in time.

However, the methods still need to be evaluated through the whole diagnostic system behaviour. Few methods have demonstrated their influence on the phases of the entire diagnostic system, as well as effect on final patient outcomes, so the results are limited and sometimes diverse. When providing an intervention or changing one factor, a series of influences on different diagnostic phases and their final effect on patient outcomes would be hard to evaluate without viewing the whole picture. Since diagnosis links several phases together and involves a range of factors, it requires a systemic structure showing the connections among phases, factors and patient cases. What is more, a systemic model can provide a quantitative simulation, which shows a series of changes of the error numbers in each phase while the values of factors are varying.

The next chapter will illustrate a systemic qualitative model to present interrelations among key factors, based on the findings from the systematic review. The qualitative model links different phases as a whole diagnostic system, and provides a systemic look at the root causes and cause-effect relationships between model variables. Both structured causal loop diagrams and hierarchical cause trees are generated in the next chapter.

Chapter 5 Qualitative System Dynamics Modelling

5.1 Introduction

To provide a systemic view and better understanding of diagnostic errors with their factors, a qualitative model of system dynamics is developed in this chapter. This qualitative system dynamics model, reflecting the findings of the systematic literature review, links model variables together, graphically presents variable interrelations, and provides clear traces of different hierarchical causes.

The chapter starts with introductions of how the qualitative system dynamics model using causal loop diagrams presents variables and relations, and explaining the meanings of presentations used in the diagram. After identifying the purpose and the boundary of the model, the details of how to develop the causal loop diagrams for diagnostic errors are described. Finally, the qualitative model in the form of a causal loop diagram is illustrated based on the systematic reviews and expert feedbacks, and the model is explained and discussed in detail. The results of the model can also be found in paper 70.

5.2 Introducing the qualitative model

A qualitative system dynamics model, also referred as to a "Causal Loop Diagram"(CLD), graphically presents a system using two components: nodes and arrows. The model variables are represented by a collection of nodes, and arrows are used to link the nodes together and to represent the qualitative relationships of the variables, particularly the cause-effect relationships.

Furthermore, because of the cause-effect relationships, variables of the system can be called as *cause variables* and *effect variables*. The "effect variable" is adjacent to the arrowhead, and the "cause variable" is at the opposite end of the arrowhead. Arrows also use a positive polarity or a negative polarity at the arrowhead to indicate that the cause has positive effect or negative effect. A positive polarity indicates the "effect variable" changes in the same direction with "cause variable"; while a negative polarity indicates the variables move in an opposite direction. [51]

Theoretically, a CLD is created following the below outline: 1, identifying the purpose of the model; 2, identifying the key variables of the model; 3, developing the reference modes, in

other words, identifying the arrows with the polarity; 4 developing the causal map of the processes.[51]

5.3 The purpose and the boundary of the qualitative model

The aim of the qualitative model is to illustrate a systemic list of factors of diagnostic errors and show all possible routes of how these factors affect diagnosis, as well as the hierarchical structure of cause-effect relationships.

A CLD model can help to understand causes of diagnostic errors and visualise the interrelations of relevant factors. It provides the routes to observe how diagnostic accuracy is affected by the change of one variable. It also helps quickly capture hypthotheses of possible solutions by analysing causal arrows and loops.

The qualitative model focuses on diagnosis in primary care only, reflects research-level factors of diagnostic errors from literature, and presents the process covering access to healthcare service, diagnostic process as well as after-diagnosis.

5.4 Developing the initial qualitative model based on systematic review

The qualitative model is constructed based on two sources: literature and clinicians. Figure 5.1 shows the steps used in the thesis to develop the qualitative model.

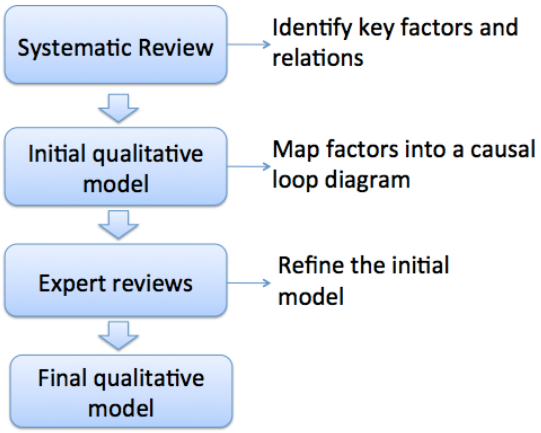


Figure 5.1 Process of developing the qualitative model

A systematic review of literature was conducted in the previous chapter. Relevant papers published between 2002 and 2012 were viewed and these papers provided the information

for model variables and relationships. According to the findings shown in the section 4.3, the following list of factors were identified: disease well researched; disease presentations; doctor's experience and knowledge background; collecting and reasoning diagnostic clues; public health awareness; patient medical history; patient-doctor communication; initial physical examinations; further diagnostic information collected from tests or consulting from other healthcare providers; bias; doctor awareness of high risk cases; close follow-up. Both causal variables and effect variables were further mapped into the patient pathway that includes access to healthcare service, diagnostic process as well as after-diagnosis, and the initial qualitative model is illustrated. The patient pathway and the initial qualitative model are shown in APPENDIX I. Following the arrows in the model, a 4-level-depth causes of diagnostic errors was displayed in APPENDIX I as well.

5.5 Amending the initial qualitative model

In this section, the initial model in APPENDIX I is further amended based on experts' feedback. The model is shown and explained to the clinicians, then clinicians' opinions and suggestions of the model are asked and discussed. The discussion covers two aspects: the suitability of variables in the model and the applicability of cause-effect interrelations. Feedbacks and suggestions from experts are collected to refine the model and increase the model acceptability and reliability. The initial model is amended in the end.

5.5.1 The process of discussions with experts

A total number of seven clinicians reviewed the model and provided feedback. Since it is an informal discussion instead of a formal study approach, all participants are clinicians who are independent clinicians and randomly selected from people with contacts from UK and Canada.

The discussion is carried out with one individual clinician at one time, and whole process mainly includes the following parts: presenting the initial model and giving an introduction of the model; explaining the model while asking relevant questions to the clinician; and collecting clinician opinions and suggestions. During the process, a step-by-step explanation of the model is given while relevant questions are asked, and the clinician can provide opinions and feedback anytime during the discussion process.

The discussions with clinicians mainly cover two aspects: the suitability of the variables which is to check whether factors included in the model were effective and reasonable, and

the applicability of cause-effect interrelations which is to check whether the arrows show correct relationships between variables. Questions, such as "Do you think this model covers all the key factors of diagnostic errors?", are asked and more details of the process can be found in APPENDIX II.

5.5.2 Feedback from experts

Clinicians provided an overall positive feedback on the initial model, and believed the model covers most of the factors that may affect errors in diagnostic decisions and arrows show clear cause-effect relationships.

At the same time, several factors are proposed to be added into the model, and relevant loops are amended. The new factors are as below:

- Continuity of care: Continuity of care indicates whether the patient sees the same doctor every time, and it will affect the quality of the outputs from the history taking and physical examinations, especially in primary care service.
- Workload for healthcare providers.[136]
- Easy access to medical service
- Patient trust to the healthcare provider: Diagnostic errors can reduce patient trust and affect patient choosing the same healthcare provider.
- Detected errors. It is the consequence of factors: "doctor awareness of an error" and "follow-up and feedback of previous diagnosis'. It contributes to a better understanding of how the two factors affect the diagnostic errors.

Relevant relationships were discussed and amended based on the initial model.

5.5.3 Final causal loop diagram for diagnostic errors

After amendment of the initial diagram, the final CLD is illustrated as Figure 5.2. Both variables and relations rely on the findings from the systematic literature review in Chapter 4 and expert reviews in Chapter 5.

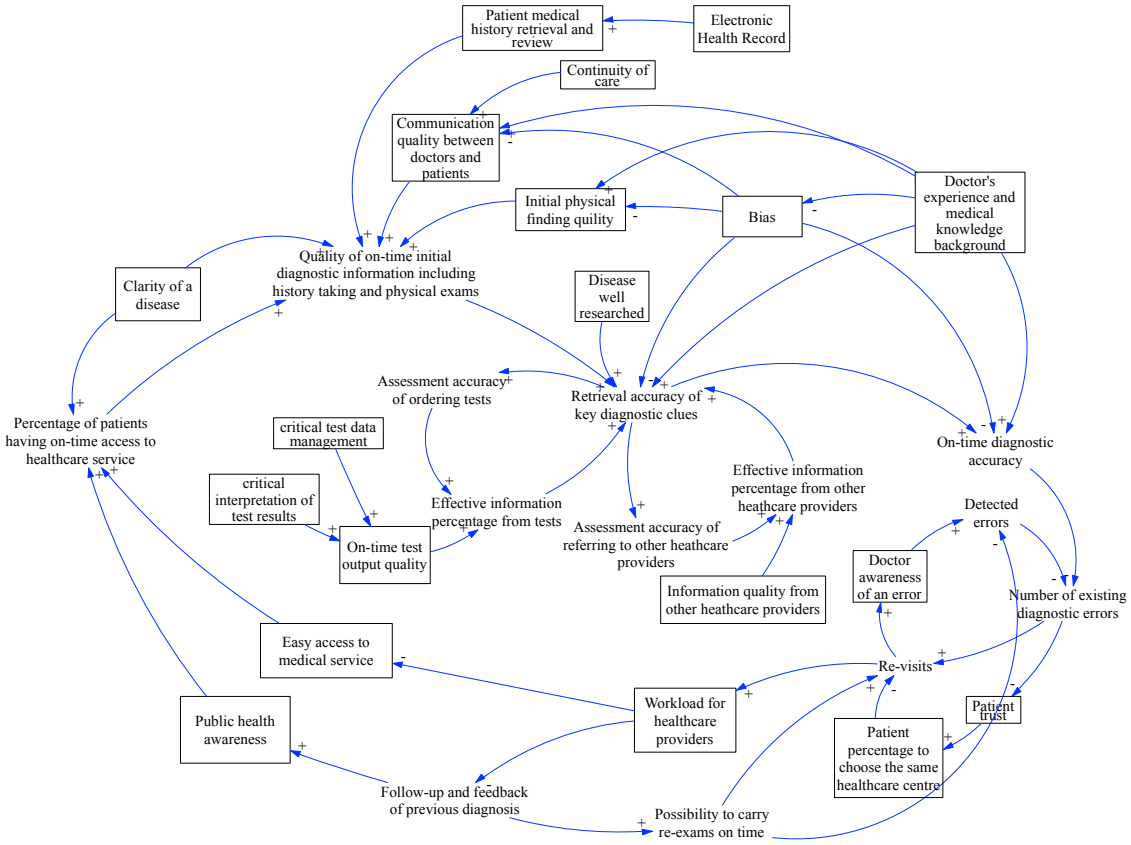


Figure 5.2 The causal loop diagram for diagnostic errors

a. Descriptions

The qualitative model illustrates a systemic look of key factors of diagnostic errors from a view of the entire healthcare system, which covers medical service access, diagnosis, and after-diagnosis.

The model shows key factors of diagnostic errors clearly and links the causal variables and effect variables together using arrows. By tracing the arrows, the model can help to discover the root causes and provide a structured and hierarchical view of the causes. Variables without boxes are internal variables in the process of diagnostic errors, and variables marked with boxes are the external variables. Following the input arrows of a variable can discover different levels of causes of the variable. The diagnostic errors are presented as the variable *"Number of existing diagnostic errors"* in the diagram. The input arrows of the variable *"Number of existing diagnostic errors"* link with two variables, and the two variables are the first-level causes that have direct causal effect on diagnostic errors. Further tracing back the input arrows of the first-level causes, the second-level causes can be found. Similarly, different levels of causes can be discovered via tracing the input arrows, and

finally form a “causes tree” that hierarchically shows the different levels of causes of diagnostic errors.

b. Causes trees

Causes trees help to summarise all causes of different levels and show the paths from the initial cause to the final effect. To get a clear view of the causal variables and effect variable of the diagnostic errors in the system, the cause trees of *“Number of existing diagnostic errors”*, showing hierarchical structure of causes of diagnostic errors, are illustrated as Figure 5.3, Figure 5.4, Figure 5.5 and Figure 5.6. More details of Figure 5.3 can be found in APPENDIX III.

The causes of the variable *“Number of existing diagnostic errors”* can be found from its input arrows. The variable has two main causal factors that are the *“On-time diagnostic accuracy*” and the “*Detected errors*”. These two variables are also affected by their input factors. The “*On-time diagnostic accuracy*” depends on *bias, doctor’s experience and medical knowledge,* and *retrieval accuracy of key diagnostic clues*. The “*Detected errors*” is influenced by *doctor awareness of an error* and *possibility to carry re-exams on time.*

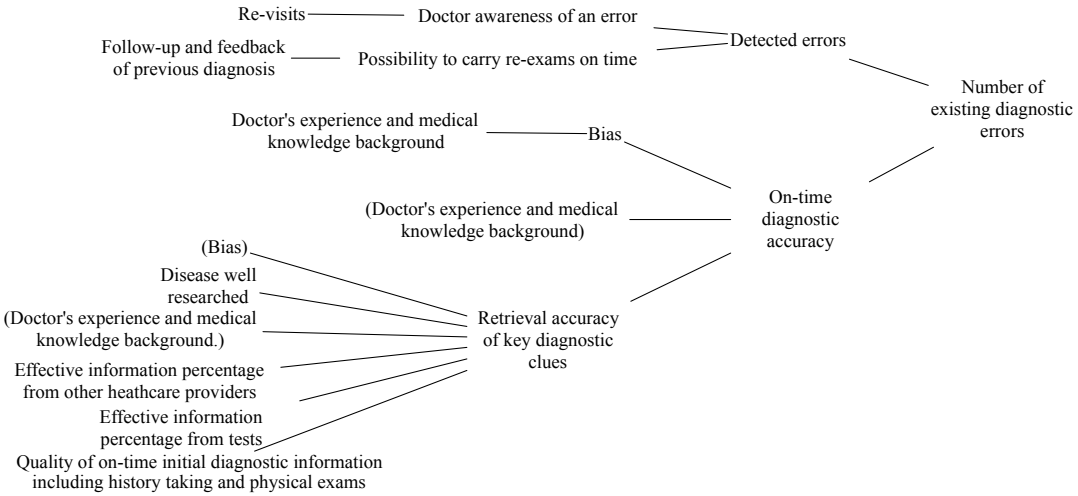


Figure 5.3 3-level of causes tree of diagnostic errors

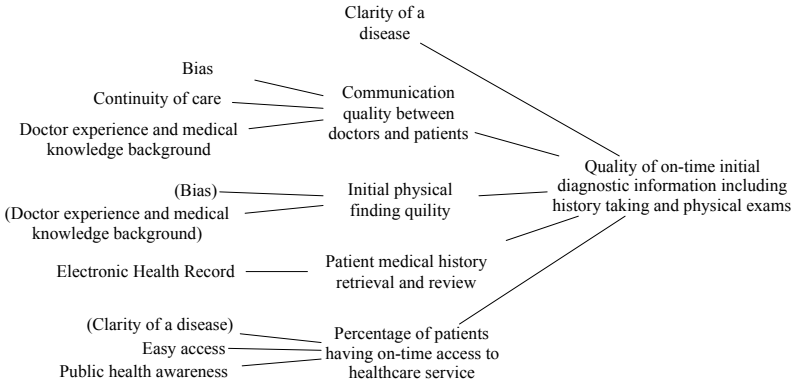


Figure 5.4 Causes tree 1 of the 3rd level factors

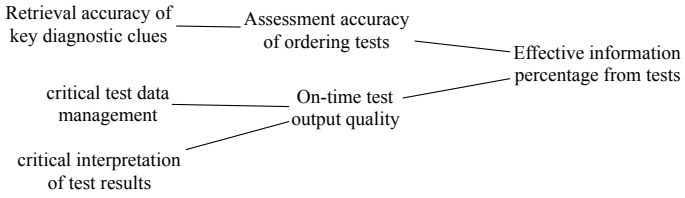


Figure 5.5 Causes tree 2 of the 3rd level factors

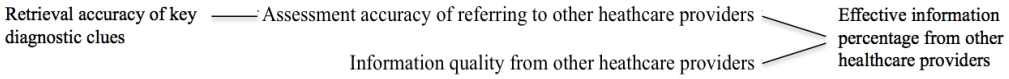


Figure 5.6 Causes tree 3 of the 3rd level factors

c. Uses tree

Similar to the causes trees, the "uses tree" of a variable shows how this variable affects other variables. Figure 5.7 describes how the diagnostic errors, "*Number of existing diagnostic errors*", works as a cause in the system.

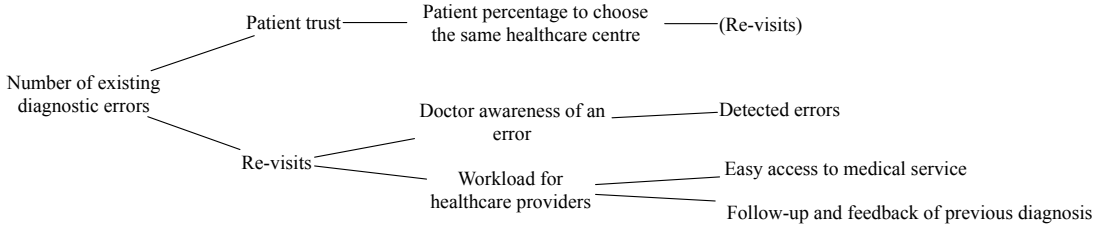


Figure 5.7 Uses tree of diagnostic errors

d. Loops

Following the arrows in the qualitative model, loops which are constructed by arrows can be found. Each loop indicates a loop of cause and effect relationships. Twelve loops that include the variable *"Number of existing diagnostic errors"* are figured out. To have a clear view of the causal loops, individual loops are retrieved out from the qualitative model, and are described in APPENDIX V.

5.6 Discussion of how to map possible strategies into the qualitative model

The qualitative model not only provides a systemic understanding of the cause-effect relationships, but also encourages the creative thinking of possible strategies. When the relevant strategies or interventions are mapped into the model and linked with relevant variables, the model can help to discover how the system can be affected.

Specifically, CLD helps to demonstrate the related variables that can be influenced, and explain the routes of how the intervention will directly or indirectly affect diagnostic errors via arrows. Figure 5.8 takes an electronic follow-up system or other close follow-up intervention as an example. After adding the intervention shown in the red box in the figure, red arrows can explain the routes of how the intervention affects the factor of *"Number of existing diagnostic errors".* At the same time, the model variables that are linked by the red arrows are the affected variables.

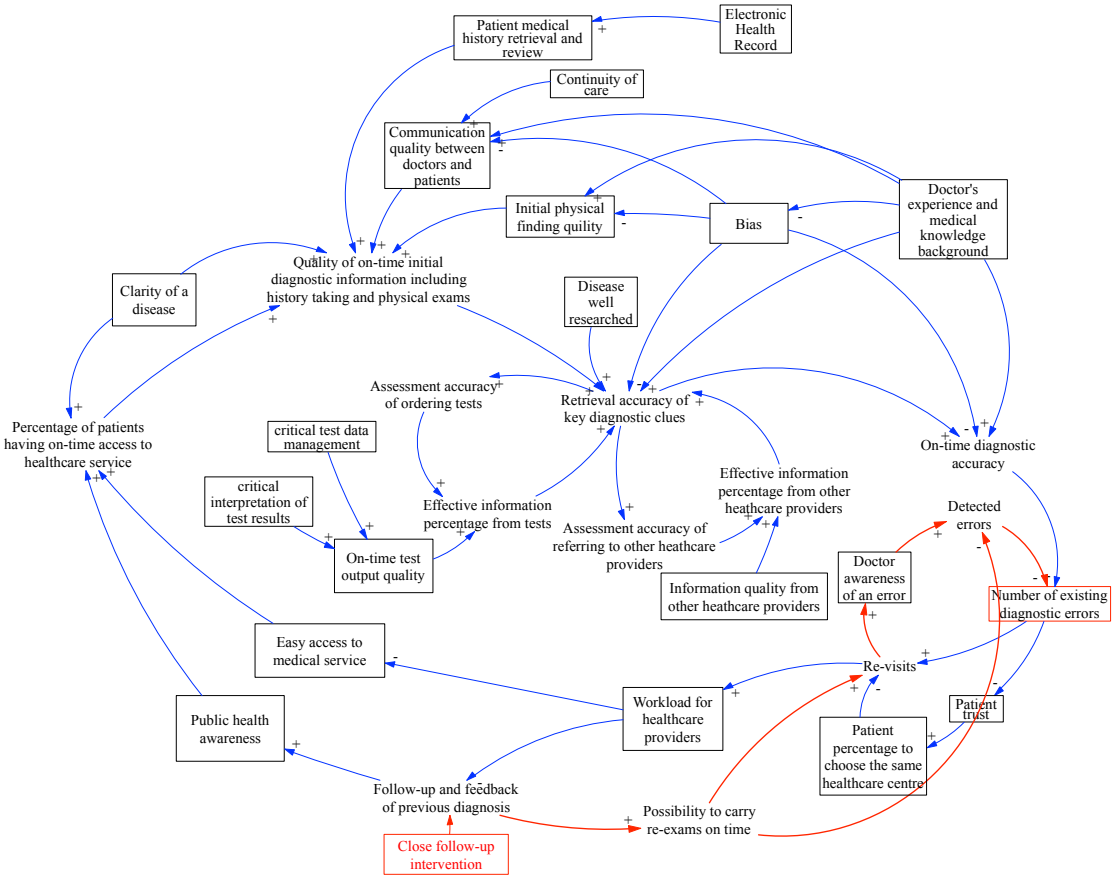


Figure 5.8 CLD of diagnostic errors adding an intervention

5.7 Summary

The qualitative model is constructed in this chapter. Based on the finding of the systematic review, an initial qualitative model is illustrated at first, and it is further revised according to the suggestions from seven clinicians. The qualitative model, using a causal loop diagram, links all relevant factors of diagnostic error together as a system, and arrows indicate interrelations among model variables. Using this model, the hierarchical structure of causes of diagnostic errors is illustrated. Furthermore, the qualitative model is also able to reflect possible strategies or interventions into the system by analysing the relevant loops, which will help to encourage possible solution thinking.

However, the qualitative model does not present the data flows of the system, and cannot reflect the quantitative changes of the system when changing model variables or mapping possible strategies into the system. In other words, it can only demonstrate which are the affected variables and routes, and it does not show how much the affected variables can change. Thus, a quantitative model is required to be developed to present the data flow of the system, then make up for the disadvantage of the qualitative model for deeper

understanding the system behaviours. Next chapter explains how to develop the quantitative model using system dynamics modelling, and then develops the quantitative model structure.

Chapter 6 Quantitative System Dynamics Model Development

6.1 Introduction

The qualitative model, the CLD, helps to understand the interrelations of the variables of the diagnostic system. By observing these factors and loops, it encourages thinking regarding possible strategies. However, causal loop diagrams cannot show quantitative changes of the variables of the system. To observe quantitative system behaviours and have a deeper look into system response to the changes of variables, a quantitative model is required to be developed. In order to present the error cases, patient cases and patient outcomes, which are all not included in the CLD, the quantitative model structure is designed in this chapter.

This chapter starts with the introduction of how to use "stock and flow diagrams" to quantitatively present variables in the diagnostic process, and lists the steps of developing a quantitative model. Then, it discusses each phase of the diagnostic process separately and constructs model structures for each phase. Relevant factors and interrelations from the CLD are mapped into the model structures. Variables in the model structures are explained in each section. In the end, the structure of the quantitative model is developed, by combining all phases together. The quantitative model structure result has been published in paper 137.

6.2 How quantitative system dynamics works in terms of diagnostic errors

In this section, how to use system dynamics to present diagnostic errors is explained in detail. System dynamics generally uses "stocks" to represent the quantity of a variable and uses "inflows"/"outflows" to represent the flows of quantities into/out of the stock at every time unit, which is referred to as "a stock and flow diagram" in system dynamics.

The flow chart below shows the basic theory of how system dynamics quantitatively presents diagnostic errors.

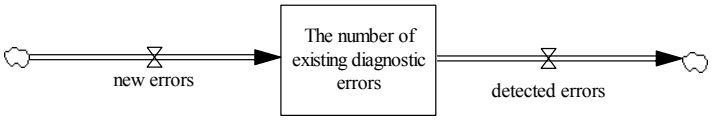


Figure 6.1 Interpreting diagnostic errors in a system dynamics model

The stock, namely "the number of existing diagnostic errors", in Figure 6.1 denotes the quantity of diagnostic errors or the number of diagnostic errors in the system. Also, the stock presents the accumulated results of its inflows and outflows over time.

The input arrow "new errors" means the newly happening diagnostic errors per time unit. It is the inflow that increases the level of the stock. The output arrow "detected errors" means the diagnostic errors being detected after diagnosis per time unit, and it is the outflow which reduces the level of stock.

The level of "*the number of existing diagnostic errors*" from time t_0 to time t can be explained using an integral equation in Eq. (6.1) and a differential equation in Eq. (6.2).

$$\begin{aligned} &the\ number\ of\ existing\ diagnostic\ errors(t) \\ &\quad = \int_{t_0}^{t} [new\ errors(s) - detected\ \ errors(s)]ds \\ &\quad + the\ number\ of\ exisiting\ diagnostic\ errors(t_0) \end{aligned} \tag{6.1}$$

where s represents any time between the initial time t_0 and the current time t.

$$\begin{aligned} &\frac{d(number\ of\ existing\ diagnostic\ errors)}{dt} \\ &\quad = Inflow{:}\ new\ errors(t) - Outflow{:}\ detected\ \ errors(t) \end{aligned} \tag{6.2}$$

Generally speaking, stocks represent the number of patient cases. Flow arrows are related to relevant actions of processing patient cases during the diagnostic process.

Considering the whole picture of the diagnostic system, the number of diagnostic errors could be taken as a stock, which remained at a certain level in the current clinical system. The factors with positive links with diagnostic errors, which were found in the CLDs, worked like input flows to the stock, and they would increase the level inside the stock, which means the number of the diagnostic errors would increase because of the input flows. Meanwhile, the negative factors for the diagnostic errors could be linked to the output flows or discharges. They helped to decrease the level of the stock, which means they helped to decrease the number of diagnostic errors. If it could be shown that the stock level decreases after an external intervention or a strategy is implemented, that would indicate an intervention or strategy which could reduce diagnostic errors.

6.3 Steps

The steps of creating a system dynamic model are as follows:

1. Identify the purpose of the model:

This quantitative model aims to quantitatively model the relations among the diagnostic errors, diagnostic process errors and relevant factors.

2. Identify boundaries and variables:

The quantitative model focuses on the diagnostic process and patient outcomes in the primary care service only. Compared to the previous qualitative model, the quantitative model only concentrates on the clinical diagnostic process in healthcare service. Thus, factors outside the system such as *public health awareness* are not included.
Besides the factors outside the system, the factor "*bias*" is not discussed in the quantitative model. This factor refers to many types of *bias* and each type requires to be studied individually. Therefore, it is not included in the quantitative model due to the time constraint.
Overall, the model is limited to mapping the diagnostic process in general practice as a system and it shows the relevant internal factors, errors as well as patient outcomes. Factors and their relations in the quantitative model are developed from the previous qualitative model, and the quantitative model also combines the phases of the diagnostic process in order to show the error flows at each stage of the diagnosis.

3. Develop the structure of the quantitative model:

This step includes mapping the CLD into stock-flow diagrams and further constructing model structure. Because the previous CLD does not present the diagnostic process errors, patient case flows and patient outcomes, simply transferring the CLD into stock-flow diagrams is not enough. Thus, the quantitative model structure is required to be designed so that the model can describe patient outcomes, all possible error flows and patient flows during the diagnostic process.

4. Identify equations of the stock and flow diagrams.
5. Run the model simulations and observe the behaviours of the diagnostic system.

6.4 Develop the structure of the quantitative model

The quantitative model follows the diagnostic process as a frame. The diagnostic process, diagnostic reasoning and diagnostic sensitivity are introduced in this section to better understand diagnosis and the model.

6.4.1 Model frame

When a patient is seeing a doctor, the doctor will take the following actions to conduct a diagnostic decision.[136] It starts with "taking a history", which refers to reviewing patient history, communicating with the patient, performing physical examinations and taking notes. All information is gathered as diagnostic information. Then more clinical examinations and tests may be ordered to gather additional information.

At some stage the clinician develops a list of some diagnostic possibilities, which are called the *differential diagnosis*. Then the clinician gathers more information to refine the list, maybe via asking more questions or undertaking more examinations. Finally, he or she arrives at a provisional diagnosis.

Generally, when clinicians face each new clinical situation in real life, they match the new situation to the past experience and act in accordance with what to do in that situation from past training and experience. Figure 6.2 uses a recognition-primed decision (RPD) model, illustrated by Klein in 1999 [138], to explain how expert clinicians make decisions in medicine.

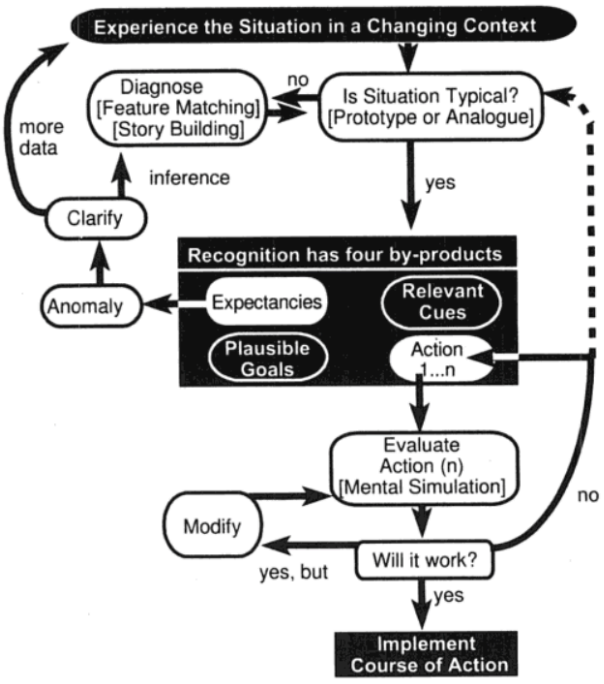


Figure 6.2 Recognition-primed decision model[138]

It is shown in Figure 6.3 that the whole diagnostic process can be divided into: pre-decision making, decision making and after-decision making. Pre-decision making is the only period when new errors in diagnostic clues happen. It is further divided into three phases for an insight view of where new errors may happen during this period, which are *"phase1 history taking and physical examination", "phase2 tests"* and *"phase3 referring to other healthcare providers".*

When a patient visits a clinician, the first and essential phase for the clinician to initiate is the phase1 history and physical examinations. It includes taking a full history of patient and conducting physical examinations if necessary. The second phase is assessment tests which cover all relevant diagnostic tests, both biochemical and radiological, to confirm or establish the diagnosis[139], for example, a laboratory test or a computed tomography (CT) scan. The third phase is about referring the patient to other healthcare providers, which is to hand over the patient's care to other healthcare providers.

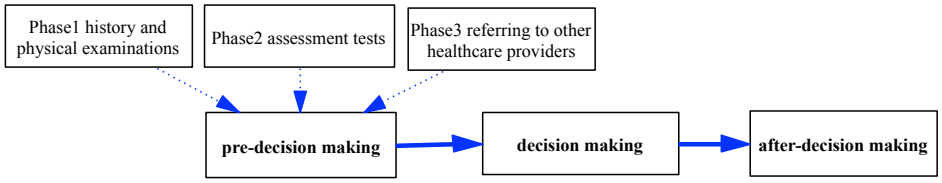


Figure 6.3 Diagnostic phases

However, all of these three phases may or may not be conducted to make a diagnosis during a patient visit. Figure 6.4 shows a total of four possible routes of patient flows during the diagnostic system, which is used as the frame for the quantitative model structure. The same numbered arrows with the same colour indicates the four different routes for patient flows. The route marked with number 1 means that the patient case begin from phase 1 and then directly goes to decision-making phase "*Treatment plan; patient discharged*". The route marked with number 2 represents that the patient case starts from phase 1 and then experience phase 2 before finally reaching the decision-making phase. Similarly, the route marked with number 3 show the patient flow that passes phase 1, phase 2 and phase 3 before arriving at the decision-making phase. Also, the route marked with number 4 implies that the patient flow goes to phase 3 directly after experiencing phase 1.

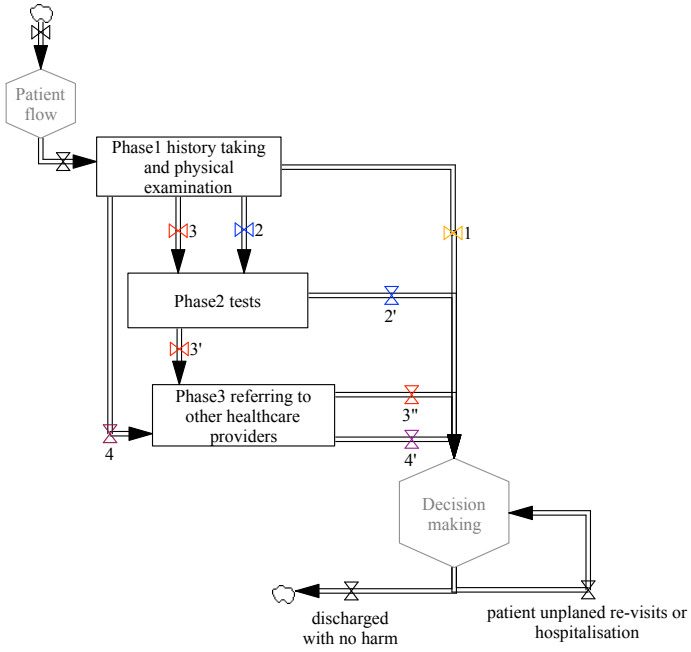


Figure 6.4 The frame of the quantitative model structure

6.4.2 How diagnostic reasoning works

Clinical diagnostic reasoning, sometimes referred to as clinical cognitive modelling, helps to understand how clinicians arrive at a diagnosis. They mainly have three types[136]: pattern recognition or feature matching, hypothetic deductive reasoning, which is either to confirm or to exclude the hypothesised diagnosis, information gathering which is used when information is not sufficient to generate a hypothesis.

In general, information gathering run through the entire pre-diagnosis period, and diagnostic reasoning is being conducted while the clinician is gathering the diagnostic information in every step. During each step, diagnostic information is collected and it is used to support pattern recognition, and then one or more hypothesis is generated when the information is sufficient. Furthermore, more information is needed if deductive reasoning is required in order to confirm or exclude the hypothesised diagnosis.

6.4.3 Diagnostic sensitivity

Sensitivity is commonly applied to the measurement of the results of diagnostic assessments, particularly including *laboratory tests* and *history taking and physical examinations*. It can directly affect diagnostic performance. Although more researches about laboratory test sensitivity are observed, rather than the sensitivity of history taking and physical examinations, both types of assessment have the same methods of representing the sensitivity and assessment performance.

Diagnostic results use “positive” or “negative” to interpret the diseased or non-diseased subjects. However, test results sometimes do not represent the true disease state. A positive result does not always represent a true positive for a diseased subject, and a negative result sometimes does not indicate a non-diseased subject. The relations of test results and disease state are listed in Table 6.1. Sometimes, the disease subject can still be shown as “negative”, which is called as “false negative”. It is the same with the non-disease subjects shown as “positive”, which is called “false positive”.

	Disease State Present	Disease State Absent
Test Outcome Positive	True Positives (TP)	False Positives (FP)
Test Outcome Negative	False Negatives (FN)	True Negatives (TN)

Table 6.1 Relations of test outcome and disease state

False positive or false negative are associated with in most tests. This can be explained in Figure 6.5 adopted from Shapiro in 1999 [140], where X indicates diseased and non-diseased subjects. The vertical line at X= γ indicates the decision limit for a positive test. The shaded area to the right of γ is the False Positive Rate (FPR), and the shaded area to the left of γ is the False Negative Rate (FNR).[140]

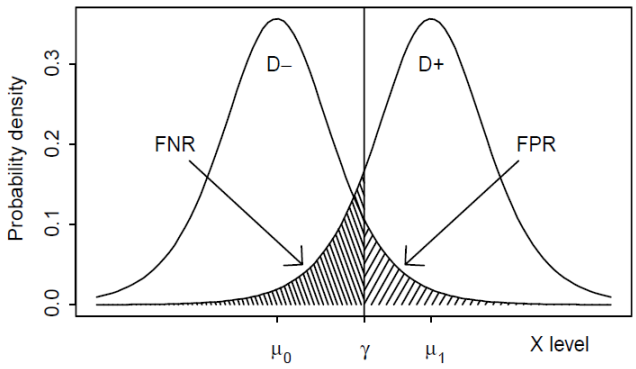


Figure 6.5 Hypothetical distributions of diagnostic test results[140]

Thus, diagnostic performance or diagnostic accuracy is usually analysed using the following terms: prevalence, sensitivity, specificity, efficiency, and predictive value. They are summarised by Wians in 2009 [141], shown as below:

Prevalence (p) = No. of individuals with disease/No. of individuals in population to be tested

Sensitivity = percentage of individuals with disease who have a positive test result = TP/(TP + FN)

Specificity = percentage of individuals without disease who have a negative test result = TN/(TN + FP)

Efficiency =percentage of individuals correctly classified by test results as being either positive or negative for the disease = (TP + TN)/(TP + FP + FN + TN)

Positive Predictive Value (PPV) = percentage of individuals with a positive test result who truly have the disease = TP/(TP + FP), or PPV = (sensitivity)(p)/[(sensitivity)(p) + (1 - specificity)(1 – p)

Negative Predictive Value (NPV) = percentage of individuals with a negative test result who do not have the disease = TN/(TN + FN), or NPV =(specificity)(1 - p)/[(specificity)(1 - p) + (1 - sensitivity)(p)]

In general, assessment tests have requirements for both *sensitivity* and *specificity*, thus, receiver- (or relative-) operator characteristic (ROC) curves are commonly used as a useful graphical tool in assessing diagnostic test or examination accuracy, shown in Figure 6.6 [140]. The ROC curve illustrates FPR (False Positive Rate) and TPR (True Positive Rate) in pairs, where actually FPR (False Positive Rate) =1 – specificity; TPR (True Positive Rate) = sensitivity.

However, diagnostic errors only concern cases with disease instead of cases without disease. The percentage of cases with disease who have a negative assessment result can be represented as FN/(TP+FN) = 1- sensitivity. Thus, *sensitivity* is selected to represent the performance of diagnostic assessments.

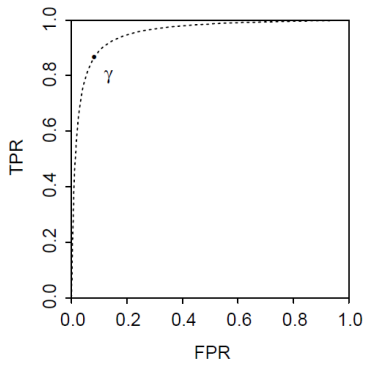


Figure 6.6 Receiver-Operator Characteristic (ROC) Curve[140]

6.5 Model for Phase1 history taking and physical examination

According to the stage where initial errors happen during pre-decision making, relevant errors are divided into three groups in this thesis, and separately discussed in three similar models for a clear explanation.

Phase 1 is an essential phase, which includes history taking and physical examination. Initial diagnostic hypotheses may be obtained after this phase. History taking covers *the history of the presenting complaint* and *review of systems*. It starts with *the history of the presenting complaint*, which is about patient presenting complains. The next step is called *the review of systems*. During this step, the clinician asks the patient questions which may or may not be related to the presenting problem, although the final aim of this step is to collect more information to find out what may be related to the patient problem.[136] A relevant study[142] shows the history taking contributes more to diagnosis than physical examinations and laboratory tests. Physical examination follows a standard order: *inspect, palpate, percuss* and *auscultate*. It has the same aim, which is to gather more information to make a diagnosis of patient's problem. At the end of this phase, the clinician arrives at a provisional diagnosis or a hypothesis list, or it has yet to be generated.

The errors occurring in this phase are referred to as the "*missed or wrong diagnostic clues*" that are information of patient symptoms and signs collected during history taking and physical examinations. The missed or wrong diagnostic clues can directly impact on the hypotheses that are named as "*missed or wrongly hypothesised diagnosis*". In particular, "*missed or wrongly hypothesised diagnosis*" indicates that the patient's health condition is not the doctor's hypotheses as a result of either not yet generating a diagnosis or wrong diagnosis.

According to the qualitative CLD model results shown by Figure 5.2 in Chapter 5, the relevant key factors that contribute to the *missed or wrong diagnostic clues* and *missed or wrongly hypothesised diagnosis* are illustrated in Figure 6.7.

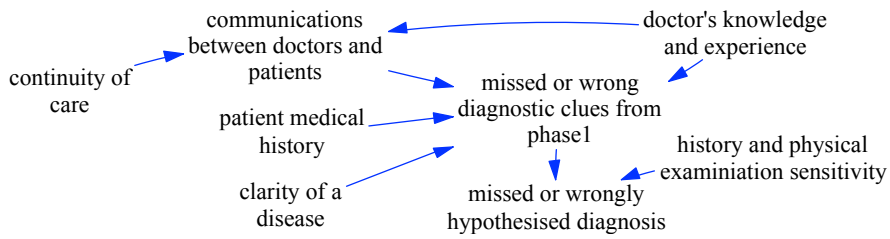


Figure 6.7 Factors and errors in diagnostic hypotheses

Continuity of care means seeing the same individual doctor which allows the doctor to know the patient, their environment and their family which contributes to the better understanding of health priorities and values[136]. *Communications between doctors and patients* should be treated as a factor. It affects data collection from patients[143]. It is suggested that it should be a two-way understanding and patients should feel that their concerns have been addressed[136]. At the same time, records of *patient medical history* are also important ways of providing patient information to different healthcare providers.

Factors, including *communications between doctors and patients, patient medical history* and *clarity of a disease,* contribute to the dependent variable "*missed or wrong diagnostic clues or diagnostic information*". This dependent variable means the diagnostic information errors and indicates that symptoms or signs are not observed or found during phase1. After phase 1, clinicians arrive at a diagnosis or a hypothesis list, or clinicians are still not yet able to generate a hypothesis. The errors of the diagnostic outputs are named as the *"missed or wrongly hypothesised diagnosis".*

To describe the error flows of the cases *"missed or wrongly hypothesised diagnosis"*, the stock and flow diagram is further developed as shown in Figure 6.8. The initial input arrow is the patient cases accessing the healthcare service, and after the phase1, patient cases become two types: cases with phase1 errors and cases without phase1 errors. Cases without phase1 errors proceed to next step via three types of outflows: cases proceeding to decision making directly, cases proceeding to phase2(tests), and cases proceeding to phase3(referrals). In terms of cases with phase1 errors, cases also have the same three types of outflows. However, it is believed that all cases with error can potentially be detected in the next phases before they proceed into decision-making. Thus, the model, instead of directly illustrating the three types of outflows for the error cases, shows that all of the error cases potentially go through phase2 and phase3. In particular, those error cases, discharged directly without being sent to phase2 or phase3, are considered to fail to order tests or referrals. The system reflects this type of outflows by showing that the error detection rate in phase2 or phase3 is zero. In the end, the missed or wrong hypotheses are shown as three types of outflows: errors detected in phase2, errors detected in phase3, and errors not detected which covers cases neither experiencing or failing to be detected in the phase2 or/and phase3. Moreover, percentages shown as the variables with brackets in Figure 6.8 indicate the administration's rates of processing the patient cases from one step to the next step.

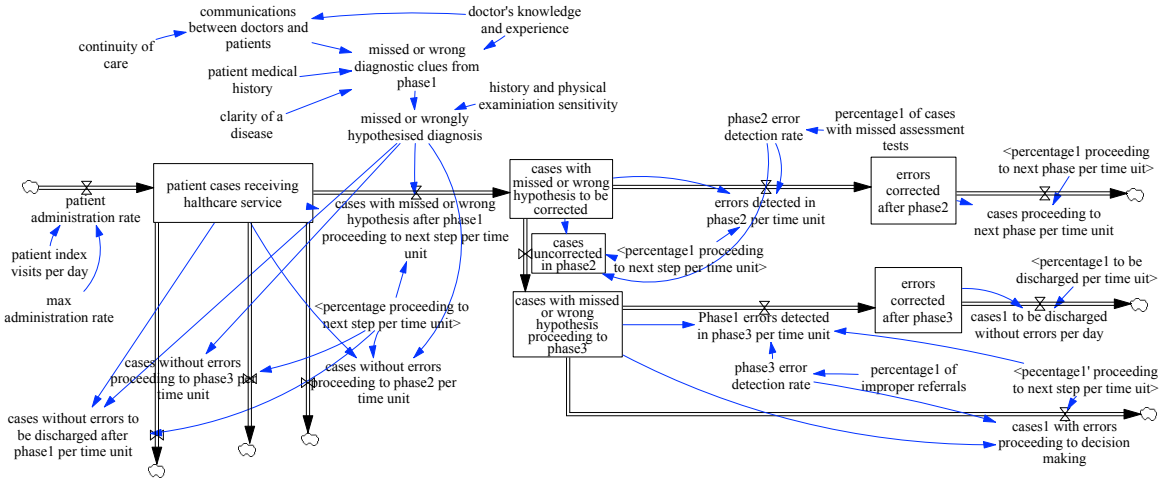


Figure 6.8 Model for errors from phase1 history taking and physical examination

In summary, the phase1 error model transfers relevant key factors from the previous CLD into the stock and flow diagram. Furthermore, it shows the case flows going through phase 1, and then the possible flows for cases with errors after phase1 and cases without errors are illustrated separately. Since error cases have the potential possibility of being corrected in the next phases - phase 2 or phase 3, the routes of the model lead all error cases to phase 2 then phase 3. Still, the model can also present error cases that do not experience phase2 or/and phase3, by setting the corresponding error detection rate to zero. These error cases are covered by the "*cases uncorrected in phase2*" and "*case1 with errors proceeding to decision making*".

6.6 Model for Phase2 tests

"Phase2 tests" includes both biochemical and radiological diagnostic tests. Errors from phase2, the biochemical test errors and radiological test errors, are usually considered as a particular type of medical errors, and are always discussed separately from diagnostics errors. However, they have a close relation with diagnosis and have a considerable impact on diagnostic decisions. Thus, in order to understand how diagnostic errors happen, it is necessary to study its relations with biochemical and radiological tests.

Errors from phase2, the biochemical test errors and radiological test errors, usually lead to "failure to diagnose". Errors in biochemical tests are often classified as: pre-analytical errors, analytical errors, and post-analytical errors.[172] Pre-analytical errors make up a large percentage of laboratory errors, and mainly including the errors happen in pre-test management, such as tube filling errors or patient ID errors. Analytical errors are the errors

randomly caused by instruments or by analytical inaccuracy. Post-analytical errors refer to the errors from the post-test management, such as result from communication breakdown. Errors in radiology are generally classified as: observer errors, errors in interpretation, failure to suggest the next appropriate procedure and failure to communicate in a timely and clinically appropriate manner. [144] [145]

Therefore, according to the factors in the previous CLD model, the factors during the period of biochemical tests and radiological tests can be summarised as ordering tests, test process management and test result interpretation. Ordering tests is an action where failure in ordering proper tests may happen. At least 10% of all diagnoses are not considered final until clinical laboratory testing is complete.[146] At the same time, the clinician's confidence in their diagnosis is not reflected in their use of tests.[142] In other words, the ordering of tests by the clinician will not be affected by whether the clinician feels confident about their diagnosis hypothesis or not. Test process management is another factor that affects diagnosis. It covers the management of test data and test workflow, such as post-analytical errors in the laboratory tests. Test result interpretations directly affect test outcomes and then impact on the diagnostic decision-making.

Figure 6.9 illustrates the stock and diagram model showing factors and cases in phase2.

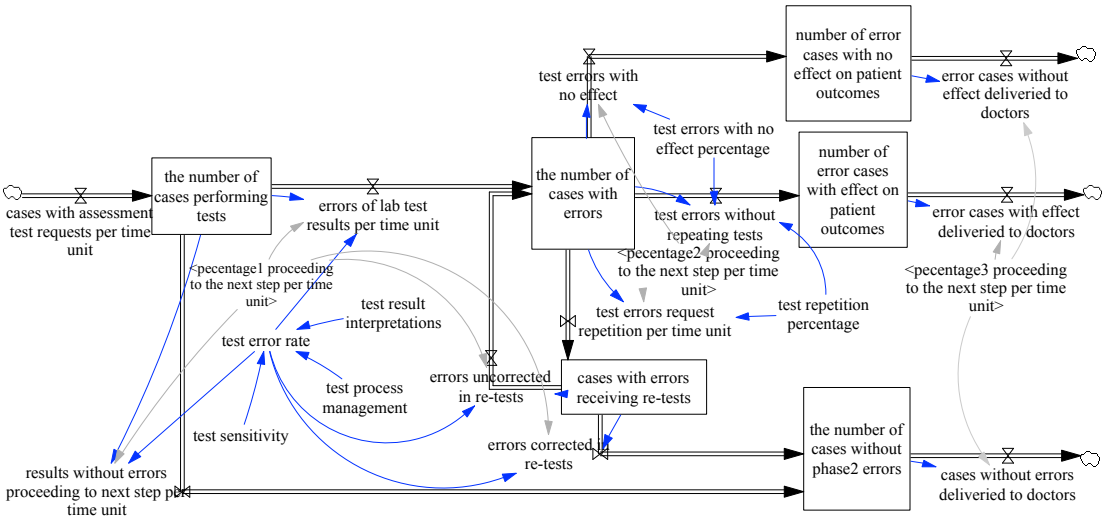


Figure 6.9. Model for errors from Phase2 tests

The initial input arrow of phase2 is the *cases with assessment test requests per time unit*, which is the number of cases ordering tests per time unit. Test error rate represents the factors during the test process: *test result interpretations, test process management* and *test sensitivity*. The phase2 model further reflects the repetition of tests, which usually

happens when clinicians are aware that errors may occur in the test results. Input cases are delivered out of phse2 and proceed into the next step, via three types of outflows: cases with errors having no effect on decision making (or no effect on patient outcomes), cases with errors having effect on decision making, and cases without errors.

In summary, the error model for phase2 shows the case flows and maps the factors related to diagnostic tests from the CLD to the stock and flow diagram. Still, the model can be further developed to discuss the deeper level of factors during this phase, and can be used to describe laboratory tests and radiology tests separately for a more specific analysis.

6.7 Model for Phase3 referrals

Phase3 is the stage of referring patients to specialists or other healthcare providers, and it is the last possible phase before decision-making. The errors of this phase mainly come from improper referrals and decision-making errors from other healthcare providers. However, improper referrals can be found and corrected when patients are delivered to the specialists. Similarly as the other two phases, phase 3 has two routes of patient case flows, which represent cases with errors from phase3 and cases without errors. Figure 6.10 shows the model of errors from phase3, and displays the case flows as well as factors.

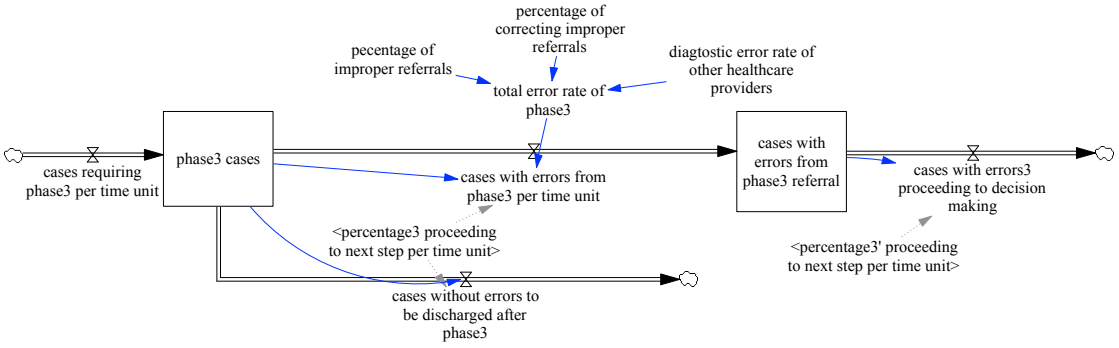


Figure 6.10 Model for errors from phase3 referrals

6.8 Model for error effects

The model for error effects displays the after-diagnosis phase including the follow-up period. It aims to explain the effects of errors on the final patient outcomes. It illustrates the possible routes of error cases and presents different patient outcomes at the end of the routes.

Combining the outflows of the three phases shown in the last three sections, there are in total three patient case flows from the three phases coming into this error-effect model:

- Case flow from phase 1, which are the cases with missed or wrong hypothesis from the initial diagnosis. In other words, the cases are the diagnostic error cases from the diagnosis of phase 1.
- Case flow from phase 2, which are the phase2 error cases with effect on decision-making. During the period of phase 2, error cases are divided into two flows, representing *errors with effect on decision-making* and *errors with no effect on decision-making* respectively. In particular, only error cases with effect on decision-making are represented in this error-effect model.
- Case flow from phase 3, which are the error cases from clinical referral stage.

The three case flows are the input flows and these proceed to clinical decisions such as treatment plans. As we can see, all of the three types of errors indicates errors in decision-making, so the sum of these three types of errors forms the diagnostic errors. However, in order to represent the current number of diagnostic errors in the system, a stock should be used to represent the current diagnostic errors in the system and the outflows have to be identified.

Figure 6.11 shows the stock and flow diagram for error effects. The three case flows are represented as a single input arrow named as "*cases with errors proceeding to decision making*" in the figure. The input arrow goes into the stock "*cases with decision-making errors*" that represents current diagnostic errors. Three output arrows emerge from the error case stock. Among these, two arrows indicate the negative patient outcomes, and they are cases with unplanned hospitalisation and cases with unscheduled visits. The third output arrow indicates the control cases. The control cases are either recovering cases or the cases with follow-up, which are the remaining cases without unplanned hospitalisation or unscheduled visits. Cases with error can take scheduled revisits suggested by follow-up service. During the revisiting period, the error detection rate is related to whether clinicians get discordant data with the initial visit that is also named as the index visit. Patients with no revisit in the control patient group are taken as the recovering patients.

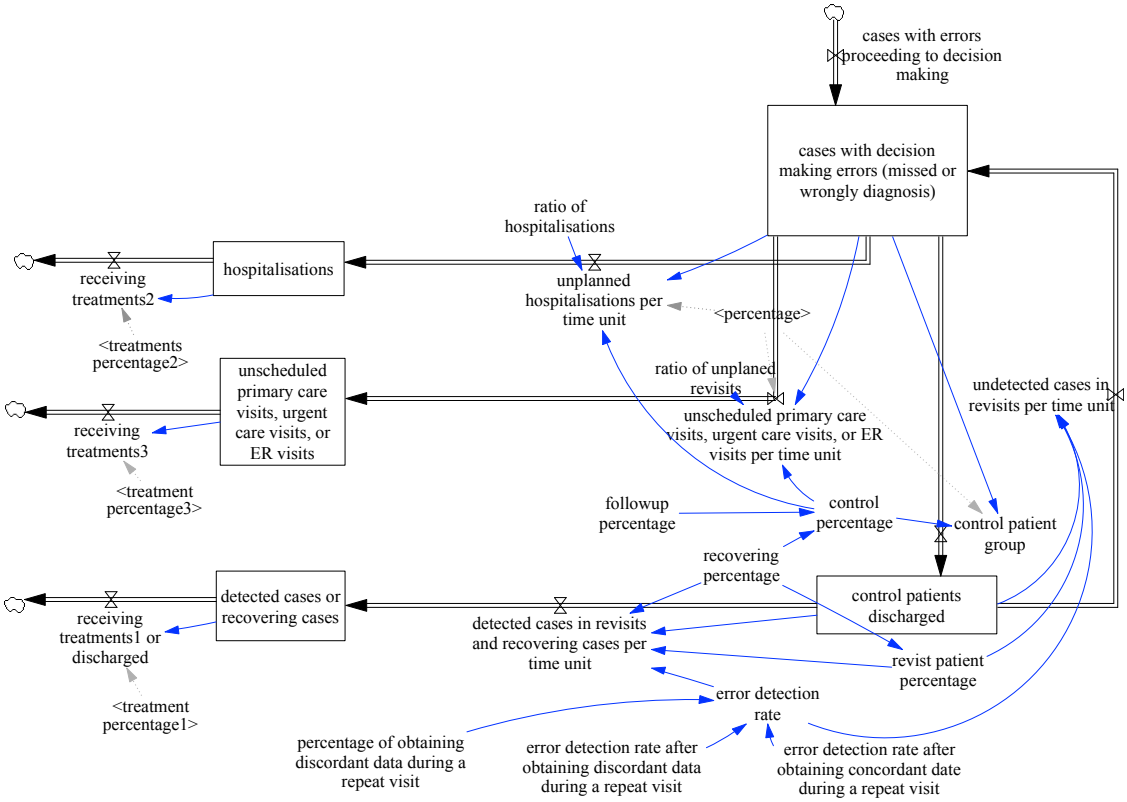


Figure 6.11 Model for error effects

6.9 Quantitative model structure of diagnostic errors in the diagnostic process

By combining the above four phase models together, the quantitative model structure of diagnostic errors in the entire diagnostic process is illustrated as Figure 6.12. The model maps the factors and links all model variables together. Thus, it consists of all possible patient case flows in the system, and analyses two types of cases separately: the cases without errors and the cases with errors. The model demonstrates possible patient paths for patient cases, and divides the error cases to analyse the possible paths and the possibilities of correcting errors at each phase.

Overall, compared with the qualitative CLD model, the quantitative model further modifies the relation loops in the CLD according to diagnostic process, so that the model can keep the consistency of quantitative variables and represent the number of patient cases either with errors or without errors. It also maps the key factors from the CLD, modifies the factors in order to fit quantitative representation, and adds more variables to the process of representing the case flows.

Still, a few factors from the CLD are not included in the quantitative model, which are *"whether the disease is well research", "bias", "workload", "easy access to medical service"* and *"public health awareness".* Specifically, *"whether the disease is well research*" is highly related to the clinician's subjective recognition of another factor "*the clarity of disease symptoms/signs*". At the same time, different diseases show distinct values in term of this factor, and the model does not aim for a specific medical disease area. Bias is also diverse and different in different situations, and a significant number of types of bias can is associated with the level of the factor of "*doctor's knowledge and experience*". Workloads require relevant data and need to further expand the current model. "*Easy access to medical service"* and *"public health awareness"* are factors related to the stage before the patient accesses to the healthcare service, and they are not included because of where the model boundary is located.

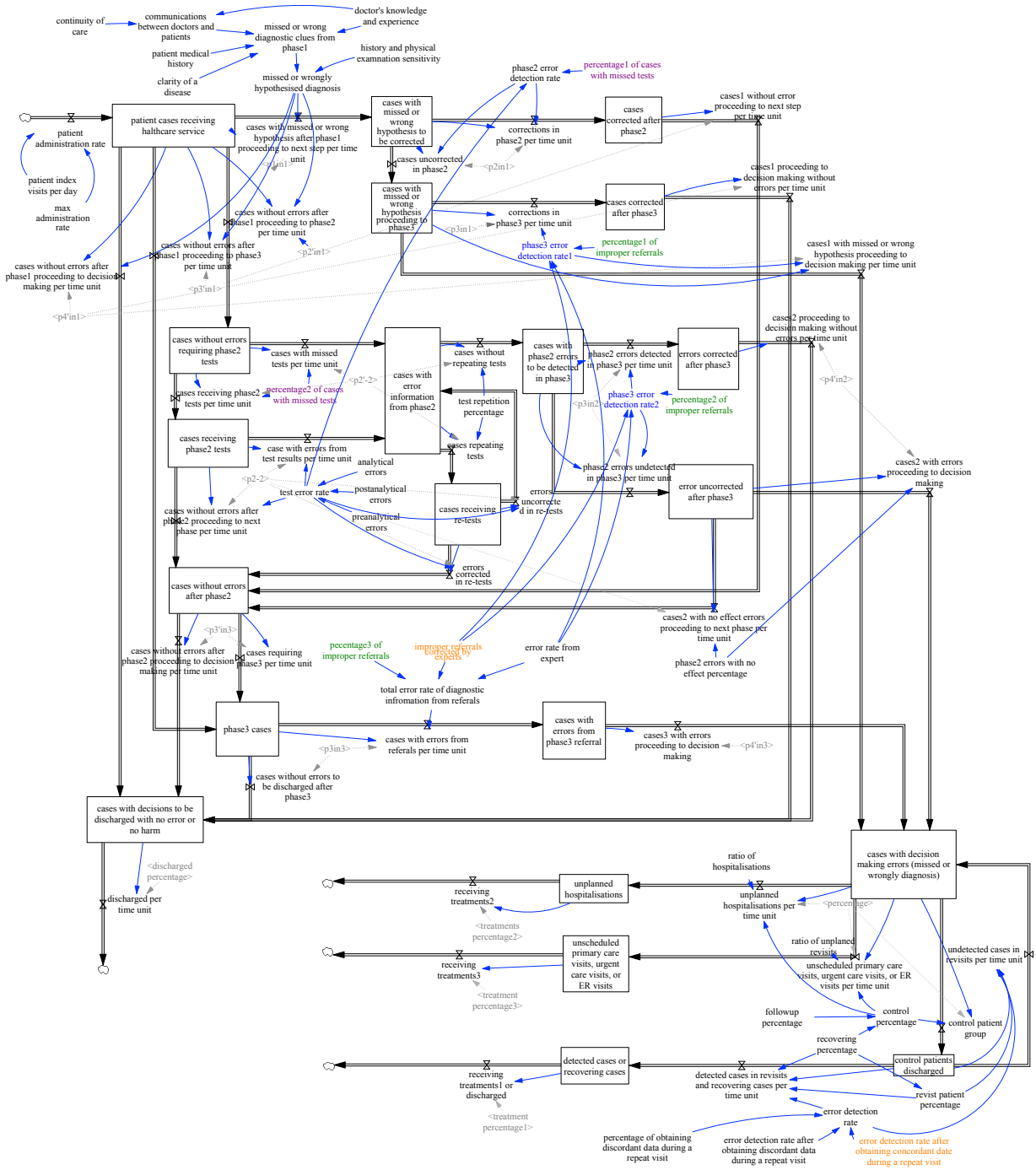


Figure 6.12 Quantitative model structure for diagnostic errors

6.10 Summary

This chapter further transfers the factors and relations in the CLD into the structure for the quantitative model. In order to present the number of cases with or without errors during diagnosis and to keep the consistency of quantitative variables, it modified the relations and factors in the CLD according to the diagnostic process. In the end, the quantitative model structure presents different case numbers at each phase as well as relevant factors. It illustrates the phases of the diagnostic process, and provides patient case flows during the diagnostic process as well as different patient outcomes. It applies inflow cases and outflow cases to present current system case numbers. Moreover, relevant key factors linked with the number of cases are mapped into the model. It helps to understand where errors occur and the corresponding relevant factors. Also, error case flows can be observed, and their outcomes are also presented.

However, every arrow in the model indicates a relationship function and represents the quantitative relationship between the variables from each end of the arrow. Although, most of the relationship functions can be either identified easily or represented using the accumulation of the inflow cases and out flow cases over time, there are several interrelations whose quantitative functions are still unclear. Specifically, the functions between the factors in phase1 and their dependent variables remain unclear. Therefore, the next chapter implements regression analysis to determine the unclear relationship functions for relevant variables.

Chapter 7 Regression Modelling

7.1 Introduction

System dynamics modelling requires that the interrelations of system variables are known or identified. However, unlike the simple interrelations in the remaining parts of diagnostic process, the correlations between relevant factors and dependent variables remain unclear in phase1. To present how the factors quantitatively affect the dependent variables, this chapter adopts regression analysis to identify their relationship functions.

This chapter firstly prepares the variables by quantifying non-numerical variables. Then, it uses data collection from questionnaires for the regression analysis. Afterwards, regression analysis is carried out, and relationship functions are identified and evaluated.

7.2 Steps of regression modelling

The regression modelling process follows four main steps: cohort construction, where aims and targets are defined; feature engineering, which includes data preparation, feature construction and feature selection; regression modelling; and model evaluation.

Specifically, the aim of the model is to identify relationship functions between the factors in phase1 and their dependent variables. The relevant variables are originally from the selected literature, which includes non-numerical variables. Thus, this chapter continues to prepare the variables. Relevant non-numerical variables are quantified. Also, questionnaires are used to collect data for the regression analysis. Afterwards, regression analysis is carried out. Relationship functions are identified and evaluated.

7.3 Methods used to quantify non-numerical variables

Regression analysis using data of relevant variables identifies correlation functions, and all variables should be able to be quantitatively presented. This section prepares the relevant variables to make sure that all variables can be quantitatively presented.

There are two types of variables involved in phase1: numerical variables and non-numerical variables. Numerical variables can be quantitatively represented using numbers, and numerical variables in phase1 are: history and physical examination sensitivity, missed or wrong diagnostic clues from phase1, and missed or wrongly hypothesised diagnosis. Also,

there are several non-numerical variables, and these variables require to be quantified first so that they can be used as model input variables. The non-numerical variables in phase1 and their relevant dependent variables are listed in Table 7.1.

Non-numerical variables:

1. Continuity of care
2. Communications between doctors and patients
3. Patient medical history
4. Clarity of disease symptoms/signs
5. Doctor's knowledge and experience

Relevant dependent variables:

6. Missed or wrong diagnostic clues from phase1
7. Missed or wrongly hypothesised diagnosis

Table 7.1 List of the non-numerical variables and relevant dependent variables

7.3.1 Likert scaling of non-numerical variables

Quantifying variables involves finding a way of measuring the variables. In this section, non-numerical variables are taken as ordinal variables, so that they can be applied by many methods in terms of measurement. The Likert scale, which was developed in 1932 and initially used for measuring attitudes, is widely applied in scaling ordinal variables. This section adopts Likert scales as the scaling method, and measures each non-numerical variable using three levels, namely, three categories. Moreover, each category is given a score to quantify the variables.

Likert scaling uses several levels/categories to measure the variable, and Table 7.2 shows how to use scales to present non-numerical variables by measurable ordinal variables. For convenience, all non-numerical variables are given a unique variable number. Likert scaling measures each variable into three levels, namely, categories c1, c2 and c3 to represent three different levels respectively. Each category indicates a "good", "average" or "bad" level or an "expert", "senior" or "junior" level.

Variable number	Variable name (Likert item)	Scaling _ level/categories(c)
Variable 1 (v1)	Continuity of care	c1.Good—c2.Average—c3.Poor
Variable 2 (v2)	Communications between doctors and patients	c1.Good— c2.Average— c3.Poor

Variable 3 (v3)	(Access to) patient medical history	c1.Good—c2.Average—c3.Poor
Variable 4 (v4)	Clarity of disease symptoms/signs	c1.Good—c2.Average—c3.Poor
Variable 5 (v5)	Doctor's knowledge and experience	c1.Level 3—c2.Level 2—c3.Level 1

Table 7.2 Non-numerical variables and scales

The dependent variables of the non-numerical variables are the number of "*missed or wrong diagnostic clues from phase1*" that indicates the number of clues or information collected from patients, and the number of cases with "*missed or wrongly hypothesised diagnosis*" that means the initial hypothesised diagnosis after phase1, as shown in Table 7.3.

Variable number	Variable name
Variable 6 (y1)	Missed or wrong diagnostic clues from phase1
Variable 7 (y2)	Missed or wrongly hypothesised diagnosis

Table 7.3 Correlated dependent variables

7.3.2 Data representation of categories

Theoretically, for an individual doctor, a descriptive table of patient cases can display the category (c1, c2 or c3) of the individual case for each variable. Taking variable1 (v1) as an example, the descriptive table should look similar to Table 7.4, where "1" indicates the patient case is in the corresponding category and "0" means the patient case is not in the corresponding category. Similarly, the descriptive table can be built for v2, v3, v4, and v5 as well.

Patient Case No.	Variable1 (v1)		
	c1	c2	c3
PatientCase 1	1	0	0
PatientCase2	0	0	1
PatientCase 3	0	1	0
: :			
PatientCase n	1	0	0

Table 7.4 A descriptive table format showing patient cases and its category

However, building these descriptive tables is based on the information regarding to individual patient case, which requires access to the details of each patient case. This study does not have access to the information of individual patient cases due to time and resource

limitations. Thus, this study is designed to ask doctors to do relevant estimations. Each doctor is asked to estimate the case distribution of an individual variable within the three categories. Again, taking variable1 as an example, doctors are asked to estimate how many patient cases are under the category1 for v1, based on the 100 patient cases they receive, how many patient cases are under the category2, and how many patient cases are under the category3. After the estimations from individual doctors are obtained, a table similar to Table 7.5 can be built.

Doctor No.	Variable1 (v1)		
	c1	c2	c3
Doctor 1	80(out of 100)	10(out of 100)	10 (out of 100)
Doctor 2	70	10	20
Doctor 3	70	30	0
…	…	…	…
Doctor n	60	10	30

Table 7.5 A table format showing individual doctors and his patient distribution within 3 categories

The numbers that the doctors provide are actually the accumulation of the 100 patient cases in Table 7.4.

7.3.3 Weighted scores

To quantify the ordinal variables, each category is assigned a score. The score weights the effect of the individual category on the overall score of an individual variable, or implies their relative importance to the overall score of a variable. By assigning the weighted score, each doctor can obtain the overall score of the individual variable, which is the outcome of the doctor's performance of the variable.

Categories	c1	c2	c3
Weighted Score	β_1	β_2	β_3

Table 7.6 The categories and its weighted score

The categories and the weighted scores can be represented by Table 7.6, where β_1 is the score for c1, β_2 is the score for c2, and β_3 is for c3. For an individual variable, the overall gained score of individual doctors can be represented as follows:

$$\text{overall score} = \beta_1 \times (c1) + \beta_2 \times (c2) + \beta_3 \times (c3)$$

where $\beta_1, \beta_2, \beta_3$ are the weighted scores or the impact parameters of the overall gained score.

During this study, it is assumed that "good" cases are equally as important as "poor" cases, which means that $|\beta_1 - \beta_2| = |\beta_3 - \beta_2|$. At the same time, the "good" cases increase the overall score of the variable, "poor" cases decrease the score with equal effort, and "average" cases neither increase nor decrease the score. Therefore, for v1, v2, v3 and v4, it can be assumed that $\beta_1 = 1, \beta_2 = 0.5, \beta_3 = 0$ during this experiment, so that the range of the overall score of variable can lie between 0 and 1. And v5 is discussed separately, since its categories measure the level of clinicians instead of patient cases. It is assumed that $\beta_1 = 3, \beta_2 = 2, \beta_3 = 1$ in this experiment, because the three categories of v5 indicates level 3, level 2 and level 1 respectively and this assumption can make the score for v5 is 1, 2 or 3.

Variable number	Variable name (Likert item)	Categories(c) with scores
Variable 1 (v1)	Continuity of care	c1.Good—c2.Average—c3.Poor 1 — 0.5 — 0
Variable 2 (v2)	Communications between doctors and patients	c1.Good— c2.Average— c3.Poor 1 — 0.5 — 0
Variable 3 (v3)	Patient medical history	c1.Good—c2.Average—c3.Poor 1 — 0.5 — 0
Variable 4 (v4)	Clarity of disease symptoms/signs	c1.Good—c2.Average—c3.Poor 1 — 0.5 — 0
Variable 5 (v5)	Doctor's knowledge and experience	c1.Level 3—c2.Level 2—c3.Level 1 3 — 2 — 1

Table 7.7 Non-numerical variables with assigned categories and scores.

Table 7.7 summarises the variables with categories and scores used in the experiments, and the definitions of the categories are further discussed in detail.

The v1, v2, v3 and v4 use the same three categories. The category 1 (c1) means a group of cases with *"Good" performance in the relevant variable,* category 2 (c2) indicates case group with *"Average" performance in the relevant variable* and category 3 (c3) refers to relevant "*Poor*" performance, where the weighted score for c1 is 1, for c2 is 0.5 and for c3 is 0. The "*continuity of care*" is scaled according to the frequency of the patient's visit to the same doctor. If a patient often sees the same general practitioner (GP), it is considered as a "*Good*" case, on the other hand, if the patient rarely visits the same GP, it is measured

"*Poor*". Neither "*Good*" or "*Poor"* cases are in the middle level "Average". "*Communications between doctors and patients*" is divided on the basis of the quality of the communications. Good communications during the visit are "*Good*" cases, and if doctor and patient hardly or rarely have communications during the visit, it is considered as "*Poor*". The "*patient medical history* " is scaled according to whether the GP can effectively access and review patient medical history. An effective access and review of patient medical history is "*Good*", while, if the GP can not access and review medical history or can only review very limited history information, it is considered as "*Poor*". The "*clarity of disease symptoms/signs*" shows the clarity level of disease symptoms/signs presented by patients. "*Good*" cases are for patients who present clear typical symptoms/signs, and "*Poor*" cases are for patients whose symptoms/signs are unclear. For example, an early stage of disease may affect the clarity of disease symptoms.

For the better understanding and explanation of the variable v5, "*Doctor's knowledge and experience*", is discussed separately. It is scaled according to the doctor's title, and is divided as "Level 1" "Level 2" and "Level 3". The standard for dividing each level is identified according to the qualifications of the three titles of health physicians: resident physician, attending physician and professor. Resident physicians are taken as "Level 1", who usually have less than 5 years of relevant work experience. Attending physicians taken as "Level 2", which are related to a 5-10 year work experience and relevant training qualifications. Professors including associate professors are in "Level 3", which indicates an over 10-year work experience with relevant qualifications. Matching details are listed in Table 7.8.

Variable 5 (v5): Doctor's knowledge and experience			
Doctor title	Resident physician	Attending physician	Professor
Category	Level 1	Level 2	Level 3
Score	1	2	3

Table 7.8 Doctor's knowledge and experience scales and scores

7.3.4 Computing the non-numerical variables

The overall scores of the variables for each clinician can be computed, after the clinician provides the numbers of the cases under different categories and the weighted scores of each category are known. For each variable, the individual clinician is asked to provide the case numbers under the different categories, which actually represents the case distributions of the clinician for the variable. Furthermore, using the case distribution and combining category scale score in Table 7.7, new data can be generated for each non-

numerical variable. The generated data presents the actual overall score of the non-numerical variable from the corresponding clinician. Take variable 1 as an example. Figure 7.1 illustrates the process of quantifying the variables. Table 7.9 demonstrates the transformation of the data from the original data table to the computed data, where some extreme examples are tested. Variables v1, v2, v3, v4 are generated from the data of the categories, while v5 remains as the ordinal format "1, 2 or 3", since v5 represents the level of doctor's knowledge and experience. Overall, a higher final score of a factor, v1, v2, v3, v4 or v5, indicates a better performance on the past of the clinician on the variable.

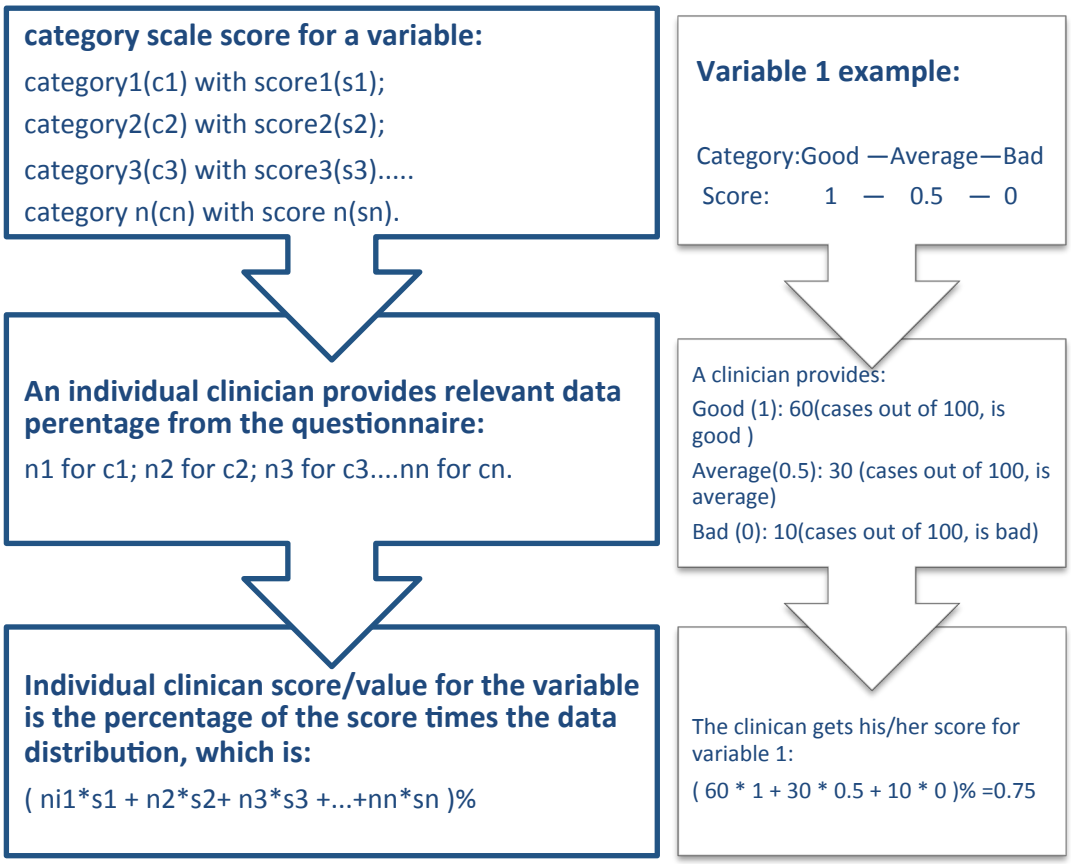


Figure 7.1 Process flow of quantifying non-numerical variables

Original data table after collecting data from clinicians:

Doctor No.	Variable1 (v1)			Variable2 (v2)		
	c1	c2	c3	c1	c2	c3
Dr.1	60 *(out of 100)*	30	10	…	…	…
Dr.2	70	10	20	…	…	…
Dr.3	99	0	1	…	…	…
Dr.4	0	100	0	…	…	…
…						
Dr.n	50	0	50	…	…	…

Computed data table by combining scores:

Doctor No.	Variable1 (v1)	Variable2 (v2)
	Total score	Total score
Dr.1	75% =(60*1+30*0.5+10*0)%	…
Dr.2	75%	…
Dr.3	99%	…
Dr.4	50%	…
…		…
Dr.n	50%	…

Table 7.9 Data transformation

7.3.5 Further discussion

Three points in the process of quantifying the non-numerical variables are further discussed in this subsection.

- Scaling method:

  The three-point scale method implemented in this study properly reflects the overall level of patient distribution, although more accurate measurement may require either dividing the patients into more groups, or reviewing and tracing the patient history medical records which can not be performed due to time and access limitations.

- The definition for different categories or levels:

  During the study, relevant categories are divided depending on the description of “often”, “hardly” or ”rarely” for the variable, without providing an exact number to classify the cases. It mainly relates to the fact that the data of each category will be further collected from the clinician estimates, rather than from reviewing individual patient medical record. The more accurate the implemented measurement is, the harder it is for the clinicians to make the estimates.

- Weighted scores:

Weighted scores provide the relation between clinician estimates of the categories and the overall score obtained. However, the study can also be conducted without assigning weighted scores to categories.

Assuming that the estimated data for the three categories of v1 are n1, n2, n3, and the y is the dependent variable of v1, instead of finding the relationship function of y and v1, the study can obtain the relationship function between y and n1, n2, n3 using the same statistical method. The parameters of n1, n2, n3 in the relationship function indicate the impact of each category on y.

However, this method is not implemented in the thesis for two reasons. Firstly, this method cannot explain the direct relations of the variables/factors and their dependent variables. Secondly, this method covers more dependent variables, and it requires more data in order to obtain the relationship functions. Thus, the weighted score method is applied in the thesis to provide a clear explanation of the experiment results.

7.4 Data for regression modelling

7.4.1 Data acquisition

Questionnaires are chosen as the method to collect the relevant data in this step. Using questionnaires, a wide range of participants can be conveniently approached and a large enough quantity of feedback can be received. This is the main reason of choosing questionnaires.

The steps below provide a detailed description of how the questionnaire study is conducted:

1) Definition of study objectives:

This questionnaire is designed to collect the quantitative data under each of the scales or categories of each ordinal variable and the quantitative data of their correlated variables. Data collected from the questionnaires are used for the further regression analysis.

Objective variables: Categories of five non-numerical variables shown in Table 7.7 need to be quantified, and their correlated variables shown in Table 7.3 also are required to provide the corresponding value to observe the inter-relations and build relationship functions.

2) Identify participants:

General practitioners are the participants for the data collection. A minimum number of 50 participants are required, because there are five independent factors to be analysed in this study. Due to ethical issues, all participants are clinicians based in Beijing, China. Participants are from two sources: hospitals and participants in medical meetings, and are randomly chosen according to the availability and interest in the study. To ensure a balance of views, there is no criterion relating to gender and levels. The intention is to have 50% female and 50% male paticipants, and the numbers of the three different levels of clinicians are equal.

3) Design questionnaire documents:

According to the objective variables to be assessed, questionnaires show a list of relevant questions. For each question, clinicians are asked to provide a number or a percentage under different scale categories for each variable. The names of categories remove all subjective words, such as “good”, or “poor”, and are named as “group 1 “, “group 2” or “group 3”, which is designed to help participants provide more objective answers.

4) Pilot test:

One doctor was chosen to engage in a face-to-face discussion to test the questionnaire and to provide comments and feedback in relation to the questionnaire. The feedback mainly focuses on whether the given questions can be easily understood and whether the given questions have an accurate or proper word description.

5) Revised questionnaire according to the feedback:

The questionnaire is further revised based on the feedback from the pilot test. The original questionnaire and its English translation are attached in APPENDIX VI.

6) Sending and collecting questionnaires:

A total of 60 questionnaires were printed out and sent to the doctors in the hospitals and the doctors attending the medical meetings. All questionnaires are anonymous. In the end, all 60 questionnaires were collected.

7) Data analysis and documentation:

A total of 60 sets of data from questionnaires are recorded in Excel documents for further analysis.

7.4.2 Data preparation

This step is to prepare the collected data for analysis. This part contains two aspects of data preparation: data cleansing and imputation, and data computing for non-numerical variables.

1) Data cleansing and imputation:

After data are collected from the questionnaires, original collected data may contain unclear data or missing data, and this step is to process the unclear or missing data and to prepare the datasets for the analysis. Unclear data or "dirty" data are very likely to happen in the manually input data, such as questionnaire data. Generally speaking, the common problems of the "dirty" data include: non-standard data format, out of range values, multiple units for one variable, or unstructured content. At the same time, besides the "dirty" data, the datasets may contain missing values. There are three methods below that can be used for data imputation. Table 7.10 summarises the three common methods for data imputation.

Common methods for data imputation:
1) Fill the value by inferring from other variables if possible. This method is particularly appropriate for dependent or overlapping variables so that the value can be inferred from other variables. 2) Fill the value with the mean of the remaining data for numerical variables, median for ordinary variables or mode for nominal variables. This method can reduce the impact of the missing values. 3) Discard the entire dataset with many missing instances, such as any with over 40% data missing.

Table 7.10 Methods for data imputation

During this study, a total of 60 datasets are reviewed, and overall data information is clear and readable.

During the data cleansing process, a model is built for systematically cleansing the data using SPSS modeler[147]. Although the quantity of data in this study is small and the data can be cleaned manually, this model provides a way of using SPSS modeler to systemically clean the data, which is helpful in analysing a large quantity of data.

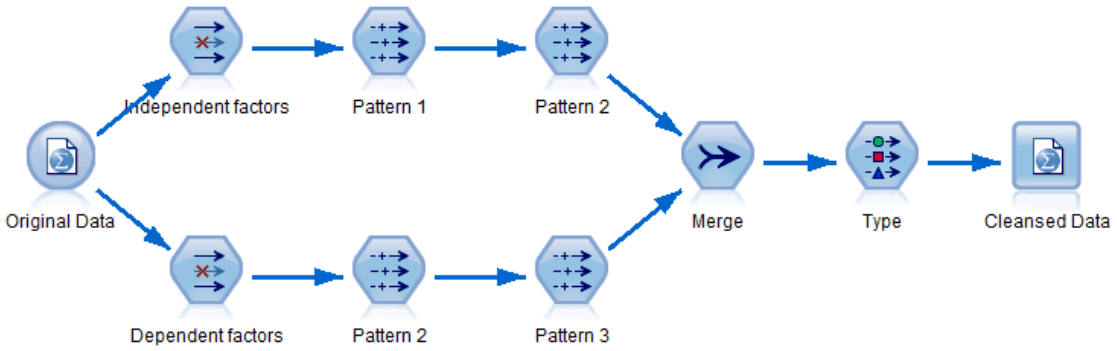


Figure 7.2 Data cleansing using SPSS modeler

Figure 7.2 is the data-cleansing model created by SPSS modeler. At the beginning, the original data table can be divided into several threads, and each thread contains different variables. This step is to separate variables and to process different variables separately. Figure 7.2 shows that the original data are separated into two threads, which represent independent variables and dependent variables respectively. Secondly, the common patterns of the improper data should be found, and cleansing patterns are performed to process the data. For example, independent variables are the number of cases, but several values with the percent symbol "%"appear many times in the collected data. This is a common pattern, so cleansing pattern "*Pattern 1*" should be removing the "%" symbol. The cleansing pattern uses a relevant formula to implement the data cleansing, and the codes of the formula for *Pattern 1* can be: *replace ("%", "", @FIELD),* which means using "" replaces all "%" in the current input data. Finally, different threads are merged together into one data table *"Cleansed Data"* again, after being processed separately.

Besides data cleansing, data imputation is also conducted, because four questionnaires contain improper data that are out of range data and missing data. The entire data sets from the questionnaire with out of range values or missing values are discarded, since the values are the information for key factors and related with other variables. This method does not result in any negative impact of the improper values on the analysis of relationship models, and it is the cleanest and simplest way.

In the end, a total of 56 sets of data from the questionnaires are ready for analysis after four sets with improper values were discarded.

2) Computing data for non-numerical variables:

The reason and the method for computing data are introduced in subsection 7.3.4. Questionnaires do not directly provide data for the non-numerical variables. For each variable, questionnaires ask clinicians to individually estimate the case numbers under the three categories. The case numbers require to be further computed to generate the data for the non-numerical variables. Thus, after the step of data preparation, the datasets from questionnaires are transformed into a new data table for non-numerical variables using the method introduced in subsection 7.3.4. The new table shows the computed data for v1, v2, v3, v4, v5 with y1, y2 being unchanged. The new data table is ready for further analysis of determining the relationship functions at the end of this step.

7.5 Methods used to build relationship functions

Relevant factors were selected from the literature review and from clinical feedback, which have been reported in the previous two chapters. Then, relevant variables were quantified to make sure variables can be measured in a quantitative way. Also, data collection using questionnaires was undertaken. After data preparation, data analysis is carried out in this section in order to identify the relationship functions.

Regression analysis is implemented as the method for building the relationship functions in this section, because the dependent variables are numerical variables and the values are all continuous. Regression modelling determines the relationship functions by finding the best-fit curve for sample data. Also, this thesis implements regression analysis with system dynamics modelling. The first part of system dynamics modelling has selected the theoretical features/factors and potential qualitative interrelations using literature knowledge and clinical experience, while regression modelling depends totally on the collected data. Thus, regression analysis in this thesis, based on the data, will test and modify the factors and interrelations again.

Specifically, this thesis conducts regression modelling following the steps below:

1) Variable correlation analysis:

After the previous system dynamics model provides theoretical factors and potential cause-effect qualitative interrelations among variables, this step conducts correlation analysis to verify the theoretical factors and interrelations. It tests the

factors and correlations, and at the same time it works as a filter factor/feature selection method, and modifies factors.

2) Univariate regression analysis using curve estimations:
   If the relationship function is between a single factor and its dependent variable, univariate analysis is conducted to seek the best curve by testing different possible curves.
3) Multiple regression analysis:
   If the relationship function has more than one factor, multiple regression analysis is implemented. It starts with linear regression. Results are evaluated.

Models are evaluated using evaluation metrics. Because of limited data resources, modela are evaluated using several metrics such as p value, standard error and R^2, so that the predicted values from regression models can be compared with the observed values from sample data.

Relevant experiments of data analysis are conducted using SPSS as the tool.

7.6 Variable correlation analysis

Correlation analysis is used to identify the association between two variables and measures the strength of the association. It is widely used as a filter in order to do feature/factor selection. There are total seven variables from phase1 for regression analysis, and the variables with their data types are listed in Table 7.11.

Variable ID	Variable Name	Data Type
v1	Continuity of care	Interval
v2	Communications between doctors and patients	Interval
v3	Patient medical history	Interval
v4	Clarity of disease symptoms/signs	Interval
v5	Doctor's knowledge and experience	Ordinal
y1	Missed or wrong diagnostic clues from phase1	Interval
y2	Missed or wrongly hypothesised diagnosis	Interval

Table 7.11 Summary of the variables for the regression analysis

There are two data types: interval and ordinal. The variable v5 is on an ordinal scale and indicates the level number of the "*doctor's knowledge and experience*", such as 1, 2 or 3.

The remaining variables are on an interval scale with the value between 0 to 1. Also, the previous qualitative model, using arrows, provides the correlation links among these variables, which indicates that there exists a cause-effect correlation among the variables as follows:

- v1 and v5 are causes of v2.
- v2, v3, v4 and v5 are causes of y1.
- y1 is a cause of y2.

Correlation analysis further verifies these hypotheses, which specifically represents cause-effect relations in the model.

There are two methods that are widely used to do the correlation analysis: Pearson's correlation and non-parametric measures such as Spearman's correlation or Kendall's correlation.

The Pearson's correlation coefficient, represented by "r", is a measure of the strength of a linear association between two variables, which is specifically for a linear correlation. It requires the assumption that the relationship between the variables is linear and the variables to be measured are on interval scales.[149]

Non-linear correlations can be assessed using Spearman's correlation or Kendall's correlation. Compared with Pearson correlation, Spearman's correlation does not require the assumption that the relationship between the variables is linear, and it also does not require the variables to be measured on interval scales.[149] It can be used for variables measured on ordinal scales. [148] Spearman's correlation coefficient "r_s" represents the strength of a monotonic association between two variables, which assesses how well an arbitrary monotonic function can describe a relationship between two variables, without making any assumption about the frequency distribution of the variables.[149]

Kendall's correlation coefficient "τ" is similar to Spearman's correlation coefficient "r_s", but Kendall's correlation can be used with smaller samples or when there are many values for the same score.[150]

These correlation coefficients fall between +1 and -1. A correlation coefficient of +1 indicates a perfect positive correlation, while a coefficient of -1 indicates a perfect negative correlation. [151] A coefficient of 0 indicates the absence of an association between the two

variables. A coefficient that is closer to +1 or -1 indicates a stronger positive or negative relation. A positive coefficient means that when one variable increases, the other variable has a corresponding increase. Similarly, a negative coefficient means that as one variable increases, the other variable decreases by a set amount.

Along with the correlation coefficients, the corresponding *statistical significance* is also used to assess the association of the variables. *Statistical significance* uses the *p value* to represent the probability to obtain an effect equal to or more extreme than the one observed.[152] Thus, the significance of the correlation coefficient represents how likely the coefficient that we would obtain from the sample data is the same value as that coefficient obtained by chance. Generally speaking, a p-value that is equal to or smaller than a 0.05 significance level is acceptable, which indicates that the correlation is significant at the significance level.

This study works out the values of correlation coefficients among the total of seven variables, as well as the significance of the correlation coefficients. Except for v5 which is on a ordinal scale, the rest of the variables are continuous data. Also, the relations are more likely to be linear according to a cause-effect association. Thus, Pearson's correlation is chosen to implement the correlation analysis first. Furthermore, Kendall's correlation is also used, because v5 is measured at the ordinal level and so does not show a strong association with other variables in the results of Pearson's correlation analysis. Also, Kendall's correlation helps to test possible monotonic relationships between variables. Table 7.12 and Table 7.13 show Pearson's correlation results and Kendall's correlation results respectively.

		v1	v2	v3	v4	v5	y1	y2
v1	Pearson Correlation	1	.611**	.346**	.470**	.271*	-.281*	-.242
	Sig. (2-tailed)		.000	.009	.000	.043	.036	.073
	N	56	56	56	56	56	56	56
v2	Pearson Correlation	.611**	1	.559**	.591**	.297*	-.359**	-.229
	Sig. (2-tailed)	.000		.000	.000	.026	.007	.090
	N	56	56	56	56	56	56	56
v3	Pearson Correlation	.346**	.559**	1	.377**	.062	-.431**	-.167
	Sig. (2-tailed)	.009	.000		.004	.652	.001	.218
	N	56	56	56	56	56	56	56
v4	Pearson Correlation	.470**	.591**	.377**	1	.283*	-.386**	-.176
	Sig. (2-tailed)	.000	.000	.004		.035	.003	.196
	N	56	56	56	56	56	56	56
v5	Pearson Correlation	.271*	.297*	.062	.283*	1	-.054	-.025
	Sig. (2-tailed)	.043	.026	.652	.035		.692	.852
	N	56	56	56	56	56	56	56
y1	Pearson Correlation	-.281*	-.359**	-.431**	-.386**	-.054	1	.462**
	Sig. (2-tailed)	.036	.007	.001	.003	.692		.000
	N	56	56	56	56	56	56	56
y2	Pearson Correlation	-.242	-.229	-.167	-.176	-.025	.462**	1
	Sig. (2-tailed)	.073	.090	.218	.196	.852	.000	
	N	56	56	56	56	56	56	56

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7.12 Pearson's correlation results

			v1	v2	v3	v4	v5	y1	y2
Kendall's tau_b	v1	Correlation Coefficient	1.000	.471**	.205*	.339**	.223*	-.200*	-.198*
		Sig. (2-tailed)	.	.000	.034	.000	.047	.044	.050
		N	56	56	56	56	56	56	56
	v2	Correlation Coefficient	.471**	1.000	.346**	.377**	.201	-.216*	-.216*
		Sig. (2-tailed)	.000	.	.000	.000	.075	.031	.033
		N	56	56	56	56	56	56	56
	v3	Correlation Coefficient	.205*	.346**	1.000	.193*	-.036	-.230*	-.067
		Sig. (2-tailed)	.034	.000	.	.047	.751	.022	.508
		N	56	56	56	56	56	56	56
	v4	Correlation Coefficient	.339**	.377**	.193*	1.000	.225*	-.333**	-.135
		Sig. (2-tailed)	.000	.000	.047	.	.046	.001	.183
		N	56	56	56	56	56	56	56
	v5	Correlation Coefficient	.223*	.201	-.036	.225*	1.000	-.047	-.115
		Sig. (2-tailed)	.047	.075	.751	.046	.	.688	.328
		N	56	56	56	56	56	56	56
	y1	Correlation Coefficient	-.200*	-.216*	-.230*	-.333**	-.047	1.000	.358**
		Sig. (2-tailed)	.044	.031	.022	.001	.688	.	.001
		N	56	56	56	56	56	56	56
	y2	Correlation Coefficient	-.198*	-.216*	-.067	-.135	-.115	.358**	1.000
		Sig. (2-tailed)	.050	.033	.508	.183	.328	.001	.
		N	56	56	56	56	56	56	56

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Table 7.13 Kendall's correlation results

A total of 56 sets of data are used for the analysis. As we have seen from the table, Pearson's correlation and Kendall's correlation indicate similar association results, and the results obtained from the sample data match most of the initial qualitative relations from the previous system dynamics modelling. To be specific, the relations are:

- v1 is significantly correlated with v2, where the r is 0.661 with a *p value* 0.00.

v5 shows less linear correlation with v2 at the 0.05 level with the r 0.271, and does not show significant correlation in Kendall's results.

- y1 shows significant correlation with v2, v3, and v4 separately at the 0.01 level. However, the 56 sets of sample data do not show correlation between v5 and y1 in either Pearson's analysis or Kendall's analysis.
- y1 and y2 are significantly correlated, where the r is 0.462 with a *p value* 0.00

In addition to the previous qualitative relations, the sample data also show the following information:

- v1 is not only highly related to v2, but also related to v3, v4 separately at the same time, although the relations with v3 or v4 are less correlated than v2.
- The correlations among three factors v2, v3, v4 are significant as well, especially for the correlation between v2 and v3, and the correlation between v2 and v4.

Although all factors are initially selected according to the literature and clinician feedback, the experiment results that are on the basis of sample data also further modify the factors and relations.

The variable v5 is removed from the correlation assumptions, because v5 does not show significant correlation in v2 and y1 according to the data results, and factors with the significance less than 0.01 are selected to remain. The other reason why v5 is removed is that the sample data of v5 is not uniformly distributed in its range. Data collected from the questionnaires for v5 "Doctor's knowledge and experience" cover an uneven distribution of three levels of clinicians. Only three questionnaires are collected from level-1 clinicians, and the rest are the datasets from level-2 and level-3 clinicians. Thus, this may be also the reason why v5 is not significantly correlated with v2 and y1. The experimental results also indicate that level-2 and level-3 clinicians do not show significant impact on the dependent variables. Overall, v5 "Doctor's knowledge and experience" is removed from the analysis.

Moreover, two new added qualitative relations between v1 and v3 and between v1 and v4 are required to be tested and evaluated. According to the data correlation results, the following relations are to be determined and evaluated:

- v1 is a cause of v2
- The relations between v1 and v3
- The relations between v1 and v4

- v2, v3, v4 are causes of y1
- y1 is a cause of y2

7.7 Univariate regression analysis

This step is to discover the relationship between a single factor and its dependent variable, and to determine the best quantitative relationship models, which cover the relations between v1 and v2, v1 and v3, v1 and v4.

SPSS provides a wide range of 11 models that can be used to estimate the best-fit curve. They are listed as below[153]: "

- ***Linear.*** *Model whose equation is Y = b0 + (b1 * t). The series values are modelled as a linear function of time.*
- ***Logarithmic.*** *Model whose equation is Y = b0 + (b1 * ln(t)).*
- ***Inverse.*** *Model whose equation is Y = b0 + (b1 / t).*
- ***Quadratic.*** *Model whose equation is Y = b0 + (b1 * t) + (b2 * t**2). The quadratic model can be used to model a series that "takes off" or a series that dampens.*
- ***Cubic.*** *Model that is defined by the equation Y = b0 + (b1 * t) + (b2 * t**2) + (b3 * t**3).*
- ***Power.*** *Model whose equation is Y = b0 * (t**b1) or ln(Y) = ln(b0) + (b1 * ln(t)).*
- ***Compound.*** *Model whose equation is Y = b0 * (b1**t) or ln(Y) = ln(b0) + (ln(b1) * t).*
- ***S-curve.*** *Model whose equation is Y = e**(b0 + (b1/t)) or ln(Y) = b0 + (b1/t).*
- ***Logistic.*** *Model whose equation is Y = 1 / (1/u + (b0 * (b1**t))) or ln(1/y-1/u) = ln (b0) + (ln(b1) * t) where u is the upper boundary value. After selecting Logistic, specify the upper boundary value to use in the regression equation. The value must be a positive number that is greater than the largest dependent variable value.*
- ***Growth.*** *Model whose equation is Y = e**(b0 + (b1 * t)) or ln(Y) = b0 + (b1 * t).*
- ***Exponential.*** *Model whose equation is Y = b0 * (e**(b1 * t)) or ln(Y) = ln(b0) + (b1 * t). "*

Since dependent variables or predicted variables are continuous values, regression analysis is used and five models were selected and compared in this section, which are linear model, quadratic model, s-curve model, logistic model and exponential model.

Each model is also evaluated at the end of each section. Since the quantity of data is small, the analysis of variance (ANOVA) is also employed, and several metrics are used to evaluate the adopted model. The metrics include:

1) Significance or P-value: It means the probability that the results observed in a study could have occurred by chance.[154] Generally speaking, a *p value* of 0.05 or below is taken as being statistically significant.
2) Standard error: It is the standard deviation of the regression line, which indicates how much the observed data or real data differ from the values on the regression line. [155] The value of standard error is less than 0.1.
3) R-squared, the coefficient of determination: It is the coefficient of determination. It is the correlation coefficient squared, which indicates the proportionate amount of variation in the response variable explained by the independent variables in the regression model.
4) Adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors or degrees of freedom in the model. R-squared increases when a new variable is added to a model every time. The adjusted R-squared, taking account of the degrees of freedom, increases only if the new predictor/factor improves the model by more than would be expected by chance, and it decreases if the new predictor is not related to the dependent variable. During this study, both R-squared and adjusted R-squared are used as a reference metric, and a larger adjusted R-squared is preferred.

7.7.1 Relationship between v1 and v2

This subsection is to determine the relational equation for v1 and v2, where v1 is the factor and v2 is the dependent variable. It first conducts the curve estimations. By comparing the statistical results of different curves or models, the best model type is chosen. Furthermore, the details of the model results are described and relevant quantities in the results are explained.

7.7.1.1 Curve estimation

At first, the five possible curves including linear model, quadratic model, s-curve model, logistic model and exponential model were tested and compared in order to represent the relationship function between v1 and v2. Figure 7.3 illustrates the graphs of the five curves for the 56 sample data. As seen in Figure 7.3, the five models show no significant difference in using graphs to represent the sample data. Thus, five models are further compared using statistical data and the results are summarised in two tables: the model summary table-Table 7.14 and the standard error summary table–Table 7.15.

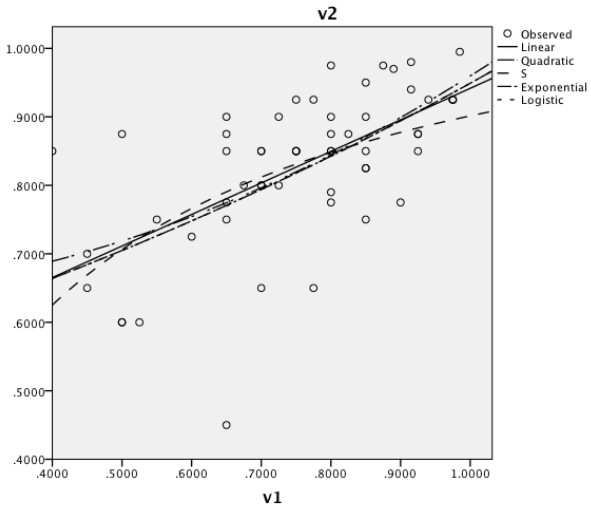


Figure 7.3 Different graphs of v2 as a function of v1

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	0.373	32.163	1	54	0.000	0.481	0.461	
Quadratic	0.378	16.113	2	53	0.000	0.634	0.013	0.312
S	0.295	22.562	1	54	0.000	0.140	-0.244	
Exponential	0.343	28.215	1	54	0.000	0.524	0.595	
Logistic	0.343	28.215	1	54	0.000	1.910	0.552	

Table 7.14 Curve estimation for v2 as a function of v1

Model information summary and parameter estimation:

Table 7.14 describes details of the model summary as well as parameters, including:

- *R-squared*;
- *F*, *the F-statistic*: It expresses the ratio of mean squares, and the F in the table can be represented as given below[156]:

$$F = \frac{mean\ square\ of\ the\ regression}{mean\ square\ of\ the\ residuals}$$

  Thus, a higher F indicates a significant effect.
- *df*, the degrees of freedom: It is the number of values that are free to vary. The regression degrees of freedom is equivalent to the number of coefficients estimated minus 1.

- *Sig*: significance level, or p value;
- *Parameter estimation*: "*b*" is the coefficient of the independent variable, and "*Constant*" is the constant in the model equation.

where sig, *F and R-squared* are the metrics to be used for curve selection.

Standard errors and p values:

Table 7.15 shows more information about standard errors. Specifically, it has two sections, and individually displays the following quantities:

- The first section: model R-squared, model adjusted R-squared, and standard error of the model.
- The second section: relevant coefficient standard errors with corresponding p values, which can be used for representing confidence intervals

Equation	Model Summary			Unstandardised Coefficients (C)-C1		Unstandardised Coefficients (C)-C2		Unstandardised Coefficients (C)-C3	
	R square	Adjusted R Square	Std. Error of the Estimate	Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
Linear	0.373	0.362	0.089	0.081	0.000	0.062	0.000	--	
Quadratic	0.378	0.355	0.090	0.699	0.985	0.485	0.523	0.245	0.013
S	0.295	0.282	0.128	0.051	0.000	0.073	0.062	--	--
Exponential	0.343	0.331	0.123	0.112	0.000	0.045	0.000	--	--
Logistic	0.343	0.331	0.123	0.062	0.000	0.164	0.000	--	--

Table 7.15 Curve standard error summary for v2 as a function of v1

To sum up the statistical results in the above two tables Table 7.14 and Table 7.15, the results show that all the curves have a *p value* = 0, which indicates that all curve results are significant. However, the linear model has the largest F. At the same time, linear model has the largest *adjusted R-squared* with the lowest *standard error*, and all coefficient standard errors are less than 0.1 with a *p value* = 0 in Table 7.15. Therefore, combining the *p value, F, R-squared* and *standard errors*, a linear model is chosen for the relationship between v1 and v2.

7.7.1.2 Model result summary and evaluation

After a linear relationship is chosen, the relationship model for v1 and v2 can be obtained using regression analysis. The regression equation is shown as below:

$$v2 = 0.461 * v1 + 0.481 \quad (7.1)$$

In addition to this equation, more detailed information regarding the model for v1 and v2 can be achieved to describe and evaluate the model fit:

1) Model type: linear model
2) R = 0.611; R-squared=0.373; Adjusted R-squared = 0.362;
   It means 37.3% of the variation in v2 "*communications between doctors and patients*" is explained by its relationship with v1 "*continuity of care*".
3) Standard error of the estimation: 0.0895

At the same time, an analysis of variance (ANOVA) is conducted to evaluate the model. It includes the "regression" that variance can be explained by the independent variables, and the "residual" that the variance is not explained by the independent variables. In the ANOVA result table, Table 7.16, the *sum of squares* is described first. The *total sum of squares* represents the total variation, which is the sum of the *regression sum of squares* and the *residual sum of squares*. The *sum of squares* provides a measure of variation from the mean, and furthermore, the *mean squares* can be obtained by *sum of squares* divided by the *degrees of freedoms (df)*. *F*, *the F-statistic*, is a ratio of mean squares, which is the *mean square of the regression* divided by the *mean square of the residual*. A high F-statistic indicates a significant effect. Also, "Sig.", p-value, is equal to zero, which is much smaller than 0.05. This indicates that the regression line is unlikely to have occurred by chance. In other words, the regression line is significantly better at predicting the dependent variable v2 from the factor v1 than using the mean of v2 every time. The *total sum of squares*, *mean square* and *F* can be represented respectively as below:

$$Total\ sum\ of\ squares = \sum(y - \bar{y})^2 = \sum(\hat{y} - \bar{y})^2 + \sum(y - \hat{y})^2,$$

where $\hat{y}$ is the predicted values and $\bar{y}$ is the mean value.

$$Mean\ square = \frac{sum\ of\ squares}{df}$$

$$F = \frac{mean\ square\ of\ the\ regression}{mean\ square\ of\ the\ residuals}$$

Model	Sum of Squares	df	Mean Square	F	Sig.
1Regression	0.257	1	0.257	32.163	0.000
Residual	0.432	54	0.008		
Total	0.690	55			

Table 7.16 ANOVA for the relational equation in v1 and v2

Figure 7.4 displays the scatterplot with the regression line. It reflects the statistical results above, and shows that there is a clear positive relationship between v1 and v2.

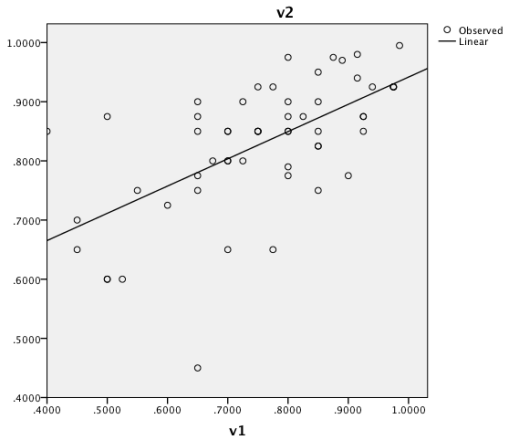


Figure 7.4 v1-v2 scatterplot with the regression line

Overall, the regression equation is summarised as below:

$$communications\ between\ doctors\ and\ patients = 0.461 * continuity\ of\ care + 0.481$$

with R^2 =0.373, adjusted R^2 = 0.362, F = 32.163, p= 0.00.

7.7.2 Relationship between v1 and v3

This subsection is to find the relationships between v1 and v3, where v1 is the factor and v3 is the dependent variable. Following the same procedure ass pointed out in the previous subsection, this subsection first estimates the best fit curve type, and then explains the model results.

7.7.2.1 Curve estimation

Five curves are compared and results are listed in Table 7.17 and Table 7.18. As shown in these two tables, linear, s, exponential and logistic models have *p values* less than 0.05.

Although the linear model has a lower R^2 and F than the rest of the models, it performs better in terms of standard error results for both model and coefficients. Considering that a simpler model is always preferred, the linear relationship is chosen to represent the association between v1 and v3.

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	0.119	7.327	1	54	0.009	0.480	0.393	
Quadratic	0.123	3.708	2	53	0.031	0.291	0.946	-0.386
S	0.123	7.598	1	54	0.008	0.103	-0.279	
Exponential	0.135	8.430	1	54	0.005	0.458	0.659	
Logistic	0.135	8.430	1	54	0.005	2.183	0.517	

Table 7.17 Curve estimation for v3 as a function of v1

Equation	Model Summary			Unstandardised Coefficients (C)-C1		Unstandardised Coefficients (C)-C2		Unstandardised Coefficients (C)-C3	
	R square	Adjusted R Square	Std. Error of the Estimate	Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
Linear	0.119	0.103	0.160	0.145	0.009	0.111	0.000	--	
Quadratic	0.123	0.090	0.161	1.252	0.453	0.869	0.658	0.439	0.511
S	0.123	0.107	0.252	0.101	0.008	0.145	0.477	--	--
Exponential	0.135	0.119	0.250	0.227	0.005	0.080	0.000	--	--
Logistic	0.135	0.119	0.250	0.117	0.000	0.380	0.000	--	--

Table 7.18 Curve standard error summary for v3 as a function of v1

7.7.2.2 Model result summary and evaluation

The linear regression is further adopted to provide details of the relationship model of v1 and v3. Also, ANOVA is performed to evaluate the regression model, and the corresponding result is shown in Table 7.19.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.187	1	0.187	7.327	0.009
Residual	1.379	54	0.026		
Total	1.566	55			

Table 7.19 ANOVA for the relational equation in v1 and v3

The regression equation for v3 on v1 is obtained as below:

$$v3 = 0.393 * v1 + 0.480 \quad (7.2)$$

which means:

$$Patient\ medical\ history = 0.393 * Continuity\ of\ care + 0.480$$

At the same time, the results provided the following information:

1) Model type: linear model. It indicates that v1 "continuity of care" positively linearly affects doctor's accessing to v3 "patient medical history".
2) R = 0.346; R-squared=0.119; Adjusted R-squared = 0.103.
   It means only 11.9% of the variation in v3 is explained by its relationship with v1. A low R-squared value also agrees with the fact that the relationship between v1 and v3 is not shown in the results from the previous qualitative CLD model.
3) Standard error of the estimation has the value 0.160, which is slightly higher than 0.1.
4) F = 7.327. The F in this model is lower, compared with the F result in the v1-v2 model.
5) p= 0.009, which is smaller than 0.05. The linear regression model is significant.

The scatterplot of sample data with the regression line is illustrated in Figure 7.5. It reflects the statistical results, and shows that there is linear relationship between v1 and v3. However, at the same time, the linear association is not very strong, since the sample data are spread widely around the line.

Overall, because v1 and v3 display a weak linear correlation and a low R^2, less than 15%, indicates a low percentage of variation explained by the relationship with v1, the relationship model of v1 and v3 is not selected for the quantitative system dynamics model, and v1 is not used for the prediction of v3 in the system dynamics model.

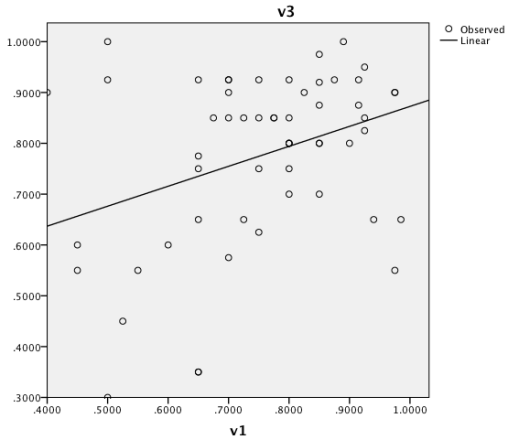


Figure 7.5 v1-v3 scatterplot with the regression line

7.7.3 Relationship between v1 and v4

This subsection determines the relationship model of v1 and v4, where v1 is the factor and v4 is the dependent variable. Similarly to the relationship between v1 and v3, the relationship between v1 and v4 is not apparent from the previous literature review or clinician feedback, but the correlation analysis indicates that the two do have an association. This subsection first finds which curve is the best to fit the relationship, and then model results are described.

7.7.3.1 Curve estimation

The model results for fives curves are described in Table 7.20 and Table 7.21. Results in Table 7.20 show that all five models are significant and have *p values* much less than 0.05. At the same time, the linear model, exponential model and logistic model have higher R squared value and F value, compared with the quadratic model and s model. Moreover, the results in Table 7.21 displays that the linear model has the lowest value for model standard error which is 0.112. Although not much difference among these results is observed, still, results show that the linear model has a better performance in terms of the coefficient standard errors that are 0.102 and 0.078 respectively with both p values being zero. Overall, the linear model is selected to express the relationship between v1 and v4.

Equation	Model Summary					Parameter Estimates		
	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	0.221	15.277	1	54	0.000	0.452	0.398	
Quadratic	0.225	7.695	2	53	0.001	0.618	-0.086	0.338
S	0.205	13.951	1	54	0.000	0.043	-0.247	
Exponential	0.237	16.795	1	54	0.000	0.472	0.599	
Logistic	0.237	16.795	1	54	0.000	2.119	0.549	

Table 7.20 Curve estimation for v4 as a function of v1

Equation	Model Summary			Unstandardised Coefficients (C)-C1		Unstandardised Coefficients (C)-C2		Unstandardised Coefficients (C)-C3	
	R square	Adjusted R Square	Std. Error of the Estimate	Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
Linear	0.221	0.206	0.112	0.102	0.000	0.078	0.000	--	
Quadratic	0.225	0.196	0.113	0.878	0.922	0.609	0.581	0.308	0.050
S	0.205	0.191	0.164	0.066	0.000	0.094	0.650	--	--
Exponential	0.237	0.223	0.161	0.146	0.000	0.053	0.000	--	--
Logistic	0.237	0.223	0.161	0.080	0.000	0.238	0.000	--	--

Table 7.21 Curve standard error summary for v4 as a function of v1

7.7.3.2 Model result summary and evaluation

The details of the linear model are discussed and the model is further analysed using ANOVA. The ANOVA results are illustrated in Table 7.22, where the independent variable is v1.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.192	1	0.192	15.277	0.000
Residual	0.680	54	0.013		
Total	0.873	55			

Table 7.22 ANOVA for the relational equation in v1 and v4

The linear regression model obtained is as below:

$$v4 = 0.398 * v1 + 0.452 \quad (3.3)$$

which represents:

$$Clarity\ of\ disease\ symptoms\ or\ signs = 0.398 * Continuity\ of\ care + 0.452$$

Additionally, the following quantities help understand and evaluate the obtained linear model:

1) Model type: linear model. It indicates that v1 "*continuity of care*" positively linearly affects doctors' findings on v4 "*clarity of disease symptoms/signs*".
2) R = 0.470; R-squared=0.221; Adjusted R-squared = 0.206.
   It explains that 22.1% of the variation in v4 "*clarity of disease symptoms/signs*" is explained by its relationship with v1 "*continuity of care*".
3) Standard error of the estimation: 0.112. It is slightly higher than 0.1.
4) F = 15.277.
5) p= 0.00, and the linear regression model is significant.

The scatterplot and the regression model are displayed in Figure 7.6. It demonstrates the same results with the above statistical results. As seen from the figure, there is a linear association between v1 and v4, although the association is not very strong. Still, according to the overall model results, it is reasonable to add the model for v1 and v4 into the quantitative system dynamics model.

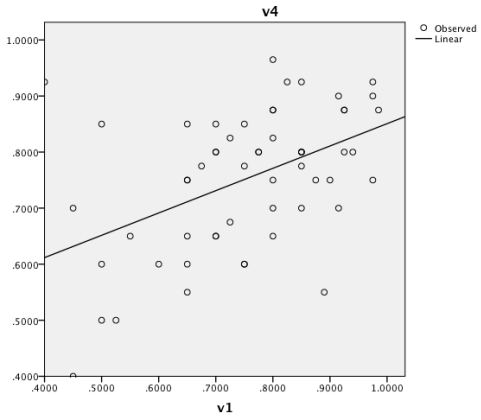


Figure 7.6 v1-v4 scatterplot with the regression line

7.7.4 Relationship between y1 and y2

Previous analysis indicates that there is an association between y1 "*missed or wrong diagnostic clues from phase1*" and the number of y2 "*missed or wrongly hypothesised diagnosis*", where the value of y1 is the number of cases with "*missed or wrong diagnostic clues from phase1*" and the value of y2 is the number of cases with "*missed or wrongly hypothesised diagnosis*". This section determines the relational equation for independent variable y1 and dependent variable y2.

7.7.4.1 Curve estimation:

Fives curves are applied to the sample data, and results are compared to choose the best fit curve. Table 7.23 and Table 7.24 display the statistical results of the five curves. From the results in Table 7.23, all five curves show their significance with p values much less than 0.05. At the same time, the linear model, exponential model and logistic model have higher R-squared value and F value. Table 7.24 further shows that the linear model has the lowest model standard error whose value is 0.1, and both of its two coefficients are significant whose standard errors are 0.125 and 0.031 respectively. Combining the results from both tables, the linear model is selected to represent the relationship of y1 and y2.

	Model Summary					Parameter Estimates		
Equation	R Square	F	df1	df2	Sig.	Constant	b1	b2
Linear	0.214	14.661	1	54	0.000	0.091	0.478	
Quadratic	0.215	7.247	2	53	0.002	0.076	0.619	-0.265
S	0.139	8.748	1	54	0.005	-1.398	-0.066	
Exponential	0.210	14.328	1	54	0.000	0.098	2.432	
Logistic	0.210	14.328	1	54	0.000	10.188	0.088	

Table 7.23 Curve estimation for y2 as a function of y1

Equation	Model Summary			Unstandardised Coefficients (C)-C1		Unstandardised Coefficients (C)-C2		Unstandardised Coefficients (C)-C3	
	R square	Adjusted R Square	Std. Error of the Estimate	Std. Error	Sig.	Std. Error	Sig.	Std. Error	Sig.
Linear	0.214	0.199	0.100	0.125	0.000	0.031	0.004	--	
Quadratic	0.215	0.185	0.101	0.509	0.229	0.926	0.776	0.061	0.219
S	0.139	0.123	0.539	0.022	0.005	0.150	0.000	--	--
Exponential	0.210	0.195	0.517	0.642	0.000	0.015	0.000	--	--
Logistic	0.210	0.195	0.517	0.056	0.125	1.602	0.000	--	--

Table 7.24 Curve standard error summary for y2 as a function of y1

7.7.4.2 Model result summary and evaluation

After the model type is chosen, the details of the model are further analysed. Also, the ANOVA is conducted and results are shown in Table 7.25.

	Sum of Squares	df	Mean Square	F	Sig.
Regression	0.148	1	0.148	14.661	0.000
Residual	0.545	54	0.010		
Total	0.693	55			

Table 7.25 ANOVA for the relational equation in y1 and y2

The relationship equation for y1 and y2 is obtained as below:

$$y2 = 0.478 * y1 + 0.091 \quad (7.4)$$

which means:

$$missed\ or\ wrong\ hypothesized\ diagnosis = 0.468 * missed\ or\ wrong\ diagnostic\ clues\ from\ phase1 + 0.091$$

Moreover, the model also shows the following information:

1) Model type: linear model. The model explains that there is a positive linear relationship between y1, that is the number of cases with "*missed or wrong diagnostic clues from phase1*", and y2, that is the number of cases with "*missed or wrongly hypothesised diagnosis*".
2) R = 0.462; R-squared=0.214; Adjusted R-squared = 0.199; It shows that there is 21.4% of the variation in y2 "*missed or wrongly hypothesised diagnosis*" explained by its relationship with y1 "*missed or wrong diagnostic clues from phase1"*.
3) Standard error of the estimation: 0.100.
4) F = 14.661.
5) p= 0.00, and the regression model is significant.

From the previous literature review, there is also another factor affecting the value of y2, which is the "*history and physical examination sensitivity*". Although this factor may also vary in different situations, it is taken as a constant during the study. This is because all sample data are collected from one area, and one research investigation in 2008 [157] shows that this factor is a constant whose value may lie around 66%. Therefore, this study considers the "*history and physical examination sensitivity*" as a constant, and constructs the y1-y2 relationship model with y1 as the factor. At the same time, this assumption may explain the slightly low R^2 value.

The scatterplot Figure 7.7 also shows that y2 and y1 remain as a linear relation, and at the same time, the data are spread widely around the regression line in the chart.

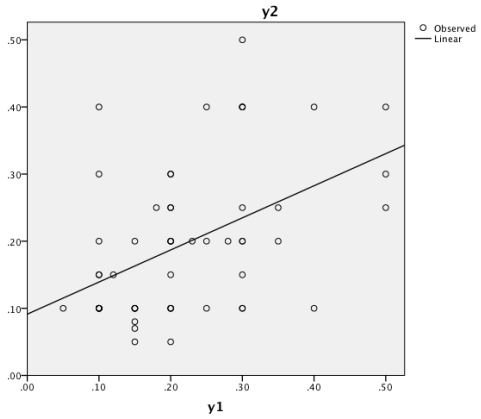


Figure 7.7 y1-y2 scatterplot with the regression line

7.8 Multiple regression analysis of the relationship between y1 and v2,v3,v4

Multiple regression analysis is conducted in this section, considering that there are more than one factor involved. The intention is to identify the relationship between y1 and its relevant factors v2, v3, v4, and determine the quantitative relationship function for the system dynamics model.

Generally speaking, multiple regression analysis starts with the estimation of a linear model format, and the relationship can be represented as:

$$y_1 = b_0 + b_1 * v_2 + b_2 * v_3 + b_3 * v_4$$

where b_0, b_1, b_2, b_3 are the coefficients to be obtained. The individual coefficient represents the independent contribution of the corresponding factor to the dependent variable y1. The correlation between an individual factor and the dependent variable is referred to as a partial correlation, because the correlation is observed with the condition of controlling the remaining two factors.

Although real practical data may show some deviation from a linear model type, a multiple linear regression model is still preferred since the procedures of the regression analysis are not greatly affected by minor deviations from this assumption.[158] However, when the data show a significant large deviation from a linear regression model, it is considered that transformation of the variables is required to allow for nonlinear components in the model. Therefore, the following subsection conducts a bivariate analysis between the individual factor and the dependent variable to observe whether there is a significant deviation from

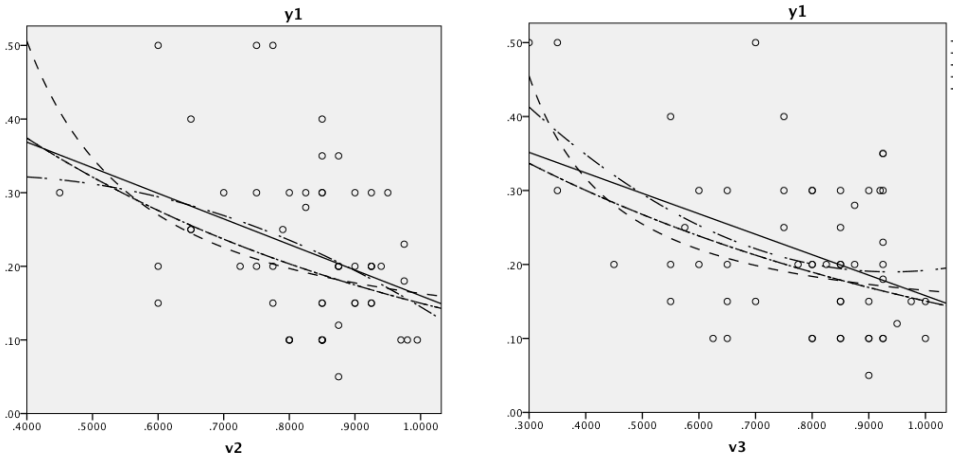
the linear model.

7.8.1 Bivariate scatterplot

Bivariate analysis is exploited to observe the relationship between an individual factor and dependent factor y1. A simple way of observing the relationship is using bivariate scatterplots. Figure 7.8 displays the graphs for y1 on v2, y1 on v3, y1 on v4 separately, together with graphs illustrating the bivariate scatterplots with five different regression curves, including linear, quadratic, s, exponential, and logistic curves. As seen in Figure 7.8, the linear line does not show significant deviation from the remaining four different curves, although there is substantial variation between the sample data and each regression curve. In addition, statistical results further support the same conclusion and show that the linear relationship is the best model in the bivariate analysis, whose details are listed below:

- Bivariate analysis of the relationship between y1 and v2:
  Linear model has the lowest p value=0.007 and the largest F=7.98. Also, it has the second largest R^2=0.127 and the largest adjusted R^2=0.113 considering the degrees of freedom.
- Bivariate analysis of the relationship between y1 and v3:
  Linear model has the lowest p value=0.001, the largest F=12.286, the second largest R^2=0.185, and the second largest adjusted R^2=0.170
- Bivariate analysis of the relationship between y1 and v4:
  Linear model has the lowest p value=0.003, the largest F=9.431, the second largest R^2=0.149, and the largest adjusted R^2=0.133

Overall, it is reasonable to employ a multiple linear regression model to represent the relationship between y1 and its factors v2, v3, v4.



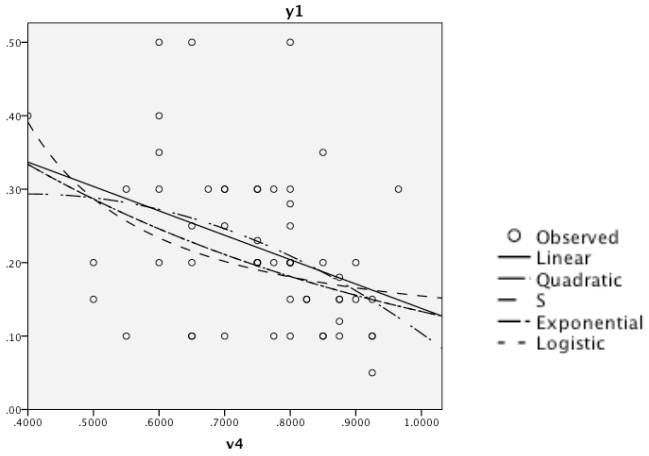


Figure 7.8 Bivariate scatterplots for y1 on its individual factor

7.8.2 Multiple linear regression results

Following the multi-variable linear function format, regression analysis is conducted to obtain the function coefficients. The regression results are shown below, where the dependent variable is y1 and the factors are v2, v3, v4:

$$y_1 = 0.566 - 0.036 * v_2 - 0.204 * v_3 - 0.210 * v_4 \quad (7.5)$$

The relational equation stands for:

$$\begin{aligned} &Missed\ or\ wrong\ diagnostic\ clues\ from\ phase1 \\ &= 0.566 - 0.036 * "Communications\ between\ doctors\ and\ patients" \\ &- 0.204 * "Patient\ medical\ history" - 0.210 \\ &* "Clarity\ of\ disease\ symptoms/signs" \end{aligned}$$

The model also shows:

1) All three factors v2, v3, v4 have a negative association with y1.

2) The model has R = 0.494; R-squared=0.244; Adjusted R-squared = 0.201; and 24.4% of the variation in y1 can be explained by its relationship with the three factors v2, v3, v4.

3) Standard error of the estimation = 0.09698 which is lower than 0.1.

4) p= 0.02 shows that the above multiple linear regression model is significant.

Residual statistical results for y1 as a function of v2, v3 and v4 shown in Table 7.26 list the residual values which are the deviations between the predicted values from the regression line and the observed values from sample data.

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	0.1547	0.3629	0.2198	0.05360	56
Residual	-0.14761	0.27185	0.00000	0.09430	56
Std. Predicted Value	-1.214	2.670	0.000	1.000	56
Std. Residual	-1.522	2.803	0.000	0.972	56

Table 7.26 Residuals statistics for y1 as a function of v2, v3 and v4

At the same time the ANOVA is conducted, and coefficients are further analysed. ANOVA results in Table 7.27 show that the model is significant but F=5.600 which is not a high value. Table 7.28 shows the details of the coefficients. All standard errors of coefficients are lower than 0.2, and in particular, the coefficient of v2 has the highest standard error =0.162. At the same time, v2 also has a very high p value which is 0.825. Moreover, v4 also has a high p value = 0.110. The reason that the v2 and v4 have high p values may lie in correlations between the factors. Thus, Pearson correlation analysis results are checked again to observe whether these correlations exist. Results in Table 7.29 confirm that v2 and v3, v2 and v4 have strong correlations, also, v3 and v4 are correlated. Therefore, Eq.(7.5) represents the relationship between y and v2, v3, v4, and a high p value of one coefficient is observed as the result of correlations between the variables.

Model	Sum of Squares	df	Mean Square	F	Sig.
1Regression	.158	3	.053	5.600	.002[b]
Residual	.489	52	.009		
Total	.647	55			

Table 7.27. ANOVA for the relational equation in y1 and v2, v3, v4

Model	Unstandardised Coefficients		Sig.
	B	Std. Error	
1(Constant)	.566	.101	.000
v2	-.036	.162	.825
v3	-.204	.094	.034
v4	-.210	.129	.110

Table 7.28. Coefficients analysis

		v2	v3	v4
v2	Pearson Correlation	1	.559**	.591**
	Sig. (2-tailed)		.000	.000
	N	56	56	56
v3	Pearson Correlation	.559**	1	.377**
	Sig. (2-tailed)	.000		.004
	N	56	56	56
v4	Pearson Correlation	.591**	.377**	1
	Sig. (2-tailed)	.000	.004	
	N	56	56	56

Table 7.29. Pearson correlation anlysis for v2, v3 and v4

7.9 Principal components analysis

Principal components analysis (PCA) is widely applied in the situation in which there are a large number of variables and the variance structure of these variables are to be discovered. PAC introduces new combinations of original variables as new components, and represents the maximum amount of variance using fewer new components. By interpreting the new components, the variance structure of these variables can be discovered. Furthermore, the number of variables as well as the correlation effect in the model can be reduced. This section applies PCA to the factors v2, v3 and v4, since the results from the last subsection show that the factors v2, v3 and v4 appear correlations and the correlation effect to the equation model is to be reduced. Although PCA is commonly used in cases with a large number of factors rather than just three factors, the experiment in this section is essentially a pilot and aims to demonstrate its utility in analysing the variance structure and helping with the factor correlation effect in situations where there is a large number of variables.

7.9.1 PCA introduction

PCA discovers the variance structure of a group of variables using linear combinations of these variables, and explains the maximum amount of variance with the fewest number of principal components.[159]

During the PCA procedure, a set of linearly correlated variables is converted into a set of new components that are linearly uncorrelated. To obtain the new uncorrelated components, PCA uses new components to represent original variables, by employing the following method:

$$\begin{cases} F_1 = a_{11}X_1 + a_{21}X_2 + \cdots a_{p1}X_p \\ F_2 = a_{12}X_1 + a_{22}X_2 + \cdots a_{p2}X_p \\ \cdots\cdots \\ F_p = a_{1p}X_1 + a_{2p}X_2 + \cdots a_{pp}X_p \end{cases}$$

where F_1, F_2 … F_p are the new uncorrelated components, X_1, X_2…X_p are the original factors represented using normalised data, and a_{11}…a_{pp} are the coefficients.

Thus, the first step of PCA is data preparation by normalizing the original raw data. Different variables may have different value ranges. For example, if one variable lies between 0 and 1, and another variable lies between 100 and 1000, then the variation value between any two samples in the first variable set always appears to be smaller than the value in the second one. The two variables cannot be compared directly. Therefore, normalisation is conducted to reduce such effect of different value ranges.

Then, the new components can be obtained, via determining eigenvectors with eigenvalues "λ". An eigenvector indicates an axis with direction. PCA changes the original axes, and selects the "major axis of variation" as the eigenvectors or the new axis. In other words, an eigenvector or the axis is selected to keep the minimum of the sum of squares of distances to the axis. At the same time, the second eigenvector is chosen in the orthogonal direction of the first eigenvector to avoid correlations, and keep the second minimum of the sum of squares of distances. Considering that the eigenvectors are selected from a multidimensional space of *p* dimensions, p eigenvectors are generated. At the same time, each eigenvector has a value "λ" named as eigenvalues. This value represents how much variance there is in the direction of the corresponding eigenvector. In other words, eigenvalues indicate how spread out the data are on the eigenvector line. Combining the eigenvectors and eigenvalues, new components F_1, F_2 … F_p are generated. For example, {a_{11}, a_{21}, … a_{p1}} is the first eigenvector. The eigenvector with the highest eigenvalue, which indicates it has the largest variance in this eigenvector direction, is taken as the Principal component of the data set. Figure 7.9 is an example and explains two principal components in two dimensions. The orange colour vector represents the 1st principal component with the largest variance, and the blue vector represents the 2nd Principal component.

To sum up, PCA uses eigenvectors to represent orthogonal direction in multiple dimensions, which solve the problem of linear correlations. Moreover, it uses eigenvalues to represent

the value of variance in the direction of the corresponding eigenvector. A larger eigenvalue means larger variance.

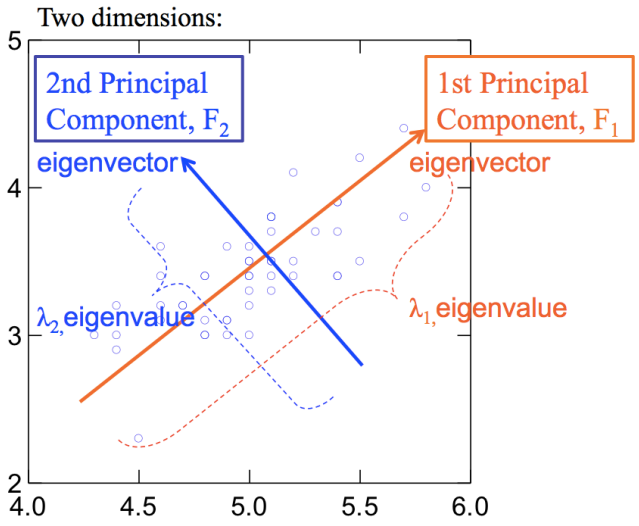


Figure 7.9 Finding principal components in PCA

7.9.2 PCA process

- Step 1. Data preparation

The first step is preparing the data. PCA analyses the normalised data instead of the original raw data of the factors, thus, the original data are required to be normalised or standardised, using the mean and standard deviation of original data.

Table 7.30 shows descriptive statistical results of v2, v3, and v4. It lists the mean value and standard deviation of the three factors. The new data, which are the normalised data, are generated according to the function below:

$$v' = \frac{v - Mean}{Std.Deviation}$$

where v are the original data and v' are the normalised new data.

	Mean	Std. Deviation	Analysis N
v2	0.828125	0.1119844	56
v3	0.775804	0.1687380	56
v4	0.752500	0.1259690	56

Table 7.30 Descriptive statistics of v2, v3, v4

- Step 2. Identify the number of principal components

This step is to determine the number of principal components from the normalised data. It first analyses the variance of three new components. Table 7.31 lists the eigenvalues of the components in descending order. The first component has the highest eigenvalue, which means the largest variance. Thus, it contributes the largest part, which is 67.44%, to the total variance. The larger the variance of the component is, the more information that component carries. Then, there is a significant decrease on the second component, which is 20.801% total variance accounted for by this component. The third contributing 11.760% to the total variance contains the least information.

Component	Eigenvalues	% of Variance	Cumulative %
1	2.023	67.440	67.440
2	.624	20.801	88.240
3	.353	11.760	100.000

Table 7.31 Component variance analysis

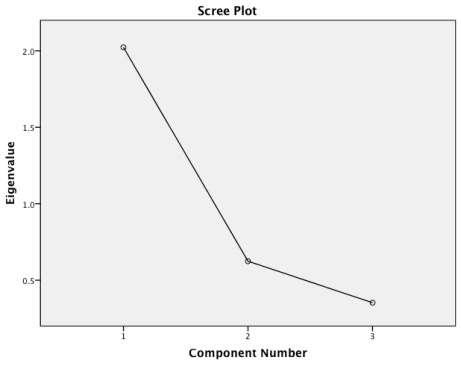


Figure 7.10 Eigenvalue graph

Generally speaking, an eigenvalue is assigned a boundary value to determine the number of principal components. Any component with its variance larger than the boundary eigenvalue is extracted as the Principal component. The number of components with eigenvalue larger than the boundary value is referred to as the number of latent roots. The setting of the boundary eigenvalue also requires considering the cumulative percentages, although the boundary value is taken as $\lambda=1$ in many cases.

As seen from Figure 7.10, one component should be extracted when the boundary eigenvalue is 1. However, the cumulative percentage shows that one component only cover 67.44% of the total variance value. At the same time, this study has three factors only, and the eigenvalue of the 2nd component is still significantly higher than that of the 3rd component, although it is much lower than that of the 1st component. Overall, considering

the eigenvalue and the cumulative percentage, first two components marked in the blue box in Table 7.31 should be extracted, and the cumulative variance percentage is up to 88.24% in this case.

7.9.3 Component rotation:

- Why to use rotation

After the new components have been chosen, they can be represented by the intersection of the original factors/features, as shown below:

$$\text{New component}_1 = a_1 * \text{original factor}_1 + a_2 * \text{original factor}_2 + \cdots + a_p * \text{original factor}_p$$

where $a_1, a_2 \ldots a_p$ are the coefficients or the eigenvectors of the new components. The coefficients identify the relative weight of each variable in the component,[159] and the component scores can be obtained by giving the coefficient matrix. The individual correlation between the new component and the factor is referred to as the "loading" of the component. Component loadings are the coefficients of linear combinations of components to predict variables. There is a relationship between the component score coefficients and component loadings, which is:

$$\text{coefficient matrix} = L\,(L^T L)^{-1}$$

where L is the matrix of component loadings.[159]

Also, squared loadings indicate that the percentage of the variance in the original variable is explained by the new component, and the sum of squared loadings of a component is its eigenvalue.[160] In other words, loadings indicate how much an original factor can be explained by the new component.

After the step above, it can be shown that fewer new components are used to represent the original factors. The new components can be understood as higher-level classifications of the original factors. For example, if a new component is related to a large loading of "electronic patient records" and a large loading of "advanced equipment", then the new component could indicate the clinical IT level.

However, the new components may not be represented by significant loadings at the current stage, which makes it difficult to interpret or explain the new components. Thus,

rotations are performed here. It is to maximize the loading of a variable on one component while minimising its loading on all other factors, so that components can be more meaningful and can be explained more simply.

To achieve the aims above, the axes of the factors, which are the components, are rotated. The purpose of rotating axes of the factors is to make the clusters of previous factors align as closely as possible to the axes or the component lines. It also indicates the reason why PCA is applicable to a large number of factors.

- Two ways of rotation

There are two approaches to rotation: orthogonal rotation and oblique rotation. Both rotation methods are illustrated in Figure 7.11. In the example in Figure 7.11, blue points and green points form two clusters, and stand for two clusters of previous factors. Before any rotation, the axes of the factors are the original black lines, which indicate the new components “component 1” and “component 2”. Orthogonal rotation means that the axes rotate while they are kept orthogonal, and oblique rotation allows two axes to rotate in different directions. However, both rotation methods make the two clusters of items fall as closely as possible to the axes, so that the components can be represented using factors with significantly different loadings.

Using the example in Figure 7.11, before the rotation, components are related with both factors in the blue cluster and factors in the green cluster. After a rotation, which is either the orthogonal rotation or the oblique rotation, the component1 line goes through the factors in the green cluster in order to fall as closely as possible with the axis line of component1. Due to this rotation, component 1 has a weak effect on the factors in the green cluster. In other words, component1 loads least to the factors in the green cluster. Similarly, component2 is rotated to go through the blue factor cluster, which makes the factors in the blue cluster are minimally loaded on to component 2. Therefore, different loadings help to interpret the extracted components. For example, new component1 and new component2 can be mainly interpreted as a new feature by the factors in the blue cluster and in the green cluster respectively.

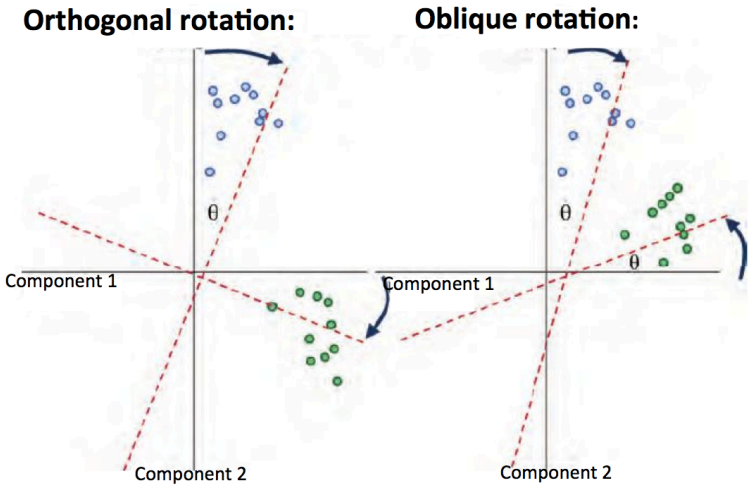


Figure 7.11 Orthogonal rotation and oblique rotation

- Choosing the rotation type

As seen in Figure 7.11, two rotation methods result in the new components having different characteristics, which indicates that two rotation methods may be appropriate for different applications.

The main difference is set out below:

- Orthogonal rotation: The new components after the rotation, sometimes refered to as new factors, are kept orthogonal, which means the new components are uncorrelated. Thus, the loadings and the correlations between the variables and factors are the same.
- Oblique rotations: The new components are not orthogonal, which indicates the components are allowed to correlate. Thus, the loadings and correlations are different and described in two different tables.

Generally speaking, an orthogonal method might be preferred to an oblique method[161].

- Results of the PCA experiment

After determining two components to be extracted by eigenvalues, the loading values before any rotation are described in Table 7.32. Table 7.32 is a matrix showing the component underlying construction or the loading values before the rotation. As seen in the table, component 1 largely loads all three factors v2, v3 and v4, while loadings of component 2 display significantly less than loadings of component 1. Component 2 has a very small loading from v2, which is -0.036, and some loadings from v3 and v4, which are

0.585 and -0.530 respectively. It is very difficult to explain the meanings of the two components. Thus, rotations are carried out in this situation to generate two new rotated components.

	Component	
	1	2
v2	0.885	-0.036
v3	0.777	0.585
v4	0.798	-0.530

Table 7.32 Component pattern matrix before rotation

In this study, both rotation methods were tested and compared. The *direct oblimin rotation*[162] method is chosen to perform an oblique rotation, and the *varimax* rotation[162] method is conducted to present an orthogonal rotation.

- Oblique rotation (direct oblimin rotation) results

Oblique rotation allows new component lines to be un-orthogonal, which indicates the new components may be correlated. Thus, for a clear description of the new components, the loading values and the correlations between original factors and new components are displayed in two separate tables. Loading results are shown in Table 7.33, which are significantly different from the results in Table 7.32. Loadings are the linear regression coefficients, thus the extent to which the two reproduced components explain the original factor can be interpreted as:

$$v_4 = 0.993 * component\ 1 - 0.089\ component\ 2$$

$$v_2 = 0.603 * component\ 1 + 0.438\ component\ 2$$

$$v_3 = -0.027 * component\ 1 + 0.984\ component\ 2$$

where v4, v2 and v3 are standardised original observed factors. v4 "*Clarity of disease symptoms/signs*" shows the largest loading to component1, while v3 "*Patient medical history*" is least related to component1. Component2 loads most on v3 "*Patient medical history*".

	Component	
	1	2
v4	0.993	-0.089
v2	0.603	0.438
v3	-0.027	0.984

Table 7.33 Pattern/loading matrix - Oblimin with Kaiser normalization

Table 7.34 is the structure matrix, which shows the results of the correlations between the original observed factors and the new components. Component1 shows a close correlation with v4 and Component2 is observed to have the closest relationship with v3.

	Component	
	1	2
v4	0.954	0.341
v2	0.793	0.699
v3	0.399	0.972

Table 7.34 Structure/correlation matrix –Oblimin with Kaiser normalization

- Orthogonal rotation (varimax rotation) results:

Orthogonal rotation assumes that the new components are orthogonal or un-correlated. Thus, the pattern matrix and the structure matrix are the same and the results are shown in Table 7.35. Table 7.35 lists the component loadings, which are also the correlations between the variable and the component. The original factor can be explained by two reproduced components after Varimax rotation as follows:

$$v_4 = 0.947 * \text{component } 1 + 0.142 * \text{component } 2$$

$$v_2 = 0.680 * \text{component } 1 + 0.567 * \text{component } 2$$

$$v_3 = 0.184 * \text{component } 1 + 0.955 * \text{component } 2$$

The result shows that the reproduced component1 is correlated with v4 and v2, and not significantly correlated with v3. Component 2 is correlated with v3 and partial v2, and it has a low correlation with v4.

	Component	
	1	2
v4	0.947	0.142
v2	0.680	0.567
v3	0.184	0.955

Table 7.35 Component/loading matrix - Varimax with Kaiser normalization

7.9.4 Regression model after PCA

This subsection sets out to discover whether y1 can be better predicted using the new components instead of the original v2, v3 and v4.

1) Components from Oblimin rotation and Varimax rotation

Previously, the relational equation between y1 and its three factors v2, v3, v4 has been identified. This step is to conduct experiments and find out the relations between y1 and the new components generated from v2, v3 and v4. Two different rotation methods produce two different sets of new components. Thus, the regression models are analysed separately.

Again, five curves, including linear, quadratic, s, exponential, and logistic curves, are tested for bivariate analysis of relationship between the individual component and dependent factor y1. Bivariate scatterplots show no significant variation between the linear line and the other four curves. Thus, multiple linear regression is carried out, and Table 7.36 compares the model results of different rotation methods. Varimax rotation shows a better performance in using its two components to predict y1, because all its coefficients have a p value lower than 0.05.

	Model summary				**Coefficient summary**			
	R^2	**Std. Error**	**F**	**Sig.**	**Coefficient**	**B**	**Std. Error**	**Sig.**
Oblimin rotation	0.228	0.097	7.829	0.001	(Constant)	0.220	0.013	0.000
					FAC1_oblimin	-0.028	0.015	0.061
					FAC2_oblimin	-0.033	0.015	0.026
Varimax rotation	0.228	0.097	7.829	0.001	(Constant)	0.220	0.013	0.000
					FAC1_varimax	-0.034	0.013	0.012
					FAC2_varimax	-0.039	0.013	0.004

Table 7.36 Regression model of y1 and new components

2) Compare the model before and after PCA:

The regression model that is obtained after PCA with Varimax rotation is compared with the original regression model for y1 and v2,v3,v4. Table 7.37 lists the comparison results. Both models are significant. R^2 and adjusted R^2 are reduced after PCA, because original factors v2, v3, v4 carry more information than the two new components. The variance is lost during the process of choosing the number of components, and the two components bring 88.24% of the total variance from the original data set. Although more information is lost, the F value is improved. Moreover, the coefficients in the after-PCA model show a significant improvement in both standard errors and p values. It confirms that the newly reproduced components are not correlated and their coefficients are significant in the regression model.

	Model Summary					**Coefficients Summary**		
	R square	Adjusted R Square	Std. Error of the Estimate	F	Sig.	Coefficients list	Standard error	P value
Before-PCA model: y1 and v2,v3,v4	0.244	0.201	0.09698	5.600	0.02	(Constant)	0.101	0.000
						V2	0.162	0.825
						V3	0.094	0.034
						V4	0.129	0.110
After-PCA model: y1 and 2 new components	0.228	0.199	0.09708	7.829	0.01	(Constant)	0.013	0.000
						Component1	0.013	0.012
						Component2	0.013	0.004

Table 7.37 Comparison of two regression models

Overall, compared with the before-PCA model that is for y1 and v2,v3,v4, the after-PCA model, which is for y1 and two new components, shows improvement in model coefficient significance, but loses variance during the process of PCA. Since the PCA is more suitable for a larger number of variables, the results of the new model will not be implemented in the quantitative model.

7.10 Discussion

This chapter first quantifies the non-numerical factors, and implements regression analysis, and then successfully determines the relevant relationship functions.

Since regression analysis relies on the data, the model results depend largely on the quantity and quality of the data. It is observed that there are substantial variations in the observed points or sample data around the fitted regression lines in the scatterplots. Because of the limitations in the data sources, data in the study are collected from questionnaires and the data are subjective in terms of the doctors' estimations. Also, the size of the dataset could be larger so that these could have been better coverage for all three levels of doctors, and different evaluation methods, such as cross validation, could have been implemented. However, several methods have been exploited in the study in order that the data collected might be as objective as possible. For example, the questionnaires are anonymous during the entire process and the questionnaire document substitutes all subjective words, such as "good" or "poor", by "group 1 ", "group 2" or "group 3". Moreover, the data used in the study could be found in better quality from other sources such as patient medical records or by recording patients' visits over a period.

PCA has also been implemented to demonstrate that it helps to understand the variance structure of factors and helps with factor correlations especially for a group of large quantity of factors, although the new components produced from PCA are not adopted in the quantitative system dynamics model.

7.11 Summary

Overall, this chapter successfully quantifies the non-numerical variables and carries out data collection from questionnaires for regression analysis. Finally, it identifies the required relationship functions using regression analysis and relationship functions are evaluated at the end of each section. The relationship functions obtained are used as the equations in the quantitative system dynamics model. The next chapter will introduce the simulation data, summarise model equations, and then conduct simulation experiments under different scenarios.

Chapter 8 Model Simulation

8.1 Introduction

This chapter commences with presenting data collection for simulation. In particular, an expert elicitation study is conducted to collect data of the estimates from clinicians. Then, it provides a summary of assumptions, equations and data used in the simulation studies. Finally, simulation experiments are carried out. System behaviours are observed under different scenarios. Parts of the simulation experiments can also be found in paper 163.

8.2 Data for model simulation

Simulation requires reliable data to be available. The quantitative model covers a variety of key factors, constants and other variables, and thus it requires a large data range for the wide range of system elements. At the same time, Brailsford in 2008 [164] points out that system dynamics modelling does not depend on large quantities of high-quality data, and its data requirement is generally weaker than that of other discrete-event simulations. It is usually higher-level and more aggregated than other models, and it can still illustrate the outputs based on highly simplified data.[164] Overall, the quantitative model needs a wide range of data for the simulation, but it has weaker data requirements in terms of data quantities and quality.

There is no available data source that covers all the variables of the system. Thus, data for different variables are collected from different sources in different ways. Overall, two main sources are used in this thesis: public data and estimated data from experts.

8.2.1 Public data

Public data are the published literature from relevant researches or published information from relevant societies. Published literature of relevant researches can be conveniently retrieved and easily accessed. Generally, it focuses on one or two factors, and provides output data from experiments or review analysis. Thus, this source provides high-level processed data, and it has the advantage of easy adoption and is time saving.

There are still several societies that focus on patient safety and diagnostic errors. They usually have richer data sources, and data are usually mixed with statistical data and higher-level findings from data. However, it has also the drawback that the data are much more

specific and it is more difficult to search out relevant data from a large quantity of data sheets. Relevant societies were found as listed below:

1) CRICO[165], owned by and serving the Harvard medical community in evidence-based risk management;
2) The commonwealth fund[166], working towards a high performance health system;
3) Agency for healthcare research and quality[167];
4) National information centres: such as NHS(*National Health Service)* information centre and clinical trials[168]; CIHI(*Canadian Institute for Health Information)*[169]*.*

Literature reviews for public data are conducted in this step. To provide more specific data, an individual literature review is performed for each phase of the model, following the model structure described in Chapter 6. Relevant literature published from 1994 to 2014 is retrieved and reviewed from relevant databases such as PubMed and EMbase and other relevant society databases.

Results show that few data were found in the society databases, and compared with other phases, a larger quantity of data evidence about laboratory tests was found. In detail, the data relating to the following variables are found from the literature, listed as below:

1) History and physical examinations:

The history and physical examination sensitivity is around 66% according to the relevant document[157].

2) Radiology tests:

The blended error rate for a wide range of modalities is 4.4%, with a possible range of errors between 0.8% and 9.2% depending on the type of studies interpreted, modality mix and subspecialty expertise of the radiologist. [170]

3) Laboratory tests:

Most studies discuss relevant laboratory errors under three headings: *pre-analytical errors, analytical errors,* and *post-analytical* errors. Although differences in laboratory error rate among study areas were witnessed in this study, results show *pre-analytical errors* take the largest percentage, 55%-77% for a 60% likelihood, of the laboratory errors in all areas, compared with the other two types: *analytical errors* and *post-analytical errors*. A relevant study[172] also describes the top three causes of pre-analytical errors as: tube filling error

(13.1% of the total pre-analytical errors), patient ID error (8.8%), and inappropriate container (8.1%). Table 8.1 summarises the relevant literature data for laboratory tests.

Papers	Year	Study Area	Laboratory test error rate	Pre-analytical error rate (ppm[b])	Analytical error rate (ppm[b])	Post-analytical error rate (ppm[b])
Abdollahi et al [171]	2014	Iran	6.30%	41007	14616	7358
Carraro & Plebani [172]	2007	Italy	0.31%	1914	463	715
Wiwanitkit [173]	2001	Thailand	ND[a]	1100	58	147
Stahla et al [174]	1998	Germany	0.61%	4575	976	549
Plebani & Carraro[175]	1997	Italy	0.47%	3183	621	863
Nutting et al [176]	1996	North America	0.11%	612	146	330
Lapworth & Teal [177]	1994	UK	0.05%	158	158	154

[a] *ND: Not identified;* [b] *ppm: parts per million*

Table 8.1 Data for laboratory test errors from literature

The simulation data on the three variables *pre-analytical errors, analytical errors,* and *post-analytical* errors are randomly selected in the circa 70% likelihood range. Specifically, for a 70% likelihood, the percentage of pre-analytical errors lies in a range of 0.087% to 0.254%, analytical error percentage lies from 0.0258% to 0.0627%, and post-analytical error percentage is around 0.0393% to 0.0674%. Moreover, test repetition rate is selected as 16.9% according to a relevant study[172].

4) After diagnosis and patient outcomes:

According to the data results of a relevant study[14] that uses an electronic health record-based method to detect the diagnostic errors, the patient outcome flows, shown in Table 8.2, were the observed data based on a total of 212,165 first visits with a 14 day observation period after first visits or index visits. The 14-day cut-off observation period is chosen, because longer intervals show that return visits are less clearly linked with errors in the index visits.[14]

	Total cases with errors (190 cases)		
	Unscheduled patient cases (177 cases)		control patient cases (13 cases)
Patient outcome flows:	A primary care visit followed by an **unplanned hospitalisation** that occurred between 24 h and 14 days after the visit.	A primary care visit followed by one or more **unscheduled primary care visits, an urgent care visit, or an ER visit** that occurred within 14 days	**Recovering cases** & **re-visiting cases scheduled** by follow-up service
Ratios:	141 out of 177 unscheduled cases	36 out of 177 unscheduled cases	13 out of 190 total error cases

Table 8.2 Patient outcome flows of diagnostic error cases

8.2.2 Estimated data from experts or "expert elicitation"

Expert elicitation is another source of data especially when the data are unattainable because of time, physical constraints or lack of resources. Figure 8.1 shows seven steps that are commonly implemented in an expert elicitation study.

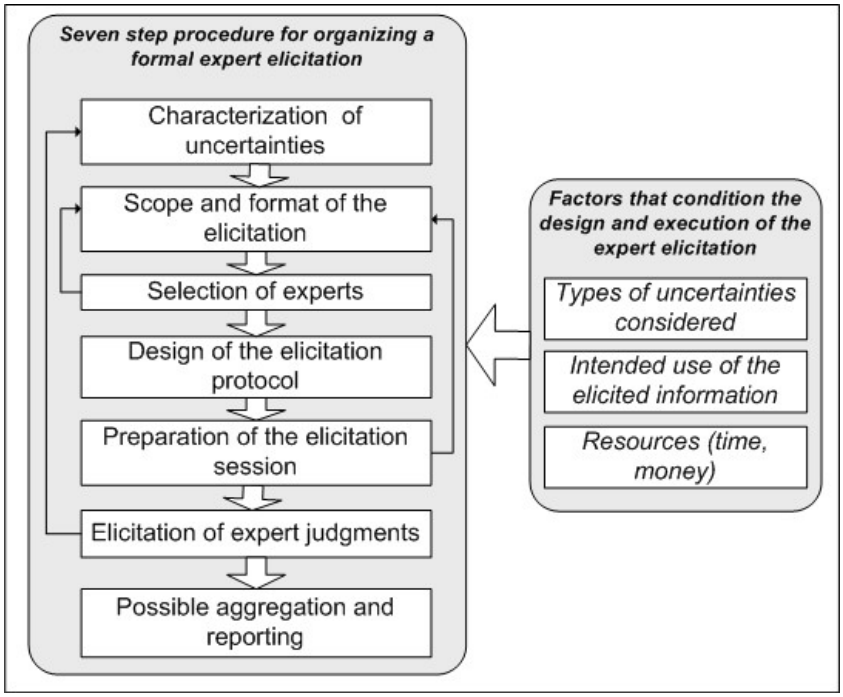


Figure 8.1 Seven steps for a formal expert elicitation[178]

8.2.2.1 Study design

An expert elicitation study is designed based on the seven steps shown in Figure 8.1. In detail, this thesis conducts the study following the same methods employed in the expert judgment study[181] by Van der Fels-Klerx et al in 2005. The steps are described below:

a. Define case structure document including uncertainties identification:

This step identifies the objectives and types of uncertainty considered. This study collects estimates of relevant variables from experts, and the main uncertainties involved are: doctor's medical knowledge and experience; doctor's preference; different working areas and patient groups.

The model is overviewed again in this step. The target variables, which are to be assessed by the experts, are identified. At the same time, two variables, whose values are known from the literature, are selected as the "seed variables". Seed variables are used to weight the individual expert's performance and then to reduce the effect of uncertainties on the data results. A summary of target variables to be estimated and seed variables to weight expert performance is listed below in Table 8.3:

Target Variables:		
Variable ID	Variable name in the model	*Explanation of the variable*
1	percentage1 of cases with missed tests	*the percentage of missing tests, for a case with errors*
2	percentage2 of cases with missed tests	*the percentage of missing tests, for a case without errors*
3	percentage1 of improper referrals; percentage2 of improper referrals	*they taken as the same value, and represented by the percentage of improper referrals for case with errors*
4	pecentage3 of improper referrals	*the percentage of improper referrals for case without errors*
5	improper referrals corrected by expert	*the percentage of improper referrals corrected by expert*
6	percentage of obtaining discordant data during a repeat visit	*during a repeat visit, the percentage of information which is discordant with the data from the first visit*
7	error detection rate after obtaining concordant data during a repeat visit	*during a repeat visit, the error detection rate if new data are concordant with the data from the first visit*

Seed Variables:	
Variable ID	Variable Name in the Model
e1	history and physical examination sensitivity
e2	error detection rate after obtaining discordant data during a repeat visit

Target Variables:	
Variable ID	Variable Description
ratio1	patient flows after phase1: *patients to receive final diagnosis and be discharged* : *patients to undergo tests* : *patients to be referred*
ratio2	patient flows after phase2: *patients to be discharged* : *patients to be referred to other experts*

Table 8.3 A summary of target variables and seed variables

b. Scope and format of the elicitation:

Since resource is limited, it is suggested that six to 12 experts are needed to conduct the study.[178] At least six experts should be included, and the benefit of including additional experts beyond 12 experts begins to drop off.[178] Thus, six to 12 experts are sufficient for this study.

Personal face-to-face interviews were chosen as the method to conduct the expert elicitation. Although interviews may be more time consuming than questionnaires, the number of participants in this study is small and data quality is of greater concern. Compared with questionnaires, face-to-face interviews provide the opportunity of communicating with participants. Interview questions can be further explained during the process and participants can give more detailed explanation in their answers. Moreover, participants in interviews are generally more motivated. For participants who are not able to engage in face-to-face interviews due to location constraints, interviews can be conducted via online video applications, such as Skype[179].

c. Identify experts and select experts:

Experts are randomly selected from the list of names based on pervious working contacts. At the same time, the selected experts need to meet the following criteria: experts should be independent clinicians with at least five years of general practice experience, and the

relevant work experience should be in countries, where the diagnostic process or patient pathway is similar to that of the UK. The number of male participants and female participants is aimed to be half and half in order to provide a balance of views. In the end, participants are identified based on the interests of the study and availability.

d. Design the elicitation format document:

The details of interviews are designed in this step. Questions to be assessed are documented.

The interviews are conducted in three main steps, and documents required for each step are generated:

- Step 1: Show participants information relating to the study and ask them to sign consent form. This is to make participants understand their rights when participating in the study and to understand that the study is confidential.
- Step 2: Introduce the background to the study. Study background information and the model framework illustrating the diagnostic process are shown and explained to the experts. The background information provides a quick introduction to the types of data that the expert should provide. More information is attached in APPENDIX VII.
- Step 3: Experts are asked to provide their estimates, following a list of semi-structured questions.

  A total of nine target variables and two seed variables in Table 8.3 are included in the questions. For each variable, experts are required to provide the median of the variable with a 90% confidence band, as well as their rationale for the probability assessments if possible. The document listing the semi-structured questions is attached in APPENDIX VII.

e. Dry-run session:

One clinician is selected and is given the elicitation format document. The clinician is asked to provide comments on the document. If needed, the document is revised based on feedback from the clinician.

f. Elicitation of expert judgements, and possible feedback communication:

Interviews are conducted this step, and data are collected. In some relevant studies, assessment may be carried out over many rounds. The experts will be provided with

feedback in terms of discrepancy analysis of the relevant variables with rationales at the end of the first round assessment or interview, and then experts are encouraged to revise their earlier answers. This step may be repeated several times until the results meets a pre-defined stop criterion which may be the number of rounds or stability of results.[180]

However, due to the study time limit and expert availability, it is not possible to ask experts to answer questions in two or more rounds in this study. At the same time, the study also aims to reduce the uncertainty impact to a minimum and obtain estimates with confidence. Therefore, this study amends the feedback session at the end of each interview. The individual expert will receive feedback on the assessment immediately at the end of the interview by comparing the assessment results with the results from the previous experts' results. To avoid misleading correct answers or to avoid possible bias, the expert will not be informed that the feedback is based on the other expert members, and will only be asked their rationale in relation to relevant variables which show discrepancy and mention of rationale from other experts at the same time. Then the expert will be asked whether he/she wants to change their answer.

g. Data analysis and documentation:

Collected data are analysed using Cooke's classical model. Cooke's method first weights the individual expert's assessment performance by scoring the likelihood that expert distributions over the set of seed items correspond to the known-observed/measured results. [181] Then, for each variable, the individual experts' assessments are aggregated to one combined probability distribution function(PDF), named the decision maker's (DM) distribution[181] which can be taken as an expected distribution function. The DM distributions reflect the weighted assessments, and are documented as the output data.

8.2.2.2 Data results

A total of six clinicians participated in the study, and completed all the estimations. Data were collected, and this includes the estimates of two ratios and nine variables. For each variable, three values are estimated, which are the values of the 5th percentile, the 50th percentile and the 95th percentile.

There are a number of software packages available to support data analysis of expert elicitation studies[182] [183]. The software package EXCALIBUR[184] is used to process the data mainly due to its function of weighting expert assessment performance. The measure used to weight expert performance in this study is called global weights. Global weights are

defined as the measures of expert performance on seed variables[184]. In other words, the global weight indicates the probability that the expected distribution takes the random individual expert distribution. For each expert, global weights are the same for all variables.

For each variable, the three percentile values from the individual expert provide the individual subjective PDF. By comparing the true values of the seed variables with the expert assessment results, the expert is assigned a weighted score, the global weight. The final weighted estimation results of the variable, namely the DM distribution function of a variable is the combination of weighting individual PDF, which can be represented as:

$$\text{DM distribution function}[\text{f}] = \frac{\sum_{i=1}^{n} f_i w_i}{\sum_{i=1}^{n} w_i},$$

where $\text{i} = 1, \dots, \text{n}$, f_i is the probability distribution function from an individual expert and w_i is the global weight (performance-based weighting) or the probability.

The data results are listed in Table 8.4 and Table 8.5, which are used as the simulation data for the quantitative model. Table 8.4 shows the DM distribution functions based on global weights. Table 8.5 shows the mean of the ratios from experts. The column of “realisation” in Table 8.4 describes the true values of the seed variables. It is witnessed that the true values of the seed variables lie in the ranges of the estimates provided by the experts.

Number	ID	Percentiles			Realisation	Full name
		5th	50th	95th		
1	1	4.968	33.73	78.07	-	percentage1 of cases with missed tests
2	2	1.51	19.14	39.98	-	percentage2 of cases with missed tests
3	3	2.196	9.611	72.53	-	percentage2 of improper referrals
4	4	1.85	10	19.42	-	pecentage3 of improper referrals
5	5	9.972	84.94	98.76	-	improper referrals corrected by expert
6	6	2.067	32.33	79.05	-	percentage of obtaining discordant data during a repeat visit
7	7	1.358	61.74	96.26	-	error detection rate after obtaining concordant date during a repeat visit
8	e1	45.57	84.44	95.16	66	history and physical examination sensitivity
9	e2	28.13	78.9	98.73	76.3	error detection rate after obtaining discordant data during a repeat visit

Table 8.4 Resulting solution (combined DM distribution of values assessed by experts)

ID	Mean	Variable Description
ratio1	65% : 30.2% : 13.8%	patient flows after phase1: *patients to receive final diagnosis and be discharged* : *patients to undergo tests* : *patients to be referred*
ratio2	86% : 14%	patient flows after phase2: *patients to be discharged* : *patients to be referred to other experts*

Table 8.5 Results of ratios

8.3 Model simulation results

This section conducts model simulation experiments. It first summarises model equations, data and assumptions used in the simulation experiments. Then, three different scenarios are simulated and relevant results are illustrated. The first scenario provides a one-year overview of the system as it relates to the current situation and assumptions. The second scenario changes one factor, and observes changes in system outputs. Two factors are selected as an example, and individually demonstrate the changes of the system outcomes. Model sensitivity is conducted to estimate model outputs as a factor changes. The third

scenario changes two factors at the same time, and simulates the patient outcomes caused by diagnostic errors.

8.3.1 Model summary of equations and data:

Relational equations, simulation data and assumptions are summarised below. The model is further refined based on the assumptions, and the simulation model is shown in Figure 8.2.

- **Assumptions used during simulation experiments:**

During the simulation experiments, some assumptions are made. Because the quantitative model reflects the whole diagnostic process and covers many variables, assumptions are made in order to simplify the simulation situation and to reduce the time and resources.

1) Assumption 1: During the simulation, phase2 represents the laboratory tests only. There are three reasons for this:
- Focusing on one type of diagnostic test can provide more detailed information. A specific simulation is preferred to a more general sum of two types of diagnostic tests.
- GPs do many more laboratory tests than radiology tests.
- There are more data available for laboratory tests than for radiology tests.

2) Assumption 2: The waiting lists are not considered in simulation experiments, and all administration rates involved in the model are taken as 100% because of insufficient data regarding administration rates.
3) Assumption 3: During this simulation, errors arising after being referred to experts are not accounted for and hence the error rate from experts is assumed to be zero.
4) Assumption 4: Two variables are not included in the simulation model.
- "*doctor's knowledge and experience*" is not included in the simulation model, because it is not identified during the regression modelling process. Regression modelling is based on the sample data, and this factor is not observed to have strong correlations with other variables.
- "*history and physical examination sensitivity*" is not shown in the simulation model, because this factor is considered as a constant according to the relevant study[157], and its dependent variable is determined as the function of relevant variables using regression modelling.

- **Summary of equations used in the simulations:**

The equations of the model are constructed in three ways: integration relations, regression relations and simple relations.

1) Integration relations are in the form of integration functions for the "stock" variables in the stock and flow diagrams. The "stock" variables can be represented using integration functions of input flows and output flows, which were introduced in Chapter 6. The dependent variables of the model are mainly the "stock" variables, illustrated as the variables in the blocks in Figure 8.2.
2) Regression relations are the relational equations identified by regression analysis, and these equations are difficult to be determined using system dynamics modelling. Especially for the variables in Phase1, regression equations are constructed using data collected from questionnaires, as shown in Chapter 7.
3) The rest are simple relations, and the relationship functions are easily identified.

- **Summary of data used in the simulations:**

The relevant factors such as the input and output arrows used in the integration equations are required to collect relevant data.

1) Data from experts: The variables, which are simulated based on the data from experts, are highlighted in different colours (purple, yellow and green) in Figure 8.2. Data collected from the experts are randomly plotted with a 90% confidence band following a normal probability distribution.
2) Data from the literature: The variables based on the data from literature are simulated based on the data results in subsection 8.2.1.
3) Simulation experiments include some soft variables, and are conducted using what-if scenarios: assumptions simplify the simulation situation. Especially for the variables whose data are very difficult to achieve, it also helps to save time and resource. Moreover, simulations under different assumption scenarios help to explore the assumptions, and observe and compare different system behaviours. The variables "*continuity of care"* and *"patient medical history"* are assumed to be at an average level. An average level indicates that the relevant score of the variable is 0.5, according to the quantifying method described in section 7.3.
4) Control patient cases include recovering cases and re-visited cases scheduled by the follow-up service. During this simulation, it is assumed that the percentage of recovering cases in the total error cases is zero.

5) Error rate from experts is assumed as zero.
6) No time delays. It is assumed that the administration rate/percentage of each step is 100%. In other words, there is no delay of administration, and 100% of the cases are administrated so that these cases proceed to the next step at every time unit.

For ease of reference, a full list of the variable equations in the simulation is attached in APPENDIX VIII.

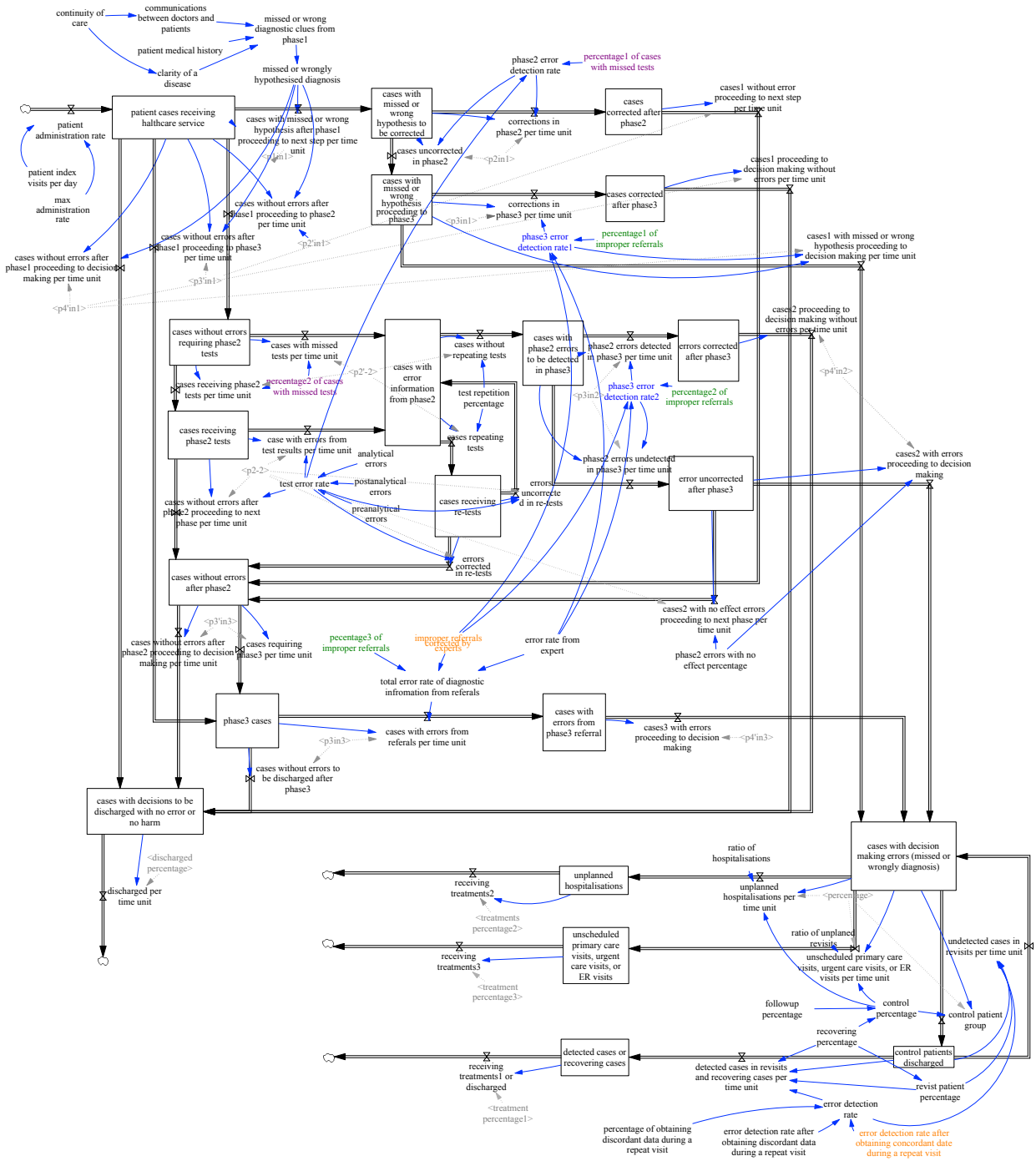


Figure 8.2 Simulation model

8.3.2 Scenario 1: one year overview

Assuming there is a constant 100 patient cases per day receiving healthcare service, the model is simulated and observed over a one-year period from day 0 to day 365, and system behaviours reflect the outcomes under the current assumptions and conditions of the system.

Figure 8.3 plots the number of cases with decision-making errors over 365 days, which is the number of diagnostic error cases in the system from day 0 to day 365. The density of the graph indicates that the data are plotted every day for 12 months. The initial data of the error case are set as 0 at the beginning, which indicates that there is no error in the system. Because it is assumed that each step takes one time unit to be processed, the proceeding time from receiving healthcare service to decision-making is reflected as a time delay on the variable of cases with decision-making errors. Thus, the variable turns out to be its first non-zero value 1.01 on day 4. As seen in this figure, the number of "cases with decision making errors" floats in a range of 0 to 7.482, where the density mainly lies around 1 and the value remains under 2.5 for the most of the days. The somewhat wide floating range of model output is because many factors in the model are chosen to plot randomly within a confidence range.

In order to provide a more detailed analysis of simulation results and to avoid the effect of the initial setting values on statistical results, a period from day51 to day350, which covers continuous 300 data points or "counts", is selected to display the statistical results. Specifically, Table 8.6 provides the statistical results of number of cases with decision-making errors over the 300 days. During the 300 points, the "min" shows the smallest value, while the "max" displays the largest value. Also, it shows the "mean" value and the "median" value. "Mean" value is the arithmetic average of all points and "median" value indicates the value that the variable is larger than one half of the time and smaller than for half the time[185]. Moreover, "StDev" is the standard deviation, and "Norm" means the normalised standard deviation, that is the standard deviation divided by the mean. Results show that the average number of *cases with decision-making errors* is 0.9156 every day over the 300 days if there is a constant 100 patient cases per day and these case visits starts on day1. It is worth noting that 0.9156 represents the number of the error cases still in the system, and does not include the number of patients discharged. The outcomes of the discharged patients are represented by other variables and will be discussed later.

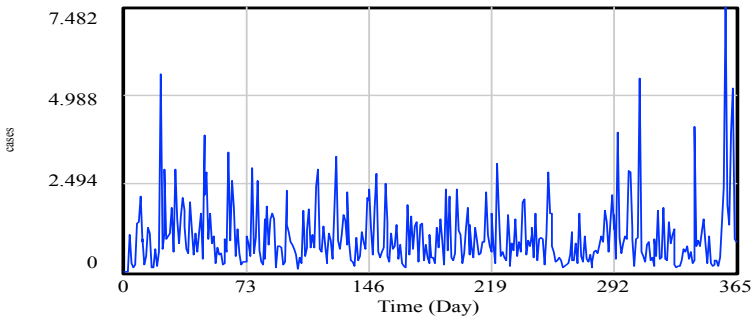


Figure 8.3 Cases with decision making errors (missed or wrongly diagnosis)

Variable (per 100 cases)	Count[a]	Min[b]	Max[c]	Mean[d]	Median[e]	StDev[f]	Norm[g]
cases with decision making errors (missed or wrongly diagnosis)	300 *(from day51 to day350)*	0.0701	5.444	0.9156	0.7063	0.7621	0.8323

a. *the total data points considered*
b. *the smallest value during all data point*
c. *the largest value during all data point*
d. *the arithmetic average of all points*
e. *the number which the variable is bigger than one half the time and smaller than on half the time*
f. *the standard deviation over all points*
g. *the normalised standard deviation, which is the standard deviation divided by the mean*

Table 8.6 Statistical results of number of cases with decision-making errors over 300 days

The variable of *cases with decision-making errors* has four input error flows. The input error flows are listed below, with the first three error flows being from three diagnostic phases.

- Flow1 - errors from phase1:

  Phase1 is history taking and physical examinations, where initial errors occur. During the current simulation, the factors "*continuity of care"* and *"patient medical history"* are assumed to be at an average level, which indicates both values are 0.5, and the error rate in the initial hypothesis is shown to be up to 22.01% under current assumptions. Also, only 2.102% of these error cases lead to the errors in the final decision-making.
- Flow2 - errors from phase2:

The total laboratory test error rate is around 0.262% on average, which agrees with the range 0.195% to 0.42% from the literature. The error cases from phase2 are delivered into two flows: "*cases2 with errors proceeding to decision making"*, which indicate that the errors have a direct effect on decision-making errors, and "*cases2 with no effect errors proceeding to next phase per time unit",* which means that although the case has error, the error has no effect on decision-making. The ratio of the two flows is 1:32.28, which indicates the cases with no-effect errors take a larger percentage in the total laboratory errors. Also, only 1.073% of the initial laboratory errors contribute to the decision-making errors. Table 8.7 illustrates the details of all three outflows of phase2.

Outflows from phase2 (per 100 cases)	**Count**	**Min**	**Max**	**Mean**	**Median**	**StDev**	**Norm**
Flow1: cases2 proceeding to decision making without errors (cases/day)	300	0.2852	6.903	3.3018	3.221	1.6983	0.5144
Flow2: cases2 with errors proceeding to decision making (cases/day)	300	0.00012	0.6869	0.04514	0.0166	0.0751	1.6639
Flow3: cases2 with no effect errors proceeding to next phase (cases/day)	300	0.00038	2.128	0.13985	0.0513	0.2327	1.6638

Table 8.7 Three outflows from phase2

- Flow3 - errors from phase3: The average rate of errors during referrals is about 2.554%.
- Flow4 – errors from revisits: A flow of error cases may revisit the healthcare service, and the undetected cases are counted as errors in revisits and go back into system diagnostic errors

Table 8.8 compares the three input error flows from three phases. Compared with errors from the other two phases, errors from Phase 1 contribute significantly more to the final decision-making errors.

Input error flows to decision making (per 100 cases)	Count	Min	Max	Mean	Median	StDev	Norm
Phase1: cases1 with missed or wrong hypothesis proceeding to decision making per time unit	300	0.0032	4.569	0.4627	0.1653	0.6575	1.4209
Phase2: cases2 with errors proceeding to decision making	300	0.00012	0.6869	0.0451	0.0166	0.0751	1.6639
Phase3: cases3 with errors proceeding to decision making	300	0.01413	2.192	0.3560	0.2074	0.3569	1.0024

Table 8.8 Comparison of three input error flows

There are a total of four patient outflows coming out of the model, namely there are four patient outcomes. They can be divided into two groups which can be discussed separately. One is patient outcome for the cases with no error or no harm. The other is outcomes for the cases with errors. Specifically, the results are as show below:

- **Outcomes for cases with no error or no harm:**

  The outcome is that patients are discharged with no error or no harm. These cases are the majority of all input cases, and the number is 99.123 on average per day from day51 to day 350 over 300-day counts if there is a constant 100 cases per day making the first visit of healthcare service since day1. Its first non-zero value occurs on day2. The sum of the mean of "cases with no error or no harm" and the mean of "cases with errors" equals 100.

  The **"**cases with no error or no harm" comes from five input flows, which are: "*cases without errors after phase1 proceeding to decision making"*; "*cases without errors after phase2 proceeding to decision making*"; "*cases without errors to be discharged after phase3*"; "*cases1 proceeding to decision making without errors*" that are the corrected errors from phase1; "*cases2 proceeding to decision making without errors*" that are the corrected errors from phase2.

- **Outcomes for cases with errors:**

- Control patients discharged. Control patients are the patients with errors, but who are either recovering or under close follow-up. It indicates that for these patients there is the chance to detect errors during follow-up. Simulation results show that there are 61.753% errors in control patients that can be detected during follow up. Control patients take 6.84% of total error cases under current simulation.
- Unplanned hospitalisations. This represents patients experiencing an unplanned hospitalisation during 14 days following a first visit, and it happens in 18.95% of error cases.
- Unscheduled primary care visits, an urgent care visit, or an emergency room (ER) visit. It includes the patients experiencing an unscheduled re-visit in 14 days after first visit, which could be a primary care visit, an urgent care visit, or an ER visit. It makes up 74.21% of the total error cases.

Among all patient outcomes, "*unplanned hospitalisations*" and "*unscheduled primary care visits, an urgent care visit, or an ER visit*" are the two main negative patient outcomes caused by relevant errors. Table 8.9 summarises the statistical results of the three patient outcomes after a diagnostic error.

Outcomes	Percentage	Count	Min	Max	Mean	Median	StDev	Norm
Control patients	6.84%	300	0.0048	0.3724	0.0626	0.0483	0.0521	0.8323
Unplanned hospitalisations	18.95%	300	0.0133	1.032	0.1735	0.1338	0.1444	0.8324
Unscheduled primary care visits, urgent care visits, or ER visits	74.21%	300	0.0520	4.04	0.6795	0.52415	0.5656	0.8323

Table 8.9 Four patient outcomes

8.3.3 Scenario 2: changing one factor

Two factors are randomly chosen during the simulation described in this subsection, and used to demonstrate how relevant model behaviour change when changing one factor. Simulation experiments are conducted separately to observe system behaviours responses to the individual factor. The same period from day51 to day350 is chosen to display the statistical simulation results. Furthermore, model sensitivities to different factors are analysed and compared.

8.3.3.1 Changing *the continuity of care*

The factor "*the continuity of care*" is selected in this section to demonstrate relevant system behaviour changes.

The score for "*the continuity of care*" is changed from 0.5 to 1 during the simulation. The score 0.5 is the current score which has been simulated in scenario 1. A score of 0.5 indicates an average level of "*the continuity of care*", and a score of 1 means the highest level, under which all patients see the same doctor all the time.

The variable "*missed or wrongly hypothesised diagnosis*", namely, the relevant error rate in phase1 is a dependent variable that is affected by the factor "*the continuity of care*". The changes of this variable are shown in Table 8.10. Table 8.10 also displays the number of cases with decision making errors, as well as the changes of the three types of patient outcomes: "*unplanned hospitalisations", "unscheduled primary care visits, urgent care visits, or ER visits"* and *"detected cases or recovering cases" which is a flow from control patients.*

Score for the factor: *the continuity of care*	Count	Min	Max	Mean	Median	StDev	(Norm)
missed or wrongly hypothesised diagnosis (in phase1) *(ratio)*							
0.5 (before)	300	0.2352	0.2352	0.2352	0.2352	-	-
1 (after)	300	0.2113	0.2113	0.2113	0.2113	-	-
cases with decision making errors (missed or wrongly diagnosis) *(unit: case)*							
0.5 (before)	300	0.07011	5.444	0.9156	0.7063	0.7621	0.8323
1 (after)	300	0.0677	4.974	0.8708	0.6917	0.7073	0.8123
patient outcome1_ unplanned hospitalisations *(unit: case)*							
0.5 (before)	300	0.01328	1.032	0.1751	0.1342	0.1457	0.8318
1 (after)	300	0.01283	0.9425	0.1665	0.1314	0.1351	0.8113
patient outcome2_ unscheduled primary care visits, urgent care visits, or ER visits *(unit: case)*							
0.5 (before)	300	0.05203	4.04	0.6860	0.5255	0.5705	0.8318
1 (after)	300	0.05024	3.692	0.6520	0.5146	0.5290	0.8113
patient outcome3_ detected cases or recovering cases from control patients *(unit: case)*							

0.5 (before)	300	0.00197	0.2399	0.0389	0.0280	0.0341	0.8759
1 (after)	300	0.00188	0.2192	0.0370	0.0272	0.0318	0.8577

Table 8.10 System outputs when changing "*the continuity of care*"

It is shown from the results in Table 8.10 that by improving the score of "*the continuity of care*" from 0.5 to 1, the rate of "*missed or wrongly hypothesised diagnosis"* of phase1 is predicted to reduce by 10.16%, and the *"cases with decision making errors"* can be reduced by 4.893% which is 0.0448 cases on average. Furthermore, two negative patient outcomes, which are "*unplanned hospitalisations", "unscheduled primary care visits, urgent care visits, or ER visits",* drop by 4.911% and 4.956% respectively. The change of *"cases with decision making errors"* is illustrated in Figure 8.4. The blue line indicates its values *when* "*the continuity of care*" is 0.5, while red line illustrate its new values after "*the continuity of care*" is changed to 1. To illustrate clearly the changes, data are plotted once in every 30 days in Figure 8.5. It is shown that the value of *"cases with decision making errors"* is apparently reduced.

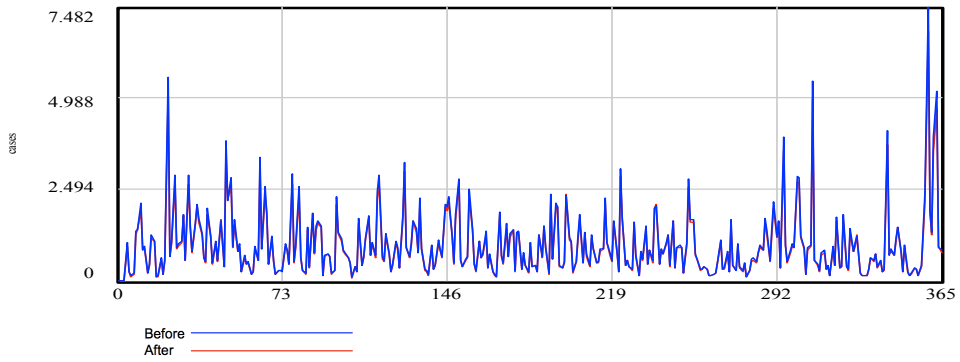


Figure 8.4 The change of "cases with decision making errors" plotted every day

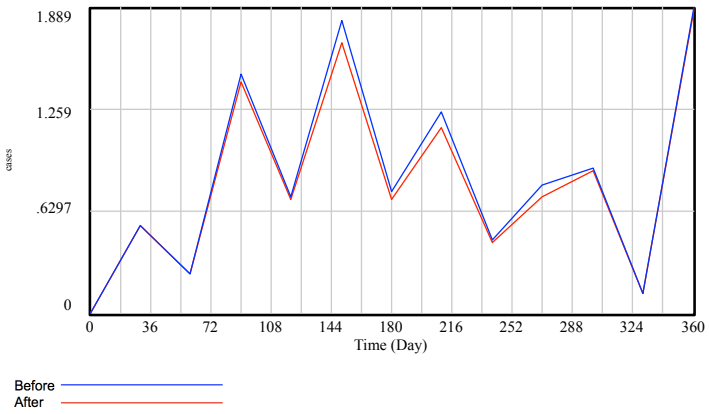


Figure 8.5. *The change of "cases with decision making errors" plotted every 30 days*

8.3.3.2 Changing the patient medical history

The factor "*patient medical history*" is selected in this section to demonstrate its impact on system behaviours.

The score for "*patient medical history*" is also chosen to change from 0.5 to 1 during the simulation. The score 0.5, which means an average score for clinicians accessing "*patient medical history*", is the current score which has been simulated in the scenario 1. Then, the score is improved to 1, which is the highest score. It is observed that relevant system outputs have changed, and some results are listed in Table 8.11. The rate of "*missed or wrongly hypothesised diagnosis"* of phase1 is observed to reduce by up to 20.75%, and the mean of *"cases with decision making errors"* can be remarkably reduced by 9.950%. The variable "*unplanned hospitalisations", "unscheduled primary care visits, urgent care visits, or ER visits"* are reduced by 10.051% and 10.045% respectively.

Score for the factor: *patient medical history*	Count	Min	Max	Mean	Median	StDev	(Norm)
	missed or wrongly hypothesised diagnosis (in phase1) *(ratio)*						
0.5 (before)	300	0.2352	0.2352	0.2352	0.2352	0	0
1 (after)	300	0.1864	0.1864	0.1864	0.1864	0	0
	cases with decision making errors (missed or wrongly diagnosis) *(unit: case)*						
0.5 (before)	300	0.07011	5.444	0.9156	0.7063	0.7621	0.8323
1 (after)	300	0.0652	4.49	0.8245	0.6402	0.6536	0.7928
	patient outcome1_ unplanned hospitalisations (unit: case)						
0.5 (before)	300	0.01328	1.032	0.1751	0.1342	0.1457	0.8318
1 (after)	300	0.01235	0.8507	0.1575	0.1228	0.1246	0.7912
	patient outcome2_ unscheduled primary care visits, urgent care visits, or ER visits *(unit: case)*						
0.5 (before)	300	0.05203	4.04	0.6859	0.52545	0.5705	0.8318
1 (after)	300	0.04839	3.332	0.6170	0.48105	0.4882	0.7912
	patient outcome3_ detected cases or recovering cases from control patients *(unit: case)*						
0.5 (before)	300	0.00197	0.2399	0.0389	0.0280	0.0341	0.8759
1 (after)	300	0.00179	0.1978	0.0351	0.0264	0.0295	0.8405

Table 8.11. System outputs when changing "*patient medical history*"

8.3.3.3 Sensitivity analysis

The results of model sensitivity analysis are described in this subsection. It has two aims. For an individual factor, probabilistic sensitivity analysis helps to quantify the confidence level of a variable for decision-makers. Also, the model sensitivities to different factors are compared so that suggestions for decision-makers can be provided. Based on the simulation tests of the continuity of care and the patient medical history, model sensitivity for the individual factor is analysed and compared.

- Model sensitivity to *continuity of care*

It is assumed that the impact of a changing range from 0 to 1 of the *continuity of care* is to be observed. The value distribution of *continuity of care,* which is normally distributed in a band from 0 to 1, is simulated. Its impact on relevant system variables is observed. Sensitivity graphs are shown. Figure 8.6 shows the sensitivity graph of "*missed or wrongly hypothesised diagnosis* in phase1". Since this variable has a linear regression relationship with "*continuity of care*", its sensitivity graph shows a band with a minimum value 21.14% and a maximum value of 25.90%. Figure 8.7 illustrates the sensitivity graphs of *"cases with decision making errors (missed or wrongly diagnosis)"* and *"unplanned hospitalizations"* separately. Data are plotted once every 30 days for a clear overview. As seen in Figure 8.7, both graphs display very narrow bands.

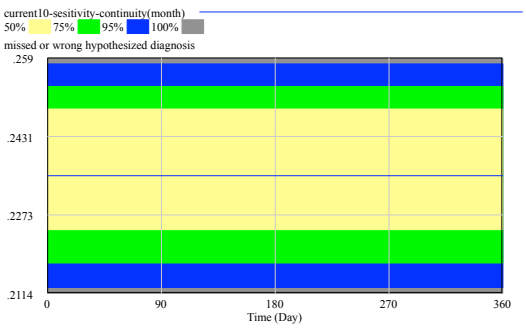


Figure 8.6 The *continuity of care* impact on *missed or wrongly hypothesised diagnosis*

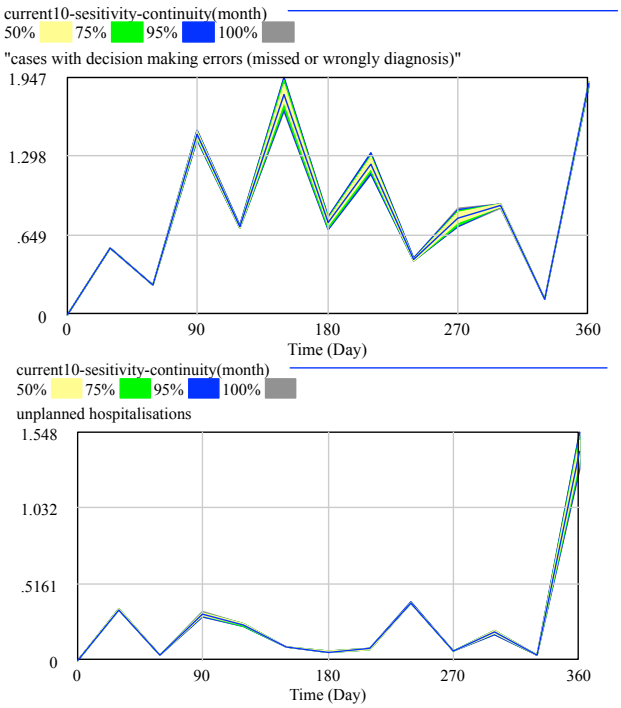


Figure 8.7 Sensitivity graphs for *continuity of care*

- Model sensitivity to *patient medical history*

Assuming that the impact of *patient medical history* from 0 to 1 is to be observed, *patient medical history* with a normal distribution from 0 to 1 is simulated.

Still, sensitivity graphs for the two variables "*missed or wrongly hypothesised diagnosis* in phase1" and *"cases with decision making errors (missed or wrongly diagnosis)"* are illustrated in Figure 8.8 and Figure 8.9. Comparing with the sensitivity graph for factor "*continuity of care"*, Figure 8.8 shows a wider band that is from 0.1874 to 0.2827 for a 100% confidence band. Comparing with Figure 8.7, the sensitivity graphs of *"cases with decision making errors (missed or wrongly diagnosis)"* and *"unplanned hospitalizations"* in Figure 8.9 are remarkable wider than that in Figure 8.7. It is clear that that the three variables are more sensitive to the factor *patient medical history* than the *continuity of care.*

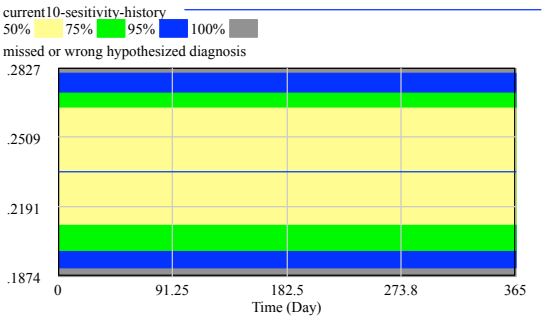


Figure 8.8 The *patient medical history* impact on *missed or wrongly hypothesised diagnosis*

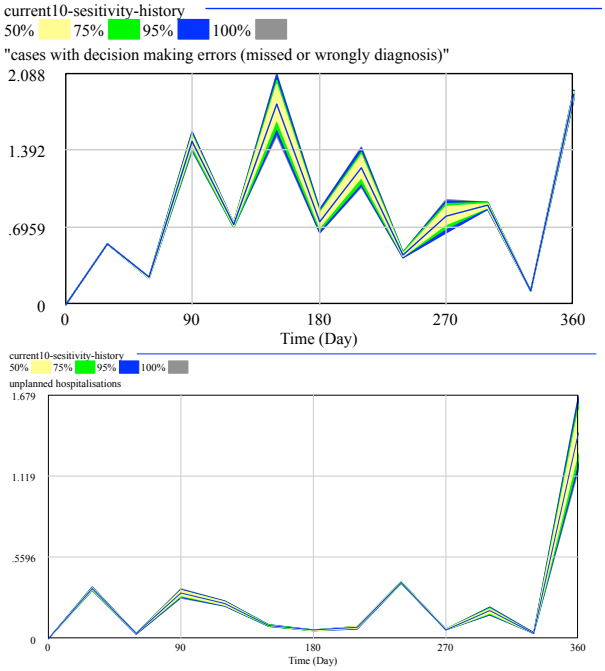


Figure 8.9 The *patient medical history* impact on *cases with decision making errors*

8.3.4 Scenario 3: changing two factors at the same time

Scenario2 demonstrates the effect of two factors from phase1 and scenario 3 chooses two factors from phase2 and after-diagnosis phase as examples. Scenario 3 first simulates the individual impact of the two factors in the first two experiments, and then the third experiment demonstrates the patient outcomes when the factors are changed together.

"Test repetition percentage" from phase2 and *"follow-up percentage"* from after-diagnosis phase are selected as the examples. The two variables, *"unplanned hospitalisations"* and *"unscheduled primary care visits, urgent care visits, or ER visits"*, directly reflect the negative impact of diagnostic errors on patient outcomes, and their simulation outputs are demonstrated.

During the first experiment, "*test repetition percentage"* is reduced from its original 16.9% to 15.4%, and "*follow-up percentage"* remains its original value at 6.84%. The simulation results of the two patient outcomes from day51 to day 350 are listed in Table 8.12. "*Unplanned hospitalisations"* displays a small increase by 0.000123 cases that is 701.76 parts per million (ppm) of the original outcome. "*Unscheduled primary care visits, urgent care visits, or ER visits"* is shown to increase by 683.74ppm.

Test repetition percentage	Count	Min	Max	Mean	Median	StDev	(Norm)
	Patient outcome1: unplanned hospitalisations *(cases)*						
before: 16.9%	300	0.01328	1.032	0.17513	0.1342	0.1457	0.8318
after: 15.4%	300	0.01332	1.032	0.17525	0.1342	0.1457	0.8315
	Patient outcome2: unscheduled primary care visits, urgent care visits, or ER visits *(cases)*						
before: 16.9%	300	0.05203	4.04	0.68593	0.52545	0.5705	0.8318
after: 15.4%	300	0.05218	4.04	0.68640	0.52535	0.5707	0.8315

Table 8.12 *Test repetition percentage* impact on patient outcomes

During the second experiment, "*follow-up percentage"* improves from 6.84% to 8.34% instead, and "*test repetition percentage"* remains at its original value at 16.9%. Simulation results in Table 8.13 show that both negative outcomes have decreased, where the two patient outcomes decline by 3625.69ppm and 3616.02ppm respectively.

Follow-up percentage	Count	Min	Max	Mean	Median	StDev	(Norm)
	Patient outcome1: unplanned hospitalisations *(cases)*						
before: 6.84%	300	0.01328	1.032	0.17513	0.13415	0.1457	0.8318
after: 8.34%	300	0.01377	1.017	0.17450	0.1341	0.1436	0.8227
	Patient outcome2: unscheduled primary care visits, urgent care visits, or ER visits *(cases)*						
before: 6.84%	300	0.05203	4.04	0.68593	0.52545	0.5705	0.8318
after: 8.34%	300	0.05394	3.982	0.68345	0.52535	0.5622	0.8227

Table 8.13 *Follow-up percentage* impact on patient outcomes

The third experiment is simulated to observe the model outcomes when the above two factors are changed together. Originally, "*test repetition percentage"* is 16.9% and "*follow-*

up percentage" is 6.84%. Afterwards, "*test repetition percentage" reduces to* 15.4%, while "*follow-up percentage" rises to* 8.34%. Results are listed in Table 8.14. It is apparent, as two factors are changed together, both negative patient outcomes have reduced by 2921.82ppm and 2932.71ppm respectively.

Two factors	Count	Min	Max	Mean	Median	StDev	(Norm)
	Patient outcome1: unplanned hospitalisations *(cases)*						
Before: *Test repetition percentage* 16.9% *Follow-up percentage* 6.84%	300	0.01328	1.032	0.17513	0.13415	0.1457	0.8318
After: *Test repetition percentage* 15.4% *Follow-up percentage* 8.34%	300	0.01381	1.017	0.17462	0.13415	0.1436	0.8224
	Patient outcome2: unscheduled primary care visits, urgent care visits, or ER visits *(cases)*						
Before: *Test repetition percentage* 16.9% *Follow-up percentage* 6.84%	300	0.05203	4.04	0.68593	0.52545	0.5705	0.8318
After: *Test repetition percentage* 15.4% *Follow-up percentage* 8.34%	300	0.05408	3.982	0.68392	0.5254	0.5624	0.8224

Table 8.14. Outcomes when changing two factors together

8.4 Discussion and summary

The model is a representation of the system itself, whereas the simulation presents the operation of the system over time. Simulation is a way of showing the eventual real effects of alternative conditions and courses of action.[186] The stock and flow diagram of diagnostic errors is a quantitative model that presents the errors and factors of the diagnostic system, while its simulation shows the behaviours of the system elements over time and provides a deep understanding of system behaviour and potential strategies.

This chapter summarises the relational equations of the quantitative model, collects the simulation data from literature and expert elicitation, and successfully conducts the model simulation experiments under different scenarios. The changes of model behaviour outputs are observed and described when varying one or more factors. This helps to understand the relations between the patient outcomes and different factors. The changes in the factors

may reflect the effect of an external intervention or policy change. Thus, simulation results provide suggestions for decision makers.

The next chapter summarises the evaluation methods for both regression models and system dynamics models, and further implements a list of evaluation assessments to comprehensively evaluate the system dynamics model.

Chapter 9 Model Evaluation

9.1 Introduction

This chapter starts with a summary of the evaluation methods used for regression modelling and system dynamics modelling in the thesis. Then, it specifically introduces the evaluation approaches for system dynamics models. Finally, a list of evaluation assessments for system dynamics models is conducted, and assessment results are provided.

9.2 Evaluation for regression models and system dynamics models

Theoretically, model evaluation involves comprehensively evaluating model performance and effectiveness, and evaluation approaches can be diverse according to different models, goals and resources. This thesis mainly contains two types of model: regression models and system dynamics models. The approaches used for the two types of models are introduced separately.

Approaches for regression model evaluation mainly involve cross validation, external validation or using metrics to analyse outputs, such as variance and R^2. Cross validation is suitable for a large data set. It randomly separates the dataset into two parts. One set is used for training the model, and the other one for testing the model. The whole process may repeat many times. It largely avoids the overfitting problem of the regression model. External validation is similar to cross validation. It uses the original dataset for training, but uses new external data for testing the model. Using metrics to evaluate outputs is a simple way of avoiding the need of introducing any new data, and is suitable for a small quantity of sample data. It evaluates the model by comparing the discrepancy between observed values and the values expected[187].

Chapter 7 has implemented metrics to conduct the evaluation of regression models. At the end of each section, each model is evaluated by the metrics after it is determined. Because the data set for the regression modelling study has a small quantity of data, which is 56 sets of data, and the study has no external data source, cross validation and external validation are not applicable. Thus, variance analysis as well as other metrics are employed, which include significance/p-value, standard error, R-squared and adjusted R-squared. The evaluation results have been shown in Chapter 7 after each model has been determined.

Several evaluation methods for system dynamics models have also been adopted during the main process of model design and model simulation. For example, the qualitative CLD model has been evaluated, by means of collecting feedbacks from clinicians. Subsection 5.5 in Chapter 5 describes the details, where seven clinicians were asked whether the factors and interrelations are reasonable. Results show that the qualitative model illustrates the key factors of diagnostic errors as well as interconnections among key factors and relevant errors.

Moreover, several simulation tests including sensitivity tests have been conducted for the quantitative system dynamics model in Chapter 8. Simulation experiments indicate that the quantitative model presents relevant case flows and analyses errors through the whole picture of the diagnostic process. Also, the model variable changes, such as negative patient outcomes, have been successfully observed when varying one or more factors.

However, this chapter conducts a further set of evaluation assessments for the system dynamics models in order to comprehensively discuss and determine the extent of the model domain and its applicability.

9.3 Introduction to evaluation approaches for system dynamics models

Evaluation for system dynamics models was firstly systematically proposed in 1980 by Forrester and Senge, and 17 tests were described for evaluating model from three perspectives: model structure, model behaviour, and policy implications.[188] At the same time, Richardson and Pugh published more evaluation approaches in 1981, including deactivating feedback loops, conducting hypothesis tests, and sensitivity analysis. Also, a table of tests was provided, which includes further aspects: model suitability, model consistency, model utility, model structure and model behaviour.[189 190] Later in 1996, Barlas provided structure tests and behaviour tests, and in particular behaviour tests introduced extreme conditions tests, behaviour sensitivity tests, modified behaviour prediction, boundary adequacy, phase relationship test, qualitative features analysis, and the Turing test.[190 191]

In 2000, Sterman summarised these prior works in the form of a list of assessments for dynamic model testing in practice in the classic textbook on business dynamics[51]. The

assessment covers the following tests: 1, boundary adequacy; 2, structure assessment; 3,dimensional consistency; 4,parameter assessment; 5, extreme conditions; 6,integration error; 7,behaviour reproduction; 8, behaviour anomaly; 9, family member; 10, surprise behaviour; 11, sensitivity analysis; 12, system improvement. [51]

The next section follows these assessments and provides a more comprehensive evaluation for the current system dynamics model of diagnostic errors.

9.4 Evaluation assessments for system dynamics modelling

This section conducts the evaluation assessments defined by Sterman in 2000, and the results are listed and discussed.

1) Boundary adequacy:

Model boundaries were discussed at the beginning of the model construction. The model focuses on primary care only. It contains key factors from the literature and the main phases of the diagnostic process. Time delays, as well as workload of healthcare providers, are not reflected in the current model. The model is reviewed again and the boundary is deemed to be appropriate.

2) Structure assessment:

Because performing structure assessment while developing the modelling is highly recommended[51], structure assessment of the system dynamics model has been carried out while constructing the model. Specifically, the approaches adopted include acquiring information from the literature and obtaining comments from clinicians. Seven clinicians provided feedback on the correlation of system variables, and then the model is transformed according to the diagnostic process, during which error cases are separated into different flows in order to highlight error flows and to observe the outcomes of the errors. Results show that the model clearly illustrates the case flows from accessing the healthcare service to the final patient outcomes. Moreover, the model structure has passed the “model check” provided by the software Vensim, and there is no flow failure or structure failure.

3) Dimensional consistency:

There is no arbitrary scaling factor involved in the model. All variables keep the dimensions consistent. The model equations keep the consistent use of units. The units in the model have passed the "units check" provided by the software Vensim.

4) Parameter assessment:

Parameters in the model have real life meaning. Parameters are estimated according to published literature or from estimates made by clinicians. Although there are uncertainties in the data collection from doctors, parameters are selected with a 90% confidence band.

5) Extreme conditions:

This testing is conducted along with the sensitivity testing. Several variables are randomly selected from each phase, including phase1, phase2, phase3, and the after-diagnosis phase. Extreme values of each variable are put into the system to observe system behaviours.

Selected variables with their corresponding boundaries are listed in Table 9.1. Results show that there is no irrational system behaviour when variables are tested at their extreme values, except the variable *follow-up percentage.* When *follow-up percentage* is set as "1", the system shows irrational outcomes of "*hospitalization*" and "*unscheduled primary care visits, urgent care visits, or ER visits*". Also, when letting the *follow-up percentage* be "0.999", the system outputs are observed to be normal, and no irrational behaviour is found.

Variable	lower boundary	higher boundary	Irrational behaviour
Continuity of care	0	1	No
Patient medical history	0	1	No
Percentage1 of improper referrals	0	1	No
Test repetition percentage	0	1	No
Improper referrals corrected by experts	0	1	No
Follow-up percentage	0	1	Yes, when the value is set as "1"
Error detection rate after obtaining discordant data during a repeat visit	0	1	No

Table 9.1 Extreme testing results

6) Integration Error:

Their integration method adopted in the previous simulation is the "Euler" integration method. However, there are a total of three integration methods: Euler integration, Difference integration and Runge-Kutta integration.

- Euler integration assumes that the rates in the model, which are the input and output arrows for the integrated variables, computed at a given time are constant through the time interval or one time step. [192]
- Difference integration is similar to Euler integration, but it records the value results before the new rates have been computed instead of recording values after determining the levels of the integrated variables. In other words, Euler integration reports levels and the values that result from those levels, whereas difference integration reports the level and the values that resulted from those levels.[192]
- Runge-Kutta integration is an extension of Euler integration. It steps into the time interval, evaluates derivatives, and then provides more accuracy without imposing a severe computational burden. [192]

The other two integration methods were tested. When the model interpretation setting is changed to "Difference" or "RK2 Auto", the outputs of system variables are compared with previous "Euler" integration outputs, it turns out that no model behaviour changes are observed.

7) Behaviour reproduction:

Simulation experiments under different scenarios were conducted in Chapter 8. The simulation results show that the negative patient outcomes will decrease when either improving "*the continuity of care*" or encouraging easier access to "*patient medical history*". The results agree with the real system expectations.

8) Behaviour anomaly:

Behaviour anomaly tests are done at an early stage of the testing. For example, initially, phase2 did not reflect the cases that require re-testing, so the model has to be modified and the flow that represents *laboratory test repetation* is added.

9) Family member:

Family member means how well the model "scales" to other members within the same class of systems. The current model not only explains how errors happen during the diagnostic process of primary care in the UK, but also the error case flows in most systems of healthcare that are similar to the system in the UK. However, the individual phase can be further developed to provide more detailed information and to analyse more factors that may be involved. Overall, the model can be potentially changed and developed to make it easier to scale.

10) Surprise behaviour:

Surprise behaviour means model behaviour that does not match expectations. System behaviours including supervise behaviours are discussed during the main process of testing the model, such as sensitivity analysis and extreme condition testing. When supervise behaviours appear, the model will be analysed again to find the reason. If the supervise behaviours can be explained, this will lead to new understanding of the model. Otherwise, the model will be amended.

11) Sensitivity analysis:

Although some sensitivity analysis has been implemented in Chapter 8, more tests are conducted in this section to provide further comprehensive results. A total of six factors are randomly selected from each phase of the diagnostic process in order to undertake the sensitivity analysis testing. "*Unplanned hospitalizations*", as one of the important negative patient outcome, is selected to demonstrate the sensitivity results. The sensitivity tests simulate the effect of each factor being changed between its lower boundary and upper boundary with normal distribution. Test results are described in Table 9.2. Different factors have varying impacts on the outcome. The changing percentages of "*unplanned hospitalizations*" are notable with the lowest 0.48% and the highest 190.38%. The factor changes could be the result of an external intervention or policy changes, and sensitivity results provide important information for healthcare decision makers.

Variable	Changing range	Mean value of *unplanned hospitalisations*				
		testing lower boundary		original	testing higher boundary	
Continuity of care	*0-1*	increase 0.1952	(4.95%)	0.186	decrease 0.1759	(05.43%)
Patient medical history	*0-1*	increase 0.2052	(10.32%)	0.186	decrease 0.166	(10.75%)
Percentage1 of improper referrals	*0-1*	decrease 0.09047	(51.36%)	0.186	increase 0.5401	(190.38%)
Test repetition percentage	*0-1*	increase 0.1869	(0.48%)	0.186	decrease 0.1787	(3.92%)
Improper referrals corrected by experts	*0-1*	increase 0.1887	(1.45%)	0.186	decrease 0.00075	(99.60%)
Follow-up percentage	*0-0.999*	increase 0.1952	(4.95%)	0.186	decrease 0.1759	(5.43%)

Table 9.2 Sensitivity test results

12) System improvement:

System simulations provide a way of exploring risk-free experiments, and results are able to suggest relevant interventions or policies via identifying that the number of errors varies with the changes of relevant factors. However, assessing the impact of a model in practice can be extremely difficult[51]. The keys to successful assessment of a modelling intervention are suggested as: prospective evaluation, use of multiple data sources, and proper experimental protocols[51].

9.5 Discussion and Summary

Overall, the model has passed units check and model structure check, and it works well under different integration methods. Moreover, it has performed well in the extreme tests,

and it shows distinct sensitivities to different model factors. Also, It provides a good representation of the current UK primary care pathway.

At the same time, the model can be improved in many ways for better performance. Firstly, providing a comprehensive list of all possible factors can help to explore more possible relations with diagnostic errors. Although the factors of the model that are selected from the literature help to remove unrelated variables and are easier for data collection, these factors are limited in the published findings, which may be not helpful for discovering new factors. Thus, if there were sufficient data sources available, all possible factors could be applied to the feature selection process before regression analysis, and classic feature selection methods such as Chi-squared filter feature selection or correlation feature selection, could be implemented.

Secondly, more data sources could help to provide a more reliable model. Data resources related to diagnostic errors are generally limited. Medical data include sensitive personal data, which are highly protected and can only be accessed with approval of the appropriate authority. Also, there is no sufficient data about diagnostic errors. Diagnostic errors are difficult to identify and report, and there is still a lack of reliable data resources currently. Data used in the studies are derived mainly from three sources: literature, questionnaires from clinicians, and semi-structured interviews with clinicians, namely expert elicitation. In particular, both questionnaires and interviews ask clinicians to provide relevant estimates. Thus, the data provided may not be objective. Their answers may differ from what they actually do and are sometimes subject to personal preference.

Nevertheless, many methods in the thesis have been implemented to reduce the problem, including anonymous answers, replacing subjective words in the questionnaires, Cook's classic mode, and data input with a confidence band. However, if there was greater access to patient medical records or there was the ability to follow clinical visits for a period of time, the model results could be significantly improved and the confidence band of the outputs could be narrowed down as well. Moreover, the model can be further developed and modified to reflect real-time data. The simulation studies choose the values of variables randomly within their confidence band at every time unit and administrative capabilities at each stage are assumed to be 100%. Thus, if real-time data could be collected, the model could provide real-time outcomes. Furthermore, the relevant clinician workloads as well as patient waiting lists could be reflected using real-time data, and the model could be enhenced by adding relevant variables.

The next chapter will summarise the contributions of the thesis to knowledge and opportunities for further research.

Chapter 10 Discussion

This chapter examines the challenges in conducting the research described in this thesis relating to diagnostic errors.

First, there has been the need to overcome the limitations of system dynamics modelling in identifying relational equations for model variables. There are several interrelations that are required to be determined before system dynamics modelling can be performed. In order to overcome this problem, this thesis implements regression analysis to determine the relational equations, and successfully applies the results of regression analysis to the quantitative system dynamics modelling.

Secondly, there are insufficient data for conducting the study of diagnostic errors. Not many data are available from the literature, and at the same time, there is no access to patient health records. In order to successfully conduct the study, the research has involved in adopting three methods to carry out the data collection: literature review, questionnaires and expert elicitation. In the methods of questionnaires and expert elicitation, the data obtained from clinicians' estimates are collected. Data from questionnaires are used for the regression analysis due to their larger quantity, and data from the expert elicitation study are used to carry out the quantitative model simulation experiments. In order to achieve objective data results, several methods have been implemented during the study, which include that replacing subjective words in the questionnaires into neural words and requiring anonymous answers of the questionnaires during the data collection period. Also, Cook's classic mode is implemented during the data analysis period to reduce the bias from expert elicitations.

Thirdly, the factors of diagnostic errors include some non-numerical variables, such as "communications between doctor and patient", and these variables cannot be quantified directly. This thesis implements the Likert scaling method and weighted scores to overcome the difficulty from such non-numerical variables. Using Likert scaling, an individual non-numerical variable is measured in terms of three levels, namely, categories, and then each category is given a weighted score. At the end of this process, non-numerical variables can be computed using the value under each category, provided by the clinician's estimates, and weighted scores.

Chapter 11 Conclusion

This chapter discusses the extent to which each of the objectives of the thesis has been achieved, states the contributions to knowledge made by the research described in this thesis, and describes the recommendations for future research.

11.1 A discussion of the extent to which each of the objectives has been achieved

The research described in the thesis aims to conduct systemic analysis and modelling of diagnostic errors. The objectives have been successfully achieved. Specifically, the objectives have been achieved to the following extent:

- Conducting a systematic review – This objective has been fully achieved as a systematic review of the relevant literature published between 2002 and 2012 has been conducted and key factors of diagnostic errors and methods of reducing diagnostic errors have been identified.
- Designing a qualitative model – This objective has been fully achieved as a quantitative model of diagnostic errors has been constructed based on the findings of the systematic review, which presents a structured picture of the causes of diagnostic errors.
- Collecting feedback from clinicians – This objective has been fully achieved as feedback has been successfully collected from seven clinicians, and the quantitative model has been refined accordingly. Collecting feedback from more clinicians could potentially further refine the model.
- Representing the qualitative model into the structure of a quantitative model – This objective has been achieved for the diagnostic process. The factors outside of the diagnostic process are out of the scope of the quantitative model. The quantitative model could be further developed and cover more factors if there was sufficient time.
- Conducting regression analysis and applying the algorithms in the system dynamics modelling – This objective has been achieved as regression analysis has been conducted based on the collected questionnaire data and the quantitative interrelations between model variables are determined. Collecting more objective data could improve the outputs of regression modeling.
- Conducting simulation experiments and observing model outputs – This objective has been achieved as the quantitative model has been simulated based on the data

from literature under different scenarios and model behaviours have been successfully observed.

- Adopting an evaluation assessment for the quantitative model – This objective has been achieved as a comprehensive evaluation assessment defined by Sterman in 2000 is conducted.

11.2 Contributions to knowledge

The main contribution to knowledge of the thesis is its new application of system dynamics modelling combined with regression modelling to the clinical diagnostic errors. Before this thesis was conducted, errors during the diagnostic process had been discussed, but it was lack of a systemic way that could present the causal-effect relations and quantitative relations between the variables in the diagnostic process. This thesis has provided the interrelation structure of the variables in the diagnostic error model, and has identified quantitative relations of the variables. Simulation results of the model show that the system dynamics model of diagnostic errors can help to understand diagnostic errors and the factors relating to diagnostic errors. It has provided a way of observing model behaviours while one or more factors are varied. In details, its main contributions to knowledge are listed as follows:

- Summarising the characteristics of diagnostic errors in Chapter 2

Five characteristics of diagnostic errors are summarised, which helps to understand the key issues of diagnostic errors that researchers are facing, and to explain the special requirements for researchers.

- Introducing two methods in a systemic analysis of diagnostic errors in Chapter 3

System dynamics modelling is introduced and proposed as a method of modelling diagnostic errors in the diagnostic process. Regression analysis is proposed as a method of identifying the interrelations of factors and dependent variables, which can make up for the limitation of system dynamics modelling.

- Discovering key factors of diagnostic errors and providing a summary of current error reduction methods in Chapter 4

A literature systematic review is conducted, and 65 relevant publications are reviewed. In the light of the findings from these reviews, key factors of diagnostic errors are

summarised into six categories and each category is explained. In addition, it lists both electronic and non-electronic methods used in researches aimed at reducing diagnostic errors.

- Providing a systemic view of the factors involved in diagnostic errors in Chapter 5

The qualitative model is developed. It connects the factors and effect together, illustrates a systemic view of relevant factors, and provides a hierarchical structure of causes of diagnostic errors.

- Presenting a systemic view of factors and errors in the diagnostic process in Chapter 6

The structure of the quantitative model is developed, and key factors and errors are mapped into the diagnostic process. The model structure presents where errors initially occur and how errors are delivered out of the system. Diagnostic phases are discussed in detail.

- Quantifying non-numerical factors and applying regression modelling to identifying relational equations in Chapter 7

Methods to quantify non-numerical factors are applied. The relationship functions of the factors and dependent variables are identified using regression analysis. Models are evaluated. PCA is conducted to demonstrate its utility in discovering variable variance and correlations.

- Conducting expert elicitation for data collection, simulating and observing system behaviours under different scenarios in Chapter 8

Data for model simulation is collected from the literature and from expert elicitation. In particular, Cook's classic model is implemented to analyse the data from the expert elicitation study. Also, risk-free simulation experiments are performed. Simulation results show system behaviours under different scenarios.

- Adopting a list of assessments for model evaluation in Chapter 9

Evaluation methods for both regression models and system dynamics models are summarised. Furthermore, more tests are conducted following an assessment list in order to provide a more comprehensive evaluation for system dynamics models.

11.3 Recommendations for future research

The research described in the thesis has a guide to future research. In the recent years, system dynamics modelling has been widely used in healthcare and more research of diagnostic errors is being conducted. The future research involves the following aspects:

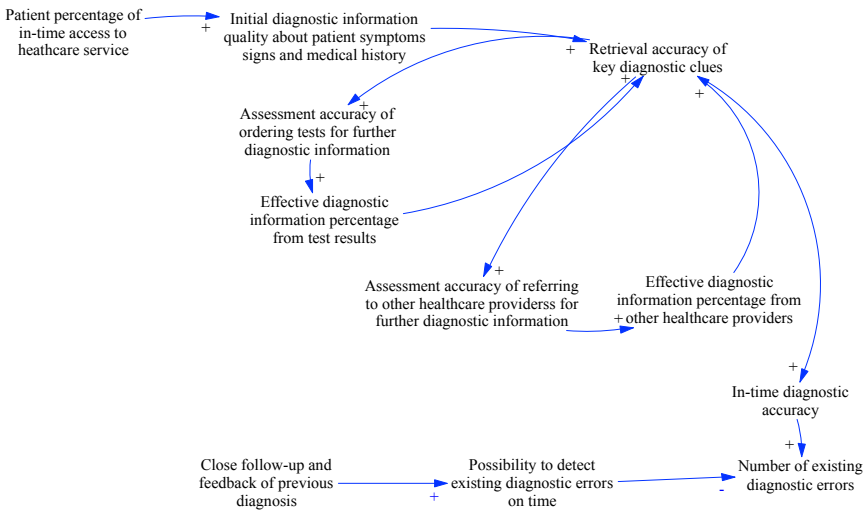
First, the future research involves collecting more objective real-time data and further modifying the models to fit different purposes and to provide more specific suggestions and strategies. This thesis provides a platform for future diagnostic error studies, and the model can be further modified to fit different purposes to provide more specific suggestions and strategies.

Additionally, the future research includes seeking a way of merging the models in the thesis with other existing models that also focus on the process of diagnosis in order to further develop the models. Since system dynamics modelling has been applied in different areas and systems of healthcare, it provides the opportunity for linking the system dynamics model in the diagnostic error context with other models, such as the model for patient waiting list.

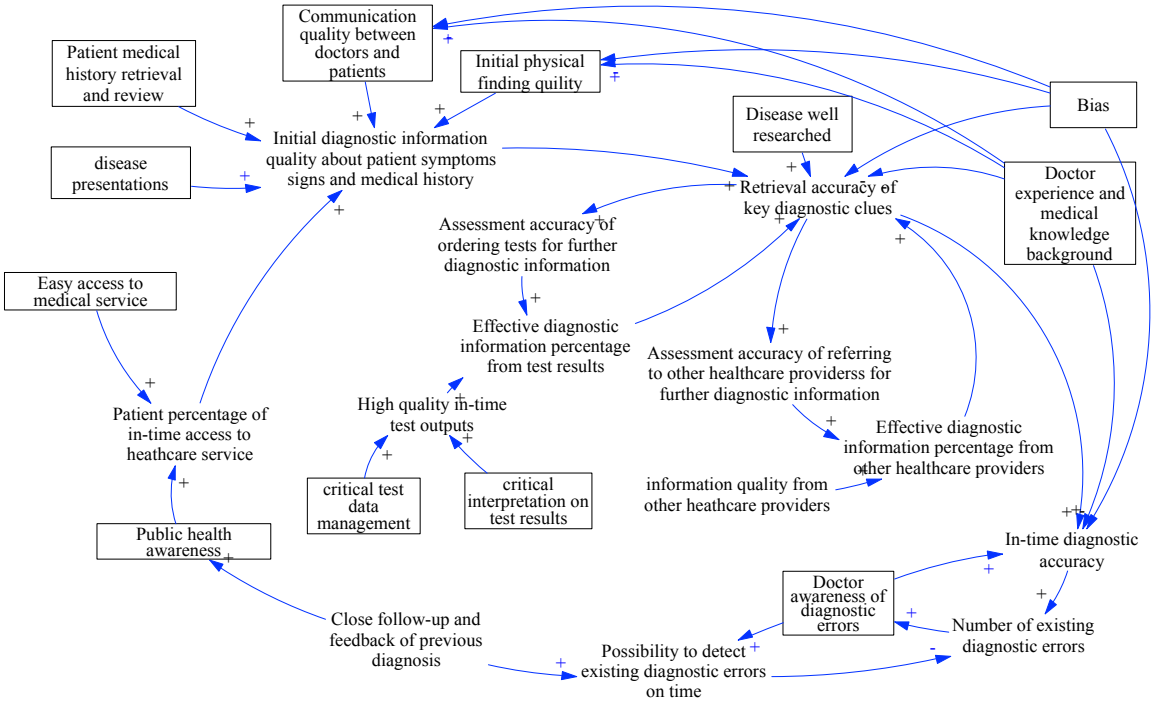
Furthermore, much research remains to be done on implementations of similar statistical approaches to diagnostic error analysis, which will contribute to a better coverage of possible factors and improvement of the quantitative modelling results. This thesis demonstrates the successful application of regression modelling in diagnostic errors in healthcare, similar statistical approaches can be implemented in the area of diagnostic error analysis or other related healthcare areas. When sufficient data are available, relevant machine learning technologies can help to discover the relationships between factors and outcomes. Patients can be classified according to different factors using unsupervised machine learning technologies, and supervised machine learning technologies such as regression modelling can be further applied to predicting the risk of diagnostic error or different patient outcomes.

APPENDIX I Initial quantitative model based on the systematic review

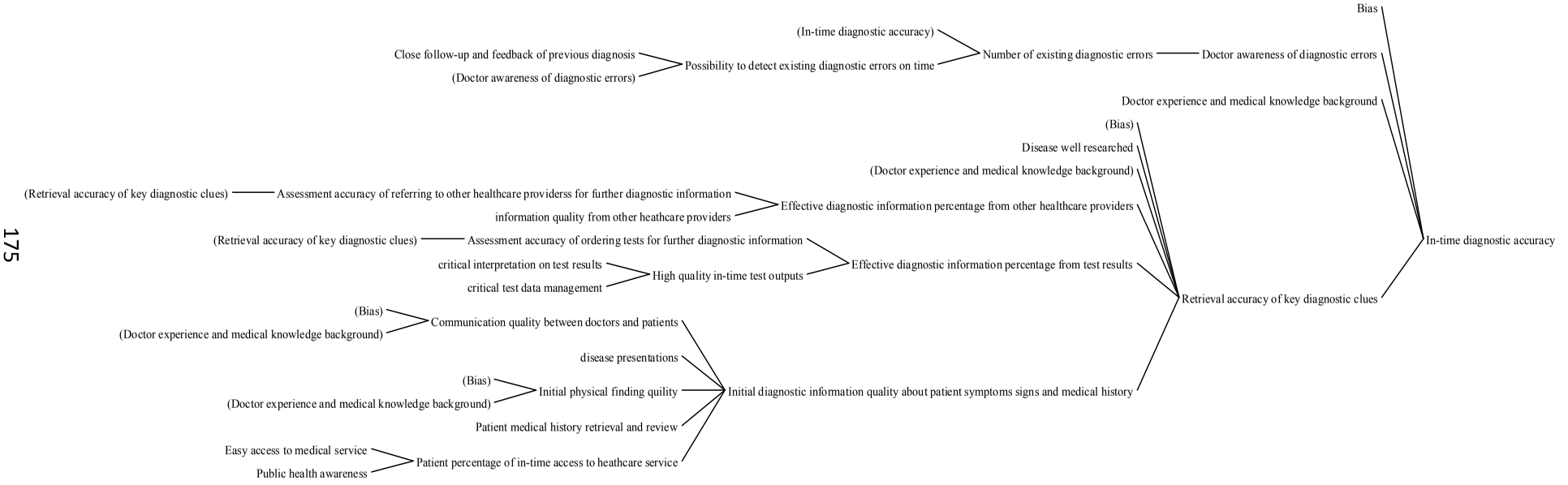
a. Patient path loops



b. Initial quantitative model based on literature systematic review



c. 4-level-depth causes tree


Bias
(In-time diagnostic accuracy)
Close follow-up and feedback of previous diagnosis
(Doctor awareness of diagnostic errors)
Possibility to detect existing diagnostic errors on time
Number of existing diagnostic errors
Doctor awareness of diagnostic errors
Doctor experience and medical knowledge background
(Bias)
Disease well researched
(Doctor experience and medical knowledge background)
(Retrieval accuracy of key diagnostic clues)
Assessment accuracy of referring to other healthcare providerss for further diagnostic information
information quality from other heathcare providers
Effective diagnostic information percentage from other healthcare providers
(Retrieval accuracy of key diagnostic clues)
Assessment accuracy of ordering tests for further diagnostic information
critical interpretation on test results
critical test data management
High quality in-time test outputs
Effective diagnostic information percentage from test results
In-time diagnostic accuracy
Retrieval accuracy of key diagnostic clues
(Bias)
(Doctor experience and medical knowledge background)
Communication quality between doctors and patients
disease presentations
(Bias)
(Doctor experience and medical knowledge background)
Initial physical finding quility
Patient medical history retrieval and review
Initial diagnostic information quality about patient symptoms signs and medical history
Easy access to medical service
Public health awareness
Patient percentage of in-time access to heathcare service

APPENDIX II Discussions with experts

The purpose of the model is to analyse the factors affecting diagnosis and to illustrate the cause-effect interrelations. The current model reflects the findings of a literature systematic review of relevant papers from 2002 to 2012.

The initial model is explained step by step, and the questions mainly cover the following aspects:

1, Relationship checking:

Such as: Do the arrows show reasonable cause-effect relationships between the two linked variables? Do all arrows work? Any suggestions for adding or deleting any arrows?

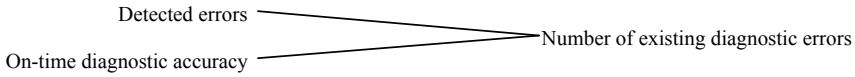
2, Variable checking:

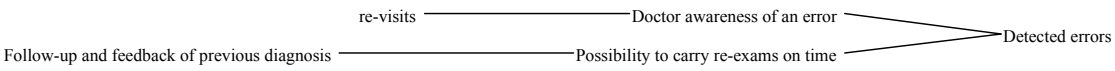
Such as: Do you think the model contains all the variables that affect the number of diagnostic errors? Any suggestions for adding or deleting any factors?

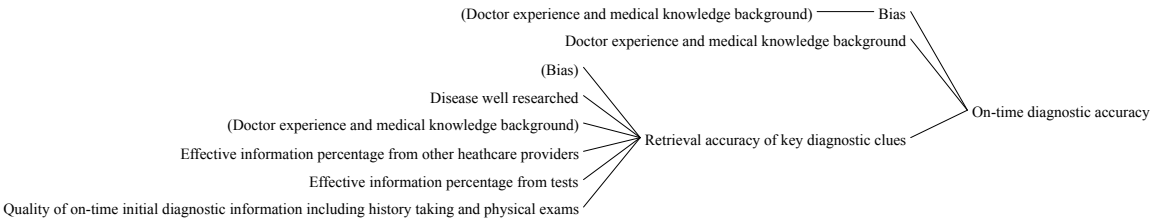
Experts are encouraged to ask questions if they are not clear or have some comments at any stage of the explanation.

APPENDIX III Causes trees of diagnostic errors

The 3-level causes tree of diagnostic errors cane be separated into the three following figures:







APPENDIX V Loops of diagnostic errors in the qualitative model

Loop Number 1 of length 3

Number of existing diagnostic errors

Re-visits

Doctor awareness of an error

Detected errors

Loop Number 2 of length 5

Number of existing diagnostic errors

Re-visits

Workload for healthcare providers

Follow-up and feedback of previous diagnosis

Possibility to carry re-exams on time

Detected errors

Loop Number 3 of length 5

Number of existing diagnostic errors

Patient trust

Patient percentage to choose the same healthcare centre

Re-visits

Doctor awareness of an error

Detected errors

Loop Number 4 of length 7

Number of existing diagnostic errors

Re-visits

Workload for healthcare providers

Easy access to medical service

Percentage of patients having on-time access to healthcare service

Quality of on-time initial diagnostic information including history taking and physical exams

Retrieval accuracy of key diagnostic clues

On-time diagnostic accuracy

Loop Number 5 of length 7

Number of existing diagnostic errors

Patient trust

Patient percentage to choose the same healthcare centre

Re-visits

Workload for healthcare providers

Follow-up and feedback of previous diagnosis

Possibility to carry re-exams on time

Detected errors

Loop Number 6 of length 8

Number of existing diagnostic errors

Re-visits

Workload for healthcare providers

Follow-up and feedback of previous diagnosis

Public health awareness

Percentage of patients having on-time access to healthcare service

Quality of on-time initial diagnostic information including history taking and physical exams

Retrieval accuracy of key diagnostic clues

On-time diagnostic accuracy

Loop Number 7 of length 9

Number of existing diagnostic errors

Patient trust

Patient percentage to choose the same healthcare centre

Re-visits

Workload for healthcare providers

Easy access to medical service

Percentage of patients having on-time access to healthcare service

Quality of on-time initial diagnostic information including history taking and physical exams

Retrieval accuracy of key diagnostic clues

On-time diagnostic accuracy

Loop Number 8 of length 10

Number of existing diagnostic errors

Patient trust

Patient percentage to choose the same healthcare centre

Re-visits

Workload for healthcare providers

Follow-up and feedback of previous diagnosis

Public health awareness

Percentage of patients having on-time access to healthcare service

Quality of on-time initial diagnostic information including history taking and physical exams

Retrieval accuracy of key diagnostic clues

On-time diagnostic accuracy

APPENDIX VI Questionnaires for regression modelling data

a. English translated document:

Questionnaire

Please tick the box for your relevant work experience:

☐less than 5 years ☐6~10 years ☐more than 10 years

Part A (total 4 tables)

Assuming 100 patients visited you, please fill the estimates of the number of the patients who are in the corresponding scale level in the blank underlines:

	Group 1 (During the 100 patients, please estimate: the number of patients who **often** see the same GP)	**Group 2** (During the 100 patients, please estimate: the number of patients who have **middle-level frequency** of seeing the same GP)	**Group 3** (During the 100 patients, please estimate: the number of patients who **rarely** see the same GP)
According to the **continuity of care**, whether the patient sees the same GP every time	____patients	____patients	____patients

[Three numbers need to add up to 100.]

	Group 1 (During 100 visits, please estimate: how many times the doctor and patient have **good** communications during the visit)	**Group 2** (how many times the doctor and patient have **middle-level** of communications during the visit)	**Group 3** (how many times the doctor and patient **hardly/are not able** to have communications during the visit)
According to the **communications between doctors and patients**	____times	____times	____times

[Three numbers need to add up to 100.]

	Group 1 (During 100 visits, please estimate: how many times you can **effectively** access and review patient medical history)	**Group 2** (how many times you can access and review **limited** patient medical history)	**Group 3** (how many times you can **not access** patient medical history or can only review **very limit** history information)
According to **accessing and reviewing patient medical history**	____times	____times	____times

[Three numbers need to add up to 100.]

	Group 1 (During 100 patients, please estimate: the number of the patients who present **clear** typical symptoms/signs)	**Group 2** (the number of the patients who present **a few symptoms/signs, but not clear or typical enough** at the present stage)	**Group 3** (the number of the patients whose symptoms/signs are **unclear**, for example early stage of disease may affect it)
According to the **clarity of disease symptoms/signs** presented by patients	____patients	____patients	____patients

[Three numbers need to add up to 100.]

Part B (total 2 questions)

- Comparing with patient's final correct diagnostic result, about _____% of patients was not presenting/ was found one or more typical symptoms/signs during physical examinations. (Hint: all 4 aspects in Part A may affect doctors collecting relevant information)

- Assuming 100 patients visited you, you may come up with one or more hypothesised diagnosis (although may not the final diagnosis) after physical examinations. About _____% patients' real health problems were **not** under your hypothesis.

b. Original document:

调查问卷:

职称: （初级，中级，高级）

工作年限: (小于 5 年，6～10 年，10 年以上)

第一部分（共 4 个表）

假设有 100 名患者前来就诊，请估计对应级别下的患者数目，填入空格横线上:

	一类 （100 名患者中：多少名**经常**就诊于同一位医生）	二类 （100 名中：多少名**不是很少但也不经常**就诊于同一位医生频率）	三类 （100 名中：多少名**很少**就诊于同一位医生）
根据是否**患者每次就诊找同一位医生**	___名	___名	___名

[3 组数字之和需为 100]

	一类 （100 例中：多少例医患间可以有**很好**的相互沟通 ）	二类 （多少例医患间有一般的沟通）	三类 （多少例医患**很难或几乎没有**相互沟通）
根据**医生和患者之间的沟通情况**	___例	___例	___例

[3 组数字之和需为 100]

	一类 （100 例中：多少例医生可很**有效**提取和浏览较完整的患者病历档案）	二类 （多少例医生可提取和浏览一部分患者病历档案）	三类 （多少例医生**很难或不能**提取和浏览患者病历档案）
根据能否**有效提取和浏览病病历档案**	___例	___例	___例

[3 组数字之和需为 100]

	一类 （100 例中：多少例患者有**明显**的症状表现）	二类 （多少例患者有症状表现，但**不特别明显**）	三类 （多少例患者症状表现**不明显**，比如疾病初期）
根据**疾病在患者身上呈现的清晰度**	___例	___例	___例

[3 组数字之和需为 100]

第二部分（共 2 题）

1. 初步症状体征检查结果中 (不含实验室／CT 等)，对比患者最终正确的确诊结果，有____% 的患者**未被发现**一个或多个典型的症状体征 。（提示: 第一部分表格中的 4 个方面都可能影响医生收集信息）
2. 初步症状体征检查结果后 (不含实验室／CT 等)，您可能会直接得到一个诊断结论，或者得到一个或多个初步猜想，____% 的患者最终真实病情**未在**您的结论或初步猜测之内。

APPENDIX VII Interview Documents

a. Introduction information

This study, called “expert elicitation”, is designed to provide data to analyse the relevant factors of making a diagnosis, via asking clinicians to provide estimates of relevant variables.

A list of variables will be given that are to be estimated.

For some variables, you are asked to provide estimates for the median value of the variable with a 90% confidence band. In other words, you are asked to provide three values based on your experience. The three values of a variable are:

- A: the value at the 5^{th} percentile. The 5^{th} percentile indicates that 5% of the cases have the value lower than the value at the 5^{th} percentile.
- B: the value at the 50^{th} percentile, which is the median value.
- C: the value at the 95^{th} percentile. The 95^{th} percentile indicates that 5% of the cases have the value larger than the value at the 95^{th} percentile.

b. Semi-structured questions

[Note: Assuming each patient attends with a single diagnosis or problem that needs to be elicited during the consultation]

Scenario 1: After first phase - the history taking and physical examinations

When a doctor is seeing a patient, the first phase is the history taking and physical examinations.

After this first phase, you may come up with one or more hypothesised diagnosis (although may not the final diagnosis).

And patient flow will go in three ways: to receive final diagnosis directly and to be discharged, to do assessment tests, or to be referred to other experts.

Please estimate:

Variable ID:	ratio1
Variable Description:	Assuming 100 patients visited you, after the first phase, what is the ratio of **patients to receive final diagnosis and to be discharged**, **patients to undergo tests**, and **patients to be referred**?
Estimates:	*(discharged : tests : referrals)* _____:_____:_____
Rationales (optional):	

Normally, for the right hypothesis (patient's health condition right under your hypothesis), please estimate the above two variables again:

Variable ID:	2
Variable Description:	What do you think is the percentage of failing to order tests, or ordering improper tests?
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	
Variable ID:	4
Variable Description:	What do you think is the percentage of failing to make a referral, or referring to improper experts?
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	

Assuming a patient is actually under a health condition, which is **not** under your initial hypothesis,

Variable ID:	1
Variable Description:	How much do you think it will affect ordering the right tests (what is the percentage of cases in which the tests are not ordered, or improper tests are ordered)?
Estimates:	*A :* _____ *B:* _____ *C:*_____
Rationales:	

Scenario 2: After assessment tests

After gathering more information from the tests, you may arrive at one or more hypothesised diagnosis.

Please estimate:

Variable ID:	ratio2
Variable Description:	After the test phase, what is the ratio of **patients to be discharged** and **patients to be referred to other experts**.
Estimates:	*(discharged : referrals)* _____:_____
Rationales:	

Assuming the patient is under a health condition, which is still **not** under your hypothesis after tests,

Variable ID:	3
Variable Description:	How much do you think it will affect referrals (percentage of cases in which a referral is not made, or improper expert is referred to)?
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	
Variable ID:	5
Variable Description:	What is the chance that the expert will find it improper and correct it?
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	

Scenario 3: Revisits

Assuming there were 100 patients are scheduled to pay an revisit,

Variable ID:	6
Variable Description:	What is the percentage of the patients who has discordant data(data that supports different diagnoses in discordant cases), compared with the first visit.
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	
Variable ID:	e2
Variable Description:	Assuming you found the discordant data during the revisit, what is the percentage of the case in which you are aware of an improper diagnostic decision and change the decision.
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	
Variable ID:	7
Variable Description:	Assuming you still found concordant date, what is the percentage of the case in which you are aware of an improper diagnostic decision and change the decision.
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	

The last variable to be estimated based on your experience:

Variable ID:	e1
Variable Description:	What is the result sensitivity of the history and physical examination?
Estimates:	*A:* _____ *B:* _____ *C:* _____
Rationales:	

APPENDIX VIII Variable Equation List

analytical errors= RANDOM UNIFORM(0.000258 , 0.000627 , 0)

Units: Dmnl

case with errors from test results per time unit=

cases receiving phase2 tests*"p2-2"*test error rate

Units: cases/Day

cases corrected after phase2= INTEG (

corrections in phase2 per time unit-

cases1 without error proceeding to next step per time unit,

0)

Units: cases

cases corrected after phase3= INTEG (

corrections in phase3 per time unit-cases1 proceeding to decision making without errors per time unit,

0)

Units: cases

cases receiving phase2 tests= INTEG (

cases receiving phase2 tests per time unit-cases without errors after phase2 proceeding to next phase per time unit-case with errors from test results per time unit,

0)

Units: cases

cases receiving phase2 tests per time unit=

cases without errors requiring phase2 tests*

(1-percentage2 of cases with missed tests)*

"p2'-2"

Units: cases/Day

"cases receiving re-tests"= INTEG (

cases repeating tests-"errors uncorrected in re-tests"-"errors corrected in re-tests",

0)

Units: cases

cases repeating tests=

cases with error information from phase2*

test repetition percentage*

"p2'-2"

Units: cases/Day

cases requiring phase3 per time unit=

cases without errors after phase2*

0.14*

p3'in3

Units: cases/Day

cases uncorrected in phase2=

cases with missed or wrong hypothesis to be corrected*

p2in1*

(1-phase2 error detection rate)

Units: cases/Day

"cases with decision making errors (missed or wrongly diagnosis)"= INTEG (

undetected cases in revisits per time unit+

cases1 with missed or wrong hypothesis proceeding to decision making per time unit+

cases2 with errors proceeding to decision making+

cases3 with errors proceeding to decision making+undetected cases in revisits per time unit-

control patient group-

unplanned hospitalisations per time unit-

"unscheduled primary care visits, urgent care visits, or ER visits per time unit",

0)

Units: cases

cases with decisions to be discharged with no error or no harm= INTEG (

cases without errors after phase1 proceeding to decision making per time unit+

cases without errors after phase2 proceeding to decision making per time unit+

cases without errors to be discharged after phase3+

cases1 proceeding to decision making without errors per time unit+

cases2 proceeding to decision making without errors per time unit-

discharged per time unit,

0)

Units: cases

cases with error information from phase2= INTEG (

case with errors from test results per time unit+cases with missed tests per time unit+"errors uncorrected in re-tests"-cases repeating tests-cases without repeating tests,

0)

Units: cases

cases with errors from phase3 referral= INTEG (

cases with errors from referals per time unit-cases3 with errors proceeding to decision making,

0)

Units: cases

cases with errors from referals per time unit=

phase3 cases*p3in3

*total error rate of diagnostic infromation from referals

Units: cases/Day

cases with missed or wrong hypothesis after phase1 proceeding to next step per time unit=

patient cases receiving halthcare service*p1in1*missed or wrongly hypothesised diagnosis

Units: cases/Day

cases with missed or wrong hypothesis proceeding to phase3= INTEG (

cases uncorrected in phase2-cases1 with missed or wrong hypothesis proceeding to decision making per time unit-corrections in phase3 per time unit,

0)

Units: cases

cases with missed or wrong hypothesis to be corrected= INTEG (

cases with missed or wrong hypothesis after phase1 proceeding to next step per time unit-

corrections in phase2 per time unit-

cases uncorrected in phase2,

0)

Units: cases

cases with missed tests per time unit=

cases without errors requiring phase2 tests*

percentage2 of cases with missed tests*

"p2'-2"

Units: cases/Day

cases with phase2 errors to be detected in phase3= INTEG (

cases without repeating tests-phase2 errors detected in phase3 per time unit-phase2 errors undetected in phase3 per time unit,

0)

Units: cases

cases without errors after phase1 proceeding to decision making per time unit=

patient cases receiving halthcare service*

(1-missed or wrongly hypothesised diagnosis)*

0.65*

p4'in1

Units: cases/Day

cases without errors after phase1 proceeding to phase2 per time unit=

patient cases receiving halthcare service*

(1-missed or wrongly hypothesised diagnosis)*

0.302*

p2'in1

Units: cases/Day

cases without errors after phase1 proceeding to phase3 per time unit=

patient cases receiving halthcare service*

(1-missed or wrongly hypothesised diagnosis)*

0.138*

p3'in1

Units: cases/Day

cases without errors after phase2= INTEG (

cases1 without error proceeding to next step per time unit+

cases without errors after phase2 proceeding to next phase per time unit+

cases2 with no effect errors proceeding to next phase per time unit+

"errors corrected in re-tests"-

cases without errors after phase2 proceeding to decision making per time unit-

cases requiring phase3 per time unit

,

0)

Units: cases

cases without errors after phase2 proceeding to decision making per time unit=

cases without errors after phase2*

0.86*

p3'in3

Units: cases/Day

cases without errors after phase2 proceeding to next phase per time unit=

cases receiving phase2 tests*"p2-2"*(1-test error rate)

Units: cases/Day

cases without errors requiring phase2 tests= INTEG (

cases without errors after phase1 proceeding to phase2 per time unit-cases receiving phase2 tests per time unit-cases with missed tests per time unit,

0)

Units: cases

cases without errors to be discharged after phase3=

phase3 cases*p3in3

Units: cases/Day

cases without repeating tests=

cases with error information from phase2*

(1-test repetition percentage)*

"p2'-2"

Units: cases/Day

cases1 proceeding to decision making without errors per time unit=

cases corrected after phase3*p4'in1

Units: cases/Day

cases1 with missed or wrong hypothesis proceeding to decision making per time unit=

cases with missed or wrong hypothesis proceeding to phase3*

(1-phase3 error detection rate1)*

p4'in1

Units: cases/Day

cases1 without error proceeding to next step per time unit=

cases corrected after phase2*p3'in1

Units: cases/Day

cases2 proceeding to decision making without errors per time unit=

errors corrected after phase3*

p4'in2

Units: cases/Day

cases2 with errors proceeding to decision making=

error uncorrected after phase3*

(1-phase2 errors with no effect percentage)*

p4'in2

Units: cases/Day

cases2 with no effect errors proceeding to next phase per time unit=

error uncorrected after phase3*

phase2 errors with no effect percentage*

"p2-2"

Units: cases/Day

cases3 with errors proceeding to decision making=

cases with errors from phase3 referral*p4'in3

Units: cases/Day

clarity of a disease=

0.398*continuity of care+0.452

Units: Dmnl

communications between doctors and patients=

0.461*continuity of care + 0.481

Units: Dmnl

continuity of care=0.5

Units: Dmnl

control patient group=

"cases with decision making errors (missed or wrongly diagnosis)"*

control percentage*

percentage

Units: cases/Day

control patients discharged= INTEG (

control patient group-detected cases in revisits and recovering cases per time unit-undetected cases in revisits per time unit,

0)

Units: cases

control percentage=

recovering percentage + followup percentage

Units: Dmnl

corrections in phase2 per time unit=

cases with missed or wrong hypothesis to be corrected*p2in1*phase2 error detection rate

Units: cases/Day

corrections in phase3 per time unit=

cases with missed or wrong hypothesis proceeding to phase3*

phase3 error detection rate1*

p3in1

Units: cases/Day

detected cases in revisits and recovering cases per time unit=

control patients discharged*

revist patient percentage*

error detection rate +

control patients discharged*

recovering percentage

Units: cases/Day

detected cases or recovering cases= INTEG (

detected cases in revisits and recovering cases per time unit-receiving treatments1 or discharged,

0)

Units: cases

discharged per time unit=

cases with decisions to be discharged with no error or no harm*discharged percentage

Units: cases/Day

discharged percentage=1

Units: Dmnl/Day

error detection rate=

percentage of obtaining discordant data during a repeat visit*error detection rate after obtaining discordant data during a repeat visit

+(1-percentage of obtaining discordant data during a repeat visit)

*error detection rate after obtaining concordant date during a repeat visit

Units: Dmnl

error detection rate after obtaining concordant date during a repeat visit= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.01358),(0.5,0.6174),(0.95,0.96)))

Units: Dmnl

error detection rate after obtaining discordant data during a repeat visit= 0.763

Units: Dmnl

error rate from expert=0

Units: Dmnl

error uncorrected after phase3= INTEG (

phase2 errors undetected in phase3 per time unit-cases2 with no effect errors proceeding to next phase per time unit-cases2 with errors proceeding to decision making,

0)

Units: cases

errors corrected after phase3= INTEG (

phase2 errors detected in phase3 per time unit-cases2 proceeding to decision making without errors per time unit,

0)

Units: cases

"errors corrected in re-tests"=

"cases receiving re-tests"*

(1-test error rate)*

"p2-2"

Units: cases/Day

"errors uncorrected in re-tests"=

"cases receiving re-tests"*

test error rate*

"p2-2"

Units: cases/Day

followup percentage= 0.0684

Units: Dmnl

improper referrals corrected by experts= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.09972),(0.5,0.8494),(0.95,0.98)))

Units: Dmnl

max administration rate=120

Units: cases/Day

missed or wrong diagnostic clues from phase1=

0.566-0.036*communications between doctors and patients-0.204*patient medical history-0.21*clarity of a disease

Units: Dmnl

missed or wrongly hypothesised diagnosis=

0.478*missed or wrong diagnostic clues from phase1+0.091

Units: Dmnl

p1in1=1

Units: Dmnl/Day

"p2'-2"=1

Units: Dmnl/Day

p2'in1=1

Units: Dmnl/Day

"p2-2"=1

Units: Dmnl/Day

p2in1=1

Units: Dmnl/Day

p3'in1=1

Units: Dmnl/Day

p3'in3=1

Units: Dmnl/Day

p3in1=1

Units: Dmnl/Day

p3in2=1

Units: Dmnl/Day

p3in3=1

Units: Dmnl/Day

p4'in1=1

Units: Dmnl/Day

p4'in2=1

Units: Dmnl/Day

p4'in3=1

Units: Dmnl/Day

patient administration rate=

IF THEN ELSE(patient index visits per day<=max administration rate, patient index visits per day , max administration rate)

Units: cases/Day

patient cases receiving halthcare service= INTEG (

patient administration rate-cases with missed or wrong hypothesis after phase1 proceeding to next step per time unit-cases without errors after phase1 proceeding to decision making per time unit-cases without errors after phase1 proceeding to phase2 per time unit-cases without errors after phase1 proceeding to phase3 per time unit

,

0)

Units: cases

patient index visits per day=100

Units: cases/Day

patient medical history=0.5

Units: Dmnl

pecentage3 of improper referrals= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.0185),(0.5,0.1),(0.95,0.1942)))

Units: Dmnl

percentage=1

Units: Dmnl/Day

percentage of obtaining discordant data during a repeat visit= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.02067),(0.5,0.32),(0.95,0.7905)))

Units: Dmnl

percentage1 of cases with missed tests= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-
(1,1),(0.05,4.9),(0.5,33.7),(0.95,78.07)],(0.05,0.04968),(0.5,0.3373),(0.95,0.7807)))

Units: Dmnl

percentage1 of improper referrals= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.02196),(0.5,0.09611),(0.95,0.7253)))

Units: Dmnl

percentage2 of cases with missed tests= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.0151),(0.5,0.1914),(0.95,0.39)))

Units: Dmnl

percentage2 of improper referrals= WITH LOOKUP (

RANDOM NORMAL(0.05 , 0.95 , 0.5, 0.367, 0),

([(0,0)-(1,1)],(0.05,0.02196),(0.5,0.09611),(0.95,0.7253)))

Units: Dmnl

phase2 error detection rate=

(1-percentage1 of cases with missed tests)*(1-test error rate)

Units: Dmnl

phase2 errors detected in phase3 per time unit=

cases with phase2 errors to be detected in phase3*

phase3 error detection rate2*

p3in2

Units: cases/Day

phase2 errors undetected in phase3 per time unit=

cases with phase2 errors to be detected in phase3*

(1-phase3 error detection rate2)*

p3in2

Units: cases/Day

phase2 errors with no effect percentage=0.756

Units: Dmnl

phase3 cases= INTEG (

cases requiring phase3 per time unit+

cases without errors after phase1 proceeding to phase3 per time unit-

cases with errors from referals per time unit-cases without errors to be discharged after Phase3,

0)

Units: cases

phase3 error detection rate1=

(1-percentage1 of improper referrals + percentage1 of improper referrals*improper referrals corrected by experts)*(1-error rate from expert)

Units: Dmnl

phase3 error detection rate2=

(1-percentage2 of improper referrals+percentage2 of improper referrals*improper referrals corrected by experts)*(1-error rate from expert)

Units: Dmnl

postanalytical errors= RANDOM UNIFORM(0.000393, 0.000674, 0)

Units: Dmnl

preanalytical errors= RANDOM UNIFORM(0.00087, 0.00254 , 0)

Units: Dmnl

ratio of hospitalisations=36/177

Units: Dmnl

ratio of unplaned revisits=141/177

Units: Dmnl

receiving treatments1 or discharged=

detected cases or recovering cases*treatment percentage1

Units: cases/Day

receiving treatments2=

unplanned hospitalisations*treatments percentage2

Units: cases/Day

receiving treatments3=

"unscheduled primary care visits, urgent care visits, or ER visits"*treatment percentage3

Units: cases/Day

recovering percentage=0

Units: Dmnl

revist patient percentage=1- recovering percentage

Units: Dmnl

test error rate=

analytical errors+preanalytical errors+postanalytical errors

Units: Dmnl

test repetition percentage=0.169

Units: Dmnl

total error rate of diagnostic infromation from referals=

pecentage3 of improper referrals*(1-improper referrals corrected by experts)+

(1-pecentage3 of improper referrals+pecentage3 of improper referrals*improper referrals corrected by experts)*error rate from expert

Units: Dmnl

treatment percentage1=1

Units: Dmnl/Day

treatment percentage3=1

Units: Dmnl/Day

treatments percentage2=1

Units: Dmnl/Day

undetected cases in revisits per time unit=

control patients discharged*revist patient percentage*(1-error detection rate)

Units: cases/Day

unplanned hospitalisations= INTEG (

unplanned hospitalisations per time unit-receiving treatments2,

0)

Units: cases

unplanned hospitalisations per time unit=

"cases with decision making errors (missed or wrongly diagnosis)"*

(1-control percentage)*ratio of hospitalisations*

percentage

Units: cases/Day

"unscheduled primary care visits, urgent care visits, or ER visits per time unit"=

"cases with decision making errors (missed or wrongly diagnosis)"*

(1-control percentage)*ratio of unplaned revisits*

percentage

Units: cases/Day

"unscheduled primary care visits, urgent care visits, or ER visits"= INTEG (

"unscheduled primary care visits, urgent care visits, or ER visits per time unit"-

receiving treatments3,

0)

Units: cases

REFERENCE

1 National Library of Medicine - Medical Subject Headings, Available from http://www.nlm.nih.gov/cgi/mesh/2013/MB_cgi.

2 Graber ML, Franklin N, Gordon R. Diagnostic error in internal medicine. Archives of internal medicine. 2005;165(13):1493–9. doi:10.1001/archinte.165.13.1493.

3 Shojania KG, Burton EC, McDonald KM, Goldman L. Changes in rates of autopsy-detected diagnostic errors over time. JAMA. 2003;289:2849- 2856.

4 Goldman L, Sayson R, Robbins S, Cohn LH, Bettmann M, Weisberg M. The value of the autopsy in three different eras. N Engl J Med. 1983;308:1000- 1005.

5 Ely JW, Graber ML, Croskerry P. Checklists to Reduce Diagnostic Errors. Academic Medicine. 2011;86(3):307-13. doi: 10.1097/ACM.0b013e31820824cd.

6 Sevdalis N, Jacklin R, Arora S, Vincent CA, Thomson, RG. Diagnostic error in a national incident reporting system in the UK. Journal of Evaluation in Clinical Practice. 2010;16(6):1276–81. doi:10.1111/j.1365-2753.2009.01328.x.

7 Kuhn GJ. Diagnostic errors. Academic Emergency Medicine. 2002;9:740–750.

8 Weeks WB, Foster T, Wallace AE, Stalhandske E. Tort Claims Analysis in the Veterans Health Administration for Quality Improvement. The Journal of Law, Medicine & Ethics. 2001;29(3-4):335-345.

9 Chandra A, Nundy S, Seabury SA. The Growth Of Physician Medical Malpractice Payments: Evidence From The National Practitioner Data Bank. Health Aff. 2005;Suppl Web Exclusive:W5-240-W5-249. doi: 10.1377/hlthaff.w5.240.

10 Controlled Risk Insurance Company (CRICO), Available from http://www.rmf.harvard.edu/About-CRICO

11 Schaefer M. Overview of CRICO Cancer-related Diagnosis Claims 1992–2001. RMF Forum. 2002;22:4–9.

12 Singh H, Graber M. Reducing Diagnostic Error Through Medical Home−Based Primary Care Reform. JAMA. 2010;304(4):463–464. doi:10.1001/jama.2010.1035.

13 Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? Academic medicine: journal of the Association of American Medical Colleges. 2002;77(10):981–992.

14 Singh H, Giardina TD, Forjuoh SN, Reis MD, Kosmach S, Khan MM, Thomas EJ. Electronic health record-based surveillance of diagnostic errors in primary care. BMJ Qual Saf. 2012;21(2):93-100.

15 Coderre S, Wright B, McLaughlin K. To think is good: Querying an initial hypothesis reduces diagnostic error in medical students. Acad Med. 2010;85(7):1125-9.

16 Eva KW. Diagnostic error in medical education: where wrongs can make rights. Advances in health sciences education : theory and practice. Advances in health sciences education. 2009;14 (Suppl 1):71-81.

17 Sherbino J, Yip S, Dore KL, Siu E, Norman GR. The effectiveness of cognitive forcing strategies to decrease diagnostic error: an exploratory study. Teach Learn Med. 2011;23(1):78-84.

18 Kleit AN, Ruiz JF. False positive mammograms and detection controlled estimation. Health Serv Res. 2003; 38(4):1207–1228.

19 Institute of Medicine. To Err is Human; Building a Safer Health System. Washington, DC: National Academy Press; 1999.

20 Schiff GD, Kim S, Abrams R, Cosby K, Lambert B, Elstein AS, et al. Diagnosing Diagnosis Errors : Lessons from a Multi-institutional Collaborative Project. Advances in Patient Safety. 2014. Available from: http://www.ahrq.gov/downloads/pub/advances/vol2/schiff.pdf

21 Institute of Medicine, Available from: http://iom.nationalacademies.org

22 Medical Errors, Scope Note, Medical Subject Headings Descripor Data. US National Library of Medicine. http://www.nlm.nih.gov/mesh/

23 Levine AI, DeMaria J S, Schwartz AD., Sim AJ. The Comprehensive Textbook of Healthcare Simulation. New York, NY: Springer Science & Business Media; *2013.*

24 Kassirer JP, Kopelman RI. Cognitive errors in diagnosis: instantiation, classification, and consequences. Am J Med. 1989;86(4):433-441.

25 Neale, G. Learning from litigation : an analysis of claims for clinical negligence. J Eval Clin Pract; 2006;12(6):665-674.

26 Zegers M, de Bruijne MC, Wagner C, Hoonhout LHF, Waaijman R, Smits M, van der Wal G. Adverse events and potentially preventable deaths in Dutch hospitals: results of a retrospective patient record review study. Quality & Safety in Health Care. 2009;18(4):297–302. doi:10.1136/qshc.2007.025924.

27 Kostopoulou O. Do GPs report diagnostic errors? Family practice. 2008;25(1): 1-2. doi:10.1093/fampra/cmn014.

28 Shojania KG, Burton EC, McDonald KM, Goldman L. Overestimation of clinical diagnostic performance caused by low necropsy rates. Quality & safety in health care. 2005;14(6):408-413. doi:10.1136/qshc.2004.011973.

29 Nolan TW. System changes to improve patient safety. BMJ. 2000;320: 771–773.

30 Leape LL, Lawthers AG, Brennan TA, Johnson WG. Preventing medical injury. QRB Qual Rev Bull. 1993;19(5):144-149.

31 Reason, J. Human error: models and management. BMJ. 2000; 320(7237): 768–770.

32 Reason J. Managing the Risks of Organizational Accidents. Brookfield, Vt: Ashgate Publishing Co; 1997.

33 Soto S, Maturana V. Diagnostic errors. Revista Medica de Chile. 2002;99(4): 289–296.

34 Newman-toker DE, Pronovost PJ. Diagnostic Errors — The Next Frontier. JAMA. 2009;301(10):1060-2. doi: 10.1001/jama.2009.249.

35 Maria A. Introduction To Modeling And Simulation. Proceedings of the 1997 Winter Simulation Conference. 1997:7–13.

36 Sommerville, I. System Models. Available from link: https://www.google.co.uk/url?sa=t&rct=j&q=&esrc=s&source=web&cd=18&cad=rja&uact=8&ved=0CH4QFjAR&url=http%3A%2F%2Fcgi.csc.liv.ac.uk%2F~igor%2FCOMP201%2FLectures%2FSE_L6.ppt&ei=dm90VcjwIln_UvTxgfAP&usg=AFQjCNG7uuG5kqKbBQ0JkFzCBeNK1nUNEw&bvm=bv.95039771,d.d24

37 Forrester JW. Industrial Dynamics. Waltham, MA: Pegasus Communications; 1961.

38 Hoyle R. Structural Equation Modeling: Concepts, Issues, and Applications. Sage Publications; 1995.

39 https://en.wikipedia.org/wiki/Structural_equation_modeling

40 Kline RB. Principles and Practice of Structural Equation Modeling. New York, NY: Guilford press; 1998.

41 Suhr D. The Basics of Structural Equation Modeling. Available from: http://www.lexjansen.com/wuss/2006/tutorials/TUT-Suhr.pdf

42 Tanaka M, Tanaka K, Takano T, Kato N, Watanabe M, Miyaoka H. Analysis of risk of medical errors using structural-equation modelling: a 6-month prospective cohort study. BMJ Qual Saf. 2012;21(9):784-90. doi: 10.1136/bmjqs-2010-048330.

43 Myers JA, Dominici F, Morlock L. Bayesian Hierarchical Models for Analysis of Medication Errors. 2010; Available from: http://www.biostat.jhsph.edu/~jamyers/causCont.pdf

44 Available from http://en.wikipedia.org/wiki/System_dynamics

45 Available from http://www.systemdynamics.org/DL-IntroSysDyn/origin.htm.

46 Meadows DH, Meadows DL, Randers J, Behrens III WW. The Limits to Growth. Universe Books; 1972

47 Senge P. The Fifth Discipline: The Art and Practice of the Learning Organization. Currency; 1990.

48 Richardson GP, Andersen DF. Teamwork in group model building. System Dynamics Review. 1995;11 (2):113–137

49 Richmond B. The strategic forum: aligning objective, strategy, and process. System Dynamics Review. 1997;13 (2):131–148.

50 Vennix J. Group model building . New York: Wiley; 1996

51 Sterman JW, Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill Education; 2000.

52 Forrester JW. The Beginning of System Dynamics. The international meeting of the System Dynamics Society Stuttgart; July 13, l989; Germany.

53 Wolstenholme EF, McKelvie D, Monk D, Todd D, Brad C. Emerging Opportunities for System Dynamics in UK Health and Social care - The Market-Pull for Systemic Thinking. Proceedings of the 2008 System Dynamics Conference. Athens. 2008

54 Hirsch G, Homer J, McDonnell G, Milstein B. Achieving Health Care Reform in the United States: Towards a Whole System Understanding. Proceedings of the InternationalSystem Dynamics Conference. Boston, USA. 2005

55 Lane DC, Monefeldt C, Rosenhead JV. Looking in the Wrong Place for Healthcare Improvements: A system dynamics study of an accident and emergency department. Journal of the Operational Research Society. 2000;51(5): 518-531.

56 Wolstenholme EF. A Case Study in Community Care using Systems Thinking. Journal of the Operational Research Society. 1993;44(9):925-934.

57 Dangerfield B, Roberts C. Optimisation as a statistical estimation tool: an example in estimating the AIDS treatment free incubation period distribution, System Dynamics Review. 1999;15(3).

58 Wolstenholme EF, Repper D, Monk D, Todd D, McKelvie D. Reforming Mental Health Services in the UK - Using System dynamics to support the design and implementation of a stepped care approach to depression in north west England. Proceedings of the 2006 System Dynamics Conference. Netherlands. 2006

59 Wolstenholme E, Mckelvie D, Smith G, Monk D. Using System Dynamics in Modelling Health and Social Care Commissioning in the UK. OLM Consulting. 2004:1–20.

60 Perelson AS. Modelling viral and immune System Dynamics. Nature Reviews Immunology. 2002;2(1):28–36. doi:10.1038/nri700.

61 Royston G, Dost A, Townshend J, Turner H. Using system dynamics to help develop and implement policies and programmes in health care in England. System Dynamics Review. 1999;15(3):293–313. doi:10.1002/(SICI)1099-1727(199923)15:3<293::AID-SDR169>3.0.CO;2-1.

62 Homer JB, Hirsch GB. System Dynamics modeling for public health: background and opportunities. American journal of public health. 2006;96(3):452–8. doi:10.2105/AJPH.2005.062059.

63 Wolstenholme EF. System enquiry: a system dynamics approach. New York, NY: Wiley; 1990.

64 Roberts N, Anderson D, Deal R, Garet M, Shaffer W. Introduction to computer simulation, a System Dynamics modelling approach. Reading, MA: Addison-Wesley; 1983.

65 Adopted from https://en.wikipedia.org/wiki/Goodness_of_fit

66 Vensim; Available from http://vensim.com/

67 SPSS; Available from http://www-01.ibm.com/software/uk/analytics/spss/

68 Devilee, JLA, Knol, AB. Software to support expert elicitation An exploratory study of existing software packages. RIVM official reports. 2011.

69 Carson ED, Cobelli C, Finkelstein L. The mathematical modeling of metabolic and endocrine systems: model formulation, identification, and validation. New York, NY: Wiley; 1983

70 Guo S, Roudsari A, Garcez A. A causal loop approach to the study of diagnostic errors. Studies in health technology and informatics. 2014;205:73-77.

71 Definition Available from https://www.nlm.nih.gov/mesh/

72 Vaos G, Skondras C. Treves' field congenital hernias in children: an unsuspected rare cause of acute small bowel obstruction. Pediatric Surgery International. 2007;23(4): 6. doi:http://dx.doi.org/10.1007/s00383-007-1877-y.

73 Usher-Smith JA, Thompson MJ, Sharp SJ, Walter FM. Factors associated with the presence of diabetic ketoacidosis at diagnosis of diabetes in children and young adults: a systematic review. BMJ. 2011;343:d4092. doi: http://dx.doi.org/10.1136/bmj.d4092.

74 Trindade MA, Varella TC, Cisneros CG, Bottini V, Moura AK. Delayed diagnosis of multibacillary leprosy: A report of eight cases. Braz J Infect Dis. 2009;13(2):155-7.

75 Parish LC. Hereditary angioedema: Diagnosis and management-a perspective for the dermatologist. J Am Acad Dermatol. 2011 Oct;65(4):843-50. doi: 10.1016/j.jaad.2010.09.715. Epub 2011 May 6.

76 Hughes AJ, Ben-Shlomo Y, Daniel SE, Lees AJ. What features improve the accuracy of clinical diagnosis in Parkinson's disease,A clinicopathologic study. Neurology. 1992;42(6):1142-6

77 Basnet R, Hinderaker SG, Enarson D, Malla P, Mørkve O. Delay in the diagnosis of tuberculosis in Nepal. BMC Public Health. 2009;9:236. doi: 10.1186/1471-2458-9-236.

78 Golub JE, Bur S, Cronin WA, Gange S, Baruch N, Comstock GW, Chaisson RE. Patient and health care system delays in pulmonary tuberculosis diagnosis in a low-incidence state. Int J Tuberc Lung Dis. 2005;9(9):992-8.

79 Bertz,Rebecca Ann Sherwood, First do no harm: A preliminary diagnostic error prevention case study conducted in the technical support department of an in vitro medical diagnostic instrumentation company. 2010

80 Eva KW. Diagnostic error in medical education: where wrongs can make rights. Advances in health sciences education : theory and practice. 2009;14(Suppl 1):71–81. doi:10.1007/s10459-009-9188-9.

81 Van Alboom E, Louis P, Van Zandijcke M, Crevits L, Vakaet A, Paemeleire K. Diagnostic and therapeutic trajectory of cluster headache patients in Flanders. Acta Neurol Belg. 2009;109(1):10-7.

82 Singh T, Rajput M. Misdiagnosis of Bipolar Disorder. Psychiatry (Edgmont). 2006;3(10): 57–63.

83 Kleit AN, Ruiz JF. False positive mammograms and detection controlled estimation. Health Serv Res. 2003;38(4):1207–1228. doi:10.1111/1475-6773.00170

84 Klein RD, Mahoney MJ. Medical legal issues in prenatal diagnosis. Clin Perinatol. 2007;34(2):287-97.

85 Wood DS, Tracey TJG. A brief feedback intervention for diagnostic overshadowing. Training and Education in Professional Psychology. 2009;3(4):218-225.

86 Wahls TL, Cram P. Proposed interventions to decrease the frequency of missed test results. Adv Health Sci Educ Theory Pract. 2009;14(Suppl 1):51-6. doi: 10.1007/s10459-009-9180-4. Epub 2009 Aug 11.

87 Allen LK, Hatfield JM, DeVetten G, Ho JC, Manyama M. Reducing malaria misdiagnosis: the importance of correctly interpreting ParaCheck Pf (R) "faint test bands" in a low transmission area of Tanzania. BMC Infect Dis. 2011;11:308.

88 Guly HR. Diagnostic errors in an accident and emergency. Emergency Medicine Journal. 2001;18(4):263–269.

89 Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, Elstein AS, Hasler S, Kabongo ML, Krosnjar N, Odwazny R, Wisniewski MF, McNutt RA. Diagnostic error in medicine: analysis of 583 physician-reported errors. Archives of Internal Medicine. 2009;169(20):1881–7. doi:10.1001/archinternmed.2009.333.

90 Ely JW., Kaldjian LC, D'Alessandro DM. Diagnostic errors in primary care: lessons learned. Journal of the American Board of Family Medicine : JABFM. 2012;25(1):87–97. doi:10.3122/jabfm.2012.01.110174.

91 Macdonald OW. Physician Perspectives on Preventing Diagnostic Errors. Waltham, MA: Quantia MD; 2011.

92 Morris AH. Decision support and safety of clinical environment. Qual Saf Health Care. 2002;11:69-75. *doi:10.1136/qhc.11.1.69.*

93 McGee RG, Neuen BL, Mitchell RL, Craig JC, Webster AC. Diagnostic test studies in nephrology: quantity, quality, and scope. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2011;58(6):921–7. doi:10.1053/j.ajkd.2011.06.019.

94 Feldman MJ, Hoffer EP, Barnett GO, Kim RJ, Famiglietti KT, Chueh H. Presence of key findings in the medical record prior to a documented high-risk diagnosis. Journal of the American Medical Informatics Association. J Am Med Inform Assoc. 2012;19(4):591-6. doi: 10.1136/amiajnl-2011-000375. Epub 2012 Mar 19.

95 Arzy S, Brezis M, Khoury S, Simon SR, Ben-Hur T. Misleading one detail: a preventable mode of diagnostic error. J Eval Clin Pract. 2009;15(5):804-6. doi: 10.1111/j.1365-2753.2008.01098.x.

96 Roy E. Cognitive factors. Encyclopedia of Behavioral.2013:447-448.

97 Phua DH, Tan NC. Cognitive aspect of diagnostic errors. Ann Acad Med Singapore. 2013;42(1):33-41.

98 Kahnemann D, Slovic P, Tversky A. Judgment Under Uncertainty: Heuristics and Biases. New York, NY: Cambridge University Press; 1982.

99 Mamede S, van Gog T, van den Berge K, Rikers RM, van Saase JL, van Guldener C, Schmidt HG. Effect of Availability Bias and Reflective Reasoning on Diagnostic Accuracy Among Internal Medicine Residents. JAMA: Journal of the American Medical Association. 2010;304:1198–1203.

100 Knottnerus JA. Interpretation of diagnostic data: an unexplored field in general practice. J R Coll Gen Pract. 1985;35(275):270-4.

101 Sherbino J, Yip S, Dore KL, Siu E, Norman GR. The effectiveness of cognitive forcing strategies to decrease diagnostic error: an exploratory study. Teach Learn Med. 2011;23(1):78-84. doi:10.1080/10401334.2011.536897.

102 Goodson WH, Moore DH. Causes of physician delay in the diagnosis of breast cancer. Arch Intern Med. 2002;162(12):1343-8.

103 Croskerry. The Importance of Cognitive Errors in Diagnosis and Strategies to Minimize Them. Acad Med. 2003;78(8):775-80.

104 Berner ES., Graber ML. Overconfidence as a Cause of Diagnostic Error in Medicine. Am J Med. 2008;121(5 Suppl):S2-23. doi: 10.1016/j.amjmed.2008.01.001.

105 Mendel R, Traut-Mattausch E, Jonas E, Leucht S, Kane JM, Maino K, Kissling W, Hamann J. Confirmation bias: Why psychiatrists stick to wrong preliminary diagnoses. Psychol Med. 2011;41(12):2651-9. doi: 10.1017/S0033291711000808. Epub 2011 May 20.

106 Sibbald M, Cavalcanti RB. The biasing effect of clinical history on physical examination diagnostic accuracy. Medical education. 2011;45(8): 827–34. doi:10.1111/j.1365-2923.2011.03997.x

107 Brigo F. We should not treat the EEG, but we should read it blind to the patient's history. Epilepsy Behav. 2011;20(1):146. doi: 10.1016/j.yebeh.2010.10.026. Epub 2010 Dec 3.

108 Trotter MJ, Bruecks AK. Interpretation of skin biopsies by general pathologists: diagnostic discrepancy rate measured by blinded review. Arch Pathol Lab Med. 2003;127(11):1489-92.

109 Coderre S, Wright B, McLaughlin K. To think is good: Querying an initial hypothesis reduces diagnostic error in medical students. Acad Med. 2010;85(7):1125-9. doi: 10.1097/ACM.0b013e3181e1b229.

110 de Lusignan S, Khunti K, Belsey J, Hattersley A, van Vlymen J, Gallagher H, Millett C, Hague NJ, Tomson C, Harris K, Majeed A. A method of identifying and correcting miscoding, misclassification and misdiagnosis in diabetes: a pilot and validation study of routinely collected data. Diabet Med. 2010;27(2):203-9. doi: 10.1111/j.1464-5491.2009.02917.x.

111 Parnes B, Fernald D, Quintela J, Araya-Guerra R, Westfall J, Harris D, Pace W. Stopping the error cascade: a report on ameliorators from the ASIPS collaborative. Qual Saf Health Care. 2007;16(1):12-6.

112 Saiz Díaz RA, Martínez Bermejo A, Gómez Alonso J. Diagnosis of epilepsy over the course of the disease. Neurologist. 2007;13(6 Suppl 1):S11-9. doi: 10.1097/NRL.0b013e31815bb146.

113 Youk JH, Kim EK, Kim MJ, Lee JY, Oh KK. Missed breast cancers at US-guided core needle biopsy: how to reduce them. Radiographics. 2007;27(1):79-94.

114 Wahls TL, Peleg I. Patient- and system-related barriers for the earlier diagnosis of colorectal cancer. BMC Fam Pract. 2009;10:65. doi: 10.1186/1471-2296-10-65.

115 Brennan PC, McEntee M, Evanoff M, Phillips P, O'Connor WT, Manning DJ. Ambient lighting: effect of illumination on soft-copy viewing of radiographs of the wrist. AJR Am J Roentgenol. 2007;188(2):W177-80.

116 Schiff GD, Hasan O, Kim S, Abrams R, Cosby K, Lambert BL, McNutt RA. Diagnostic error in medicine: analysis of 583 physician-reported errors. Archives of Internal Medicine. 2009;169(20):1881–7. doi:10.1001/archinternmed.2009.333

117 Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care--a systematic review. Family Practice. 2008;25(6):400–13. doi:10.1093/fampra/cmn071.

118 Schiff GD. Finding and fixing diagnosis errors : can triggers help ? BMJ Qual Saf. 2012;21(2):89-92. doi: 10.1136/bmjqs-2011-000590. Epub 2011 Nov 30.

119 Schiefer TK, Parker WL, Anakwenze OA, Amadio PC, Inwards CY, Spinner RJ. Extradigital glomus tumors: A 20-year experience. Mayo Clin Proc. 2006;81(10):1337-44.

120 Cree C, Pillai A, Jones B, Blyth M. Bilateral patellar tendon ruptures: a missed diagnosis : case report and literature review. Knee Surg Sports Traumatol Arthrosc. 2007;15(11):1350-4. Epub 2007 Jun 29.

121 Lovelock CE, Redgrave JN, Briley D, Rothwell PM. The SCAN rule: a clinical rule to reduce CT misdiagnosis of intracerebral haemorrhage in minor stroke. J Neurol Neurosurg Psychiatry. 2010;81(3):271-5. doi: 10.1136/jnnp.2008.169227. Epub 2009 Aug 16.

122 Bairati I, Jobin E, Fillion L, Larochelle M, Vincent L. Determinants of delay for breast cancer diagnosis. Cancer Detect Prev. 2007;31(4):323-31.

123 Kleit AN, Ruiz JF. False positive mammograms and detection controlled estimation. Health Serv Res. 2003;38(4):1207-28.

124 Ely JW, Graber ML, Croskerry P. Checklists to reduce diagnostic errors. Acad Med. 2011;86(3):307-13. doi: 10.1097/ACM.0b013e31820824cd.

125 Singh H, Petersen LA, Thomas EJ. Understanding diagnostic errors in medicine: a lesson from aviation. Qual Saf Health Care. 2006;15(3):159-64.

126 A H Morris. Decision support and safety of clinical environments. Qual Saf Health Care. 2002;11(1):69–75. doi:10.1136/qhc.11.1.69.

127 McGee RG, Neuen BL, Mitchell RL, Craig JC, Webster AC. Diagnostic test studies in nephrology: quantity, quality, and scope. American journal of kidney diseases : the official journal of the National Kidney Foundation. 2011;58(6):921–7. doi:10.1053/j.ajkd.2011.06.019.

128 Huang Z, Zuo C, Guan Y, Zhang Z, Liu P, Xue F, Lin X, Misdiagnoses of 11C-choline combined with 18F-FDG PET imaging in brain tumours. Nucl Med Commun. 2008;29(4):354-8. doi: 10.1097/MNM.0b013e3282f4a21e.

129 Wahls TL, Cram P. Proposed interventions to decrease the frequency of missed test results. Adv Health Sci Educ Theory Pract. 2009;14 (Suppl 1):51-6. doi: 10.1007/s10459-009-9180-4. Epub 2009 Aug 11.

130 Blaya JA, Shin SS, Yale G, Suarez C, Asencios L, Contreras C, Rodriguez P, Kim J, Cegielski P, Fraser HS. Electronic laboratory system reduces errors in National Tuberculosis Program: a cluster randomized controlled trial. Int J Tuberc Lung Dis. 2010;14(8):1009-15.

131 Chen YW, Larbani M, Hsieh CY, Chen CW. Introduction of affinity set and its application in data-mining example of delayed diagnosis---- Also see: Using Affinity Set on Finding the Key Attributes of Delayed Diagnosis. Expert Systems with Applications. 2009;36(8):10883-10889. Retrieved from http://www.sciencedirect.com/science/article/pii/S095741740900178X.

132 De Simone R, Coppola G, Ranieri A, Bussone G, Cortelli P, D'Amico D, d'Onofrio F, Manzoni GC, Marano E, Perini F, Torelli P, Beneduce L, Ciccarelli G, Mea E, Penza P, Ripa P, Sancisi E, Bonavita V. Validation of AIDA Cefalee, a computer-assisted diagnosis database for the management of headache patients. Neurol Sci. 2007;28(Suppl 2):S213-6.

133 Tang H, Ng JHK. Googling for a diagnosis--use of Google as a diagnostic aid: internet based study. BMJ (Clinical research ed.). 200;333(7579):1143-5. doi:10.1136/bmj.39003.640567.AE.

134 Bertaud-Gounot V, Duvauferrier R, Burgun A. Ontology and medical diagnosis. Inform Health Soc Care. 2012;37(2):51-61.

135 Singh H, Graber ML, Kissam SM, Sorensen AV., Lenfestey NF., Tant EM, Henriksen K, et al. System-related interventions to reduce diagnostic errors : a narrative review. BMJ Qual Saf. 2012;21(2):160-70. doi: 10.1136/bmjqs-2011-000150. Epub 2011 Nov 30.

136 Mar CD, Doust J, Glasziou P. Clinical thinking : evidence, communication and decision-making, Oxford: Blackwell; 2006.

137 Guo S, Roudsari A, Garcez A. Modelling clinical diagnostic errors: a system dynamics approach. Studies in health technology and informatics. 2015;208:160-164.

138 Klein, Gary A. Sources of power: how people make decisions. Cambridge, MA: MIT press; 1999.

139 Chatten K, Howe M, Marks G, Smith T, Noble L. MBBS Year 4 GUIDE TO HISTORY TAKING AND EXAMINATION 2012-2013. Retrieved from http://www.ucl.ac.uk/pcph/undergrad/cbt/year4/history-and-examination.

140 Shapiro DE. The interpretation of diagnostic tests. Statistical Methods in Medical Research. 1999; 8:113-134.

141 Wians FH. Clinical Laboratory Tests: Which, Why, and What Do The Results Mean? Laboratory Medicine. 2009;40(2):105–113. doi:10.1309/LM4O4L0HHUTWWUDD.

142 Hampton JR, Harrison MJ , Mitchell JR., Prichard JS., Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. British Medical Journal. 1975; 2(5969):486–489. doi:10.1136/bmj.2.5969.486.

143 Beckman HB, Frankel RM The effect of physician behavior on the collection of data. Ann Intern Med. 1984;101(5):692-6.

144 Pinto A, Brunese L. Spectrum of diagnostic errors in radiology. World Journal of Radiology. 2010;2(10):377–383. doi:10.4329/wjr.v2.i10.377.

145 Brady A, Laoide RÓ, McCarthy P ,McDermott R. Discrepancy and error in radiology: Concepts, causes and consequences. Ulster Medical Journal. 2012;81(1): 3–9.

146 Epner PL, Gans JE, Graber ML. When diagnostic testing leads to harm: a new outcomes-based approach for laboratory medicine. BMJ Qual Saf. 2013;22(Suppl 2):6-10. doi: 10.1136/bmjqs-2012-001621. Epub 2013 Aug 16.

147 Available from http://www-03.ibm.com/software/products/en/spss-modeler.

148 Pagano RR. Understanding statistics in the behavioral sciences (9th Edition). Belmont, CA: Wadsworth Publishing; 2009.

149 Hauke J, Kossowski T. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data. Quaestiones Geographicae. 2011;30(2):87-93. doi: 10.2478/v10117-011-0021-1.

150 Field A. Discovering statistics using SPSS (3rd Edition). London: Sage Publications; 2009.

151 Available from http://www.sussex.ac.uk.

152 Biau DJ, Jolles BM, Porcher R. P value and the theory of hypothesis testing: an explanation for new researchers. Clin Orthop Relat Res. 2010;468(3):885-92. doi: 10.1007/s11999-009-1164-4.

153 IBM Knowledge Center, Available from http://www-01.ibm.com/support/knowledgecenter/#!/SSLVMB_20.0.0/com.ibm.spss.statistics.help/curve_estimation_models.htm.

154 Available from http://www.medicine.ox.ac.uk/bandolier/booth/glossary/pvalue.html.

155 Available from http://www.statistica.com.au/regression.html.

156 Winter, B, The F distribution and the basic principle behind ANOVA. Available from http://www.bodowinter.com/tutorial/bw_anova_general.pdf.

157 Simel DL, Rennie D., Bossuyt PMM. The STARD statement for reporting diagnostic accuracy studies: Application to the history and physical examination. Journal of General Internal Medicine. 2008;23(6):768–774. doi:10.1007/s11606-008-0583-3.

158 Access via http://www.statsoft.com/Textbook/Multiple-Regression.

159 Available from http://support.minitab.com/en-us/minitab/17/topic-library/modeling-statistics/multivariate/principal-components-and-factor-analysis/what-is-pca/.

160 Available from ftp://statgen.ncsu.edu/pub/thorne/molevoclass/AtchleyOct19.pdf.

161 Brown JD. Choosing the Right Type of Rotation in PCA and EFA. JALT Testing & Evaluation SIG Newsletter. 2009;13(3):20-25.

162 Child D. The Essentials of Factor Analysis (3rd ed.). New York, NY: Continuum International; 2006.

163 Guo S, Roudsari A, Garcez A. A System Dynamics Approach to Analyze Laboratory Test Errors. Studies in health technology and informatics, 26th European Medical Informatics Conference. 2015;210:266-70.

164 Brailsford SC. System Dynamics: What's in it for healthcare simulation modelers. 2008 Winter simulation conference. 2008.

165 Available from https://www.rmf.harvard.edu/.

166 Available from http://www.commonwealthfund.org/Maps-and-Data.aspx.

167 Available from http://www.ahrq.gov/.

168 Available from http://www.clinicaltrials.gov/ct2/about-site.

169 Available from http://www.cihi.ca/CIHI-ext-portal/internet/EN/Home/home/cihi000001.

170 Radiology Quality Inistitute. Diagnostic Accuracy in Radiology: Defining a Literature-Based Benchmark. 2012. Available from http://www.radisphereradiology.com/wp-content/uploads/Diagnostic-Accuracy-in-Radiology.

171 Abdollahi A, Saffar H, Saffar H. Types and frequency of errors during different phases of testing at a clinical medical laboratory of a teaching hospital in Tehran, Iran. N Am J Med Sci. 2014;6(5):224-8. doi:10.4103/1947-2714.132941.

172 Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem. 2007;53(7):1338-42.

173 Wiwanitkit V. Types and frequency of preanalytical mistakes in the first Thai ISO 9002:1994 certified clinical laboratory, a 6 – month monitoring. BMC Clin Pathol. 2001;1(1):5.

174 Stahla M , Lund ED. Brandslund I. Reasons for a laboratory's inability to report results for requested analytical tests. Clin Chem. 1998;44(10):2195-7.

175 Plebani M, Carraro P. Mistakes in a stat laboratory : types and frequency. Clin Chem. 1997;43(8 Pt 1):1348-51.

176 Nutting PA, Main DS, Fischer PM, Stull TM, Pontious M, Seifert M, et al. Problems in laboratory testing in primary care. JAMA. 1996;275:635–9.

177 Lapworth R, Teal TK. Laboratory blunders revisited. Ann Clin Biochem. 1994;31:78–84.

178 Knol, A.B., Slottje, P., van der Sluijs, J.P., Lebret, E. The use of expert elicitation in environmental health impact assessment: a seven step procedure. Environ Health. 2010;9:19. doi: 10.1186/1476-069X-9-19.

179 Skype application. Available from: http://www.skype.com/en/.

180 Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. International Journal of Forecasting. 1999;15(4):353-375.

181 Van der Fels-Klerx HJ, Cooke RM, Nauta MN, Goossens LH, Havelaar AH. A Structured Expert Judgment Study for a Model of Campylobacter Transmission During Broiler-Chicken Processing. Risk Anal. 2005;25(1):109-24.

182 Devilee JLA, Knol AB. Software to support expert elicitation: An exploratory study of existing software packages. RIVM Report. 2011: 630003001.

183 Relevant available software list: http://www.expertsinuncertainty.net/Software/tabid/4149/Default.aspx.

184 EXCALIBUR-Integrated System for Processing Expert Judgements http://www.lighttwist.net/wp/excalibur.

185 Vensim help. Available from https://www.vensim.com/documentation/index.html?users_guide.htm.

186 Shah C, Kumar V, Knoche C, Simulation in Medical Education. International Journal of Basic and Applied Physiology. 2012;1(1):167–170.

187 Available from https://en.wikipedia.org/wiki/Goodness_of_fit.

188 Jay F, and Senge P. Tests for building confidence in S-D models. TIMS Studies in the Management Sciences. 1980;14:208-228.

189 George R, Pugh A, Introduction to System Dynamics Modeling with DYNAMO. Cambridge, MA: MIT Press; 1981.

190 Wakeland W, Hoarfrost M. The Case for Thoroughly Testing Complex System Dynamics Models. 2005

191 Yaman B. Formal Aspects of Model Validity and Validation in System Dynamics. System Dynamics Review. 1996;13(3):183-210.

192 Vensim help. Available from https://www.vensim.com/documentation/index.html?ref_time_step.htm.