



City Research Online

## City, University of London Institutional Repository

---

**Citation:** Williams, L. R., Ali, M., VandenBerg, K., Williams, L. J., Abo, M., Becker, F., Bowen, A., Brandenburg, C., Breitenstein, C., Bruehl, S., et al (2022). Utilising a systematic review-based approach to create a database of individual participant data for meta- and network meta-analyses: the RELEASE database of aphasia after stroke. *Aphasiology*, 36(4), pp. 513-533. doi: 10.1080/02687038.2021.1897081

This is the published version of the paper.

This version of the publication may differ from the final published version.

---

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/26982/>

**Link to published version:** <https://doi.org/10.1080/02687038.2021.1897081>

**Copyright:** City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

**Reuse:** Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

---

City Research Online:

<http://openaccess.city.ac.uk/>

[publications@city.ac.uk](mailto:publications@city.ac.uk)

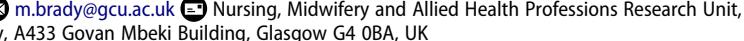
---



## Utilising a systematic review-based approach to create a database of individual participant data for meta- and network meta-analyses: the RELEASE database of aphasia after stroke

Louise R Williams <sup>a</sup>, Myzoon Ali <sup>a</sup>, Kathryn VandenBerg <sup>a</sup>, Linda J Williams <sup>b</sup>, Masahiro Abo <sup>c</sup>, Frank Becker <sup>d</sup>, Audrey Bowen <sup>e</sup>, Caitlin Brandenburg <sup>f</sup>, Caterina Breitenstein <sup>g</sup>, Stefanie Bruehl <sup>h</sup>, David A Copland <sup>f</sup>, Tamara B Cranfill <sup>i</sup>, Marie Di Pietro-Bachmann <sup>j</sup>, Pamela Enderby <sup>k</sup>, Joanne Fillingham <sup>l</sup>, Federica Lucia Galli <sup>m</sup>, Marialuisa Gandolfi <sup>n</sup>, Bertrand Glize <sup>o,p</sup>, Erin Godecke <sup>q</sup>, Neil Hawkins <sup>r</sup>, Katerina Hilari <sup>s</sup>, Jacqueline Hinckley <sup>t</sup>, Simon Horton <sup>u</sup>, David Howard <sup>v</sup>, Petra Jaecks <sup>w</sup>, Elizabeth Jefferies <sup>x</sup>, Luis MT Jesus <sup>y</sup>, Maria Kambanaros <sup>z</sup>, Eun Kyoung Kang <sup>aa</sup>, Eman M Khedr <sup>bb</sup>, Anthony Pak-Hin Kong <sup>cc</sup>, Tarja Kukkonen <sup>dd</sup>, Marina Laganaro <sup>ee</sup>, Matthew A Lambon Ralph <sup>ff</sup>, Ann Charlotte Laska <sup>gg</sup>, Béatrice Leemann <sup>hh</sup>, Alexander P Leff <sup>ii</sup>, Roxele Ribeiro Lima <sup>jj</sup>, Antje Lorenz <sup>kk</sup>, Brian MacWhinney <sup>ll</sup>, Rebecca Shisler Marshall <sup>mm</sup>, Flavia Mattioli <sup>nn</sup>, İlknur Maviş <sup>oo</sup>, Marcus Meinzer <sup>pp</sup>, Reza Nilipour <sup>qq</sup>, Enrique Noé <sup>rr</sup>, Nam-Jong Paik <sup>ss</sup>, Rebecca Palmer <sup>k</sup>, Ilias Papathanasiou <sup>tt</sup>, Brigida F Patricio <sup>uu</sup>, Isabel Pavão Martins <sup>vv</sup>, Cathy Price <sup>ww</sup>, Tatjana Prizl Jakovac <sup>xx</sup>, Elizabeth Rochon <sup>yy</sup>, Miranda L Rose <sup>zz</sup>, Charlotte Rosso <sup>aaa</sup>, Ilona Rubi-Fessen <sup>bbb</sup>, Marina B Ruiter <sup>ccc</sup>, Claerwen Snell <sup>ddd</sup>, Benjamin Stahl <sup>eee</sup>, Jerzy P Szaflarski <sup>fff</sup>, Shirley A Thomas <sup>ggg</sup>, Mieke Van De Sandt-Koenderman <sup>hhh</sup>, Ineke Van Der Meulen <sup>hhh</sup>, Evy Visch-Brink <sup>iii</sup>, Linda Worrall <sup>f</sup>, Heather Harris Wright <sup>jjj</sup>, Marian C Brady <sup>a</sup> and The RELEASE Collaborators

<sup>a</sup>Nursing Midwifery and Allied Health Professions Research Unit, Glasgow Caledonian University, Glasgow, UK; <sup>b</sup>Usher Institute, University of Edinburgh, Edinburgh, UK; <sup>c</sup>Department of Rehabilitation Medicine, The Jikei University School of Medicine, Tokyo, Japan; <sup>d</sup>Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, and Sunnaas Rehabilitation Hospital, Nesoddtangen, Norway; <sup>e</sup>Geoffrey Jefferson Brain Research Centre, Manchester Academic Health Science Centre, Northern Care Alliance & University of Manchester, Manchester, UK; <sup>f</sup>School of Health and Rehabilitation Sciences, the University of Queensland, Brisbane, Australia; <sup>g</sup>Department of Neurology with Institute of Translational Neurology, University of Muenster, Muenster, Germany; <sup>h</sup>St Mauritius Rehabilitation Centre, Meerbusch, Germany; <sup>i</sup>Clinical and Cognitive Neurosciences, Department of Neurology, Medical Faculty, RWTH Aachen University, Aachen, Germany; <sup>j</sup>Division of Neuroscience and Experimental Psychology, School of Biological Sciences, University of Manchester, UK; <sup>k</sup>Department of Communication Sciences and Disorders, Eastern Kentucky University, Richmond, Kentucky, USA; <sup>l</sup>Division of Neurorehabilitation, Department of Clinical Neurosciences, University Hospitals and University of Geneva, Geneva, Switzerland; <sup>m</sup>School of Health and Related Research, University of Sheffield, Sheffield, UK; <sup>n</sup>Nursing Directorate, NHS Improvement, London, UK; <sup>o</sup>Azienda Sanitaria Unica Regionale - ASUR, Marche Polytechnic University, Ancona, Italy; <sup>p</sup>Department of Neuroscience, Biomedicine and Movement Sciences, University of Verona, Verona, Italy; <sup>q</sup>EA 4136 Handicap Activity Cognition Health, University of Bordeaux, France; <sup>r</sup>Department of Physical Medicine and Rehabilitation, CHU De Bordeaux, France; <sup>s</sup>School of Medical and Health Sciences, Speech Pathology, Edith Cowan University and Sir Charles Gairdner Hospital, Perth, Australia; <sup>t</sup>Institute of Health & Wellbeing, University of Glasgow, Glasgow, UK; <sup>u</sup>Centre for Language and Communication Science Research, City, University of London, London, UK; <sup>v</sup>Nova Southeastern University, Fort Lauderdale, Florida, USA; <sup>w</sup>School of Health Sciences, University of East Anglia,

**CONTACT** Marian C Brady  [m.brady@gcu.ac.uk](mailto:m.brady@gcu.ac.uk) 

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.  
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Norwich, UK; <sup>v</sup>School of Education Communication and Language Sciences, Newcastle University, Newcastle-Upon-Tyne, UK; <sup>w</sup>Faculty of Linguistics and Literary Studies, Bielefeld University, Bielefeld, Germany; <sup>x</sup>Department of Psychology, University of York, York, UK; <sup>y</sup>School of Health Sciences (ESSUA) and Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro, Aveiro, Portugal; <sup>z</sup>Department of Rehabilitation Sciences, Cyprus University of Technology, Limassol, Cyprus; <sup>aa</sup>Department of Rehabilitation Medicine, Gangwon National University Hospital, Gangwon-do, Republic of Korea; <sup>bb</sup>Department of Neurology, Assiut University Hospital, Assiut, Egypt; <sup>cc</sup>School of Communication Sciences and Disorders, University of Central Florida, Orlando, USA; <sup>dd</sup>ENT/Department of Phoniatriy, Tampere University Hospital, Tampere, Finland; <sup>ee</sup>Faculty of Psychology and Educational Science, University of Geneva, Geneva, Switzerland; <sup>ff</sup>MRC Cognition and Brain Sciences Unit, University of Cambridge, Cambridge, UK; <sup>gg</sup>Department of Clinical Sciences, Karolinska Institutet, Danderyd Hospital, Stockholm, Sweden; <sup>hh</sup>Neurorééducation, Département Des Neurosciences Cliniques, Hôpitaux Universitaires De Genève, Geneva, Switzerland; <sup>ii</sup>Department of Brain Repair and Rehabilitation, Institute of Neurology, UCL, London, UK; <sup>jj</sup>Department of Speech Language Pathology, Educational Association Bom Jesus – IELUSC, Santa Catarina, Brazil; <sup>kk</sup>Department of Psychology, Humboldt-Universität zu Berlin, Berlin, Germany; <sup>ll</sup>Department of Psychology, Carnegie Mellon University, Pittsburgh, USA; <sup>mm</sup>Department of Communication Sciences and Special Education, University of Georgia, Georgia, USA; <sup>nn</sup>Neuropsychology Unit, ASST Spedali Civili of Brescia, Brescia, Italy; <sup>oo</sup>Department of Speech and Language Therapy, Anadolu University, Eskişehir, Turkey; <sup>pp</sup>Department of Neurology, University Medicine Greifswald, Greifswald, Germany; <sup>qq</sup>Department of Speech Therapy, University of Social Welfare and Rehabilitation Sciences, Tehran, Iran; <sup>rr</sup>NEURORHB-Hospitales Vithas, Valencia, Spain; <sup>ss</sup>Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul National University Bundang Hospital, Seongnam, Republic of Korea; <sup>tt</sup>Department of Speech and Language Therapy, University of Patras, Patras, Greece; <sup>uu</sup>Department of Speech Therapy, School of Health, Polytechnic Institute of Porto, Porto, Portugal; <sup>vv</sup>Laboratório De Estudos De Linguagem, Faculdade De Medicina De Lisboa, Universidade De Lisboa, Lisbon, Portugal; <sup>ww</sup>Wellcome Centre for Human Neuroimaging, UCL, London, UK; <sup>xx</sup>Department of Speech and Language Pathology, Faculty of Education and Rehabilitation Sciences, University of Zagreb, Zagreb, Croatia; <sup>yy</sup>Department of Speech-Language Pathology and Rehabilitation Sciences Institute, University of Toronto, and Toronto Rehabilitation Institute, Toronto, Canada; <sup>zz</sup>Centre of Research Excellence in Aphasia Recovery and Rehabilitation, School of Allied Health, Human Services and Sport, La Trobe University, Melbourne, Australia; <sup>aaa</sup>Institut Du Cerveau Et Del La Moelle Épineière, Sorbonne University, and APHP, Urgences Cérébro-Vasculaires, Hôpital De La Pitié Salpêtrière, Paris, France; <sup>bbb</sup>RehaNova Rehabilitation Hospital and Department of Special Education and Rehabilitation, University of Cologne, Cologne, Germany; <sup>ccc</sup>Sint Maartenskliniek, Rehabilitation Centre and Centre for Language Studies, Radboud University, Nijmegen, The Netherlands; <sup>ddd</sup>Warrington and Halton Teaching Hospitals NHS Foundation Trust, Warrington, UK; <sup>eee</sup>Department of Neurology, Charité - Universitätsmedizin Berlin, Berlin, Germany; <sup>fff</sup>UAB Epilepsy Centre, Department of Neurology, University of Alabama at Birmingham, Alabama, USA; <sup>ggg</sup>Division of Rehabilitation & Ageing; School of Medicine, University of Nottingham, Nottingham, UK; <sup>hhh</sup>Department of Rehabilitation, Rijndam rehabilitation centre, Rotterdam, the Netherlands AND Erasmus MC University Medical Centre, Rotterdam, The Netherlands; <sup>iii</sup>Department of Neurology and Neurosurgery, Erasmus University Medical Center, Rotterdam, Netherlands; <sup>jjj</sup>College of Allied Health Sciences, East Carolina University, Greenville, North Carolina, USA

## ABSTRACT

**Background:** Collation of aphasia research data across settings, countries and study designs using big data principles will support analyses across different language modalities, levels of impairment, and therapy interventions in this heterogeneous population. Big data approaches in aphasia research may support vital analyses, which are unachievable within individual trial datasets. However, we lack insight into the requirements for a systematically created database, the feasibility and challenges and potential utility of the type of data collated.

**Aim:** To report the development, preparation and establishment of an internationally agreed aphasia after stroke research database of individual participant data (IPD) to facilitate planned aphasia research analyses.

## ARTICLE HISTORY

Received 17 August 2020  
Accepted 28 January 2021

## KEYWORDS

Stroke; aphasia; individual participant data; speech and language therapy; rehabilitation reporting standards

**Methods:** Data were collated by systematically identifying existing, eligible studies in any language ( $\geq 10$  IPD, data on time since stroke, and language performance) and included sourcing from relevant aphasia research networks. We invited electronic contributions and also extracted IPD from the public domain. Data were assessed for completeness, validity of value-ranges within variables, and described according to pre-defined categories of demographic data, therapy descriptions, and language domain measurements. We cleaned, clarified, imputed and standardised relevant data in collaboration with the original study investigators. We presented participant, language, stroke, and therapy data characteristics of the final database using summary statistics.

**Results:** From 5256 screened records, 698 datasets were potentially eligible for inclusion; 174 datasets (5928 IPD) from 28 countries were included, 47/174 RCT datasets (1778 IPD) and 91/174 (2834 IPD) included a speech and language therapy (SLT) intervention. Participants' median age was 63 years (interquartile range [53, 72]), 3407 (61.4%) were male and median recruitment time was 321 days (IQR 30, 1156) after stroke. IPD were available for aphasia severity or ability overall ( $n = 2699$ ; 80 datasets), naming ( $n = 2886$ ; 75 datasets), auditory comprehension ( $n = 2750$ ; 71 datasets), functional communication ( $n = 1591$ ; 29 datasets), reading ( $n = 770$ ; 12 datasets) and writing ( $n = 724$ ; 13 datasets). Information on SLT interventions were described by theoretical approach, therapy target, mode of delivery, setting and provider. Therapy regimen was described according to intensity (1882 IPD; 60 datasets), frequency (2057 IPD; 66 datasets), duration (1960 IPD; 64 datasets) and dosage (1978 IPD; 62 datasets).

**Discussion:** Our international IPD archive demonstrates the application of big data principles in the context of aphasia research; our rigorous methodology for data acquisition and cleaning can serve as a template for the establishment of similar databases in other research areas.

## Background

Big data approaches to data synthesis and analysis typically involve collation of large datasets to reveal patterns, trends, and associations which would be overlooked in single trials due to a lack of statistical power. These approaches generate data "high in volume and diversity, gathered from single participants to large groups, at one or more time points" (Auffray et al., 2016), and can be applied to a range of health care settings. Big data methods have several advantages, the most important being sample size (Fan et al., 2014). Synthesising large volumes of data, facilitates greater representation of the population under investigation, whilst also reducing research waste (Rumsfeld et al., 2016). Collating data at the level of the individual, instead of at the level of aggregated, group or summary data, permits analyses that can account for individual confounders and missing data, and mitigates ecological bias, which can arise when using aggregated or group level data.

Historically, primary aphasia research studies in the stroke population have enrolled and analysed smaller sample sizes due to the heterogeneity of the population (Plowman et al., 2012), recruitment challenges (Wallace, 2010), the assessment of a variety of

outcomes (Flowers et al., 2016), with data gathered on a range of stroke, aphasia and language measurement tools, capturing information on selected language modalities and conducting assessments in a range of different languages. Additionally, there is heterogeneity in the availability of participant, intervention and outcome data across existing aphasia research (Brady et al., 2020), limiting clinical implementation of research findings. The feasibility of developing a large aphasia individual participant data (IPD) database is challenging due to the differences in study designs (e.g., randomised controlled trials 'RCTs', case studies, cohorts), data collection time points, approaches for collecting, processing and interpreting data (Brady et al., 2016). Generating high-quality data, which is representative of people affected by post-stroke aphasia, representing the different language domains (e.g., everyday communication, auditory comprehension, naming, reading comprehension and writing), time since stroke, overall aphasia severity and therapy data (e.g., approaches, frequency, intensity, duration, and dosage) in large numbers would facilitate important language recovery analyses, both within and between sub-populations. Pooling pre-existing IPD from international studies in a large, multi-disciplinary, multi-lingual database offers a cost and time-effective (Nishimura et al., 2016) approach to addressing current aphasia research priorities (Franklin et al., 2018; Pollock et al., 2012).

The specialised nature of aphasia research provides opportunities to work collaboratively with internationally renowned research hubs to collate datasets across languages, research settings, and using different measurement tools. However, this can also pose a challenge to synthesising and standardizing IPD into a uniform format for analyses, and relies on working closely with collaborators, ensuring trust in how data are to be collated, stored, used, analysed and acknowledged in subsequent publications. Each of these aspects of international collaborative work is essential in order to build an IPD resource which can be adopted internationally. Yet the feasibility of the development and aggregation of such diverse research and clinical datasets into a big-data archive is currently unknown within the field of aphasia rehabilitation.

We report here on the outcome of an international, multidisciplinary collaborative effort to systematically develop the REhabilitation and recovery of peopLE with Aphasia after Stroke (RELEASE) IPD database, highlighting the feasibility of developing a big data post-stroke aphasia research resource and providing a template for replication by researchers in other areas of neuro-rehabilitation and communication research.

## **Methods**

Full details on the search strategy, study selection, collation, data management, data extraction and synthesis are described in our published protocol (Brady et al., 2020)

### ***Ethical approval and study registration***

We received university ethical approval to collate data in the RELEASE archive (HLS/NCH/15/09). Principal investigators of included studies obtained local ethical approval for sharing their data to the RELEASE database as required. Our protocol was registered with the international prospective register of systematic reviews (PROSPERO CRD42018110947),

and the database was approved by the UK's Integrated Research Application System (IRAS; 179,505).

### ***Dataset identification & management***

We conducted a systematic search of existing literature to identify potential datasets in any language that met our eligibility criteria; IPD for a minimum of ten participants with stroke-related aphasia, time since stroke information and a measurement of language ability (Brady et al., 2020). Working copies of original datasets were given a unique identifier, standardised to SAS Inc. 9.4 software for data management, and stored with all available associated dataset documents (e.g., study protocols, papers and data dictionaries). Extracted data items were recorded in a data extraction table in accordance with the Template for Intervention Description and Replication (TIDieR) guidelines in reporting complex interventions (Hoffmann et al., 2014). We standardised the format, coding and units of measurement of variables, maximising the number of available IPD for analysis and documented this on a decision log. At primary dataset level, we described study design, inclusion/exclusion criteria, recruitment date (or publication date as a proxy when recruitment date was unavailable), sample size, country and language, data collection time point(s), blinding of assessors, dropouts and for RCTs, randomisation method and allocation concealment. Participant level data included demographic, stroke and language impairment information. Data cleaning and categorisation decisions were taken in collaboration with the wider RELEASE group. Where possible, each dataset profile summary was sent to the collaborating investigator for review and approval prior to inclusion in the final RELEASE database.

### ***Risk of bias assessment in individual studies***

We assessed methodological quality according to selection, detection, attrition and other potential sources of bias as previously detailed (Brady et al., 2020). We coded studies with potential risk of bias as low, unclear, or high risk, and appraised the impact of these potential biases.

### ***Risk of bias across studies***

We maintained an inclusive approach to dataset acceptance. Our systematic review used a broad search strategy and intentionally inclusive pre-defined eligibility criteria. Data extraction errors were minimised by two researchers independently reviewing data, and through verification of electronic contributions with primary researchers.

### ***Outcomes***

In collaboration with the contributing researchers, we categorised language measurements into the following language domains: overall language ability, spoken language production (including a sub-category for naming), auditory comprehension, reading comprehension, writing and functional communication. Data for each outcome measurement instrument were investigated for validity, for example, by examining ranges, outliers

and missing-ness of data with queries resolved (where possible) through communication with collaborators.

### **Summary statistics**

We described the contents of the database using summary statistics. Continuous variables were described using medians and interquartile ranges (IQR), while categorical variables were described using frequencies and percentages. Where possible, we described the measurement tools that were used to capture each of the language domains of interest, the number of studies that used the measurement, the IPD available and missing, and instances where the measurement tool was used but the data was not reported.

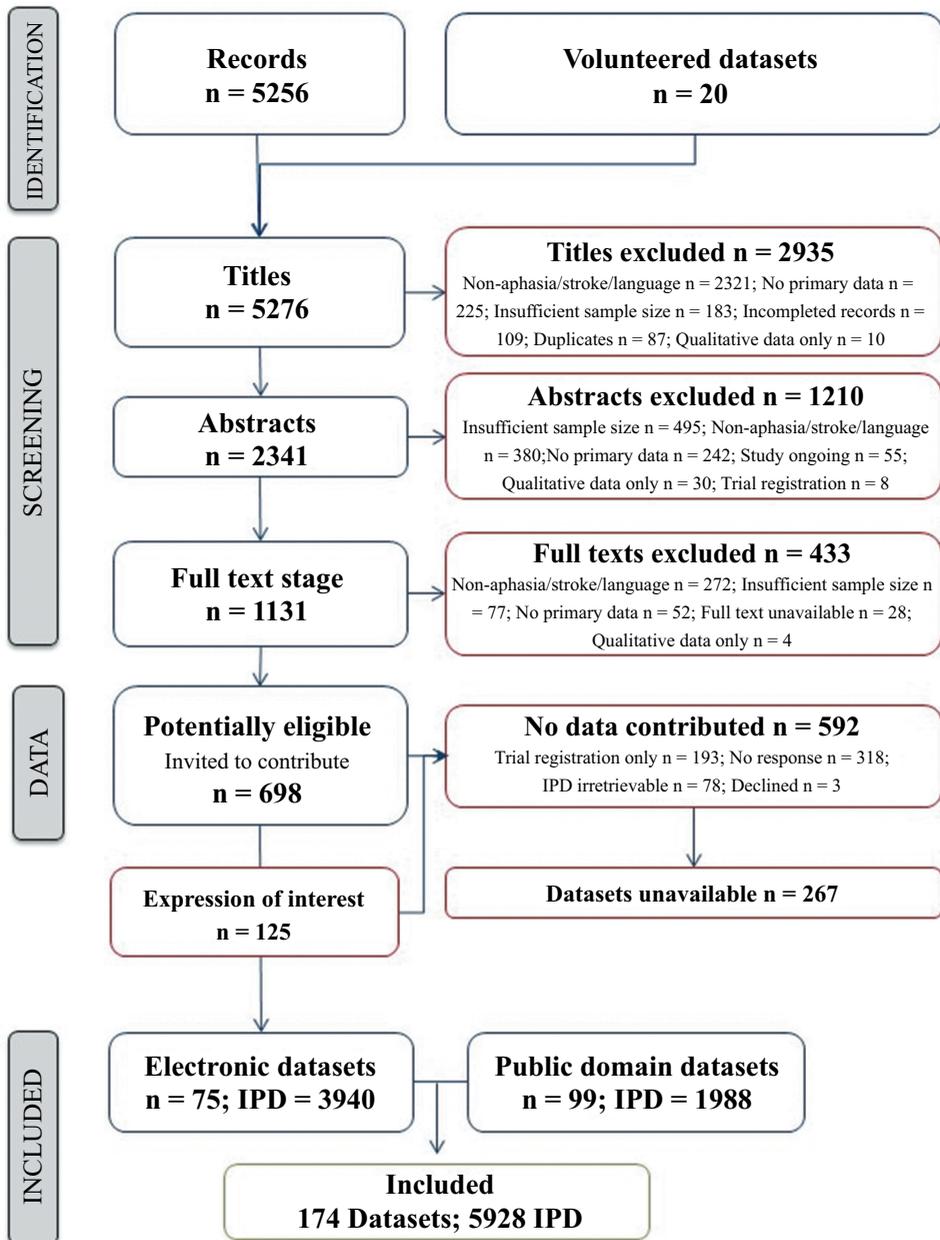
## **Results**

### **Datasets**

Following the systematic search, 11,314 records were identified. After the removal of duplicates, we screened 5256 records, excluding 2935 ineligible records. We excluded a further 1210 following abstract review and an additional 433 records after a review of the full texts (Figure 1). We invited the authors of 698 potentially eligible reports to contribute the associated IPD. Of which, 592 did not result in data contribution as 193/592 were trial registration records, no responses were received from 318/592 research teams, IPD from 78 datasets were unavailable IPD, and 3/592 declined to participate. We received 125 expressions of interest relating to eligible datasets and identified a further 99 eligible datasets in the public domain (Figure 1). Our RELEASE database thus comprised of 174 datasets with 5928 IPD from 28 countries, in 23 languages; 13 were unpublished. We received 75/174 (43.1%) datasets electronically (3940/5928, 66.5% IPD), and included 99 (56.9%) from the public domain (1988, 33.5% IPD).

### **Study types**

More than a quarter of the primary research studies 47/174 (27.0%) were RCTs (1778/5928, 30.0% IPD); over half 104/174 (59.8%) were a case series or cohort design (2886, 48.7% IPD); non-randomized controlled trial studies whereby participants were recruited to groups, but random allocation was not applied (non RCTs) accounted for 18/174 (10.3%) datasets ( $n = 411$ , 6.9% IPD). The remaining 2.9% (5/174) of datasets came from registries (853, 1.4% IPD). Data inclusion in each planned analysis was on a dataset-by-dataset basis and did not necessarily reflect the primary research study designs. For example, in some cases, participants were randomly allocated to two (non-SLT) interventions but all received the same SLT treatment. In such cases, the IPD were variously analysed as a single cohort or by group depending on the nature of the dataset and the planned analyses. Availability of data on randomised allocation also impacted the analysis; for example one primary research report described an RCT study design but IPD on group allocation was unavailable as the researchers reported the dataset as a single cohort.



**Figure 1.** PRISMA flow diagram: Eligible database identification and contribution.

### Participants

Participants had a median age of 63 (IQR 53–72) years, most 3407/5550 IPD (61.4%) were male (Table 1) and median formal education was 12 years (IQR 10–16). Median participant recruitment time was 321 days (IQR 30–1156) after stroke (5841/5928 IPD, 98.5%). Three quarters of participants (1116/1475 IPD, 75.5%) were White and a majority lived with

**Table 1.** RELEASE database participant descriptors.

Category	Datasets n (%) of total 174	IPD n (%) of total 5928	Sub-Category	Available IPD (%)
Sex	158 (90.8)	5,550 (93.6)	Male	3,407 (61.4)
			Female	2,143 (38.6)
Handedness	111 (63.8)	3,879 (65.4)	Right	3,719 (95.9)
			Left	133 (3.4)
			Ambidextrous	27 (0.7)
Ethnicity	24 (13.8)	1,475 (24.9)	White	1,116 (75.5)
			Asian	248 (16.8)
			Black	67 (4.5)
			Hispanic/Latino	9 (0.6)
			Mixed	2 (0.1)
			Other	33 (2.2)
Living Context	21 (12.1)	701 (11.8)	Living with others	473 (67.5)
			Living alone	146 (20.8)
			Formal care environment	70 (10.0)
			Mixed	12 (1.7)
Stroke	97 (55.7)	3,416 (57.6)	Ischaemic	2,795 (81.8)
			Intracerebral Haemorrhage	547 (16.0)
			Subarachnoid Haemorrhage	31 (0.9)
			Mixed	42 (1.2)
			Aneurysm	1 (0.03)
Prior Stroke	45 (25.8)	1,274 (21.5)	No	1164 (91.4)
			Yes	110 (8.6)
Affected Hemisphere	130 (74.7)	4,130 (69.7)	Left	3,965 (96.0)
			Right	81 (2.0)
			Bilateral	84 (2.0)
			-	175 (0.03)
Socioeconomic Status	4 (2.3)	175 (3.0)	-	175 (0.03)
Visual Impairment	35 (20.1)	1,494 (25.2)	No impairment	1,122 (75.1)
			Impairment present	60 (4.0)
			Corrected	312 (20.9)
Cognitive Impairment	100 (57.5)	3,945 (66.5)	No impairment	3,347 (84.8)
			Impairment present	7 (0.2)
			Score reported	591 (15.0)
Dysarthria	19 (10.9)	937 (15.8)	No dysarthria	634 (67.7)
			Dysarthria	111 (11.8)
			Mild	155 (16.5)
			No or Mild	23 (2.5)
			No severe	14 (1.5)
Apraxia	34 (19.5)	1,011 (17.1)	No Apraxia	606 (59.9)
			Apraxia	329 (32.5)
			Very mild apraxia	1 (0.1)
			No severe apraxia	36 (3.6)
			Mild-Moderate apraxia	7 (0.7)
			Moderate-Severe apraxia	2 (0.2)
			Assessment score available	30 (3.0)
Depression	62 (35.6)	2075 (35.0)	Absent	1,723 (83.0)
			Present	352 (17.0)
				<b>Median [IQR]</b>
Days since stroke	173 (99.4)	5841 (98.5)	-	321; [30–1,156]
Age (years)	169 (97.1)	5,785 (97.6)	-	63; [53–72]
Education (years)	84 (48.3)	3125 (52.7)	-	12; [10–16]
Stroke Severity	8 (4.6)	716 (12.1)	NIHSS	11; [5–17]
	6 (3.4)	489 (8.2)	mRS	4; [3–4]
	5 (2.9)	442 (7.5)	BI (0–20 scale)	16; [9–20]
			BI (0–100 scale)	60; [15–95]

Key – IPD Individual Participant Data; % percentage; IQR Interquartile range; NIHSS National Institute of Health Stroke Scale; mRS Modified Rankin Scale; BI Barthel Index.

**Table 2.** Data availability for language outcomes.

<i>Language outcome</i>	<i>Datasets</i>	<i>IPD</i>
	<i>N = 174 (%)</i>	<i>n = 5928 (%)</i>
Overall language ability	80 (46.0)	2699 (45.5)
Naming	75 (43)	2886 (48.7)
Other spoken language	9 (5.2)	380 (6.4)
Auditory comprehension	71 (40.8)	2750 (46.4)
Reading comprehension	12 (6.9)	770 (13.0)
Writing	13 (7.5)	724 (12.2)
Function communication – observer rated	29 (16.7)	1591 (26.8)
Functional communication – self rated	3 (1.7)	68 (1.1)

Key-IPD Individual Participant Data; % percentage; N = total datasets; n = total IPD

others (473/701 IPD, 67.5%, [Table 1](#)). Ischaemic stroke (2795/3416 IPD, 81.8%) predominated stroke aetiology and stroke typically affected the left hemisphere (3965/4130, 96.0%). The majority of participants were right-handed (3719/3879, 95.9%) with no visual impairment (1122/1494; 75.1% IPD); 111 participants (11.8%) were classified as also having dysarthria, and 329 (32.5%) were recorded as having aphasia along with apraxia of speech ([Table 1](#)).

Data were collected across 28 countries, in 23 languages, the majority in English (3162/5928, 53.3% IPD). Of the 79 datasets reporting date of recruitment, participants were recruited between 1973 and 2018. Of 161 published datasets, 26 (16.1% datasets; 664/4627 IPD, 14%) were published before 2000.

## Outcomes

A total of 64 language measurement tools, which captured the outcomes relevant to our database across 174 datasets were included ([Table 2](#)). The measurement tool used by the

**Table 3.** Overall language ability assessment tools (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report, or unavailable.

Overall Language Ability Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Aachen Aphasia Test (AAT) overall Severity Score	1 (12,0)	15 (537)
Afazi Dil Değerlendirme Testi (ADD)	1 (30,23)	0 (0)
Aphasia Handicap Scale (AHS)	2 (39,19)	0 (0)
Aphasia Severity Rating Scale (ASRS)	14 (441,6)	0 (0)
Boston Assessment of Severe Aphasia (BASA)	1 (15,0)	0 (0)
Comprehensive Aphasia Test (CAT)	3 (433,37)	5 (180)
Norsk Grunntest for Afaxi (NGA)	3 (62,0)	0 (0)
Porch Index of Communicative Ability (PICA)	8 (171,1)	0 (0)
Short Norsk Grunntest for Afasi (Short NGA)	2 (241,0)	0 (0)
Sprachsystemtisches Aphasie Screening (SAPS)	1 (133,9)	0 (0)
Standard Language Test of Aphasia (SLTA)	2 (24,0)	1 (36)
Western Aphasia Battery - Aphasia Quotient*	35 (733,0)	0 (0)
Western Aphasia Battery-Revised Aphasia Quotient	6 (69,0)	1 (18)
Western Aphasia Battery-Cantonese	1 (105,0)	0 (0)
Western Aphasia Battery-Japanese	1 (24,0)	0 (0)
Western Aphasia Battery- Korean	3 (125,3)	0 (0)
Western Aphasia Battery-Persian	2 (86,0)	0 (0)

Key \*Anchor Measure; IPD Individual Participant Data.

**Table 4a.** Naming assessments (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Naming Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Aachen Aphasia Test (AAT)	16 (524,14)	1 (11)
Afazi Dil Degerlendirme Testi (ADD)	1 (53,0)	0 (0)
Batteria per l'Analisi dei Deficit Afasici (BADA)	2 (12,0)	1 (6)
BETA	1 (127,134)	0 (0)
Boston Diagnostic Aphasia Examination (BDAE)	7 (0, 0)	7 (115)
Boston Naming Test (BNT)/60*	30 (831,0)	0 (0)
Boston Naming Test (BNT)/30	2 (21,0)	0 (0)
Boston Naming Test (BNT) Short Form	3 (129,0)	0 (0)
Boston Diagnostic Aphasia Examination Greek version (G-BDAE)	2 (38,0)	1 (12)
Boston Diagnostic Aphasia Examination Polish version (P-BDAE)	1 (40,0)	0 (0)
Cambridge Naming Test (CNT)	1 (70,0)	0 (0)
Comprehensive Aphasia Test (CAT)	4 (479,5)	4 (129)
Graded Naming Test	1 (11,0)	0 (0)
Greek Boston Naming Test	1 (38,0)	0 (0)
Korean Boston Naming Test	1 (10,0)	0 (0)
Lisbon Aphasia Assessment Battery (LAAB)	2 (55,0)	0 (0)
Norsk Grunntest for Afaxi (Full NGA)	1 (39,0)	2 (23)
Object Naming Test	3 (53,0)	0 (0)
PALPA 53	1 (11,0)	0 (0)
Philadelphia Naming Test (PNT)	2 (25,0)	0 (0)
Porch Index of Communicative Ability (PICA)	3 (52,0)	5 (120)
Psychological Assessment of Language (PAL)	2 (81,3)	0 (0)
Snodgrass and Vanderwort	3 (80,0)	0 (0)
Standard Language Test of Aphasia (SLTA)	2 (36,0)	1 (24)
VNT	1 (260,54)	0 (0)
Western Aphasia Battery	2 (44,0)	20 (381)
Western Aphasia Battery-Revised	6 (72,0)	1 (15)
Western Aphasia Battery-Cantonese	1 (105,0)	0 (0)
Western Aphasia Battery- Korean	3 (117,1)	1 (10)
Western Aphasia Battery- Mandarin	1 (23,0)	0 (0)

Key: IPD Individual participant data; \*Anchor Measure

**Table 4b.** Other spoken language assessments (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Other spoken language assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Boston Diagnostic Aphasia Examination Greek version (G-BDAE)	2 (38,0)	1 (12)
Communication Disability Profile (CDP)	1 (25,0)	0 (0)
Porch Index of Communicative Ability (PICA)*	2 (40,0)	6 (132)
Schuell	1 (10,0)	0 (0)
Standard Language Test of Aphasia (SLTA)	2 (24,0)	1 (36)
Sprachsystemtisches Aphasie Screening (SAPS)	1 (139,3)	0 (0)
Western Aphasia Battery-Cantonese	1 (105,0)	0 (0)
Western Aphasia Battery- Korean	1 (21,0)	2 (107)

Key: IPD Individual participant data; \*Anchor Measure

**Table 5.** Auditory comprehension assessments (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Auditory Comprehension Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Aachen Aphasia Test (AAT)	3 (194,2)	13 (353)
AAT Italian Version of Token Test -T score	1 (11,0)	0 (0)
AAT Token Test*	15 (505,0)	0 (0)
AAT Token Test T Score	2 (27,0)	0 (0)
Afazi Dil Değerlendirme Testi (ADD)	1 (53,0)	0 (0)
Batteria per l'Analisi dei Deficit Afasici (BADA)	1 (12,0)	1 (6)
Boston Aphasia Severity Assessment (BASA)	9 (15,0)	0 (0)
Boston Diagnostic Aphasia Examination (BDAE)	7 (115,0)	0 (0)
BDAE 3	1 (29,0)	0 (0)
BDAE 12	1 (11,0)	0 (0)
Greek version of BDAE (G-BDAE)	2 (50,0)	0 (0)
Communication Disability Profile (CDP)	9 (6,0)	1 (19)
Comprehensive Aphasia Test (CAT)	6 (576,6)	2 (31)
DeRenzi, Vignolo Token Test/62	2 (34,0)	0 (0)
DeRenzi, Vignolo Token Test/36	10 (639,0)	0 (0)
DeRenzi Modified Short form of the Token Test	3 (53,0)	0 (0)
Lisbon Aphasia Assessment Battery (LAAB)	2 (55,0)	0 (0)
Montreal Toulouse	1 (11,0)	0 (0)
Norsk Grunntest for Afaxi (NGA)	3 (62,0)	0 (0)
Peabody Picture Vocabulary Test (PPVT)	2 (33,3)	0 (0)
Pizzamiglion Sentence Comprehension Test (PSCT)	1 (25,11)	0 (0)
Porch Index of Communicative Ability (PICA)	3 (52,0)	5 (120)
Psychological Assessment of Language (PAL)	2 (79,5)	0 (0)
Revised Token Test Arvedson, McNeil & West/five item	1 (16,0)	0 (0)
Schuell	1 (10,0)	0 (0)
Sprachsystematisches Aphasiescreening (SAPS)	1 (136,6)	0 (0)
Standard Language Test of Aphasia (SLTA)	2 (60,0)	0 (0)
Token Test/15	1 (25,0)	0 (0)
Western Aphasia Battery	3 (54,0)	19 (371)
Western Aphasia Battery-Revised	5 (71,1)	1 (15)
Western Aphasia Battery-Cantonese	1 (105,0)	0 (0)
Western Aphasia Battery- Korean	2 (117,1)	1 (10)

Key-IPD Individual Participant Data; \*Anchor Measure.

most datasets for each language outcome became an “anchor measure” to which the scores of all other tools measuring the same domain were transformed (see Table 3–7 for anchor and minority measures of overall language and each language domain). Many datasets (80/174;46.0%) included a measure of overall language ability (2699/5928, 45.5% IPD), 75 (43.0%) used a naming measure (2886, 48.7% IPD), and 71 (40.8%) measured auditory comprehension (2750, 46.4% IPD). Other language domains of interest (functional communication, writing, reading comprehension and other spoken language) were also measured but by smaller numbers of primary research teams (Table 2). Available measurement tools were profiled by language modality (Table 3–7). For some datasets measurement tools described in accompanying study documentation, were unavailable in a format that could be profiled or analysed for each language domain and were unavailable from the primary research team. For example, the Aachen Aphasia Test (AAT), while described in reports relating to 17 datasets, only 16/17 had AAT IPD available for one or more of the language outcomes relevant to our database.

**Table 6a.** Reading comprehension and writing assessments (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Reading Comprehension Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Aachen Aphasia Test (AAT)	2 (54,2)	14 (493)
BETA	1 (126,135)	0 (0)
Boston Diagnostic Aphasia Examination (BDAE)	1 (13,0)	6 (102)
Boston Diagnostic Aphasia Examination Greek version (G-BDAE)	1 (38,0)	1 (12)
Comprehensive Aphasia Test (CAT)*	3 (460,7)	5 (146)
Norsk Grunntest for Afasi (NGA)	1 (39,0)	2 (23)
PALPA 31	1 (11,0)	0 (0)
PALPA 48	1 (11,0)	0 (0)
Schuell	1 (10,0)	0 (0)
Standard Language Test of Aphasia (SLTA)	1 (36,0)	1 (24)
Western Aphasia Battery	1 (13,0)	21 (412)
Western Aphasia Battery- Korean	1 (81,16)	2 (31)
Western Aphasia Battery- Mandarin	1 (23,0)	0 (0)

Key: IPD Individual Participant Data; \*Anchor Measure.

**Table 6b.** Reading comprehension and writing assessments (at baseline) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Writing assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Boston Diagnostic Aphasia Examination (BDAE)	1 (13,0)	6 (102)
Boston Diagnostic Aphasia Examination Greek version (G-BDAE)	1 (38,0)	1 (12)
Communication Disability Profile (CDP)	1 (24,0)	0 (0)
Comprehensive Aphasia Test (CAT)*	3 (454,13)	5 (146)
Norsk Grunntest for Afasi (NGA)	1 (39,0)	2 (23)
Porch Index of Communicative Ability (PICA)	3 (51,1)	5 (120)
Schuell	1 (10,0)	0 (0)
Standard Language Test of Aphasia (SLTA)	1 (24,0)	1 (36)
Western Aphasia Battery- Korean	1 (71,26)	2 (31)

Key: IPD Individual Participant Data; \*Anchor Measure.

## Interventions

Just over half the datasets described an SLT intervention (91/174 52.3%; 2746/5928, 46.3% IPD). Data on language measurements were available at pre- and post-intervention time points for 67/174 (38.5%) datasets, comprising 2330/5928 IPD (39.3%). These 67 intervention studies were profiled by method of delivery, theoretical approach, and language treatment targets (Table 8). In 7/174 (4%) datasets, participants were allocated across three groups (different SLT interventions, in some cases with a control comparison group), (128/5928, 0.02% IPD), and in three datasets (1.7%) across four groups (84 IPD). Five datasets (4%) employed a delayed treatment start for one group (128/5928, 0.02% IPD). Therapy regime was described according to intensity (hours per week, 1882 IPD; 60 datasets), frequency (number of sessions

**Table 7a.** Functional Communication (observer-rated) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Functional Communication Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Aachen Aphasia Test (AAT) Spontaneous Speech*	8 (402,4)	8 (147)
American Speech and Hearing Association Functional assessment of communication skills	3 (135,0)	0 (0)
Amsterdam Nijmegen Everyday Language Test (ANELT)	7 (453,142)	0 (0)
Communication Activity Log (CAL)	1 (10,0)	0 (0)
Communication Effectiveness Index (CETI)	7 (300,0)	0 (0)
Communicative Pragmatic Screening (KOPS)	1 (131,11)	0 (0)
Functional Communication Profile	4 (249,5)	0 (0)
Mini Communication Activity Log (Mini CAL)	1 (24,0)	0 (0)
Scenario Test (2010)	1 (11,0)	0 (0)
Therapy Outcome Measures (TOMS) Activity	3 (257,1)	0 (0)
Therapy Outcome Measures (TOMS) Participation	3 (257,1)	0 (0)

Key: IPD Individual Participant Data; \*Anchor Measure.

**Table 7b.** Functional Communication (self-rated) included in RELEASE – Datasets and IPD where measure is reported, available, missing from report or unavailable.

Functional Communication Assessment	Datasets	
	Reported (IPD available, missing)	Assessed but unavailable (IPD)
Assessment for Living with Aphasia (ALA)*	2 (43,2)	0 (0)
Communication Confidence Rating Scale for Aphasia (CCRSA)	1 (32,2)	0 (0)
Communication Disability Profile (CDP) Activity	1 (25,0)	0 (0)
Communication Disability Profile (CDP) Auditory Comprehension	1 (6,19)	0 (0)
Communication Disability Profile (CDP) Spoken Language	1 (25,0)	0 (0)
Communication Disability Profile (CDP) Participation	1 (25,0)	0 (0)
Communication Disability Profile (CDP) Writing	1 (24,1)	0 (0)

Key: IPD Individual Participant Data; \*Anchor Measure.

per week, 2057 IPD; 66 datasets), duration (total number of days, 1960 IPD; 64 datasets) and dosage (total hours, 1978 IPD; 62 datasets). Furthermore, interventions were delivered by professionals in 62 datasets (92.5%) and non-professionals in seven datasets (10.4%); in clinics, hospitals or rehabilitation settings in 48 datasets (71.6%) and at home in 17 datasets (23.4%).

### **Risk of meta-bias – across studies**

From the 47 RCTs identified in a previous systematic review (Brady et al., 2016), we included IPD from 17 eligible trials, in addition to IPD from five trials that were ongoing, and a further four trials that were unavailable at the time of that review. As a consequence of our systematic searching, we included 122 more datasets in our IPD database than we had identified *a priori*, increasing our original estimate from 3181 IPD to 5928. This reduced the risk of selection bias and increased the availability of IPD. Due to the nature of our dataset, generation of funnel plots to assess publication bias were not

**Table 8.** Speech and language therapy (SLT) Interventions in the RELEASE dataset.

Speech and language therapy intervention descriptor		Datasets N = 67 (%)	IPD n = 2330 (%)
SLT method of delivery	Face-to-face	60 (89.6)	1957 (84.0)
	Computer	15 (22.4)	315 (13.5)
	Telephone	1 (1.5)	15 (0.6)
	Constraint induced aphasia therapy	7 (10.4)	113 (4.8)
	Self-managed	4 (6.0)	106 (4.5)
	One-to-one	47 (70.1)	1613 (69.2)
	Group	8 (11.9)	148 (6.4)
Theoretical approach	Mixed	9 (13.4)	207 (8.9)
	Semantic	2 (1.5)	34 (1.5)
	Phonological	9 (13.4)	124 (5.3)
	Semantic and phonological	15 (22.4)	260 (11.2)
	Functional and pragmatic	8 (11.9)	246 (10.6)
	Constraint induced aphasia therapy	7 (10.4)	113 (4.8)
	Melodic intonation therapy	4 (6.0)	61 (2.6)
Target of Impairment	Conversational partner training	2	55 (2.4)
	Spoken language	41	734 (31.5)
	Auditory comprehension	4	68 (2.9)
	Auditory comprehension & spoken language	24	651 (27.9)
	Reading	1	10 (0.4)
	Writing	0	0

Note: Categories were not mutually exclusive; an intervention may span categories or appear more than once.

possible. Instead, we checked distribution between sample size and studies reporting significant results. Of the 164 published datasets, no evidence of publication bias by sample size was evident (chi-squared,  $p = 0.77$  Table 9).

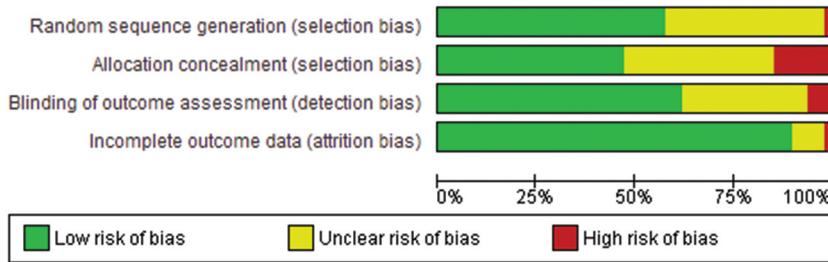
### **Risk of bias – within studies**

Across the included RCTs there was a low risk of bias (random sequence generation, allocation concealment, blinding of outcome measure and a small number with unexplained attrition) (Figure 2). Of the total 47 RCTs included, 27 reported adequate generation of a random sequence; 22 described an adequate method to ensure concealment of allocation. Within RCTs, baseline differences between participants in the treatment groups were assessed for age (3/47 showing baseline differences), time since stroke (2/

**Table 9.** Significant and non-significant results reported, by participant number.

Sample size (group)	Significant result reported	
	No	Yes % of total publications
Participants		
>100	3	5 (63%)
>50 to 100	3	8 (73%)
>30 to 50	3	15 (83%)
>20 to 30	8	25 (76%)
>10 to 20	12	57 (83%)
Up to 10	5	20 (80%)

Key % percentage; > greater than.



**Figure 2.** Randomised controlled trials included in the RELEASE database and risk of bias.

47 showing baseline differences) and language impairment (1/47 showing baseline differences) using the Wilcoxon signed rank test (Brady et al., [in press](#)). Between group baseline differences in sex were assessed using the chi-square test (1/47 showing baseline differences).

Blinding of outcome assessment was reported in 49/174 datasets (28.2%) though in many cases blinding was not relevant to the primary dataset study design (e.g., a clinical registry or observational study). An additional nine datasets (5.2%) described blinding for a subset of data and 117 datasets either did not report blinding (91/117) or had indicated that blinding was not present (25/117).

A third of the IPD were available in the public domain; some of these IPD datasets were only partially reported in those publications (for example, baseline data only) while in other cases full RCT datasets were available. Some primary research dataset electronic contributions were only partially available with published reports relating to that dataset indicating that additional data items were available at the time of publication.

## Discussion

Using an international multidisciplinary collaborative approach, robust methodology, and applying big data principles within the context of aphasia research, we systematically created an international IPD database that adhered to data sharing guidelines (Knoppers, 2014), and which ensured the feasibility of our analysis plan (Brady et al., 2020). Working with our collaborators and accessing data in the public domain, allowed us to include demographic, stroke, aphasia, language impairment and SLT intervention IPD from 28 countries and language outcomes gathered using 64 measurement tools. Over half the datasets included an SLT intervention, IPD that can be used to inform future analyses to optimise the delivery of SLT in people with aphasia. Our database is representative of available aphasia research datasets and reflects the heterogeneity in data collection practices across a range of different study designs and contexts. Collation of these data highlighted variation in the measurement tools that were used to capture data within the same language domain and strengthens the evidence to support greater consistency in the use of an aphasia core outcome set (Wallace et al., 2019) in future aphasia research. Similarly, the need for greater consistency in the collection of participant descriptors in future aphasia studies was evident (The RELEASE Collaboration, 2020), and an international

consensus initiative to address this need is currently ongoing within the Collaboration of Aphasia Trialists ([www.aphasiatrials.org/describe](http://www.aphasiatrials.org/describe)).

A key strength in our study was our international, collaborative approach, which involved multidisciplinary aphasia researchers and permitted the sharing of expertise and greater clarity on international aphasia research practice. Primary researchers were actively involved in the development and cleaning of data, thereby contributing to the validity, accuracy and robustness of the data contained within the archive.

The research landscape across a range of health settings has highlighted the importance of minimising research waste (Chan et al., 2014; Doubal et al., 2017), resulting in encouragement and recognition of the value (Taichman et al., 2016), in sharing research datasets to facilitate new analyses. Guidelines exist to support IPD sharing (Waithira et al., 2019), with appropriate approvals, accreditation and acknowledgement (Wilkinson et al., 2016), but the feasibility of such an approach in the context of aphasia rehabilitation research had not yet been examined; we set out to explore this.

Previous work has highlighted that health research protocols, full study reports, and participant-level datasets are not widely available in the public domain (Chan et al., 2014). However, recent initiatives that require the lodging of datasets in publicly available resources have changed the landscape of health research (Taichman et al., 2016). Additionally, in one study publications reporting study findings were available for only half of all studies (Chan et al., 2014), and can be subject to selective reporting or other sources of bias. A strength in our method of identifying eligible studies was that it relied not only on searches of the existing literature, but also on communication with existing international research networks such as the Collaboration of Aphasia Trialists ([www.aphasiatrials.org](http://www.aphasiatrials.org)) to identify potentially eligible datasets for inclusion. Our model builds on historical examples of in-house research activities that have included data sharing amongst specific research teams. Use of datasets based on availability and access within research groups can often bias dataset selection, impacting upon the representativeness of the data and limiting the generalisability of results.

Our study highlighted some restrictions to collaborative approaches. Despite our wide eligibility criteria and our systematic identification of, and open invitation to principal investigators with potentially eligible datasets, we were only able to acquire a subset of potentially eligible datasets; of 698 datasets identified as potentially eligible, 267 were unavailable for data sharing.

We highlighted several important methodological steps that were required to develop this international database. These included the establishment of relevant but inclusive eligibility criteria, the systematic identification of eligible studies using multiple information sources, open invitation to international researchers to contribute data, collaborative working and consensus building to agree on variable parameters and categories, cleaning and standardisation of the data, and finally transparent description of the database to facilitate informed exploratory analyses. This model can be used by investigators in other health research areas to create similar data repositories within which to conduct exploratory analyses. Beyond the RELEASE project, a long-term legacy, big data repository for aphasia datasets is under development within the Collaboration of Aphasia Trialists, using the principles set out above. We invite further contributions of eligible datasets to this aphasia archive, to facilitate future analyses for the benefit of people with aphasia.

## Acknowledgements

Open Access funding provided by the Qatar National Library.

## Disclosure statement

No potential conflict of interest was reported by the authors.

Marian Brady reports grants from Chief Scientist Office (CSO), Scottish Government Health and Social Care Directorates, grants from EU Cooperation in Science and Technology (COST) funded Collaboration of Aphasia Scientists (IS1208 [www.aphasiatrials.org](http://www.aphasiatrials.org)), and grants from The Tavistock Trust for Aphasia during the conduct of the study, and is a member of the Royal College of Speech and Language Therapists.

Audrey Bowen reports that data from her research is included within the analyses in the RELEASE report. Her post at the University of Manchester is partly funded by research grants and personal awards from NIHR and Stroke Association.

Caterina Breitenstein reports grants from the German Federal Ministry of Education and Research (BMBF) during the conduct of the study.

Erin Godecke reports Western Australian State Health Research Advisory Council (SHRAC) Research Translation Project Grants RSD-02720; 2008/2009 during the conduct of the study.

Neil Hawkins reports grants from National Institute for Health Research during the conduct of the study.

Katerina Hilari reports grants from The Stroke Association, grants from European Social Fund and Greek National Strategic Reference Framework, and grants from The Tavistock Trust for Aphasia outside the submitted work.

Petra Jaecks reports a PhD grant from Weidmüller Stiftung.

Brian MacWhinney reports grants from National Institutes of Health.

Rebecca Marshall reports grants from National Institute of Deafness and Other Communication Disorders, NIH, USA during the conduct of the study.

Rebecca Palmer reports grants from NIHR senior clinical academic lectureship, grants from NIHR HTA, and grants from Tavistock Trust for Aphasia outside the submitted work.

Ilias Papatathanasiou reports funding from European Social Fund and Greek National Strategic Reference Framework.

Elizabeth Rochon reports funding from the Heart and Stroke Foundation of Canada.

Jerzy Szaflarski reports personal fees from SK Life Sciences, personal fees from LivaNova Inc, personal fees from Lundbeck, personal fees from NeuroPace Inc, personal fees from Upsher-Smith Laboratories, Inc, grants and personal fees from SAGE Pharmaceuticals, grants and personal fees from UCB Pharma, grants from Biogen, grants from Eisai Inc, and other from GW Pharmaceuticals, outside the submitted work.

Shirley Thomas reports research grants from NIHR and The Stroke Association outside the submitted work.

Ineke van der Meulen reports grants from Stichting Rotterdams Kinderrevalidatiefonds Adriaanstichting, other from Stichting Afasie Nederland, other from Stichting Coolsingel, and other from Bohn Stafleu van Loghum during the conduct of the study.

Linda Worrall reports a grant from the National Health and Medical Research Council of Australia.

All other authors have declared no competing interests

## Funding

This project was funded by the National Institute for Health Research (NIHR) Health Services and Delivery Research [14/04/22], and The Tavistock Trust for Aphasia and will be published in full in the Health Services and Delivery Research Journal. Further information is available at [www.journal.slibrary.nihr.ac.uk/programmes/hsdr/140422#/NMAHPRU](http://www.journal.slibrary.nihr.ac.uk/programmes/hsdr/140422#/NMAHPRU). NMAHPRU and MCB is funded by the

Chief Scientist Office (CSO), Scottish Government Health and Social Care Directorates. The views expressed are those of the authors and not necessarily those of the funders.

## ORCID

Louise R Williams  <http://orcid.org/0000-0003-2430-1142>  
Myzoon Ali  <http://orcid.org/0000-0001-5899-2485>  
Kathryn VandenBerg  <http://orcid.org/0000-0001-5035-9650>  
Linda J Williams  <http://orcid.org/0000-0002-6317-1718>  
Masahiro Abo  <http://orcid.org/0000-0001-6701-4974>  
Frank Becker  <http://orcid.org/0000-0002-0857-0628>  
Audrey Bowen  <http://orcid.org/0000-0003-4075-1215>  
Caitlin Brandenburg  <http://orcid.org/0000-0002-6992-7790>  
Caterina Breitenstein  <http://orcid.org/0000-0002-6408-873X>  
Stefanie Bruehl  <http://orcid.org/0000-0003-4826-1990>  
David A Copland  <http://orcid.org/0000-0002-2257-4270>  
Tamara B Cranfill  <http://orcid.org/0000-0001-7608-6443>  
Marie Di Pietro-Bachmann  <http://orcid.org/0000-0001-8027-2337>  
Pamela Enderby  <http://orcid.org/0000-0002-4371-9053>  
Joanne Fillingham  <http://orcid.org/0000-0002-0363-8021>  
Federica Lucia Galli  <http://orcid.org/0000-0001-9244-9179>  
Marialuisa Gandolfi  <http://orcid.org/0000-0002-0877-4807>  
Bertrand Glize  <http://orcid.org/0000-0001-9618-2088>  
Erin Godecke  <http://orcid.org/0000-0002-7210-1295>  
Neil Hawkins  <http://orcid.org/0000-0003-3199-221X>  
Katerina Hilari  <http://orcid.org/0000-0003-2091-4849>  
Jacqueline Hinckley  <http://orcid.org/0000-0002-4052-1420>  
Simon Horton  <http://orcid.org/0000-0002-2133-1410>  
David Howard  <http://orcid.org/0000-0001-9141-5751>  
Petra Jaecks  <http://orcid.org/0000-0002-5878-1327>  
Elizabeth Jefferies  <http://orcid.org/0000-0002-3826-4330>  
Luis MT Jesus  <http://orcid.org/0000-0002-8534-3218>  
Maria Kambanaros  <http://orcid.org/0000-0002-5857-9460>  
Eun Kyoung Kang  <http://orcid.org/0000-0001-5315-1361>  
Eman M Khedr  <http://orcid.org/0000-0001-5679-9833>  
Anthony Pak-Hin Kong  <http://orcid.org/0000-0002-6211-0358>  
Tarja Kukkonen  <http://orcid.org/0000-0002-8189-0337>  
Marina Laganaro  <http://orcid.org/0000-0002-4054-0939>  
Matthew A Lambon Ralph  <http://orcid.org/0000-0001-5907-2488>  
Ann Charlotte Laska  <http://orcid.org/0000-0002-7330-940X>  
Béatrice Leemann  <http://orcid.org/0000-0003-2226-6777>  
Alexander P Leff  <http://orcid.org/0000-0002-0831-3541>  
Roxele Ribeiro Lima  <http://orcid.org/0000-0002-9914-4789>  
Antje Lorenz  <http://orcid.org/0000-0002-0200-1977>  
Brian MacWhinney  <http://orcid.org/0000-0002-4988-1342>  
Rebecca Shisler Marshall  <http://orcid.org/0000-0001-9313-5454>  
Flavia Mattioli  <http://orcid.org/0000-0002-4912-5520>  
İlknur Maviş  <http://orcid.org/0000-0003-3924-1138>  
Marcus Meinzer  <http://orcid.org/0000-0003-1370-3947>  
Reza Nilipour  <http://orcid.org/0000-0003-4180-7989>  
Enrique Noé  <http://orcid.org/0000-0002-2547-8727>  
Nam-Jong Paik  <http://orcid.org/0000-0002-5193-8678>  
Rebecca Palmer  <http://orcid.org/0000-0002-2335-7104>

Ilias Papanthasiou  <http://orcid.org/0000-0003-0999-696X>  
 Brigida F Patricio  <http://orcid.org/0000-0002-2619-470X>  
 Isabel Pavão Martins  <http://orcid.org/0000-0002-9611-7400>  
 Cathy Price  <http://orcid.org/0000-0003-0111-9364>  
 Tatjana Prizl Jakovac  <http://orcid.org/0000-0002-5018-9556>  
 Elizabeth Rochon  <http://orcid.org/0000-0001-5521-0513>  
 Miranda L Rose  <http://orcid.org/0000-0002-8892-0965>  
 Charlotte Rosso  <http://orcid.org/0000-0001-7236-1508>  
 Ilona Rubi-Fessen  <http://orcid.org/0000-0002-9775-3812>  
 Marina B Ruiter  <http://orcid.org/0000-0001-6147-5235>  
 Claerwen Snell  <http://orcid.org/0000-0001-8606-7801>  
 Benjamin Stahl  <http://orcid.org/0000-0003-3957-1495>  
 Jerzy P Szaflarski  <http://orcid.org/0000-0002-5936-6627>  
 Shirley A Thomas  <http://orcid.org/0000-0003-0704-9387>  
 Mieke Van De Sandt-Koenderman  <http://orcid.org/0000-0002-8104-6937>  
 Ineke Van Der Meulen  <http://orcid.org/0000-0002-6156-3873>  
 Evy Visch-Brink  <http://orcid.org/0000-0001-7833-0112>  
 Linda Worrall  <http://orcid.org/0000-0002-3283-7038>  
 Heather Harris Wright  <http://orcid.org/0000-0001-6922-6364>  
 Marian C Brady  <http://orcid.org/0000-0002-4589-7021>

## Disclaimer

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, MRC, CCF, NETSCC, the HS&DR programme, the Department of Health and Social Care, the Tavistock Trust for Aphasia, the Chief Scientist Office or the Scottish Government Health and Social Care Directorates.

## Data availability statement & Data deposition

Contributing trialists who have granted permission for extended use of their datasets beyond RELEASE have deposited their data in the Collaboration of Aphasia Trialists' (CATs) Aphasia Resource (available at <https://www.aphasiatrials.org/aphasia-dataset/>).

## References

- Auffray, C., Balling, R., Barroso, I., Bencze, L., Benson, M., Bergeron, J., Bernal-Delgado, E., Blomberg, N., Bock, C., Conesa, A., Del Signore, S., Delogne, C., Devilee, P., Di Meglio, A., Eijkemans, M., Flicek, P., Graf, N., Grimm, V., Guchelaar, H. J., Guo, Y. K., Gut, I. G., . . . Zanetti, G. (2016). Making sense of big data in health research: Towards an EU action plan. *Genome Medicine*, 8(1), 71. <https://doi.org/10.1186/s13073-016-0323-y>
- Brady, M. C., Ali, M., VandenBerg, K., Williams, L. J., Williams, L. R., Abo, M., Becker, F., Bowen, A., Brandenburg, C., Breitenstein, C., Bruehl, S., Copland, D. A., Cranfill, R. B., Di Pietro-bachmann, M., Enderby, P., Fillingham, J., Galli, F. L., Gandolfi, M., Glize, B., Godecke, E., Harris Wright, H., et al. (2020). RELEASE a protocol for a systematic review based, individual participant data, meta-and network meta-analysis, of complex speech-language therapy interventions for stroke-related aphasia. *Aphasiology*, 34(2), 137–157. <https://doi.org/10.1080/02687038.2019.1643003>
- Brady, M. C., Ali, M., VandenBerg, K., Williams, L. J., Williams, L. R., Abo, M., Becker, F., Bowen, A., Brandenburg, C., Breitenstein, C., Bruehl, S., Copland, D. A., Cranfill, R. B., Di Pietro-bachmann, M., Enderby, P., Fillingham, J., Galli, F. L., Gandolfi, M., Glize, B., Godecke, E., Harris Wright, H., et al. (in press). RELEASE: An individual participant data meta-analysis, incorporating systematic review and network meta-analysis, of complex speech-language therapy interventions for stroke related

- aphasia. *National Institute of Health Research: Health Services and Delivery Research*. <https://www.journalslibrary.nihr.ac.uk/programmes/hsdr/140422/#/>
- Brady, M. C., Kelly, H., Godwin, J., Enderby, P., & Campbell, P. (2016). Speech and language therapy for aphasia following stroke. *Cochrane Database of Systematic Reviews*, (6). Art. No: CD000425. <https://doi.org/10.1002/14651858.CD000425.pub4>
- Chan, A.-W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gotzsche, P. C., Krumholz, H. M., Ghersi, D., & Bart Van Der Worp, H. (2014). Increasing value and reducing waste: Addressing inaccessible research. *The Lancet*, 18(9913), 257–266. 383 [https://doi.org/10.1016/S0140-6736\(13\)62296-5](https://doi.org/10.1016/S0140-6736(13)62296-5)
- Doubal, F. N., Ali, M., Batty, G. D., Charidimou, A., Eriksson, M., Hofmann-Apitius, M., Kim, Y.-H., Levine, D. A., Mean, G., Mucke, H. A. N., Ritchie, C. W., Roberts, C. J., Russ, T. C., Stewart, R., Whiteley, W., & Quinn, T. J. (2017). Big data and data repurposing - using existing data to answer new questions in vascular dementia research. *BMC Neurology*, 17(1), 72. <https://doi.org/10.1186/s12883-017-0841-2>
- Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review*, 1(2), 293–314. <https://doi.org/10.1093/nsr/nwt032>
- Flowers, H. L., Skoretz, S. A., Silvers, F. L., Rochon, E., Fang, J., Flaman-Rose, C., & Martino, R. (2016). Poststroke aphasia frequency, recovery and outcomes: A systematic review and meta-analysis. *Archives of Physical Medicine and Rehabilitation*, 97(12), 2188–2201. <http://dx.doi.org/10.1016/j.apmr.2016.03.006>
- Franklin, S., Harhen, D., Hayes, M., McManus, S., & Pollock, A. (2018). Top 10 research priorities relating to aphasia following stroke. *Aphasiology*, 32(11), 1388–1395. <https://doi.org/10.1080/02687038.2017.1417539>
- Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Woods, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, 348(mar07 3), g1687. <https://doi.org/10.1136/bmj.g1687>
- Knoppers, B. M. (2014). Framework for responsible sharing of genomic and health-related data. *The HUGO Journal*, 8(1), 3. <https://doi.org/10.1186/s11568-014-0003-1>
- Nishimura, A., Nishimura, K., Kada, A., & Iihara, K., & J-ASPECT Study GROUP. (2016). Status and future perspectives of utilizing big data in neurosurgical and stroke research. *Neurologica Medico-Chirurgica*, 56(11), 655–663. [doi.org/10.2176/nmc.ra.2016-0174](https://doi.org/10.2176/nmc.ra.2016-0174)
- Plowman, E., Hentz, B., & Ellis, C. (2012). Post-stroke aphasia prognosis: A review of patient-related and stroke related factors. *Journal of Evaluation in Clinical Practice*, 18(3), 689–694. <https://doi.org/10.1111/j.1365-2753.2011.01650.x>
- Pollock, A., St George, B., Fenton, M., & Firkins, L. (2012). Top ten research priorities relating to life after stroke. *Lancet Neurology*, 11(3), 209. [https://doi.org/10.1016/S1474-4422\(12\)70029-7](https://doi.org/10.1016/S1474-4422(12)70029-7)
- Rumsfeld, J. S., Joynt, K. E., & Maddox, T. M. (2016). Big data analytics to improve cardiovascular care: Promise and challenges. *Nature Reviews Cardiology*, 13(6), 350–359. <https://doi.org/10.1038/nrcardio.2016.42>
- Taichman, D. B., Backus, J., Baethge, C., Bauchner, H., De Leeuw, P. W., Drazen, J. M., Fletcher, J., Frizelle, F. A., Groves, T., Haileamlak, A., James, A., Laine, C., Peiperl, L., Pinborg, A., Sahni, P., & Wu, S. (2016). Sharing Clinical Trial Data — A proposal from the International Committee of Medical Journal Editors. *The New England Journal of Medicine*, 374(4), 384–386. <https://doi.org/10.1056/nejme1515172>
- The RELEASE Collaboration. (2020). Communicating simply, but not too simply: Reporting of participants and speech and language interventions for aphasia after stroke. *International Journal of Speech-Language Pathology*, 22(3), 302–312. <https://doi.org/10.1080/17549507.2020.1762000>
- Waithira, N., Mutinda, B., & Cheah, P. Y. (2019). Data management and sharing policy: The first step towards promoting data sharing. *BMC Medicine*, 17(1), Article: 80. <https://doi.org/10.1186/s12916-019-1315-8>
- Wallace, G. L. (2010). Profile of life participation after stroke and aphasia. *Topics in Stroke Rehabilitation*, 17(6), 432–450. <https://doi.org/10.1310/tsr1706-432>

- Wallace, S. J., Worrall, L., Rose, T., Le Dorze, G., Breitenstein, C., Hilari, K., Babbitt, E., Bose, A., Brady, M., Cherney, L. R., Copland, D., Cruice, M., Enderby, P., Hersh, D., Howe, T., Kelly, H., Kiran, S., Laska, A.-C., Marshall, J., Nicholas, M., & Webster, J. (2019). A core outcome set for aphasia treatment research: The ROMA consensus statement. *International Journal of Stroke, 14*(2), 180–185. <https://doi.org/10.1177/1747493018806200>
- Wilkinson, M., Dumontier, M., Aalbersberg, I., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Duon, O., Edmonds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., & Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data, 3*(1), 1-9. <http://dx.doi.org/10.1038/sdata.2016.18>