

**6. Louise R Williams et al. & The RELEASE Collaborators (2021) 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*, DOI: [10.1080/02687038.2021.1897081](https://doi.org/10.1080/02687038.2021.1897081)**

*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.

*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.