# The effect of age, education, and early speech and language therapy on aphasia outcomes and recovery

Sophie Madeleine Roberts

UCL Queen Square Institute of Neurology

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Neuropsychology and Cognitive Neuroscience

Primary supervisor: Professor Cathy Price

Secondary supervisor: Dr Thomas Hope

## Declaration page

I, Sophie Roberts, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## Abstract

Post-stroke aphasia outcomes vary due to the influence of many stroke- and non-stroke-related factors. My thesis investigates how 'predictor' variables, which are known early post-stroke, influence long-term language scores and participant-reported recovery in a sample of 749 stroke survivors with heterogeneous lesions and symptoms. The effect of early clinical therapy was also investigated in a smaller sample of participants with left hemisphere lesions and severe-to-moderate initial aphasia.

Five variables had a significant influence on long-term language scores: left hemisphere lesion size, initial severity of aphasia symptoms, age at stroke, amount of formal education and provision of early therapy. The size and significance of these effects, and the proportion of variance explained by these variables depended on the sample investigated. Specifically, age was significant in both samples, whereas education was only significant in the larger and more heterogeneous sample, and early therapy was only significant in the smaller sample of patients with aphasia. Crucially, these experiments also identified how significant effects of one variable depended on others.

Unsurprisingly, larger left hemisphere lesion size and initial aphasia severity had the greatest effect on language ability. Nevertheless, the disadvantage of older age, and the benefit of more education were remarkably robust across language tasks and participants with different lesion sizes and initial language abilities. The exception was participants with large lesions and severe initial aphasia. This group had the lowest scores, irrespective of age or education. Finally, early post-stroke clinical speech and language therapy resulted in better short- and long-term language outcomes. These results have novel implications for future studies aiming to predict individual patients' speech and language outcomes after stroke, and their response to therapy. They also have implications for clinicians providing information and speech and language therapy. Ultimately, this contributes to improving patients' understanding of their recovery potential.

## Impact statement

Loss of language after a stroke, known as aphasia, affects approximately one third of the 12 million new stroke survivors around the world, every year.

Aphasia can affect all aspects of communication, resulting in a significant emotional and financial burden. Patients and their carers are often desperate for information about the likely recovery of their language. However, this personalised information cannot currently be provided, because language outcomes and recovery depend on many different variables that interact in ways that are not fully understood. With both an ageing population, and an increasingly younger population of stroke survivors, the number of stroke survivors living with aphasia is rising, and so there is a compelling need to understand the determinants of recovery, translate this understanding into meaningful information for patients and carers, and facilitate clinicians to optimise the rehabilitation pathway through personalised therapies.

The work in this thesis addresses the problem by (1) identifying which variables significantly influence language ability, and the relative size of their effects, (2) differentiating which language tasks are more, or less, influenced by those variables, (3) determining when their influence depends on other variables, and (4) investigating their effect on language recovery.

The findings advance our knowledge of the determinants of aphasia outcomes. The potential impact of this work spans both aphasia research, and clinical practice.

Regarding aphasia research: the results demonstrate that two demographic characteristics, age at stroke and amount of pre-stroke education, influence performance on language tasks from the Comprehensive Aphasia Test. As these effects spanned multiple groups of participants with varying clinical characteristics, they indicate that this commonly-used clinical assessment is sensitive to demographic variables. Future studies that investigate the determinants of aphasia recovery should (a) include measurements of age and education as prognostic variables, and (b) investigate the impact of other demographic variables on language tasks.

Regarding clinical practice, there are two areas of impact to consider. The first is for patients with very mild language impairments, in whom the interpretation of their language assessment scores may vary depending on their age or education. Given the widespread use of the CAT by speech and language therapists around the UK, there is a compelling rationale to adjust task scores for individuals' age and amount of education. This will enable (a) insight into reasons for low scores, leading to (b) more precise interpretation of language abilities, and ultimately (c) personalised rehabilitation options.

The second area of clinical impact is for patients who have more widespread stroke damage, and who thus experience the most severe symptoms. These patients were demonstrated in this thesis to make less recovery of language impairments irrespective of age or education – but, crucially, did still benefit from the provision of early therapy. This knowledge bolsters the evidence base, whilst reinforcing the clinical experience and judgement that also underlie clinical practice. Ultimately, it further compels speech and language therapists to provide realistic expectations, and individually tailored therapy - as early as can be tolerated - in order to motivate patients and maximise recovery potential.

## Acknowledgements

Thank you Cathy, for this opportunity, and for teaching me how to conduct research to an exceptionally high standard. I have so appreciated your kind and supportive supervision, your positivity, and your constant encouragement to find joy and pride in my work.

Thank you also to the many members of the Language Group, past and present, who have helped, taught, and encouraged me over the years. Thanks especially to Tom, David, Sharon, Howard, and Andrea for sharing your knowledge and expertise and always providing insightful feedback; to Diego for also patiently supporting me with statistics; and to Rachel, Storm, Kate, Hayley and Shumi for sharing your clinical knowledge and experiences, and for being a fabulous team to work with.

Thank you to the many other colleagues and friends who make the FIL such a rewarding and supportive place to work.

To all the patients who participated in this research: thank you for your willingness to share your deeply personal experiences during a particularly challenging and vulnerable time of life.

To my very dear friends, thank you for the many years of fun, support and advice, in this, and in all aspects of life. Very special thanks to Charlotte T for always going above and beyond the call of friendship.

Finally, thank you to my brilliant family – my mum, dad, sister Charlotte, brother-in-law Andy, and partner Richard. I am endlessly grateful for your unfaltering support, wisdom, humour, and kindness.

# UCL Research Paper Declaration Form

## referencing the doctoral candidate's own published work(s)

*Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:*

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

*This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.*

### 1. For a research manuscript that has already been published (if not yet published, please skip to section 2)

a) **What is the title of the manuscript?**

Click or tap here to enter text.

b) **Please include a link to or doi for the work**

Click or tap here to enter text.

c) **Where was the work published?**

Click or tap here to enter text.

d) **Who published the work?** (e.g. OUP)

Click or tap here to enter text.

e) **When was the work published?**

Click or tap here to enter text.

f) **List the manuscript's authors in the order they appear on the publication**

Click or tap here to enter text.

g) **Was the work peer reviewed?**

Click or tap here to enter text.

h) **Have you retained the copyright?**

Click or tap here to enter text.

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

Click or tap here to enter text.

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☐

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

a) **What is the current title of the manuscript?**

The impact of the amount of pre-stroke formal education on language ability and recovery after stroke.

b) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

No

c) **Where is the work intended to be published?** (e.g. journal names)

To be decided.

d) **List the manuscript's authors in the intended authorship order**

Sophie Roberts, Rachel Bruce, Thomas Hope, Sharon Geva, Kate Ledingham, Storm Anderson, Hayley Woodgate, Andrea Gajardo-Vidal, Diego Lorca-Puls, Jennifer Crinion, Alexander Leff, David Green, Cathy Price.

e) **Stage of publication** (e.g. in submission)

In preparation

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

Study conception & design: Sophie Roberts & Cathy Price

Data collection: Sophie Roberts, Rachel Bruce, Kate Ledingham, Storm Anderson, Hayley Woodgate

Data analysis & interpretation: Sophie Roberts, Cathy Price

Manuscript preparation: Sophie Roberts, Rachel Bruce, Thomas Hope, Sharon Geva, Kate Ledingham, Storm Anderson, Hayley Woodgate, Andrea Gajardo-Vidal, Diego Lorca-Puls, Jennifer Crinion, Alexander Leff, David Green & Cathy Price

4. **In which chapter(s) of your thesis can this material be found?**

   Chapter 5

5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

*Candidate*

[redacted]

*Date:*

13.03.2024

*Supervisor/ Senior Author (where appropriate)*

[redacted]

*Date*

**16/03/2024**

# UCL Research Paper Declaration Form

## referencing the doctoral candidate's own published work(s)

*Please use this form to declare if parts of your thesis are already available in another format, e.g. if data, text, or figures:*

- *have been uploaded to a preprint server*
- *are in submission to a peer-reviewed publication*
- *have been published in a peer-reviewed publication, e.g. journal, textbook.*

*This form should be completed as many times as necessary. For instance, if you have seven thesis chapters, two of which containing material that has already been published, you would complete this form twice.*

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

j) **What is the title of the manuscript?**

Better long-term speech outcomes in stroke survivors who received early clinical speech and language therapy: What's driving recovery?

k) **Please include a link to or doi for the work**

https://doi.org/10.1080/09602011.2021.1944883

l) Where was the work published?

Neuropsychological Rehabilitation

m) **Who published the work?** (e.g. OUP)

Taylor & Francis

n) **When was the work published?**

2021

o) **List the manuscript's authors in the order they appear on the publication**

Sophie Roberts, Rachel M Bruce, Louise Lim, Hayley Woodgate, Kate Ledingham, Storm Anderson, Diego L Lorca-Puls, Andrea Gajardo-Vidal, Alexander P Leff, Thomas M H Hope, David W Green, Jennifer T Crinion and Cathy J Price.

p) **Was the work peer reviewed?**

Yes

q) **Have you retained the copyright?**

No

r) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

   f) **What is the current title of the manuscript?**

   Click or tap here to enter text.

   g) **Has the manuscript been uploaded to a preprint server?** (e.g. medRxiv; if 'Yes', please give a link or doi)

   Click or tap here to enter text.

   h) **Where is the work intended to be published?** (e.g. journal names)

   Click or tap here to enter text.

   i) **List the manuscript's authors in the intended authorship order**

   Click or tap here to enter text.

   j) **Stage of publication** (e.g. in submission)

   Click or tap here to enter text.

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

   Study conception & design: Sophie Roberts & Cathy Price

   Data collection: Sophie Roberts, Rachel M Bruce, Louise Lim, Hayley Woodgate, Kate Ledingham, & Storm Anderson.

   Data analysis & interpretation: Sophie Roberts & Cathy Price

   Manuscript preparation: Sophie Roberts, Rachel M Bruce, Louise Lim, Hayley Woodgate, Kate Ledingham, Storm Anderson, Diego L Lorca-Puls, Andrea Gajardo-Vidal, Alexander P Leff, Thomas M H Hope, David W Green, Jennifer T Crinion and Cathy J Price.

4. **In which chapter(s) of your thesis can this material be found?**

   **Chapter 6**

5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

*Candidate*

[redacted]

*Date:*

13.03.2024

*Supervisor/ Senior Author (where appropriate)*

[redacted]

*Date*

**16/03/2024**

## Contents

## List of Figures

## List of Tables

**Abbreviations**

| | |
|---|---|
| AAC | Alternative and Augmentative Communication |
| ANCOVA | Analysis of Covariance |
| ANOVA | Analysis of Variance |
| ARTQ | Aphasia Recovery and Therapy Questionnaire |
| AWC | Auditory word comprehension |
| BDAE | Boston Diagnostic Aphasia Evaluation |
| BF | Bayes Factor |
| BMI | Body Mass Index |
| BNT | Boston Naming Test |
| CAT | Comprehensive Aphasia Test |
| CAUK | Communication Access UK |
| CIAT | Constraint-Induced Aphasia Therapy |
| CRN | Clinical Research Network |
| CSF | Cerebral Spinal Fluid |
| CT | Computed Tomography |
| ET | Early therapy |
| FAST | Frenchay Aphasia Screen Test |
| fMRI | functional Magnetic Resonance Imaging |
| ICF | International Classification of Functioning, Disability and Health |
| ICW | Information Carrying Word |
| LH | Left hemisphere |
| LT | Later therapy |
| MNI | Montreal Neurological Institute |
| MRI | Magnetic Resonance Imaging |
| NICE | National Institute for Health and Care Excellence |
| NIHR | National Institute for Health Research |
| NIHSS | National Institute of Health Stroke Scales |
| OR | Odds Ratio |
| PET | Positron emission tomography |
| PLORAS | Predicting Language Outcome and Recovery After Stroke |
| PNT | Philadelphia Naming test |
| PROM | Patient-Reported Outcome Measure |

| | |
|---|---|
| RCSLT | Royal Collect of Speech and Language Therapists |
| RF | Radio Frequency |
| RH | Right hemisphere |
| RS | Reading severity |
| SD | Standard deviation |
| SDOH | Social Determinants of Health |
| SLTA | Standard Language Test of Aphasia |
| SM | Semantic memory |
| SPD | Spoken picture description |
| SPM | Statistical Parametric Mapping |
| SS | Speaking severity |
| SSNAP | Sentinel Stroke National Audit Programme |
| TPS | Time post-stroke |
| US | Understanding severity |
| VBM | Voxel Based Morphometry |
| VIF | Variance Inflation Factor |
| VLSM | Voxel-Based Lesion Symptom Mapping |
| WAB-AQ | Western Aphasia Battery Aphasia Quotient |
| WAIS | Wechsler Adult Intelligence Scale |

# Chapter 1: General Introduction

## 1.1 Summary of introduction

Aphasia has a devastating and persisting impact on the lives of millions of people around the world. Recovery from aphasia is possible, but extremely variable, because of many different factors that influence outcomes and recovery – and it is not currently possible to predict who will recover, nor the speed of recovery. Speech and language therapy is thought to improve aphasic symptoms and reduce the burden of impairments – but the true impact of therapy cannot be known without knowing how patients would recover without therapy.

Prior studies of the variables that influence aphasia outcomes and recovery have not sufficiently explored their independent effects, nor how their effect depends on other variables. This is essential in order to provide accurate, individual language recovery prognoses. The overarching goal of this thesis is therefore to further our understanding of individual variation in aphasia outcomes and recovery, by: investigating (i) which variables have the biggest impact on a range of language outcomes and recovery in stroke survivors; (ii) which specific language functions are most sensitive to their effects, (iii) how their effects depend on other variables, and (iv) whether their influence facilities recovery from aphasia.

This introductory chapter is presented in five main sections. After this summary, the second section (1.2) describes relevant background information, including the prevalence and burden of aphasia (1.2.1), the earliest descriptions of aphasia, including its varying presentations and symptoms (1.2.2), the advantages and disadvantages of the Boston aphasia classification system (1.2.3), and the routine clinical assessment and diagnosis of aphasia (1.2.4). The third (and longest) section (1.3) focuses on aphasia recovery which is the main topic of my thesis experiments. This includes consideration of individual variation in recovery from aphasia (1.3.1), the biological mechanisms that underlie recovery (1.3.2), the need for accurate individual prognosis (1.3.3), variables that may, or may not, influence aphasia outcomes and recovery (1.3.4), and an evaluation of the strengths and limitations of prior studies that investigated these variables (1.3.5). Section 1.4 explains how the UCL PLORAS study attempts to overcome some of the current challenges faced by prior multivariate studies. Finally, Section 1.5 summarises how I used PLORAS data in the four Experiments reported in this thesis, which collectively aim to investigate (A) which non-lesion variables have the largest impact on a range of language outcomes, (B) whether and how their effects depend on other variables, and (C) how they contribute to aphasia recovery.

## 1.2 Aphasia background

### 1.2.1 The prevalence, and burden, of aphasia

Over 100 million people worldwide are living with stroke, with an estimated 1 in 4 people estimated to have a stroke in their lifetime (Feigin et al., 2022). In the UK alone, the Stroke Association report that 100,000 people experience a stroke every year, with an estimated 1.3 million stroke survivors in the UK. They live with a number of disabling and devastating consequences, including difficulties with movement, cognition, language, swallowing, emotions, and fatigue (NICE, 2023).

One third of stroke survivors will experience an acquired speech and language disorder called aphasia (Mitchell et al., 2021). An estimated 350,000 stroke survivors in the UK live with aphasia. Aphasia limits a person's ability to use and understand language, by affecting speaking, understanding, reading, and writing. Its impact is distressing and pervasive, affecting relationships and social interaction, psychological wellbeing, and the ability to work (Hilari, 2011, Bullier et al., 2020). In turn, not being able to work can result in loss of identity, increased financial burden on family members, and/or financial hardship (Stroke Association, 2012). Even when disability benefits and support are available, they can be difficult to negotiate at the best of times, and are thus extra challenging for people with aphasia (Parr, 2007).

Aphasia does not affect a person's intelligence, but this unfortunate assumption is often made due to low public awareness and knowledge of aphasia (Code et al., 2016, Hill et al., 2019). Consequently, people with aphasia can experience negative interactions and environmental barriers, leading to further isolation and feelings of exclusion from society (Parr, 2007). This is all compounded by the uncertainty, chaos and worry that accompanies a life-changing stroke itself.

### 1.2.2 Early descriptions of aphasia

The first cases of aphasia were identified and described in the late 1800s, by pioneering scientists Paul Broca and Carl Wernicke. They observed that patients with focal lesions to the left posterior inferior frontal region, and left posterior superior temporal regions of the brain resulted in impairments to speech production and comprehension respectively (Broca, 1861, Wernicke, 1874). This led to two primary theories: first, that language processing is lateralised to the left hemisphere, and second, that speech production and speech comprehension are functionally distinct, and localised to two discrete areas, newly coined ‘Broca’s’ and ‘Wernicke’s’ areas. Subsequent work by Ludwig Lichtheim in the late 1800s and Norman Geschwind in the mid-1900s, amongst others, continued attempting to identify the neural correlates of broad aphasic symptoms. A white matter network fibre connecting the two areas was identified - the arcuate fasciculus (Lichtheim, 1885, Geschwind, 1965). Thus arose a ‘classic’ model of language organisation in the brain, based largely on the contributions of these four researchers. The model formed the foundation of the ‘localisation’ approach, which states that specific brain regions are responsible for specific functions, and injury to those regions results in impairment to, or loss of, that function (Rao, 1994).

Throughout the 20th century, the assessment of language functions became increasingly comprehensive, enabling the identification of a number of distinct aphasia subtypes which were characterised by different patterns of impairments and abilities. The culmination of this work was the development of the Boston Classification System (Goodglass, 1993).

### 1.2.3 The Boston classification of aphasia system

Eight main aphasia subtypes are described in the Boston classification system, and are described below, with reference to Sheppard and Sebastian (2021) and Bunker and Hillis (2022). The first four subtypes are labelled as non-fluent (or expressive) aphasia; the second four subtypes are fluent (or receptive) aphasia.

**Subtypes of non-fluent (or expressive) aphasias:**

**Broca’s aphasia** is predominantly characterised by slow and effortful speech production, including poor repetition, and possibly stereotyped phrases. Comprehension is typically intact, but can be worse when sentences are complex.

**Transcortical motor aphasia** is similar to Broca's aphasia, but with preserved repetition. Speech is more fluent during repetition than spontaneous speech production.

**Global aphasia** is considered the most severe aphasia subtype, as it affects all aspects of language. There may be limited (or no) spoken output, or any speech is characterised by perseverations (i.e. multiple complete or incomplete attempts at producing a word) or stereotyped phrases. Comprehension is impaired; usually severely.

**Transcortical mixed aphasia** is similar to global aphasia, but with preserved repetition.

**Subtypes of fluent (or receptive) aphasias:**

**Wernicke's aphasia** is characterised by fluent but meaningless speech, often containing paraphasias (when the wrong word is produced, e.g. *dog* instead of *cat*), neologisms (nonwords, which have no meaning, e.g. *swanip*), and jargon (an incomprehensible mixture of real words and nonwords). Repetition abilities are also poor. Comprehension is typically impaired, and a patient will often have limited awareness of their difficulties.

**Transcortical sensory aphasia** is similar to Wernicke's aphasia, but with preserved repetition.

**Conduction aphasia** is characterised predominantly by disproportionately poor repetition, alongside relatively fluent speech production and intact comprehension. Speech may contain phonemic errors, and naming may be difficult.

**Anomic aphasia** is characterised by a specific difficulty with naming, or retrieving words. Speech is otherwise fluent, albeit with the pauses and hesitations that accompany word-finding difficulties.

The term 'anomia' also describes a symptom that can occur in all subtypes of aphasia, not just in anomic aphasia. Indeed, these word-finding difficulties are often considered a hallmark symptom of expressive aphasias.

### Advantages and disadvantages of the Boston aphasia classification

Aphasia assessment batteries such as the Boston Diagnostic Aphasia Evaluation, or BDAE, (Goodglass et al., 2001), and the Western Aphasia Battery, or WAB, (Kertesz, 2007) were designed to facilitate diagnosis of aphasia subtype. The advantages of these assessments include long-standing familiarity amongst healthcare professionals of terms like ‘Broca’s’ and ‘Wernicke’s’, ‘fluent’ and ‘non-fluent’, or ‘receptive’ and ‘expressive’ (Tremblay and Dick, 2016). For experts in the field, who are familiar with the Boston aphasia classification system, diagnostic labels such as Broca’s and Wernicke’s aphasia offer an instant and succinct impression of an individual’s communication abilities, that can be embellished with indicators of the severity level (e.g. severe expressive aphasia with mild receptive aphasia). Nonetheless, there are a number of shortcomings to this approach that have long been recognised (Caramazza, 1984, Marshall, 2010).

A major limitation of traditional classification models is that they over-simplify the language system and do not account for the multiple cognitive processes that are required to perform language tasks. Consider, for example, the object naming task. As described in Whatmough and Chertkow (2002), this requires the following cognitive processes to be intact: (i) visual perception/recognition of the picture - e.g. the object has a round shape, short stem etc., (ii) conceptual/semantic processing - e.g. it’s a type of *fruit*, (iii) lexical retrieval – it’s called an *‘apple*’ not a *‘banana*’ and (iv) articulatory planning, that coordinates the articulators for the production of the phonemes /æ/ /p/ /ɫ/. An incorrect object naming response could result from an impairment at any of these processing levels, and does not necessarily indicate a language problem or aphasia (e.g. in patients with visual perceptual impairments). This emphasises the need for neuropsychological models that differentiate which types of cognitive processing are impaired.

Another limitation to the Boston aphasia classification system is that patients’ symptoms do not necessarily fit neatly into one aphasia subtype; they can span multiple subtypes. Conversely, a patient may not necessarily experience all of the symptoms associated with a subtype. Thus, diagnosis according to aphasia subtype may overlook some symptoms (or abilities). In other words, a broad classification system can't account for individual variability between aphasia profiles (Marshall, 2010).

Finally, it is well-recognised that aphasia subtypes (e.g. Broca's and Wernicke's aphasia) are not neatly associated with damage to specific parts of the brain. For example, damage to 'Broca's area' in the left posterior inferior frontal cortex does not necessarily cause Broca's aphasia (Mohr et al., 1978, Alexander et al., 1990), and conversely Broca's aphasia can occur following damage that is distant from 'Broca's area' (Fridriksson et al., 2015, Fridriksson et al., 2007, Brais, 1992). These conundrums arose because the conclusions drawn by Paul Broca and Carl Wernicke involved single case post-mortem examinations. They can now be explained, in light of more recent neuroimaging knowledge that language processing is not localised to a few parts of the brain (Broca's and Wernicke's areas) but instead involves widely distributed sets of inter-connected regions that work together to support the multiple linguistic and non-linguistic functions that are required to perform language tasks (Hickok, 2009, Poeppel et al., 2012, Fujii et al., 2016, Price, 2018, Price, 2000).

### 1.2.4 Routine clinical assessment and diagnosis of aphasia

Clinical guidelines outlining the stroke diagnosis and rehabilitation pathway in the UK dictate that patients should be screened for communication difficulties within 72 hours of onset of stroke symptoms. Typical measures include the National Institute of Health Stroke Scale, or NIHSS (Brott et al., 1989), or the Frenchay Aphasia Screen Test, or FAST (Enderby and Crow, 1996). If a communication impairment is indicated, patients should be referred for further assessment by a speech and language therapist (NICE, 2023).

Assessment of aphasia is recommended to be conducted in line with the World Health Organisation's International Classification of Functioning, Disability and Health (ICF) framework (World Health Organisation, 2001), which specifies four categories of focus. These were interpreted in relation to aphasia by Simmons-Mackie and Kagan (2007), and are described below.

**Assessment of impairments to body structure and function** requires language impairments and abilities to be assessed using a standardised objective aphasia assessment battery (see Varkanitsa and Kiran (2024) for a recent review of impairment-based assessments). Whilst there are variations between these assessments in their specific tests and items, they all typically assess the same broad areas of language, including Naming, Repetition, Reading, Comprehension of Spoken and Written language, and Picture Description/Narrative. Given their length and comprehensive nature, they are not designed to be conducted in full, and it would be unusual for this to be done in clinical practice, particularly when patients have concomitant medical and rehabilitation needs. Instead, key individual tasks are typically selected (such as Naming, Picture description, and Comprehension of spoken words and sentences), from which further assessment targets can be identified if necessary. In summary, impairment-based assessments identify hallmark symptoms of aphasia including word-finding, syntactic, and comprehension difficulties, enabling differential diagnosis, and/or targeted therapy planning.

**Assessment of activities and participation** identifies how aphasia impacts a patient's ability to carry out their typical daily activities and participate in life situations – for example, communicating with family, carrying out personal care, socialising with friends, work, and hobbies. This is typically done through formal assessment of functional language (see Doedens and Meteyard (2020) for a review of functional communication assessments), and informal assessment, such as observation of the patient communicating in a naturalistic environment.

**Assessment of wellbeing and personal factors** involves consideration of a patient's personal identity and emotions. This does not occur via formal assessment, but rather by recognising how certain characteristics (such as demographics, cultural factors and emotions) might influence a person's wellbeing in the context of their communication difficulties.

**Assessment of environment** considers the physical, social, and attitudinal barriers and facilitators present in an individual's environment that affect their ability to communicate. Immediate environmental adjustments could include conversation partner awareness and training. Larger-scale environmental adjustments could include campaigns to increase awareness of aphasia, and provision of communication training within public services such as healthcare and transport, in order to improve accessibility of services and reduce the negative experience people with aphasia can have when communicating with others (for example, the Communication Access UK (CAUK) initiative, led by the Royal College of Speech and Language Therapists, RCSLT).

A final assessment consideration is that aphasia may occur with concomitant motor speech disorders, such as apraxia of speech and dysarthria (Mitchell et al., 2021), which must be differentially diagnosed via separate assessments. Critically, symptoms could overlap with multiple disorders. For example, a phonemic paraphasia such as ‘shaxophone’ (target *saxophone*) could arise from the patient selecting the wrong phoneme /ʃ/ (‘sh’) instead of /s/ (‘s’), or from distorting the phoneme /s/ ‘s’ due to dysarthric muscle weakness (having selected the correct phoneme).

In summary, the broad aims of aphasia assessment are to (1) identify the nature and extent of aphasic impairments, and of any residual communication abilities, (2) determine the impact of aphasia on the patient and their family, (3) gauge psychological and general wellbeing, and (4) establish a baseline communication profile from which subsequent improvement, and thus recovery, can be measured.

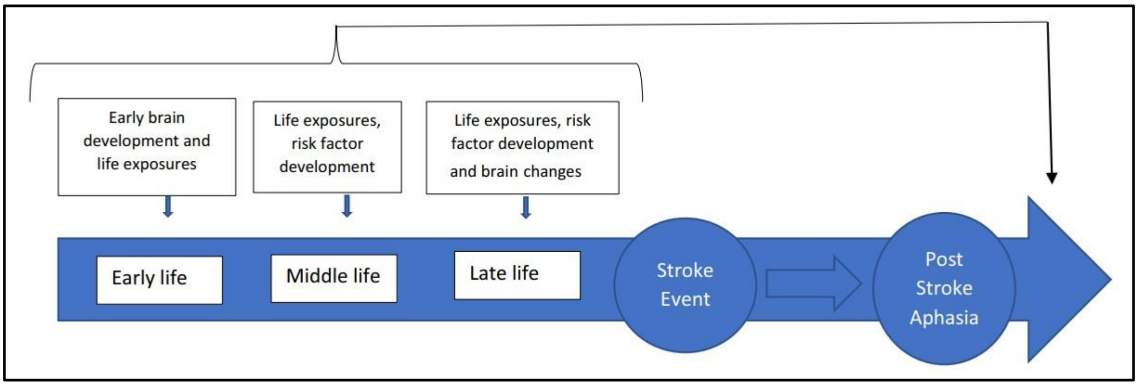
## 1.3 Recovery from aphasia

### 1.3.1 Individual variability in recovery from aphasia

Having established the historical context and current landscape of how aphasia impairments are assessed, I will now turn to the crux of the problem addressed in this thesis – aphasia recovery. Full recovery from aphasia is possible, but extremely variable. Patients’ aphasic symptoms can change over time, even years after stroke. Although most patients improve, many show no change and some deteriorate (Hope et al., 2017). An additional layer of complexity lies in the understanding that recovery can occur with, or without, therapy. Because of this variability, it is not currently possible to accurately predict who will recover.

Variability in aphasia outcomes and recovery may be due to innate individual characteristics (e.g. sex, handedness, pre-stroke language organisation in the brain). It may also arise from early and mid-life experiences (e.g. amount of education received, or occupation), or from biological changes in later life (e.g. cognitive decline due to age and normal neurodegeneration processes); see Figure 1.1 for an illustration, from Jacobs et al. (2023), of the different sources of individual variance across the lifespan. Most critically, variation in the effects of the stroke will inevitably arise from differing patterns of stroke damage between any two patients. Put together, this heterogeneity creates an enormous challenge for first understanding, and second predicting recovery.

**Figure 1.1** Individual variation throughout the lifespan.



**Legend:** Individual variation can arise from multiple factors during early, middle and late life, and after a stroke. Reprinted from Jacobs et al. (2023) with permission from Elsevier.

The overarching goal of this thesis is to understand the influence of some of the key variables that are known, or speculated, to influence aphasia outcomes and recovery. To this end, the following section (1.3.2) will describe the mechanisms of recovery (i.e. the physiological processes that occur following a stroke) and how these manifest in language abilities. Section 1.3.3 discusses the importance of understanding individual recovery with precision, both for clinical practice and for research. Section 1.3.4 will introduce the challenges of predicting aphasia outcomes and recovery, listing the variables that are known, or speculated to influence recovery. Finally, section 1.3.5 highlights the need for multivariate methods that compare the relative influence of variables on language outcomes, and reviews prior studies that have adopted this approach.

### 1.3.2 Biological mechanisms of recovery

Three main stages of language recovery after a stroke have been identified by studies that used a variety of neuroimaging techniques (fMRI, CT, PET). The exact timing and duration of each stage varies slightly across studies and contexts, however, in general: (1) the acute stage describes the hours and days immediately following stroke onset, (2) the subacute stage ranges from weeks to up to six months post-stroke, and (3) the chronic stage ranges from six months to many years post-stroke. Descriptions of these three stages are given below, with reference to prior reviews by Hillis and Heidler (2002), Marsh and Hillis (2006), Nouwens et al. (2015), Gerstenecker and Lazar (2019), and Kiran and Thompson (2019).

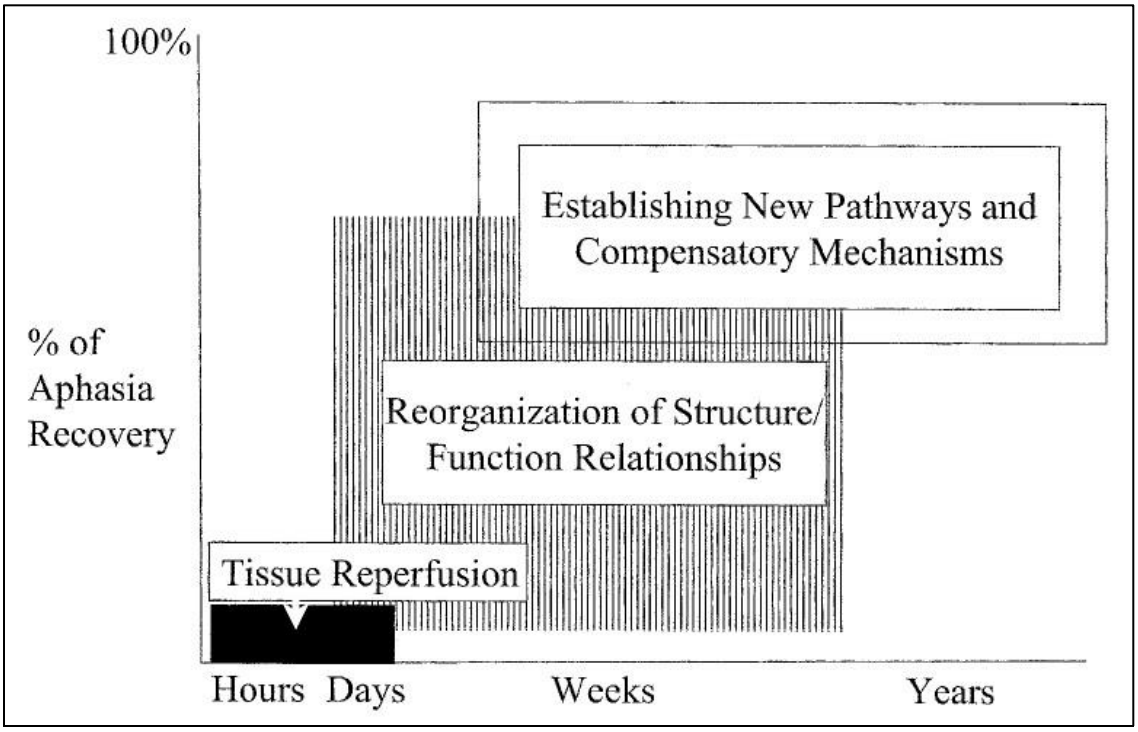
The **acute** stage is primarily characterised by restoration of blood and oxygen supply to the regions of the brain damaged by the stroke; a process called reperfusion. This can occur when cerebral oedema resolves, and tissue swelling and displacement decreases (Gu et al., 2022). Damage to the neural tissue directly surrounding the lesioned area (the ischemic penumbra) is reversible if treatment is received quickly; this is essential for early recovery of language functions (Kiran, 2012).

During the **subacute** stage, diaschisis resolves. Diaschisis is a phenomenon in which areas of the brain that are not directly implicated by a focal lesion can still be impacted if they are connected to the lesioned area by a neural pathway. Diaschisis can result in a widespread breakdown of the language system, which presents clinically as global aphasia – however the damage is temporary, and can be reversed when diaschisis resolves (Wawrzyniak et al., 2022). The subacute phase is also characterised by structural and functional reorganisation. Neuroplastic processes such as synaptogenesis (formation of new synapses), axonal sprouting, dendrite growth, and neural reorganisation result in new neural pathways being formed (neurogenesis), the utilisation of previously unused neural pathways, and recruitment of the right hemisphere; all of which expands the available neural network (Murphy and Corbett, 2009, Alia et al., 2017). This has been evidenced by studies that have observed language activation in preserved regions of the left hemisphere, and in homologous right hemisphere regions (Saur et al., 2006, Turkeltaub et al., 2011). It is during this sub-acute stage that the fastest period of language recovery occurs – even if no language therapy is provided (El Hachioui et al., 2013a). The restoration of language, without relearning, is known as ‘spontaneous recovery’.

During the **chronic** stage of recovery, the brain is typically considered to be in a stable state, although mechanisms of neuroplasticity (e.g. synaptic sprouting) remain operational, forming new neural pathways which may support language. During this time, a patient may also be learning compensatory communication strategies. These mechanisms can occur well beyond the first year post-stroke, giving rise to the notion that language recovery can continue for years (Hope et al., 2017, Smania et al., 2010, Stark, 2010, Holland et al., 2017, Johnson et al., 2019).

The three stages of recovery are illustrated in Figure 1.2.

**Figure 1.2** The three stages of aphasia recovery.



**Legend:** The timing of each stage is not definitive. Broadly, the acute stage, primarily characterised by tissue reperfusion, occurs in the hours and days immediately following the stroke. The sub-acute stage, characterised by structural and functional reorganisation, ranges from weeks to roughly six months post-stroke. The chronic stage is typically six months and more post-stroke, and is characterised by the establishment of new neural pathways and compensatory mechanisms. Reprinted from Hillis and Heidler (2002) with permission from Taylor and Francis.

### 1.3.3 The need for accurate individualised prognoses

The observation that the fastest period of recovery occurs within the first 6 months of stroke (El Hachioui et al., 2013a) has led to the unfortunate impression that recovery will *only* occur within this period and that, afterwards, language plateaus. This is not the case: a number of studies have shown that recovery can, and indeed does, continue well beyond this period (see studies cited in the previous paragraph). The misleading message that language recovery plateaus could discourage patients from engaging in further therapy or language practice opportunities; thus itself limiting recovery. Whilst the ‘plateau’ myth is slowly being debunked, it prevails amongst patients – including some of those who participated in the experiments reported in this thesis. It is therefore paramount to understand individual recovery capacity. The potential benefits span both clinical practice and research.

An accurate understanding of recovery is crucial in clinical practice, allowing healthcare professionals to provide patients and caregivers with a prognosis for their likely recovery potential. Prognostication is already widely used in many health conditions, particularly in cancers, and often considered crucial for planning rehabilitation schedules and for being able to safely resume daily activities. But prognoses are not currently possible for aphasia, despite this information being desperately sought by patients and carers (Worrall et al., 2011). A prognosis for aphasia recovery could support: realistic goal-setting, enhanced engagement with therapy, boosting mood and mental health, assisting with future planning, and managing expectations. This is of course caveated by the understanding that a prognosis may be poor. For example, a patient's language symptoms may remain stable, experiencing no further improvement; or, they may decline. Knowing this could be unmotivating at best, or damaging at worst, if it reduces engagement with therapy, and/or worsens mood (Ahalt et al., 2012). In such cases, the prognosis would need to be framed such that it manages expectations, avoids false hope, and facilitates the planning of appropriate compensatory therapies and strategies to maximise residual communication skills. Understanding and predicting individual recovery therefore comes with a responsibility to ensure any prognosis is delivered sensitively, and with appropriate support in place for the patient and their carers (Graugaard et al., 2011).

Accurate prognoses would also benefit speech and language therapists in planning and maximising treatment. Goal-setting, which usually occurs in tandem with the therapist, patient and their carers, could be aided when equipped with the knowledge of a likely recovery timeframe.

A third need to understand recovery is the potential benefit for aphasia therapy research, and for more accurately evaluating the effects of speech and language therapy. Evidence from numerous therapy trials suggest that therapy is effective (at a group level) for improving language impairments and functional communication after a stroke - however, these effects are often weak, inconsistent, and not maintained long-term (Brady et al., 2016). More critically, a benefit of therapy at the group level is not necessarily consistent in all the patients who received it (Lazar et al., 2008, Code et al., 2010, Breitenstein et al., 2017, Kristinsson et al., 2023), nor is it necessarily replicated in other studies. This is likely because of the individual variation in recovery without therapy. For example, recovery potential may be limited after extensive damage to critical language regions, irrespective of whether therapy is provided or not. In this scenario, the lack of recovery could be falsely interpreted as a lack of therapy effect (false negative). Conversely, an improvement could be falsely interpreted as a therapy effect (false positive) if a patient was likely to recover without therapy. If a benefit of therapy depends on lesion site (Aguilar et al., 2018, Fleming et al., 2020, Hope et al., 2021), studies may be unable to detect large and robust therapy effects when data are pooled from patients with heterogeneous lesions. This variance needs to be controlled in order to accurately evaluate therapy effects, namely by grouping participants according to different predictive factors (e.g. matching them for lesion site). This is almost impossible for therapy trials, because of the sheer number of factors that could influence recovery. In particular, detailed lesion data are rarely available due to the cost and logistical challenge of obtaining it, and the expertise required to interpret and analyse the data.

Understanding how an individual patient would recover without therapy is therefore critical for understanding the real effects of therapy. A vision for this approach would be a stratification system, whereby patients are grouped into prognostic categories according to key predictors of outcome. In patients with good prognoses, who are expected to recover within a certain timeframe, a benefit of therapy would need to show accelerated recovery. In patients with poor prognoses (e.g. those expected to remain stable or deteriorate), a benefit of therapy would need to show just a small improvement, or less of a decline. This then also paves the way for investigating how moderating certain therapy variables (e.g. amount, frequency, timing, and type) influences outcomes further – and further enabling therapists to select appropriate treatment approaches, and maximise therapy impact – adopting the precision rehabilitation approach advocated by Brady et al. (2022).

This approach, which has been labelled “predictive enrichment” (Simon, 2010), describes the identification of key characteristics that increase (or decrease) suitability for a certain treatment approach. Specifically, in this field, predictive enrichment could help to identify patients who are more likely to respond to a particular type of therapy, or those who would benefit from brain stimulation therapies designed to induce neuroplasticity (Harvey and Hamilton, 2022, Holland and Crinion, 2012).

### 1.3.4 Variables that may influence aphasia outcomes and recovery

Predicting recovery from aphasia has thus far not been possible because of group heterogeneity in recovery, arising from many variables whose influence on recovery is still not fully understood (Price et al., 2017). Some of these variables are known and their effects well-established; others remain under-researched, and their effects only hypothesised. The focus of this section therefore turns to identifying and understanding the influence of these variables, both on aphasia recovery and on the effects of other variables.

This section starts with an overview of the many variables that are thought to influence aphasia outcomes. In brief, these variables are grouped into 6 broad categories: (1) stroke variables, (2) demographic variables, (3) cognitive variables, (4) speech and language therapy variables, (5) social variables and (6) health variables. They have been discussed in detail in seminal review articles by Ferro et al. (1999), Lazar and Antoniello (2008), Plowman et al. (2012), Watila and Balarabe (2015), Gerstenecker and Lazar (2019), Varkanitsa and Kiran (2022), O'Halloran et al. (2023), and Varkanitsa and Kiran (2024), which cover decades of past research. The sections that follow are intended to provide a brief overview of each variable. A more detailed review of the evidence for the specific variables investigated in Experiments 1-4 will then be presented in the relevant chapter.

#### Stroke variables

**Lesion size and site** describe the precise location and extent of the lesion in the brain. If the region damaged by the stroke is responsible for some aspect of language processing, this may manifest as an impairment to that language function. The severity of the impairment may depend on (i) the degree of damage to that region, and (ii) whether non-damaged and preserved regions of the brain can adopt that function. Measurement of lesion factors requires access to neuroimaging technology (MRI or CT scans), as well as software and expertise to analyse the scan data. Much of the current focus lies in identifying ‘critical language regions’: brain regions which, when damaged, result in long-term language impairments. Lesion location and size are considered to be amongst the strongest predictors of aphasia outcomes and recovery (Hope et al., 2013).

**Brain health** describes the integrity of brain tissue. Markers of brain health include: white matter hyperintensities (leukoaraiosis), atrophy, enlarged perivascular spaces, lacunes, cerebral microbleeds and small subcortical infarcts (Hannan et al., 2023), and all may impact on neural function. Presence of these markers increases with age, and their prevalence is also greater amongst stroke survivors because of the increased cardio-vascular risk factors within this population that contribute to deteriorating brain health. Aphasia recovery depends on intact brain tissue and pathways, so if these are implicated over time by progressive deterioration, a significant negative influence of brain health on recovery is likely. The last decade of research has seen increased attention given to the role of brain health in relation to cognitive and language abilities, but thus far with mixed conclusions (Hannan et al., 2023)

**Initial aphasia severity** describes the severity of aphasia symptoms, typically in the acute or sub-acute stage after stroke (although the exact time period of ‘initial’ is not definitive). Aphasia severity is typically measured according to severity of impairments (rather than functional communication), often using a composite score measuring different language skills such as the WAB-AQ, or a brief screen such as the FAST. Participants may be assigned to a severity category; alternatively, assessment scores themselves may be used as a linear variable. Initial aphasia severity has long been established as a significant predictor of language recovery (Laska et al., 2001).

**Time post-stroke** describes the amount of time between stroke onset and a defined subsequent point – which in research is often designated by a language assessment or period of therapy. This can be hours, days, weeks, months or years post-stroke. Whilst better language abilities have been observed with longer time post-stroke, suggesting it may be predictive of better recovery (Hope et al., 2017), decline in abilities has also been observed over time (Holland et al., 2017).

Other stroke factors that have been considered to influence language recovery include aphasia type (with global aphasia thought to have the worst prognosis), stroke type, and initial stroke severity.

### Demographic variables

The demographic factors most frequently reported are age at stroke, biological sex, handedness and pre-stroke education. A detailed review of the evidence for these factors is provided in Experiments 1-3; but briefly, the influence of age and education on aphasia outcomes and recovery remains unclear and inconclusive. A more confident assertion can be made for sex and handedness, the influence of which appears to be negligible.

### Cognitive variables

Non-linguistic (or ‘domain general’) cognitive functions include attention, working memory, executive function, cognitive control, semantic/conceptual knowledge, and learning ability. When integrated with linguistic skills, they are thought to be essential for language processing. Aphasia is often (although not always) accompanied by non-linguistic cognitive impairments (Gilmore et al., 2019b, Fonseca et al., 2019, Yao et al., 2020). If such impairments are indicated after a stroke, cognitive screening and subsequent detailed assessment is recommended, particularly for visual function, memory, and attention (NICE, 2023). The impact of impaired cognition on language is particularly pertinent to the success of speech and language therapy. Therapy activities typically use pictures, written words, and auditory stimuli to support the relearning of words and their relationship to meanings. Engagement with these activities depends on certain cognitive functions being intact: attention is required to maintain focus on task instructions and stimuli. Verbal short-term memory is essential for retaining and repeating phonological sequences. Visual short-term memory is necessary for retaining key features of visual stimuli in mind while making linguistic decisions. Cognitive control is essential for switching between different tasks, and for suppressing incorrect responses. Metacognitive strategies are important for maintaining learning practice outside of clinical settings, for example for planning one’s independent learning, and identifying appropriate and effective learning strategies. Whilst small samples and individual variation contribute to mixed evidence (Diedrichs et al., 2022), it is generally thought that better non-linguistic cognitive abilities – particularly executive function, attention, and short-term memory (Gilmore et al., 2019b, Pisano et al., 2022, Varkanitsa et al., 2023) support relearning and functional communication, and therefore are associated with better therapy response and greater recovery.

### Speech and language therapy variables

In the UK, clinical guidelines dictate that speech and language therapy must be offered to individuals with post-stroke communication difficulties. The role of a speech and language therapist includes: delivering individualised therapy for specific impairments, providing communication aids and relevant training, signposting patients to support organisations, and providing appropriate information and training to the patient's multidisciplinary team and carers (NICE, 2023). Therapy for aphasia can be restorative, where the goal is to improve language impairments using the remaining neural network (targeting the 'body structure and function' element of the ICF framework). Alternatively, therapy can be compensatory, where the goal is to compensate for language impairments that can't be restored (targeting the 'activity and participation' element of the ICF framework). Compensatory therapy is likely to be centred around real-world communication and treatment goals. Therapy can be altered for amount and frequency and timing (i.e. how long since stroke onset). Furthermore, therapy content and approach typically varies between individual patients, being tailored to each patient's specific abilities and impairments, goals, and preferences. A detailed review of the evidence for speech and language therapy for aphasia is presented in Experiment 4. Briefly, and as previously discussed, the evidence from numerous therapy trials suggests it is effective (Brady et al., 2016) - but given the extreme variations in recovery, the effect of therapy cannot truly be known without an understanding of how an individual patient would recover naturally without therapy. The prior evidence for therapy therefore needs to be reviewed in light of this statement.

### Social variables

Also referred to as 'social determinants of health' (SDOH), non-medical social variables are already known to influence general health outcomes (World Health Organisation, 2008). They include early development, income, socioeconomic status, social inclusion, and access and attitudes to healthcare – amongst many more. Evidence suggests that lower socioeconomic status, social isolation, and no health insurance (in the USA) are associated with increased risk of stroke

(Reshetnyak et al., 2020). Employment and higher income are also thought to be associated with greater treatment adherence, and overall increased participation and quality of life in patients with neurological disorders (Frier et al., 2017). However, the role that social factors play in aphasia outcomes and recovery has not been investigated in depth (O'Halloran et al., 2023). Factors such as ethnicity, socioeconomic status and social support have started to be investigated in recent years, but the evidence base to date is sparse and thus inconclusive. One reason for this, proposed by O’Halloran et al., is the lack of routine data collection for social factors in aphasia research. This highlights the need for this data to be collected routinely by aphasia trials – and critically, consistently between different trials, to support data sharing and collaboration (Wallace et al., 2022).

#### Health variables

The ‘health’ factors described to be important for aphasia recovery strongly overlap with medical and lifestyle factors. A distinction needs to be made between physical health factors and mental health factors. Physical health factors include cardiovascular risk factors (e.g. hypertension, coronary disease, atrial fibrillation, and diabetes), smoking status, alcohol use, body mass index (BMI), comorbid conditions – and many more. As with the social factors discussed in the previous section, health factors are starting to be reported more in aphasia trials – but again, the research remains in its infancy. Mental health factors include depression, and emotional state (including mood and motivation). Their role in aphasia recovery is more established: post-stroke depression and low mood are thought to significantly impact motivation and response to therapy, and ultimately result in worse recovery (Code and Herrmann, 2003).

### 1.3.5 Evaluating prior literature

What follows is a review of prior multivariate analyses that report (1) how much variance in an outcome is explained by a set of variables, and (2) which of those variables had a significant effect on the outcome. The aim here was to overview the current landscape of aphasia prediction research, and to assess (i) the range of variance that has been accounted for, (ii) which variables are more, and less, commonly included, and which do and don’t contribute significantly to the outcome, and (iii) which outcomes are more, and less, commonly investigated. I will then discuss the variation between how each of these studies was conducted, and how this may limit the conclusions we can draw from these studies as a whole. Only studies that used a multivariate approach to investigating the influence of a number of variables on language outcomes or recovery in a group of participants (both with and without therapy) were considered. The variables and outcomes investigated and the amount of variance explained by each model, are summarised in Table 1.1. If multiple models were conducted within the same study, with varying combinations of variables, only the ‘final’ one (as judged by the study authors) is reported here, or failing that, the one which explained the greatest amount of variance.

**Table 1.1** Multivariate analyses that investigate the relative influence of different variables on language outcomes and/or recovery.

| Study | N | Predictor variables | Outcome measure & timing | Amount of variance explained |
|---|---|---|---|---|
| Lazar et al. (2008) | 22 | Baseline language severity*, age, education, lesion size. | Composite language score 90 days post-stroke | 33% |
| El Hachioui et al. (2013b) | 130 | Initial phonological abilities, age, education, stroke subtype, functional independence. | Functional verbal communication (BDAE) 1 year post-stroke | 56% |
| Godecke et al. (2013) | 79 | Initial aphasia severity*, speech and language therapy amount*, age, initial stroke severity (mRS)*, number of therapy sessions. | Change in aphasia severity (WAB-AQ) from baseline to 5 weeks post-stroke | 30% |
| Hope et al. (2013) | 270 | Time post-stroke*, lesion volume, lesion location | Composite ‘speech production’ score (word and sentence repetition, object naming and picture description) | 59% |
| Ramsey et al. (2017) | 132 | Initial aphasia severity*, education*, lesion volume*, age, amount of speech and language therapy. | Composite language score 3 months post-stroke | 69% |
| Kim et al. (2019) | 235 | Age*, education*, initial NIHSS score*, initial aphasia severity*, lesion volume*, obesity, lesion location. | Change in overall language ability (K-FAST) from 1 week to 1 year post-stroke | 40% |
| Nakagawa et al. (2019) | 121 | Age*, lesion location*, initial aphasia severity*, initial phonological & semantic abilities*. | Overall aphasia severity (SLTA) | 56% |
| Osa García et al. (2020) | 20 | Initial aphasia severity*, age, lesion size, lesion load to arcuate fasciculus. | Change in composite aphasia score | 73% |

| | | | | |
|---|---|---|---|---|
| | | | from baseline to 2 weeks post-stroke | |
| Shin et al. (2022) | 4443 | Stroke type*, age*, sex* education*, initial NIHSS score*, atrial fibrillation*, comorbidities*, initial aphasia severity*, initial cognitive abilities*, motor function*, hypertension, diabetes,<br>BMI, smoking, alcohol. | Composite language score (K-FAST) 5 years post-stroke | Not reported |
| Harvey et al. (2022) | 18 | Lesion volume*, self-reported burden of aphasia on activity & participation*, age, sex, education, time post-stroke, amount of speech and language therapy. | Change in aphasia severity (WAB-AQ) from baseline to 1-year follow-up | 68% |
| Gadson et al. (2022) | 85 | Lesion size*, income, education, white matter disease, race, age, sex, time post-stroke. | WAB-AQ | 51% |
| Johnson et al. (2022) | 106 | Non-linguistic cognitive ability (WAIS)*, time post-stroke*, proportion of damage to critical ROIs (superior longitudinal fasciculus & posterior insula)*, brain health (white matter hyperintensities)*, lesion volume, age at stroke. | Overall aphasia severity (WAB-AQ) | 55% |
| Kristinsson et al. (2023) | 107 | Baseline (not initial) aphasia severity (WAB-AQ)*, time post-stroke*, age, education, cognitive reserve (WAIS score). | Change in Naming 1 week 1 month and 6 months post-therapy (PNT) | 11-21% |
| Jacobs et al. (2023) | 165 | Age*, family income*, sex, race*, family size*, aphasia type*, education* time post-stroke, marital status, region of residence. | Naming (BNT) | Not reported |

**Legend:** * = statistically significant. BDAE = Boston Diagnostic Aphasia Examination. BNT = Boston Naming Test. mRS = modified Rankin Scale. PNT = Philadelphia Naming Test. SLTA = Standard Language Test of Aphasia. WAB-AQ = Western Aphasia Battery Aphasia Quotient. K-FAST = Korean version of Frenchay Aphasia Screening Test. WAIS = Wechsler Adult Intelligence Scale. NIHSS = National Institute of Health Stroke Scale. BMI = Body Mass Index. ROI = Region of interest.

The first observation, that is immediately clear from Table 1.1, is that the amount of variance in language outcomes explained by prior prediction models varies enormously. The study that explained the greatest amount of variance was by Osa García et al. (2020), who explained 73% of variance in early recovery (within the first two weeks post-stroke) in 20 participants. In this model, initial aphasia severity was the only significant predictor of language recovery, even though lesion size and load (to the arcuate fasciculus) were also included. Conversely, the study that explained the smallest amount of variance was by Kristinsson et al. (2023), which explained 11% of variance in recovery of Naming after therapy in 107 participants with chronic aphasia. This model controlled for baseline (not initial) aphasia severity, and did not include any lesion factors.

The second observation from Table 1.1 is that the combination of variables that were investigated fluctuated, and some variables are included more frequently than others. Whilst some variables are common to most studies (such as sex, lesion size and age), many variables are unique to one or a few. Lesion location was only included as a variable in 5 studies (Hope et al., 2013, Kim et al., 2019, Nakagawa et al., 2019, Osa García et al., 2020, Johnson et al., 2022), which accounted for 40-73% of the variance. This suggests that (i) non-lesion factors may explain a higher proportion of the variance than previously expected; (ii) the choice of lesion locations considered was insufficient or (iii) both. Indeed, there is ongoing debate over which lesion sites need to be considered and how this choice depends on the type of language outcome or recovery. Social and health factors have only been considered in the last two years. For example, the influence of ethnicity, income and family size on post-stroke language outcomes were investigated in Jacobs et al. (2023) – all of which were found to be significant; the impact of aphasia on activity and participation was investigated in Harvey et al. (2022); and the effect of BMI, diabetes status, smoking and alcohol use (amongst other health factors) was investigated in Shin et al. (2022).

A third observation from Table 1.1 is that the significance of the effects of the same variables differed between studies. Initial aphasia severity was significant in each of the 6 studies that included it. Lesion size was significant in 4 of the 7 studies that included it. Age was significant in 4 of the 13 studies that included it. The amount of speech and language therapy was included in 3 studies, but found to be significant in only one (Godecke et al., 2013).

A fourth observation from Table 1.1 that might explain inconsistency in the significance of each variable is that there were substantial differences in study design that particularly related to sample size (ranging from 22 in Lazar et al. to 4443 in Shin et al.), and participant characteristics. For example, (i) participants over 80 years old were excluded in Kristinsson et al. and Jacobs et al., but not in Osa Garcia et al., Gadson et al. or Shin et al.; (ii) participants were in the acute/sub-acute stage in Osa García et al. and Godecke et al., and in the chronic stage in Johnson et al. and Kristinsson et al.; (iii) the most common language tested was English but the largest study was in Korean (Shin et al.) with another substantial study in Japanese (Nakagawa et al.).

A fifth observation from Table 1.1 is that the measurement of variables varied wildly between studies. For example, the amount of education a patient received was recorded in (i) years, (ii) the highest level achieved, or (iii) year ranges. Likewise, the measurement of aphasia severity varied across studies and at the baseline timepoint.

A sixth observation from Table 1.1 is the choice of outcome measurement. The majority of studies opted for a composite measure of language ability/severity; with a few also including the component modalities contributing to this composite measure. Only two studies focused on Naming (Kristinsson et al., 2023, Jacobs et al., 2023) and only one chose a functional communication measure (El Hachioui et al., 2013b). Critically, no studies investigated a full range of language abilities (i.e. spanning different speech production and comprehension abilities, in both auditory and written modalities). All together, these multiple variations in how studies are conducted may contribute to the vast differences in the strength and significance of effects reported in Table 1.1, and ultimately the conclusions that can be drawn.

The seventh and final observation from the prior literature reported in Table 1.1 is that even when significant effects were observed, they were not explored in depth. For example, the influence of age was significant in 4 studies – and its influence was assumed to be consistent across all participants, when it may in fact (i) vary with other factors and (ii) depend on the task used to measure outcomes. Determining how effects depend on other variables and outcome measures requires a statistical comparison of effects across variables and a range of language tasks. For example, the disadvantage of poor post-stroke cognitive abilities may be reduced in patients who spent longer time in education, if their prior learning experience taught them strategies that could overcome a mild cognitive impairment. Similarly, the benefit of therapy may be greater in patients whose lesions spared brain regions necessary for relearning and recovery, or the influence of education may be more detectible in tasks that require more sophisticated vocabulary and/or grammar.

A final crucial consideration that has not yet been systematically investigated in the prior literature is the relationship between variables. For example, lesion size is related to initial aphasia severity, because a larger stroke lesion is more likely to damage language regions, resulting in poorer initial symptoms. More education is related to greater socioeconomic status, greater occupational opportunities and income – and subsequently better access to healthcare, healthier lifestyle and greater early development. The impact of this is that variance may be incorrectly attributed to a variable, if another variable with which it is closely related is not considered or controlled (Rohrer, 2018).

This need for more in-depth analysis motivated the PLORAS study (Price et al., 2010) and the experiments reported in this thesis, as described in the next two sections.

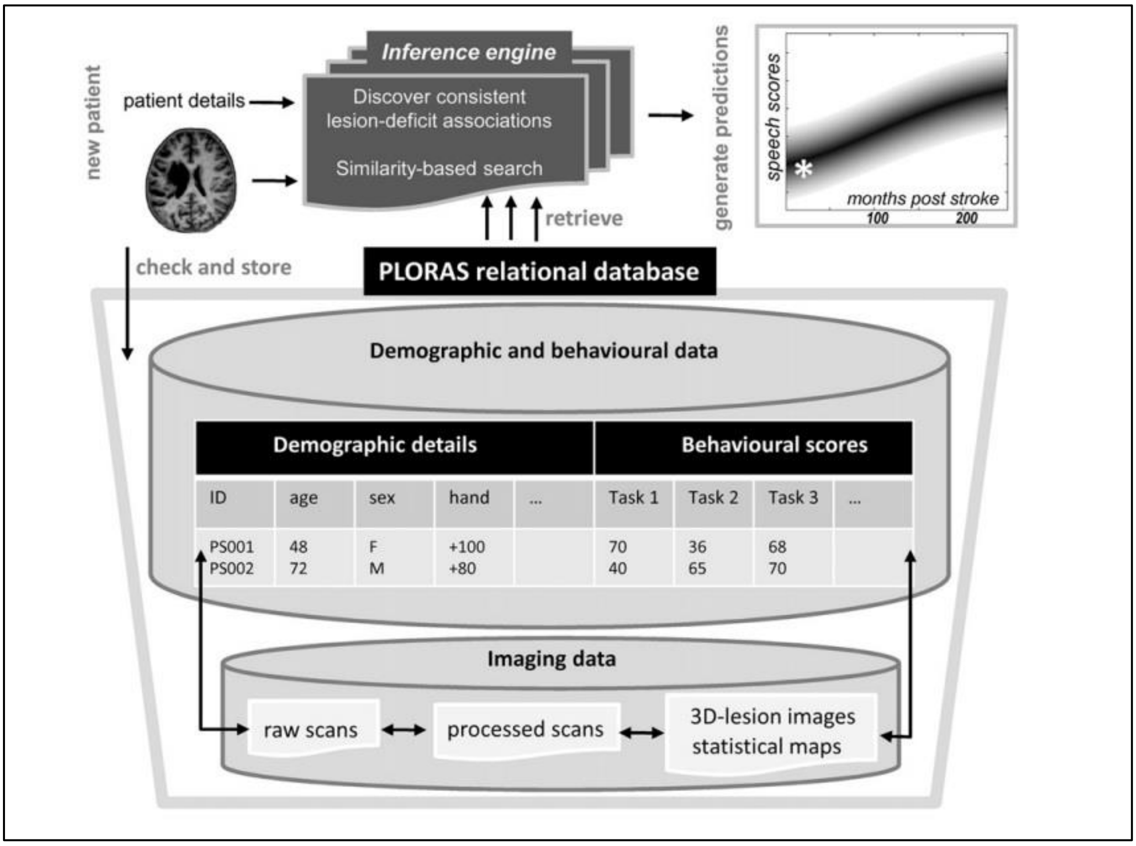
## 1.4 The goals of the PLORAS study

It is abundantly clear that the scale of the challenge of predicting recovery lies in the sheer number, and potential combinations, of predictor variables. In order to understand their independent and collective influence on language, large studies, involving hundreds of participants and complex analyses, are required to estimate the strength and direction of the relationships between each variable and the outcome measure. The use of machine learning methods is invaluable in this endeavour. For example, machine learning can be used to identify which variables contribute or don’t contribute to an optimal prediction model (known as ‘feature selection’). Regression analysis is one such machine learning tool, which is generally considered to be powerful and informative. More details about regression modelling, including its advantages and limitations, are given in Chapter 2, section 2.9.

The major challenges faced by large-scale approaches include the need for: (1) a large sample of participants, (2) data for a wide range of variables, and (3) the manpower, expertise and funding to collect, manage, and analyse such a dataset. Some data are simple to obtain, for example from participant reports, or from clinical records (e.g. sex, age, stroke type). Others are more challenging, because they require (i) specialised technology and expertise (neuroimaging data), (ii) extensive formal assessments of language and cognitive abilities, (iii) detailed records of speech and language therapy and language practice activities, and (iv) participant consent, ability, cooperation and willingness to give detailed personal information, undergo assessments and answer sensitive medical questions – while recovering from the life-changing impact of a stroke.

At UCL, the Predicting Language Outcome and Recovery After Stroke (PLORAS) study is attempting to overcome the above challenges. Recruitment commenced in 2003, and is currently funded until 2027. The overarching aim of PLORAS is to create a clinical tool for patients and clinicians that will provide confident, accurate, and individualised predictions about the most likely course of language recovery; based on the recovery profiles of previously tested patients who have ‘matching’ stroke lesions, initial symptoms, demographic and therapy data. The PLORAS approach addresses inter-patient variability in stroke and non-stroke factors by collecting, storing and analysing data from a very large number of participants, with the intention of investigating the contribution of these different factors to language outcomes, thereby explaining as much of this variance as possible. Figure 1.3 illustrates the PLORAS vision for generating personalised predictions.

**Figure 1.3** An illustration of the PLORAS approach.



**Legend:** Neuroimaging, demographic and behavioural data are entered into a database. Participants with similar characteristics (e.g. lesions in similar brain regions) are grouped, and their language outcomes are plotted. As participants are included at any time post-stroke, this generates a "probabilistic recovery curve" (top right) which can be used to predict outcomes for future patients with similar characteristics – along with a confidence in that prediction that is based on the probability of recovery in other similar patients. Reproduced with permission from Seghier et al. (2016); licensed under CC BY-NC-ND 4.0.

## 1.5 Overview of thesis experiments

Using data collected for the PLORAS study (by myself and other members of the research team; see Chapter 2 for more details), I have conducted four experiments which were designed to address the overarching aim of this thesis - which is to understand the collective and unique contributions of a set of specific variables to aphasia outcomes and recovery. Each experiment constituted a retrospective analysis. Experiment 1 considers the effect of multiple variables, and investigates which have the greatest influence on long-term overall language ability in stroke survivors. The results of this experiment guided the three subsequent experiments, which each conducted further analyses into the influence of a single variable on a range of specific language outcomes and recovery, whilst controlling for variance from a range of lesion and non-lesion factors that have not all been accounted for in previous studies. The variables selected for further analyses were: age at stroke (Experiment 2), pre-stroke education (Experiment 3), and early clinical therapy (Experiment 4).

# Chapter 2: General Methods

## 2.1 Summary

In this chapter, I describe the general materials and methods used to collect and analyse the data reported in Experiments 1-4. Each experiment's specific methods and analyses will then be detailed in the relevant experimental chapter.

I will first describe the inclusion and exclusion criteria for participants in the PLORAS research study, in which Experiments 1-4 are embedded. I will then describe the data collection methods for four types of data: (1) objective language test scores, (2) participant-reported language and therapy measures, (3) participant-reported demographics, and (4) neuroimaging data. Finally, I will describe the main statistical analyses used in each Experiment, including their advantages and limitations.

## 2.2 Ethical approval

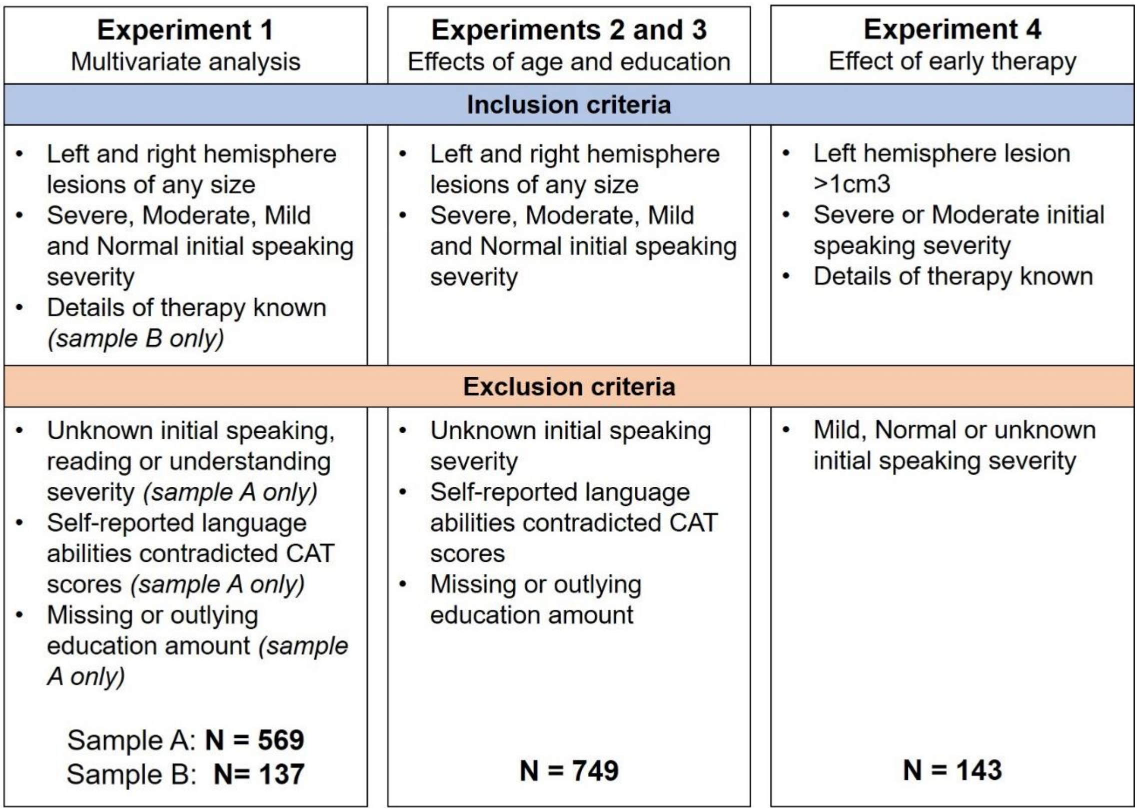
Each Experiment reported in this thesis used data obtained for the Predicting Language Outcome and Recovery After Stroke (PLORAS) study (Price et al., 2010). The PLORAS study was approved by the London Queen Square Research Ethics Committee (study reference 13/LO/1515), and all participants gave written informed consent (or assent, via a consultee) to participate.

## 2.3 Participants

Participants were stroke survivors selected from the PLORAS database (Seghier et al., 2016). Each participant travelled to the Wellcome Centre for Human Neuroimaging in the UCL Queen Square Institute of Neurology (London) for: (1) an objective language assessment, (2) a participant-reported questionnaire recording language abilities and therapy details, (3) a demographic questionnaire, and (4) a research MRI brain scan. They were recruited from a variety of locations around the UK via the National Institute for Health Research (NIHR) Clinical Research Network, stroke clubs, GP surgeries, and word-of-mouth recommendations.

Inclusion criteria for the PLORAS database include: (1) pre-stroke fluency in English prior to their stroke (although non-native English speakers are accepted), and (2) clinical diagnosis of a stroke in either hemisphere, regardless of symptoms or lesion size. Participants who had a stroke in the brainstem were excluded. Exclusion criteria for recruitment into the study include: (1) diagnosis of another neurological or psychiatric disorder which may affect speech and language abilities. Approximately 1500 participants were available for selection, from which separate samples were identified for each experiment reported in this thesis. A commonality between all four experiments was that they included only native English speakers. Additional inclusion and exclusion criteria, and the sample size for each experiment, are detailed in Figure 2.1, and in each experimental chapter.

**Figure 2.1** The inclusion and exclusion criteria for the four experiments reported in this thesis.



| **Experiment 1**<br>Multivariate analysis | **Experiments 2 and 3**<br>Effects of age and education | **Experiment 4**<br>Effect of early therapy |
|---|---|---|
| **Inclusion criteria** | | |
| • Left and right hemisphere lesions of any size<br>• Severe, Moderate, Mild and Normal initial speaking severity<br>• Details of therapy known *(sample B only)* | • Left and right hemisphere lesions of any size<br>• Severe, Moderate, Mild and Normal initial speaking severity | • Left hemisphere lesion >1cm3<br>• Severe or Moderate initial speaking severity<br>• Details of therapy known |
| **Exclusion criteria** | | |
| • Unknown initial speaking, reading or understanding severity *(sample A only)*<br>• Self-reported language abilities contradicted CAT scores *(sample A only)*<br>• Missing or outlying education amount *(sample A only)* | • Unknown initial speaking severity<br>• Self-reported language abilities contradicted CAT scores<br>• Missing or outlying education amount | • Mild, Normal or unknown initial speaking severity |
| Sample A: **N = 569**<br>Sample B: **N= 137** | **N = 749** | **N = 143** |

**Legend:** Additional criteria differentiated Samples A and B in Experiment 1. These criteria are described in Chapter 3, section 3.3.1.

The participants included in the experiments reported for this thesis were all tested between 2003 and March 2020 (the start of the Covid-19 lockdown in the UK). Data collected from participants recruited after March 2020 are not included in this thesis because (i) they have clinical brain imaging (CT or MRI) rather than homogenised high-resolution neuroimaging for research purposes, and (ii) their language assessments are based only on self-reported questionnaires rather than objective language assessments.

### 2.3.1 Author's contribution to the PLORAS database

As a research assistant in the PLORAS research study since 2015, I have primarily been responsible for collecting language, demographic and lesion data. Specifically, for almost 400 participants, I have conducted and scored over 600 objective and participant-reported language assessments, using my expertise as a registered speech and language therapist to understand each participant's language profile, and differentiate language impairments from motor speech impairments or other (non-language) difficulties. I have also facilitated the collection and processing of MRI scan data.

Paramount to all of these data collection activities is the provision of support to the participants throughout their entire research experience, facilitating their communication, advocating for any additional needs they have as a result of their stroke, and ensuring they have a positive experience of contributing to research.

## 2.4 Objective language data: Comprehensive Aphasia Test

All participants were assessed with an objective language and cognitive assessment - the Comprehensive Aphasia Test, or CAT (Swinburn et al., 2004) - by myself or another trained speech and language therapist within the PLORAS team.

The CAT is a fully standardised test battery, which consists of 27 different tasks that assess both language and cognitive abilities. The scope of the CAT is clearly defined: it assesses impairments; not functional communication abilities. Prior to starting the CAT, hearing and vision abilities are informally checked, to ensure these are corrected as best as possible, and any difficulties with the assessment stimuli are not the consequence of hearing or visual impairments.

### 2.4.1 Task details

The outcome measures for the experiments reported in this thesis included the scores from 18/27 CAT tasks that tested auditory and written comprehension, auditory repetition, reading, naming and spoken picture description. Each of these 18 tasks are described below along with the semantic picture matching task, scores from which were added as a covariate in some of the analyses.

#### Spoken comprehension tasks

**Comprehension of spoken words:** Participants hear a word, and must select the correct matching picture (e.g. *mouse*) from four options, in the presence of a phonological, semantic, and unrelated distractor (albeit the unrelated distractor is semantically related to the phonological distractor) (e.g. *house*, *rabbit*, *church*). There are 15 trials within this task.

**Comprehension of spoken sentences:** Participants hear a sentence, and must select the correct matching picture from four options (e.g. *the woman is sitting*). The sentences gradually increase in length and syntactic complexity, and include a range of sentence types designed to detect difficulties commonly associated with aphasia (e.g. passive structures, relative clauses). There are 16 trials in this task.

**Comprehension of spoken paragraphs:** Participants listen to two short stories, before then answering four questions about each story by pointing to a Yes or No response. The questions are paired; i.e. two questions essentially ask the same thing, but phrased in a slightly different way. For example, (a) *were they on time?* And (b) *were they early?*

### Written comprehension tasks

**Comprehension of written words:** Participants read a word, and must select the correct matching picture (e.g. *rocket*) from four options, again in the presence of a phonological, semantic, and unrelated distractor (e.g. *pocket*, *aeroplane*, *sleeve*). As with spoken word comprehension, the unrelated distractor is semantically related to the phonological distractor. There are 15 trials within this task.

**Comprehension of written sentences:** Participants read a sentence, and must select the correct matching picture from four options (e.g. *the man is sitting*). The exact sentences are different to those in the comprehension of spoken sentences task, to avoid any priming, although some lexical items are repeated in both tasks. As with comprehension of spoken sentences, the written sentences gradually increase in length and syntactic complexity, and there are 16 trials.

### Repetition tasks

**Word repetition:** participants hear a target word, which they must repeat (e.g. *table*). The task contains 16 trials, with differences in imageability, frequency and syllable length.

**Complex word repetition:** participants hear a morphologically complex word (i.e. containing prefixes and suffixes, e.g. *unthinkable*), which they must repeat. There are 3 trials in this task.

**Nonword repetition:** participants hear a nonword (e.g. *gart*), which they must repeat. There are 5 trials in this task, which increase in length and phonological complexity.

**Digit span:** participants hear a string of digits, and must repeat them in the same order. The digit strings increase in length from 2 to 7. If a participant fails on one trial, they may hear a second (different) string of the same length. If they repeat this second string correctly they progress to the next level; if they repeat it incorrectly, the task is stopped.

**Sentence repetition:** participants hear a series of sentences, increasing in length from 3 content words to 6 content words. As with digit span, there are two trials possible at each sentence length. If they repeat the first trial correctly, they progress to the next level. If they repeat the first trial incorrectly, they hear the second trial at the same level. If this second trial repeated correctly, they progress to the next level. If they fail on the second trial, the task is stopped.

**Reading aloud tasks**

**Reading words:** participants are visually presented with real words and must read them aloud (e.g. *chair*). There are 24 trials in this task, which vary in frequency, imageability and syllable length.

**Reading complex words:** participants are visually presented with morphologically complex words (e.g. *informative*), and must read them aloud. There are 3 trials in this task.

**Reading function words:** participants are visually presented with function words (e.g. *but*), and must read them aloud. There are 3 trials in this task.

**Reading nonwords:** participants are visually presented with novel nonwords (e.g. *fask*) and must read them aloud. There are 5 trials in this task.

**Naming tasks**

**Naming objects:** participants are presented with a simple line drawing (e.g. *car*), and must name the object. There are 24 items in this task, which differ in frequency, imageability and syllable length.

**Naming actions:** participants are presented with a line drawing of a person carrying out an action, and are asked to name what the person is doing (e.g. *eating*). There are 5 trials in this task.

**Word fluency:** this standard verbal fluency task comprises two parts. Part one involves semantic fluency, in which participants must name as many animals as possible within one minute. Part two involves letter fluency, in which participants must name as many words beginning with the letter ‘s’ as possible within one minute (excluding proper nouns and names). One point is given for each correct new word produced.

**Spoken picture description:**

Participants are visually presented with a line drawing of a complex scene, and must describe the events taking place in the scene within one minute. This task is highly valuable for assessing connected speech, and is considered the most ecologically valid of all the CAT tasks (i.e. most similar to real-word communication), and therefore crucial for understanding how performance on the task might relate to language in the real world.

**Semantic memory**

One cognitive task, ‘semantic memory’, was selected from the CAT, to be used as a covariate to control for object recognition and semantic impairments which might explain poor performance on language tasks that use picture or conceptual stimuli (i.e. any task involving pictures of words of familiar concepts but not non-words). This task measures the ability to perceive pictures and identify the correct semantic link between a target at the centre of the page (e.g. *monkey*) and four surrounding pictures which include the correct semantic link (e.g. *banana*) in the presence of three distractors (e.g. *pear*, *chocolate*, *envelope*). It does not require a verbal response, and so represents the closest measurement of pure semantic ability as was possible to obtain from the CAT. There are 10 items in this task.

### 2.4.2 Scoring CAT data

For all tasks except spoken picture description and spoken paragraph comprehension, each trial is assigned a score of 2, 1, or 0. A score of 2 is awarded for a correct response within 5 seconds. A score of 1 is given if a correct response is either (i) delayed, (ii) follows an initially incorrect response, or (iii) requires repetition of the auditory stimulus. A score of 0 is given for an incorrect response or no response. For all speech production tasks, phonemes that have been correctly selected but distorted due to dysarthria are accepted as correct. In contrast, verbal, phonemic, neologistic and dyspraxic errors are scored as incorrect for all speech production tasks, except Digit span and Sentence repetition - which assess verbal memory span, not phonological accuracy.

For Spoken picture description, speech output is recorded and subsequently analysed to determine: (i) the number of appropriate information carrying words produced, (ii) the number of inappropriate information carrying words produced, (iii) the total number of words produced (appropriate plus inappropriate), (iv) grammatical well-formedness on a scale of 0 to 6; (v) syntactic variety of the sentences on a scale of 0 to 6, and (iv) the speed of production, assessed on a scale of 0 to 3, according to whether there were any lengthy delays or hesitations. The scores are then summed to generate a ‘total score’ for Spoken picture description.

For spoken paragraph comprehension, the response to both paired questions must be correct in order to receive one point. The maximum score is therefore 2 points for each of the two stories.

### 2.4.3 Diagnosing Aphasia

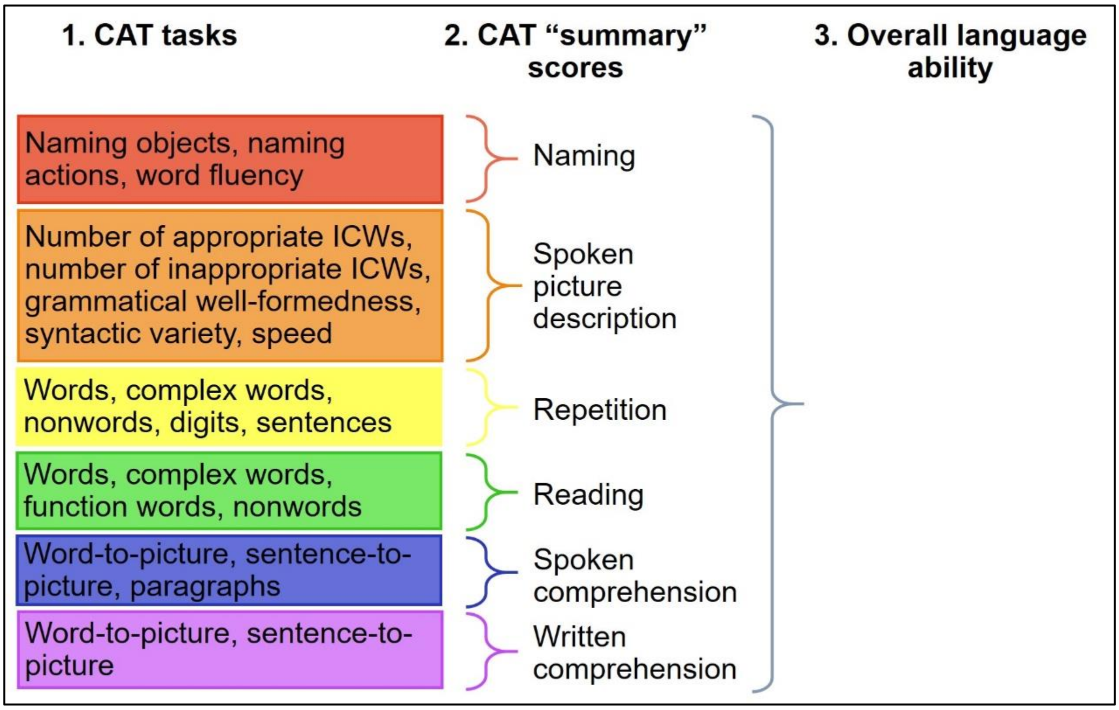
An aphasia/normal threshold for each language task, and summary measure, is built into the CAT. Raw scores for each CAT task are converted through a non-linear transformation into T-scores, which represent how well the participant performed relative to an independent sample of 113 patients with aphasia and 27 non-aphasic controls. The lowest 5% of scores were considered to be ‘below normal’, and therefore designated as ‘aphasic’ for stroke patients. For example, for Naming, T scores of 62 and lower indicate aphasia, and 63 and above indicate language within normal limits.

Summary T scores were calculated for (1) Naming, (2) Repetition, (3) Reading, (4) Comprehension of Spoken material, (5) Comprehension of Written material, and (6) Spoken picture description, by combining scores from similar tasks (see Figure 2.2). For example, five tasks involving repetition of an auditory stimulus are combined to give a summary score for ‘Repetition’. Finally, an ‘overall language ability’ score was calculated by averaging across the six summary T scores, as in Winans-Mitrik Ronda et al. (2014), see Figure 2.2. This is similar to the Aphasia severity Quotient in the Western Aphasia Battery (WAB-AQ). As the calculation of the overall language ability score is not defined in the CAT, an aphasia/normal threshold for this was not available.

The key advantages of using summary scores are that they reduce the number of statistical comparisons that needed to be considered, and avoid any a priori assumptions about which tasks would and would not be influenced by each variable.

Figure 2.2 shows the tasks that contribute to each summary score, and subsequently to the overall language ability score.

**Figure 2.2** The CAT tasks that contribute to each summary score, and to the constructed measure of overall language ability.



**Legend:** ICW = Information Carrying Word.

### 2.4.4 Advantages of the CAT

There are a number of advantages of the CAT which make it an appropriate assessment for the PLORAS study, and the experiments reported in this thesis. First, the CAT is well-constructed, with robust psychometric properties. Specifically, the authors selected stimuli within each subtest that would allow the effects of imageability, frequency and word length to be investigated. Although these variables were not of interest in Experiments 1-4, there is rich potential for future investigations that look for patterns in an individual's specific linguistic deficits, and how these vary between individuals. For example, an interaction between item frequency and education may indicate whether the advantage of more education is more apparent on lower- than higher- frequency items.

The CAT is also reasonably quick to administer (typically between 1 and 2 hours, depending on language ability), despite testing a wide range of language functions and including a cognitive screen that can help to determine whether poor language performance is a consequence of more general cognitive functions. The choice of tasks in the CAT is based on neuropsychological models of language processing. In contrast, the Western Aphasia Battery, or WAB (Kertesz, 2007) and Boston Diagnostic Aphasia Examination, or BDAE (Goodglass et al., 2001), are designed to test for classical, but arguably less clinically useful, aphasia syndromes. That said, the CAT includes many tasks (e.g. naming and repetition) that are largely shared with other popular aphasia assessments (Castro et al., 2023).

The popularity of the CAT emphasises the advantages highlighted above. The CAT is widely used for clinical assessment by speech and language therapists throughout the UK, and around the world – indeed it is the most widely-translated aphasia assessment (REF2014, 2014). Research evidence reporting tasks from the CAT have direct clinical relevance and can be instantly understood by clinicians, aiding clinical translation of research into practice.

### 2.4.5 Limitations of the CAT

As with any assessment, there are a number of limitations to the CAT. First, the stimuli used could be considered archaic. In the twenty years since the CAT's original publication in 2004, some of the stimuli are already undeniably outdated. Consider, for example, the rotary telephone (in Object Naming) and the typewriter (in Action Naming). Errors or delays in naming these items may simply arise from unfamiliarity with the image, not from any aphasic difficulty. Second, the CAT may be sensitive to individual differences in cultural background, native language, age, life experience or educational background. For example, a participant with musical experience is more likely to recognise and correctly name a '*saxophone'* (in Object Naming) compared to a participant with no experience of musical instruments. A '*leek'* (Comprehension of Spoken Words) may be more common in some world cuisines than others. While these might seem trivial issues, they may result in delays or errors that are unrelated to aphasia, and therefore an incorrectly penalised score.

Third, some CAT tasks have very few trials (e.g. 5 in Action naming) that will inevitably limit sensitivity for detecting mild impairments. To get around this problem, the CAT authors suggest that if difficulties on those tasks are suspected, they should be probed further with a more specific assessment (e.g. for verbs, the Northwestern Assessment of Verbs and Sentences (Thompson, 2012)). It should be noted, however, that is it because of the short tests with few items that the length of the assessments is kept to a minimum. Moreover, one could argue that the focus of the CAT was appropriately given to the more important tasks, such as Object naming.

### 2.4.6 Using the CAT to measure recovery

As the CAT was typically only conducted once for each participant, it was not possible to measure recovery as a change in CAT scores between two assessment timepoints. Instead, recovery was measured as change between initial severity (i.e. participant-reported ability at one week post-stroke, see section 2.5) and long-term CAT score, with the time between stroke and CAT assessment serving as an indicator of recovery time. A limitation of this method is that early recovery (within the first year) could not be assessed, because for the majority of participants in the PLORAS database, the CAT was conducted over a year post-stroke. A second limitation of this recovery measure is that the participant-reported measurement of initial severity does not correspond with the CAT measurement. Specifically, initial severity is measured in terms of whether the participant was able to produce, comprehend, read or write words and sentences (see section 2.5 below) – but there are no CAT tasks that measure this directly. Given these limitations, a participant-reported measure of recovery was also used to supplement this objective measure (see section 2.5).

## 2.5 Participant-reported language and therapy data: the Aphasia Recovery and Therapy Questionnaire (ARTQ)

All participants completed the PLORAS in-house ‘Aphasia Recovery and Therapy Questionnaire’ (ARTQ) on entry into the study. This is a retrospective, participant-reported outcome measure of language ability, on a nonlinear 7-point Likert scale. For the experiments reported in this thesis, the 7-point scale was converted into a linear scale with four severity levels: (1) Severe (unable to produce/comprehend any speech, or only able to produce automatic speech), (2) Moderate (able to produce/comprehend single words and phrases), (3) Mild (able to produce/comprehend lexically meaningful short sentences), and (4) Normal. This was repeated independently for four language modalities (speaking, understanding, reading and writing). The same participant could therefore be assigned different severity ratings for different language functions (if, for example, they had speaking difficulties, but their understanding was unaffected). The ratings are repeated for three timepoints: one week, one month, and one year post-stroke. The ARTQ therefore serves two important purposes: first, the one week post-stroke rating serves as a measure of initial aphasia severity; and second, change in ability between one week and one month/one year post-stroke can be used as a measure of language recovery.

Due to the importance of initial speaking severity as a predictor, participants were excluded from my experiments if they (1) lacked this data, or (2) were medically unwell (e.g. in a coma) one week post-stroke, because their inability to produce speech was not necessarily related to the presence of aphasia. This is crucially different from participants rated as Severe, who were conscious, and physically capable of attempting to speak – but could not produce any words, due to aphasia (and dysarthria/apraxia). If, however, only initial understanding or reading severity data were missing, these participants were included. Reasons for this missing data were typically that they could not remember those initial abilities, or because they were not rated (e.g. reading was not attempted, and so reading ability could not be assessed). Each Experiment reports when participants were excluded from specific analyses because of missing initial understanding or reading severity data.

Using the same questionnaire, participants also estimate (i) how many hours of speech and language therapy they received, (ii) when speech and language therapy was received (start and end points and frequency), and (iii) therapy activities (see Figure 2.3).

**Figure 2.3** Questions asked to participants about their speech and language therapy.

Have you received any speech and language therapy since your stroke? **Yes/No**

**If yes:**
When did therapy start?..............................................................
When did therapy end? ..............................................................
How often did you do therapy? (e.g. one hour every week)
..........................................................................................
What sort of activities did you do in therapy?
..........................................................................................

Carers were encouraged to provide their own report of participants' abilities and therapy, either to supplement the participant's report, or to substitute it, if the participant had memory or other cognitive impairments that prevent them from providing sufficient details. Finally, PLORAS research speech and language therapists supported participants and carers to provide more details where necessary, to ensure an accurate report of their difficulties (for example, differentiating fluent from non-fluent aphasia).

### 2.5.1 Advantages of the ARTQ

The ARTQ provides a measure of initial severity which would otherwise be unavailable for the majority of participants in the PLORAS database (because early language assessment data was not obtained for the study). Furthermore, assessing initial severity retrospectively may actually increase the number of participants with severe initial aphasia included in the study - because if their severe symptoms (or concomitant non-language impairments) precluded assessment at the time, they would not be included.

Another major advantage of the ARTQ is that it provides insight into participants' perspectives of their language abilities and impairments, and recovery. Participants may report difficulties (or abilities) that would not be detected on an objective assessment, as well as whether their language abilities fluctuate according to external factors (e.g. fatigue, time of day). Patient-reported outcome measures (PROMs) are widely used across a number of other health conditions (including musculoskeletal conditions, cancer and mental health, (Churruca et al., 2021), and have been shown to improve diagnosis and communication between patients and clinicians (Valderas et al., 2008).

Finally, participant reports of therapy received may offer valuable insight into their level of engagement and motivation with therapy.

### 2.5.2 Limitations of the ARTQ

The ARTQ score categories are relatively broad, and capture no detail about the specific nature of language impairments. For example, participants can indicate if they were ‘able to speak in short sentences’ (placing them in the ‘Mild’ category), but any word-finding difficulties, articulatory impairments or difficulties with grammar or syntax, could go unreported. This extra detail can of course be gained by the researcher probing the participant for more information, however this results in interrater differences in questionnaire administration. The ARTQ is therefore less sensitive than the CAT at capturing specific details of language impairments.

Finally, as a retrospective report, the questionnaire depends on participant (or carer) memories, which may be inaccurate or susceptible to recall bias. For example, a good recovery may motivate a more positive retrospective rating of one’s initial symptoms, whereas a poor recovery may motivate a more severe retrospective rating (Kress and Aue, 2017). Furthermore, the reliability of the self-report may itself differ across the language modalities. Self-reported understanding abilities are expected to be more difficult to rate than speaking abilities for both participants and carers, and thus may be less reliable.

### 2.5.3 Using the ARTQ to measure recovery

As the ARTQ is repeated at three timepoints (one week, one month, and one year post-stroke), recovery can be measured as the change in participant-reported score between two time points. The most intuitive way to report recovery is in terms of a change in raw scores between one time point and another (i.e. actual recovery). However, when participants have different baseline scores, recovery will be greatest for those with severe baseline scores, because these participants can potentially increase 3 levels (Severe to Moderate, Moderate to Mild and Mild to Normal); whereas participants with mild baseline scores are only able to improve 1 level (Mild to Normal). An alternative approach, which attempts to normalise baseline scores across individuals, is to report ‘proportional recovery’ as the change in raw scores (actual recovery) divided by how much recovery would be possible. For example, a participant who improved by 2 out of 3 possible performance increases would receive a recovery score of 0.66. In contrast, a participant who moved up 2 out of 2 possible performance increases (i.e. moderate to mild and mild to normal) would receive a recovery score of 1.00. In summary, proportional recovery will be greatest for those with milder baseline scores whereas actual recovery would be greatest for those with severe baseline scores. The choice between the two therefore depends on whether one wants to highlight the extent of improvement in absolute terms or in terms of the individuals' capacity to recover.

The experiments reported in this thesis analysed both actual and proportion recovery scores – without notable differences in the results. Experiments 2 and 3 report proportional recovery scores because these are more appropriate when comparing recovery outcomes across participants with different initial severity levels (Krakauer and Marshall, 2015). Experiment 4 reports actual recovery scores because there was more homogeneity in initial severity in the sample (all participants had either Severe or Moderate initial speaking symptoms) and the focus of the experiment was on the amount of actual improvement.

### 2.6 Demographic data

Questionnaires administered by PLORAS research staff capture demographic data for each participant on entry into the study.

**Age at stroke** is calculated from date of birth and date of stroke, and entered into analyses as a linear variable.

**Education** is reported as the number of years of formal education completed after age 14 (the least amount of education received in the sample). This includes formal qualifications (e.g. AS and A Levels, undergraduate and postgraduate degrees and diplomas) but not work-related qualifications, continuing professional development courses, or apprenticeships. 'Years of education' is entered into analyses as a linear variable.

**Sex** is entered into analyses as a binary categorical variable: 'male' or 'female'.

**Pre-stroke handedness** is reported as Right, Left or Ambidextrous. The number and proportion of participants in each handedness category is reported for each experiment in Tables detailing participant characteristics, but handedness was entered into analyses as a binary categorical variable: ‘right-handed’ or ‘not right-handed’. This was because there were insufficient participants (n=24; 3% of the sample) who were ambidextrous for robust statistical analysis.

**Vision status** is participant-reported at the point of CAT assessment, and entered into analyses as a binary categorical variable: ‘vision OK’ (which includes normal and corrected-to-normal with e.g. glasses) or ‘vision impaired’.

**Hearing status** is participant-reported at the point of CAT assessment, and entered into analyses as a binary categorical variable: ‘hearing OK’ (which includes normal and corrected-to-normal with, e.g. hearing aids) or ‘hearing impaired’.

**Developmental dyslexia status** is participant reported, and entered into analyses as a binary categorical variable: ‘formal diagnosis’ or ‘none’. Participants with suspected dyslexia without a formal diagnosis were treated as having missing data, because their dyslexia status could not be truly known.

Other demographic data collected for the PLORAS study include: occupation, post-stroke handedness and footedness, birth country, additional languages spoken, and parental languages spoken. These data were not analysed in Experiments 1-4, but are available for future analyses.

**Time post-stroke of CAT** was obtained by calculating the difference between the date of stroke, and the date CAT assessment was conducted.

## 2.7 Neuroimaging data

High-resolution (1mm x 1mm x 1mm), whole brain T1-weighted structural MRI brain images were acquired for all participants on entry into the study using research-dedicated scanners at the Wellcome Centre for Human Neuroimaging and the Birkbeck-UCL Centre for Neuroimaging. The MRI scanners used were all from Siemens Healthcare (Erlangen, Germany): 3T Trio (including Prisma upgrade), 1.5T Sonata, 1.5T Avanto, and 3T Allegra.

### 2.7.1 How structural MRI works

This section describes the basic principles of Magnetic Resonance Imaging (MRI), largely based on the explanatory article by Currie et al. (2013).

MRI is a non-invasive technique used to produce detailed 3D images of anatomical structures. It is completely safe, provided strict safety criteria are followed, and can be repeated multiple times without harm. It is an integral part of the stroke diagnosis pathway (NICE, 2022) due to the amount of detail it can differentiate in the case of small or unusually-located lesions.

In essence, MRI works by applying a very powerful magnet to harness the magnetic properties of water molecules present throughout the human body. More specifically, the hydrogen atoms in water molecules ($H_2O$) have a positive charge because they each comprise a positively charged proton and a neutron (that has no charge). Protons possess a natural ‘spin’ property, meaning they are constantly spinning on their own axis. The electrical charge that arises from this ‘spin’ property generates a small magnetic field. In normal conditions, these protons and their magnetic fields are randomly orientated. However, when an external magnetic field is applied, they align with that external field. The result is a group of aligned, spinning protons; a phenomenon known as ‘precession’. This is what occurs when a participant enters the magnetic field of an MRI scanner (i.e. by being positioned inside it): the randomly-orientated protons in their body are influenced by the strong magnetic field of the scanner. The protons essentially de-randomise; becoming aligned with the direction of the magnetic field. This occurs prior to any actual scanning beginning.

When the MRI scan begins, radiofrequency (RF) pulses are introduced into the magnetic field, from a different direction to the magnetic field. The RF pulse disrupts the alignment of the magnetic field, causing the protons in the participant to switch direction briefly. Once the RF pulses are stopped, the protons return to their original position in the MRI environment (i.e. they are re-aligned with the scanner’s magnetic field). As the protons return to this position, they release RF energy, which is measured by a device that picks up the energy signal from the body part being imaged. In the case of brain imaging, the device is called a ‘head coil’.

Protons in different tissues and structures emit different amounts of energy, depending on their fat and water content. This means that different anatomical structures can be imaged and differentiated. In the case of neuroimaging, differentiation is particularly clear between cortical grey matter, subcortical white matter structures, cerebral spinal fluid (CSF) and skeletal structures.

The strength of the magnetic field is measured in Tesla (T), with hospital scanners typically between 1.5-3T. Stronger magnetic fields provide more detailed images.

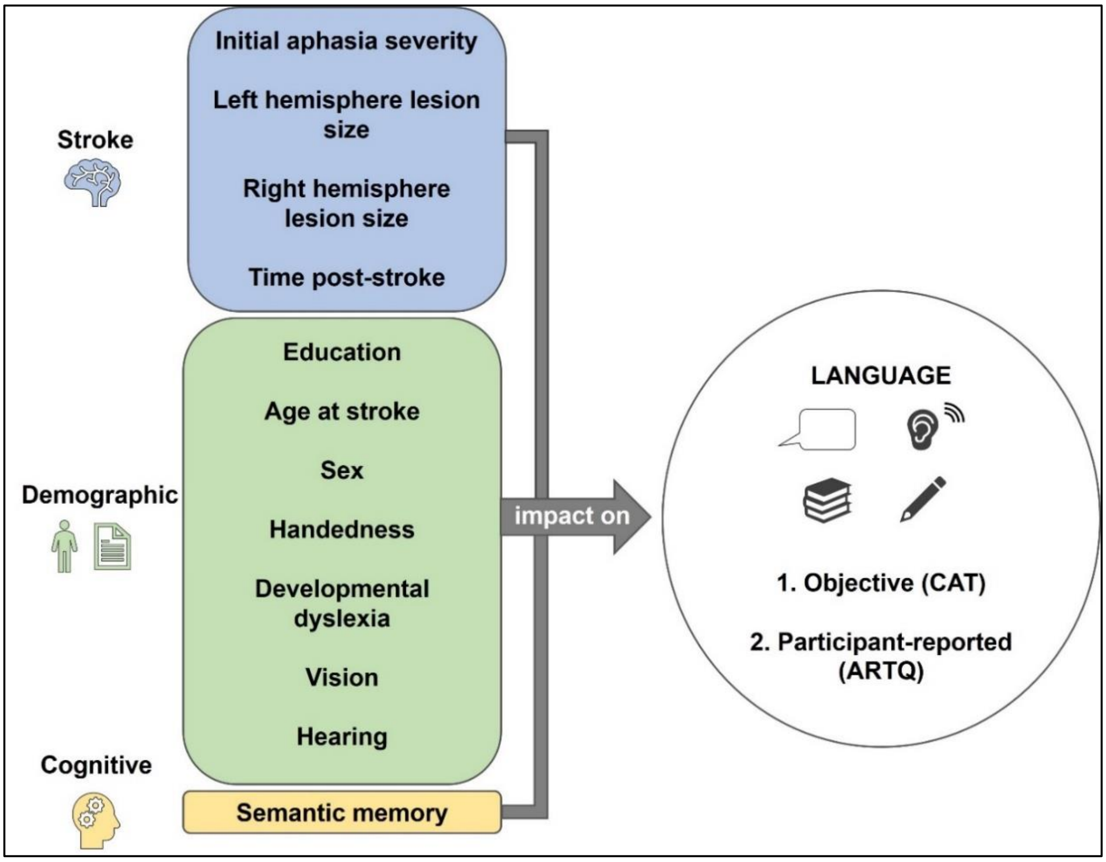
### 2.7.2 How lesion size is estimated from MRI scans

Each T1-weighted MRI image was analysed using standard procedures within SPM (Statistical Parametric Mapping) software (Wellcome Centre for Human Neuroimaging, London, UK; https://www.fil.ion.ucl.ac.uk/spm/), running in MATLAB environment (2018a Mathworks, Sherbon, MA, USA). The first step is to spatially normalise the image into a standard template that is known as the MNI template. The image is then converted into a quantitative assessment of structural abnormality that is independent of the scanner used (Seghier et al., 2008). Abnormality is measured on a continuous scale from 0 (completely normal) to 1 (completely abnormal) relative to normative data from a sample of 64 neurologically-intact controls, collected on the same scanners used for the stroke participants. The 64 controls were aged from 21-75 years, thus ensuring that any normal age-related structural changes were represented and identified as ‘normal’. This image of a continuous measure results in what is referred to as a ‘fuzzy’ lesion image. To delineate the lesions precisely, estimate lesion volume, and generate lesion overlap maps, each fuzzy lesion image was thresholded into a ‘binary’ lesion image (i.e. presence of absence of a lesion, at each voxel). The abnormality threshold used is 0.3 (U value, on the 0 to 1 scale described above), as recommended in (Seghier et al., 2008). Separate lesion size estimates were made for left and right hemisphere lesion size, first in voxels, which was then converted into $cm^3$, to ease interpretability. Left and right hemisphere lesion size were entered separately as linear variables into statistical analyses.

## 2.8 Summary of experimental design

The stroke, demographic and assessment variables described in this chapter were entered into a regression model (more details are given in the next section) as independent variables, to investigate their effect on one of the measures of language ability or recovery, also described in this chapter. This broad model is illustrated in Figure 2.4, and more details of each Experiment methods and design are given in the relevant Chapter.

**Figure 2.4** An illustration of the variables selected to investigate their impact on (1) objective (CAT) and (2) participant-reported (ARTQ) language outcomes.



## 2.9 Statistical analyses

Multiple linear regression was chosen as the primary method of statistical analysis for all four experiments. Regression is a sophisticated method for assessing the relationship between an outcome variable (often called the dependent variable) and a predictor variable (often called the independent variable). It can be seen as a more powerful and informative extension of correlation analysis, as it also enables control for the influence of many other variables.

There were three results of interest from each regression analysis' output, for each experiment. First is the correlation coefficient, also called the standardised beta (β). This provides an estimate of the size of the effect on outcome variables, when other variables are controlled for. For example, a model including both lesion size and age would indicate the relative size of both variables' effects on language outcomes.

The second result of interest is the significance of each variable's effect, expressed as a p value. In all experiments, the threshold for significance is set as p=<0.05 uncorrected, except in some analyses where the threshold for significance is altered to correct for multiple comparisons. More specifically, a Bonferroni correction was applied, by dividing the standard threshold (0.05) by the number of analyses conducted (Curtin and Schulz, 1998). For example, if 20 analyses were run, the corrected threshold would be 0.003 (0.05/20); any p value below this would thus be considered significant. This correction avoids exaggerating the number of significant effects that occur due to chance when many analyses are conducted.

The third result of interest is the '$R^2$' value which describes the overall 'fit' of a model. It can be interpreted as an estimate of the amount of variance in the outcome variable that is explained by the set of predictors included in the model. For example, an $R^2$ of 0.35 indicates that 35% of variance in the outcome variable was explained by that set of predictors. A higher $R^2$ suggests more variance has been explained, and thus a better fitting model – however, care must be taken to ensure $R^2$ is not falsely inflated (more detail is given below).

Some of the advantages of regression are described below:

(1) Regression can determine the relative influence of multiple variables on an outcome variable. A caveat to this is that it requires a sufficiently large sample, that needs to increase in proportion to the number of covariates.
(2) Interactions between variables can be investigated using 'moderator analyses'. These enable investigation of whether the effect of one variable depends on that of another variable. This was crucial for Experiments 2 and 3, which specifically questioned whether the effects of age at stroke and education depended on other variables, such as lesion size and initial severity.
(3) Different models can be compared easily to determine the best combination of predictor variables for any given outcome variable.
(4) Alternative types of regression can be selected according to the type of outcome measure. For example, multiple linear regression was appropriate when the outcome variable comprised linear CAT scores. But binary logistic regression was more appropriate when the outcome variable was a binary variable (e.g. whether participants improved or not

after therapy, in Experiment 4). The output provided in a binary logistic regression reflects an ‘odds ratio’ instead of an effect size, but the broad interpretation of the analysis remains the same.

There are also several limitations to consider when using regression:

(1) Overfitting can occur if too many covariates are included, and/or the sample size is insufficiently small. Put simply, overfitting occurs when a model is too complex, and describes not only the data itself, but also random noise (i.e. variance that can’t be explained by the predictor variables). Overfitting may result in inflated effect sizes and $R^2$. Consequently the model would not translate well to a different set of data (Harrell, 2015).

(2) Many assumptions of regression analysis must be met in order for a model to be considered reliable. These include: linearity, normality, homoscedasticity, independence, and multicollinearity. Deviations from these assumptions can occur if, for example, there are outliers in the data, or a strong relationship between two covariates. The assumptions of regression should be checked via further statistical tests, visual inspections of data, and plots.

(3) With a sufficiently large sample, the chance of detecting significant but very small effects increases (Sullivan and Feinn, 2012). For this reason, the experiments in this thesis (i) report the size of each significant effect and (ii) use Bayesian statistics to determine the strength of evidence for or against any reported effects (Hackenberger, 2019).

In summary, multiple linear regression was considered the most suitable method of analysis for all experiments reported in this thesis, which each investigated the effect of a single variable on an outcome whilst controlling for the effect of a number of other variables. The multi-factorial models were possible because of the very large sample size available. All results of interest were explored further using post hoc tests, subgroup analyses and illustrations, all of which are described in more detail in the relevant Experiment’s methods. In addition, to supplement the regression results, Bayesian statistics were also used in Experiments 2 and 3, to estimate the strength of the evidence for any results reported.

All statistical analyses of language outcomes were conducted in IBM SPSS Statistics (version 25.0).

# Chapter 3: Experiment 1. What are the main variables influencing long-term language outcomes in the PLORAS dataset?

## 3.1. Summary of experiment

**Introduction:** Identifying predictors of language outcomes and recovery after stroke is crucial in order to fully understand individual recovery trajectories. The primary goal of this study was to identify ‘predictor’ variables (which are known and measured early post-stroke) that can be used prospectively to predict overall language ability. The results from this study were expected to guide further, more detailed analyses of those predictor variables, which would investigate how the effect of each one depended on other variables.

**Methods:** The effect of stroke-related predictor variables (left and right hemisphere lesion size, initial severity) and demographic predictor variables (age at stroke, amount of education, biological sex, handedness, dyslexia status) was investigated on overall language ability measured months to years post-stroke. The analysis used multiple linear regression with data from 569 stroke survivors, whilst factoring out variance from assessment variables (semantic memory, vision and hearing status, and time post-stroke). The effect of early speech and language therapy on overall language ability was also investigated in a smaller sample of 137 participants.

**Results:** The predictor variables with the largest influence on long-term overall language ability were: left hemisphere lesion size (β=-0.465), initial speaking severity (β=0.201), initial reading severity (β=0.132), age at stroke (β=-0.125), and amount of pre-stroke education (β=0.075). Specifically, larger left hemisphere lesions, worse initial (speaking and reading) impairment severity, older age at stroke, and less education were significantly associated with worse language ability. The predictor variables with no significant influence on overall language ability were: initial understanding severity, sex, handedness, dyslexia status, right hemisphere lesion size and early therapy.

**Discussion:** The disadvantage of larger left hemisphere lesions and worse initial symptoms on overall language ability was consistent with prior evidence. The significant benefits of younger age at stroke and more education on language ability were more intriguing, as their effects have been debated in prior studies. Additionally, the benefit of education was not observed in the smaller sample with therapy data, suggesting that effects may be driven by participants with certain characteristics, and warranting further investigation. The absence of an effect of early therapy on overall language ability might be explained by the broad outcome measure which may not be sensitive to the therapy provided, and prompting further analysis of its effects on individual language skills. Overall, these results motivated the selection of age at stroke, amount of pre-stroke education, and early therapy for further investigation in subsequent experiments, whilst controlling for variance from other lesion, demographic and assessment variables.

## 3.2. Introduction

The aim of this experiment was to: (A) identify which variables, known at the time of stroke, are most likely to prospectively predict long-term overall language ability in 569 stroke survivors with a mixture of left, right and bilateral lesions, and (B) establish the influence of early speech and language therapy on overall language ability in a smaller sample of 137 stroke survivors who all had left hemisphere lesions and therapy data available. The results from these preliminary analyses were intended to motivate the choice of variables to investigate in more detailed follow-up experiments that each (i) investigate the influence of an individual variable on a wider range of more specific language functions, and (ii) determine how the influence of each variable depends on other contributing variables. For example, how the effect of education depends on lesion size. Below I describe the variables included in this experiment.

The variables of interest were those that could be established at the time of the stroke and may function as ‘predictor variables’ for long-term outcomes and therefore can be used for prospective predictions. The choice of potential predictor variables included in the study was determined by a combination of the data acquired for the PLORAS study and a review of the prior literature.

### 3.2.1. Left hemisphere lesion size:

Left hemisphere lesion size was expected to have the greatest influence on post-stroke language outcomes and recovery, based on a wealth of prior literature (Mazzoni et al., 1992, Goldenberg and Spatt, 1994, Plowman et al., 2012, Hope et al., 2013, Watila and Balarabe, 2015, Martins et al., 2017, Gadson et al., 2022, Harvey et al., 2022). Larger left hemisphere lesions are associated with worse language outcomes and recovery, because language processing is typically lateralised to the left hemisphere (see section below on the right hemisphere for exceptions), and larger left hemisphere lesions are therefore more likely to impact the regions involved in language processing. Only one study identified has reported no relationship between lesion size and language outcomes (Laska et al., 2001).

### 3.2.2. Initial aphasia severity:

Initial aphasia severity, like lesion size, is also well-established as one of the strongest predictors of language outcomes and recovery (Mazzoni et al., 1992, Pedersen et al., 1995, Laska et al., 2001, Pedersen et al., 2004, Lazar et al., 2008, Lazar et al., 2010, Godecke et al., 2013, Martins et al., 2017, Ramsey et al., 2017, Basilakos et al., 2019, Osa García et al., 2020). As discussed in Chapter 2, section 2.5.3, the influence of initial severity on recovery depends on how recovery is measured. If recovery was measured as the degree of change between Timepoint 1 (i.e. the initial severity) and Timepoint 2, then more recovery would be possible for patients who are more severe initially because there is more scope for recovery. On the other hand, if recovery was measured as the proportion of those who move into the normal range, then milder initial symptoms are more likely to result in a complete recovery. To accommodate this, recovery can be measured as the absolute change compared to the total possible change (Lazar et al., 2010). This has led to the “proportional recovery rule” which states that, across a group of participants, recovery will be approximately 70% of the lost function (Krakauer and Marshall, 2015), even in the language domain (Marchi et al., 2017). The exceptions to the rule are patients with severe symptoms who fail to make any recovery or may be too unwell to participate in research studies. These exceptions lead to observations that the predictive value of initial severity decreases when symptoms are worse (Benghanem et al., 2019).

Despite its potential advantage, the proportional recovery rule has been criticised on the grounds of “mathematical coupling” and “compression to ceiling” (Hope et al., 2019, Bonkhoff et al., 2020, Bowman et al., 2021). Mathematical coupling occurs because one variable (initial severity) is included within the other variable (difference between initial severity and severity at a later date). This can result in inflated correlation coefficients. Compression to ceiling occurs when the sample includes multiple participants who score at, or close, to the highest possible test score. There is therefore little scope for recovery, either because little improvement is needed for full recovery, or the participants improve but the test score remains at ceiling.

Another important point to note about initial severity is that it is highly correlated with left hemisphere lesion size, and indeed could be explained by left hemisphere lesion size. This is because larger left hemisphere lesions are more likely to involve critical language regions, and thus result in severe initial symptoms (Basilakos et al., 2019, Osa García et al., 2020). However, initial severity is also expected to influence outcomes independently of lesion size, because (1) small lesions can also result in severe initial symptoms if they involve critical language regions, and (2) large lesions may result in mild aphasia if they occur in other (non-critical) left hemisphere regions, that are not associated with long-term aphasia (e.g. left cerebellum, occipital and prefrontal regions). The contribution of initial severity, over and above lesion size, has been demonstrated in studies where initial severity accounts for more variance in outcomes than lesion size or location, and is significantly predictive when lesion size is not (Benghanem et al., 2019, Osa García et al., 2020).

### 3.2.3. Age at stroke:

Older age at stroke is known to be associated with reduced functional independence and poorer motor outcomes compared to younger age at stroke (Meyer et al., 2015, Yoo et al., 2020). The influence of age at stroke on language outcomes is much less established, with an almost equal balance between prior studies which report a disadvantage of older age at stroke on language outcomes and recovery (Kertesz and McCabe, 1977, Holland et al., 1989, Laska et al., 2001, Kastrau et al., 2005, Naess et al., 2009, El Hachioui et al., 2013b, Nakagawa et al., 2019, Johnson et al., 2019, Dresang et al., 2022, Shin et al., 2022, Brady et al., 2022, Kristinsson et al., 2023), and those which found no effect of age on outcomes or recovery (Keenan and Brassell, 1974, Pickersgill and Lincoln, 1983, Lendrem and Lincoln, 1985, Pedersen et al., 1995, de Riesthal and Wertz, 2004, Pedersen et al., 2004, Lazar et al., 2008, Inatomi et al., 2008, Oliveira and Damasceno, 2009, Seniów et al., 2009, Basilakos et al., 2019, Gadson et al., 2022, Johnson et al., 2022, Kristinsson et al., 2022). A more detailed review of this literature is provided in Chapter 4, but briefly, the inconsistent evidence from these numerous prior studies is potentially due to differences in (1) sample sizes, (2) the choice of language outcome measures, and (3) covariate selection – specifically the exclusion of influential variables like lesion size that might influence the contribution of age to outcome. Crucially, the dependence of an age effect on other variables has not been explored in prior studies, increasing the likelihood that a significant effect of age, could be washed out when the sample is large and heterogeneous. The effect of age on language in a stroke population must also be differentiated from the effect of age on language in normal, healthy ageing, which may arise from cognitive decline and sensory changes (Cohen et al., 2019).

### 3.2.4. Amount of education:

A greater amount of education is known to be associated with better language abilities in neurologically normal populations, and more education is also associated with a reduced risk of dementia (Borod et al., 1980, Sharp and Gatz, 2011). Greater cognitive skills learned or gained during formal education (‘cognitive reserve’) may enable compensation for clinical symptoms arising from neurodegeneration (Stern, 2009, Stern, 2012, Meng and D’Arcy, 2012, Stern et al., 2020). However, it is not yet clear whether the same is true in stroke populations, and the relationship between education and post-stroke language outcomes and recovery requires further investigation. Prior studies have yielded inconsistent evidence: Connor et al. (2001) reported a benefit of education on initial aphasia severity, but from a sample of only 39 participants. González-Fernández et al. (2011) also reported a significant benefit of more education on initial aphasia severity, but only on certain language tasks (reading aloud and written word comprehension; not auditory comprehension or naming). Conversely, Johnson et al. (2022) reported no benefit of more education on aphasia outcomes in 147 participants when controlling for other demographic, lesion and cognitive variables. There is also only limited evidence that education influences language recovery, with a significant benefit reported by Kim et al. (2019) in 235 participants. Two other studies which did not find a relationship between education and recovery, by Connor et al. (2001) and Lazar et al. (2008), may have been limited by small sample sizes (n=39 and n=22 respectively) and by how ‘recovery’ was defined (in relation to an aphasia threshold, rather than a change in score). A more detailed review of these studies is provided in Chapter 5, but to summarise here, the influence of education on aphasia outcome may vary depending on the specific language task assessed, and the covariates controlled for. The evidence for an effect of education on aphasia recovery is limited.

### 3.2.5. Early speech and language therapy:

Speech and language therapy overall is known to benefit aphasia outcomes (Brady et al., 2016), but the evidence for a benefit of early therapy is relatively weak. A significant advantage of early therapy (within the first four weeks of stroke onset) has been reported by four studies. More details are given in Chapter 6, but briefly, their conclusions are weakened by several factors, including: (1) participant groups not being matched for initial severity (Godecke et al., 2012, Godecke et al., 2014); (2) non-maintenance of the effects at 6 months post-stroke (Godecke et al., 2012); (3) small sample size (n=6 in Mattioli et al. (2014); and (4) an absence of a therapy benefit in participants who were unable to tolerate intensive therapy (Nouwens et al., 2017). Four other studies found no benefit of early therapy on aphasia outcomes. Plausibly, this could be explained by (1) the therapy targeting a different language skill (comprehension) to the language outcome measured (speech production) (Laska et al., 2011); (2) insufficient therapy amount (1.3 hours per week in Bowen et al. (2012a)); or (3) an absence of a no-therapy control group (Woldag et al., 2017, Godecke et al., 2020). In summary, the benefit of early therapy is unclear, and warrants further investigation in larger samples of participants who are matched for factors known to influence recovery, such as initial severity and lesion size.

### 3.2.6. Pre-stroke handedness and right hemisphere lesion size:

The relationship between pre-stroke handedness and post-stroke language outcomes and recovery is inherently linked to theories of language lateralisation and the role of the right hemisphere. Language is typically lateralised to the left hemisphere, and so right hemisphere lesions are not expected to result in severe aphasia. However, prior studies have shown that language lateralisation differs for right-handers compared to people who are left-handed or ambidextrous. For example, language is left-lateralised in 95% of right-handers but only 70-85% of left-handers (Knecht et al., 2000, Ocklenburg et al., 2014). Right-handers are therefore more likely than left handers to experience language problems after left hemisphere lesions. Conversely, left-handers are more likely to experience ‘crossed aphasia’ after right hemisphere lesions (Sheehy Laurie, 2006).

In Sample A (described in Methods, section 3.3.1), 86% (492/569) were right-handed before their stroke; the remainder being left-handed or ambidextrous. These proportions are roughly comparable with those of the general population (Papadatou-Pastou et al., 2020). As noted above, atypical language lateralisation should be expected in a small proportion of the sample: approximately 5% of right-handers and 15-30% of left-handers will have right-lateralised language.

Although prior studies have not shown an effect of handedness on language recovery post-stroke (Pickersgill and Lincoln, 1983, Pedersen et al., 1995), the current study controlled for (1) right hemisphere lesion size, with the expectation that larger right hemisphere lesions might be more likely to result in aphasia in participants with right-lateralised language; and (2) pre-stroke handedness, because right-lateralised language is more likely in left-handers.

### 3.2.7. Pre-stroke dyslexia status:

Developmental dyslexia is a learning difficulty which is thought to affect 10% of people in the UK (Erbeli et al., 2022). It affects reading and writing abilities, and is specifically characterised by poor phonological skills, verbal memory and verbal processing speed. No studies were identified that have specifically investigated language outcomes and recovery in participants diagnosed with developmental dyslexia. That said, its influence is worth considering, because poor phonological skills arising from developmental dyslexia may be mistaken for aphasic symptoms, and/or interfere with aphasia recovery (Hudson et al., 2007). There may also be an interactive effect between dyslexia status and education amount: a diagnosis of dyslexia may result in more time spent in education receiving specialised therapy, tuition or one-to-one learning support. Conversely (and perhaps more likely), a negative experience of education resulting from the difficulties associated with dyslexia may reduce the likelihood of pursuing further education (Rose, 2009). Dyslexia awareness, assessment and diagnosis is far greater in contemporary education settings than it was for older generations. Therefore, bias within the sample may arise from younger participants being more likely to have a formal diagnosis of dyslexia than older participants. To control for this, participants also report whether they had pre-stroke difficulties with reading, spelling and writing (suggesting undiagnosed dyslexia).

### 3.2.8. Biological sex:

There is mixed and inconclusive evidence for an effect of biological sex (sometimes referred to as gender in the literature) on aphasia incidence, type, and recovery. The possible theoretical explanations for an effect of sex on language outcomes include: (1) claims that males are more likely to be left-lateralised, whereas females are more likely to have bilateral language processing (Shaywitz et al., 1995) which, if true, would result in females having more resources for neural reorganisation and compensation; and (2) males being more likely to receive hospital treatment within three hours of stroke onset; maximising their outcomes compared to females (Smith et al., 2010). However, evidence for an effect of sex on language outcomes and recovery is overall weak. See reviews by Plowman et al. (2012), Watila and Balarabe (2015), Gerstenecker and Lazar (2019), O'Halloran et al. (2023).

### 3.2.9. Assessment variables:

Four ‘assessment’ variables may influence language scores: time post-stroke of assessment, semantic impairments (measured using the CAT), visual impairments or hearing impairments. Time post-stroke may influence language scores if a greater proportion of participants show better scores later compared to earlier post-stroke (indicating recovery) than earlier compared to later. The CAT semantic memory test assesses the ability to recognise pictures and match them according to their relationship. Poor scores on this task are expected to impair performance on language tasks that use pictures, even when participants don’t have language impairments. Impaired vision or hearing would also impact performance on all relevant tasks that rely on intact vision and/or hearing, even in participants who are not aphasic. In the current study, all of these variables were measured at the same time as the language assessments, therefore cannot be used for prospective prediction of long-term language ability.

### 3.2.10. Summary, research questions and hypotheses

The evidence for a range of predictive factors of aphasia outcomes and recovery ranges from strong and convincing for clinical variables (initial severity, lesion size), to inconsistent and weak for demographic and therapy variables (amount of education, age at stroke, and early therapy), to insufficient and/or only relevant to a small proportion of patients (pre-stroke handedness, right hemisphere lesion size, dyslexia status, biological sex, and perceptual and cognitive abilities). The inconsistent and weaker evidence appears to arise when sample sizes were small, which limits the ability to control for other predictors of outcomes. This challenge is often faced by aphasia trials due to the logistics and cost of obtaining brain scans, and the expertise required for analysis of brain imaging, in addition to that of obtaining therapy details and other demographic data. The current study was able to address these challenges by drawing participants from the PLORAS database; securing both a large sample size and a wide range of lesion, demographic, therapy, and assessment variables for each participant.

The study objectives were to identify which predictor variables were most likely to prospectively predict long-term overall language ability. The results were expected to guide further, more detailed analyses of these variables in subsequent experiments. To this end, the research questions were: A. What are the strongest predictors of overall language ability? and B. What is the effect of early speech and language therapy on overall language ability? With regard to question A, the hypotheses were as follows. First, larger left hemisphere lesions and worse initial severity were unequivocally expected to have the greatest disadvantage on language scores. Second, and less palpable: additive benefits of more pre-stroke education and younger age at stroke may be (i) strong, if their effects are consistent across all participants and tasks, but (ii) weak or absent if they are very small, driven by subsets of participants with certain characteristics (for example, small lesions), and/or only observed on certain tasks. The third hypothesis, with regard to question B, was more complex. A benefit of early therapy would be expected if those who received therapy within the first four weeks had better language ability than those who received no early therapy. However an absence of an early therapy effect may arise from (i) insufficient therapy amount, frequency or intensity, given that it is well established that therapy effects are more likely when therapy amount is greater (Bhogal et al., 2003, Godecke et al., 2013), (ii) a mismatch between the therapy content and the outcome measure (e.g. naming therapy with broad language outcomes), (iii) individual variation in response to therapy (some responding, others not) that conceals therapy effects at the group level, (iv) non-maintenance or attrition of a benefit of early therapy on long-term overall language ability, given that this was typically assessed many years post-stroke, or (v) small effects that are masked out by variance from other predictors of aphasia. To address the latter point, therapy effects in the current experiment were investigated in a smaller sample of participants, who were selected after matching lesioned hemisphere and initial severity in those who did, and did not, receive early therapy.

## 3.3. Methods

### 3.3.1 Participant selection

Participants were selected from the PLORAS database if they were native English speakers. They were excluded if (i) they had unknown initial speaking severity (n=212), (ii) they were medically unwell (e.g. in a coma) one week post-stroke, and therefore unable to attempt speech production nor rate their initial severity (n=145), (iii) their self-reported language ability mismatched their objectively assessed language abilities (i.e. those rating themselves as ‘Normal’ had aphasic speech production according to CAT scores; n=12), or (iv) the amount of pre-stroke education they received after the age of 14 was unknown (n=6) or outlying (n=5; two with 16 years, two with 17 years, and one with 18 years). These criteria identified 749 participants, from which two different samples were selected: sample A (for research question A) and sample B (for research question B). Sample A included 569 participants with full behavioural, demographic and lesion data available. 180/749 were excluded because data were not available for some covariates. For example, for 168 participants, it was not possible for them to assess their initial reading abilities because they had not attempted to read at that point.

Sample B included 137 participants in whom to analyse the effect of early therapy. This sample differed from Sample A in three ways: (1) the amount of therapy received was known for all participants in Sample B, with early therapy amount ranging from none to 30 hours; (2) all participants in Sample B had left hemisphere damage (115 with left lateralised lesions and 22 with bilateral lesions) whereas 32% of Sample A did not have left hemisphere damage, and (3) all participants in Sample B had severe or moderate aphasia symptoms in the first week post-stroke, whereas 58% of Sample A had mild or no symptoms in the first week post-stroke. Participants were not selected for this analysis if their memories of therapy amount or timing were insufficiently vague. Participants with Mild or Normal initial severity were also excluded from the therapy sample because they (i) were less likely to have needed, or received therapy, and/or (ii) were recovering sufficiently regardless of therapy, such that the benefit of therapy could not be disentangled. Participants with no left hemisphere damage were not included because the majority of them had mild or no initial aphasia (115/142), and of the 27 remaining participants with severe or moderate initial aphasia, the majority (n=18) did not receive any early therapy. There were therefore insufficient numbers of participants without left hemisphere damage in whom to assess the effect of early therapy.

Participant characteristics for samples A and B are given in Table 3.1.

**Table 3.1** Participant characteristics for Samples A (n=569) and B (n=137).

| | | **Sample A** | **Sample B** |
|---|---|---|---|
| | N | 569 | 137 |
| **Sex** | Male (%) | 401 (70) | 102 (74) |
| | Female (%) | 168 (30) | 35 (26) |
| **Age at stroke** | Mean (SD) | 58 (12.99) | 57 (12.42) |
| | Range | 18 to 99 | 30 to 85 |
| **Years of education after age 14** | Mean (SD) | 5 (3.07) | 5 (3.17) |
| | Range | 0 to 13 | 0 to 13 |
| **Lesioned hemisphere** | Left (%) | 282 (50) | 115 (84) |
| | Right (%) | 162 (28) | 0 (0) |
| | Bilateral (%) | 105 (18) | 22 (16) |
| | Undetected (%) | 20 (4) | 0 (0) |
| **Left hemisphere lesion size (cm$^3$)** | Mean (SD) | 25 (48.84) | 63 (62.51) |
| | Range | 0 to 354.53 | 1 to 355 |
| **Right hemisphere lesion size (cm$^3$)** | Mean (SD) | 17 (45.20) | 3 (15.88) |
| | Range | 0 to 357 | 0 to 174 |
| **Initial speaking severity** | Severe (%) | 145 (25) | 104 (76) |
| | Moderate (%) | 92 (16) | 33 (24) |
| | Mild (%) | 202 (36) | 0 (0) |
| | Normal (%) | 130 (23) | 0 (0) |
| **Pre-stroke handedness** | Left (%) | 56 (10) | 9 (7) |
| | Right (%) | 492 (86) | 120 (87) |
| | Ambidextrous (%) | 21 (4) | 8 (6) |
| **Years post-stroke of CAT** | Mean (SD) | 3 (4.34) | 5 (4.61) |
| | Range | 0 to 36 | 0 to 24 |
| **Years post-stroke of ARTQ** | Mean (SD) | 4 (4.66) | 5 (4.73) |
| | Range | 0 to 39 | 0 to 24 |
| **Pre-stroke dyslexia status** | Formal diagnosis (%) | 12 (2) | 3 (2) |
| | Difficulties relative to peers [a] (%) | 40 (7) | 9 (7) |
| | None (%) | 330 (58) | 75 (55) |
| | Unknown (%) | 187 (33) | 50 (36) |

**Legend:** [a] ‘Difficulties relative to peers’ is reported when participants had reading or spelling difficulties compared to their peers that may indicate developmental dyslexia, but did not receive a formal diagnosis - usually because it was not assessed. For these analyses, these participants were grouped with those whose dyslexia status was ‘unknown’.

### 3.3.2 Language outcomes

The sole outcome measure for the current experiment was ‘overall language ability’ score, which was generated by averaging across six summary T scores from the CAT – see Chapter 2 section 2.4.3 for more details.

### 3.3.3 Therapy measures

Details of speech and language therapy received are self-reported on entry into the study, via the ARTQ which captures the (i) total amount and (ii) timing of therapy, as well as (iii) the broad content of therapy (see Chapter 2 section 2.5). For these analyses, two therapy periods were distinguished, and the amount received during each period was calculated: ‘early’ therapy was the number of hours received within one month of stroke onset, and ‘later’ therapy was the number of hours received between one month and one year post-stroke. The content of therapy was largely unknown, and was assumed to naturally vary between participants because of regional variation in service provision; however, given that all participants in Sample B had severe-to-moderate initial speaking difficulties, at least some of their therapy was expected to be targeted to those speaking impairments.

### 3.3.4 Lesion size

Structural brain images were acquired for all participants. In sample A, 383 were imaged on a 3T Trio (43 of which were after Prisma upgrade), 10 were imaged on a 3T Allegra, 136 were imaged on a 1.5T Avanto, and 40 were imaged on a 1.5T Sonata. In sample B, 79 were imaged on a 3T Trio, 1 was imaged on a 3T Allegra, 36 were imaged on a 1.5T Avanto, and 21 were imaged on a 3T Sonata. Lesions were delineated and an estimate of lesion volume made for left and right hemisphere lesion size separately. See Chapter 2 section 2.7 for more details of MRI scan acquisition and lesion identification.

### 3.3.5 Initial severity

Each participant was assigned an initial severity rating, derived from a retrospective participant-reported outcome measure of language ability at one week post-stroke, the ARTQ (see Chapter 2 section 2.5). This was repeated for each language modality (speaking, understanding, reading and writing). The severity categories were: Severe (unable to produce/comprehend any speech, or only able to produce automatic speech), Moderate (able to produce/comprehend single words and phrases), Mild (able to produce/comprehend lexically meaningful short sentences), and Normal. This categorisation of initial severity for each language skill was particularly advantageous for the outcome measure ‘overall language ability’ because the independent contributions of initial speaking, reading and understanding severity could be considered. That said, these variables were not entirely independent of one another; moderate correlations were observed between them.

### 3.3.6 Demographic variables

Questionnaires capture each participant's biological sex, pre-stroke handedness, dyslexia status, age at stroke, and amount of pre-stroke education (see Table 3.1 for sample characteristics, and Chapter 2 section 2.6 for details of data collection).

### 3.3.7 Assessment variables

Semantic memory was assessed using the CAT, which was typically (but not always) administered on the same day as the brain imaging. Scores from this task were included to control for the influence of visual perceptual, object recognition and semantic impairments on overall language ability.

Hearing and vision status were self-reported on the same day as the brain imaging and CAT assessment. Participants reporting visual or hearing impairments were encouraged to wear any corrective aids that would assist them to perceive the assessment stimuli (e.g. hearing aids, glasses). Nevertheless, the potential influence of uncorrected sensory impairments on overall summary scores was controlled.

Finally, time post-stroke was calculated by subtracting the stroke date from the assessment date, and converting this into number of years. Time post-stroke was included in the analysis to control for changes in overall language ability over time, as expected for patients who continue to recover months or years post stroke (Hope et al., 2017).

### 3.3.8 Statistical analyses

The analyses were designed to answer two research questions:

***A) What are the strongest predictors of overall language ability?***

Multiple linear regression was used to investigate the effects of the six predictor variables on overall language ability. As a reminder, these six variables were: left hemisphere lesion size, initial speaking, reading and understanding severity, amount of pre-stroke education and age at stroke. The eight additional regressors (to control for variance of no interest, as described above) were: pre-stroke handedness, right hemisphere lesion size, dyslexia status, sex, time post-stroke, semantic memory, and vision and hearing status. The ratio of total number of regressors (14) to participants (569) was within the recommendation for reliable assessment (Harrell, 2015).

### B) *What is the effect of early speech and language therapy on overall language ability?*

Multiple linear regression was repeated in Sample B (n=137) with two therapy variables: early therapy (as a binary variable: ≥1 hour versus none), and amount of later therapy (as a continuous variable: number of hours). The six predictor variables included these two therapy measures, left hemisphere lesion size, initial speaking severity, pre-stroke education amount and age at stroke. The two additional regressors were: time post-stroke and semantic memory. The following eight variables were not included in the therapy analysis: initial reading and understanding severity, right hemisphere lesion size, sex, handedness, and dyslexia, hearing and vision status. The motivation for excluding these variables from the therapy analysis were three-fold. First, if they were included the ratio of participants (n=137) to covariates (16) would have been too small for reliable assessment (Harrell, 2015) and this was corrected by reducing the total number of regressors to eight. Second, the analysis with 569 participants (excluding therapy effects) did not find a significant influence of initial understanding severity, right hemisphere lesion size, sex, handedness, dyslexia, vision or hearing status. Third, initial reading severity was considered to be less important for the investigation of therapy effects, because therapy for these participants was expected to prioritise their initial speaking impairments, not reading.

Care was taken to ensure the data met the assumptions of multiple linear regression. There was independence of residuals, as assessed by a Durbin-Watson statistic of 1.917. There was minimal collinearity in the data because neither analyses A nor B yielded Tolerance values of less than 0.1, nor Variance Inflation Factor (VIF) values that were greater than 10. Moreover, none of the independent variables have correlation values greater than 0.7 (see Table 3.2 and Table 3.3). There was normality of the residuals, as confirmed by inspection of a P-P plot. Six outliers were identified, whose standardised residuals were more than 3 standard deviations. This was because 4 had very low CAT scores, and 2 had very large left hemisphere lesions. As these measurements were not considered abnormal, the participants were kept in the analyses.

**Table 3.2** Pearson's correlations between the 14 covariates included in analysis A.

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **LH** | 1 | | | | | | | | | | | | | |
| **RH** | 0.06 | 1 | | | | | | | | | | | | |
| **SS** | 0.46* | 0.14* | 1 | | | | | | | | | | | |
| **US** | 0.33* | 0.00 | 0.53* | 1 | | | | | | | | | | |
| **RS** | 0.42* | 0.06 | 0.59* | 0.64* | 1 | | | | | | | | | |
| **Age** | 0.02 | 0.00 | 0.13* | 0.14* | 0.12* | 1 | | | | | | | | |
| **Edu** | 0.02 | 0.06 | 0.02 | 0.07* | 0.01 | 0.04 | 1 | | | | | | | |
| **TPS** | 0.31* | 0.12* | 0.23* | 0.23* | 0.21* | 0.22* | 0.02 | 1 | | | | | | |
| **SM** | 0.22* | 0.10* | 0.11* | 0.13* | 0.13* | 0.13* | 0.07 | 0.04 | 1 | | | | | |
| **Sex** | 0.07* | 0.01 | 0.05 | 0.00 | 0.07* | 0.14* | 0.02 | 0.01 | 0.05 | 1 | | | | |
| **Han** | 0.04 | 0.04 | 0.04 | 0.03 | 0.01 | -0.03 | 0.05 | 0.02 | 0.02 | 0.05 | 1 | | | |
| **Dys** | 0.06 | 0.05 | 0.06 | 0.01 | 0.03 | 0.14* | 0.05 | 0.05 | 0.04 | 0.02 | 0.05 | 1 | | |
| **Vis** | 0.05 | 0.05 | 0.01 | 0.06 | 0.07 | 0.05 | 0.04 | 0.05 | 0.12* | 0.05 | 0.04 | 0.01 | 1 | |
| **Hea** | 0.03 | 0.03 | 0.09* | 0.10* | 0.09* | 0.14* | 0.07* | 0.03 | 0.03 | 0.03 | 0.04 | 0.05 | 0.00 | 1 |
| | **LH** | **RH** | **SS** | **US** | **RS** | **Age** | **Edu** | **TPS** | **SM** | **Sex** | **Han** | **Dys** | **Vis** | **Hea** |

**Legend**: LH = left hemisphere lesion size. RH = right hemisphere lesion size. SS = initial speaking severity. US = initial understanding severity. RS = initial reading severity. Age = age at stroke. Edu = education amount. TPS = time post-stroke. SM = semantic memory. Han = handedness. Dys = dyslexia status. Vis = vision status. Hea = hearing status. Blue shading = negative value. Red = shading = positive value. * = significant at p=<0.05.

**Table 3.3** Pearson's correlations between the 8 variables included in analysis B.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **LH** | 1 | | | | | | | |
| **SS** | 0.23* | 1 | | | | | | |
| **Age** | 0.08 | 0.12 | 1 | | | | | |
| **Edu** | 0.03 | 0.08 | 0.04 | 1 | | | | |
| **TPS** | 0.40* | 0.10 | 0.29* | 0.04 | 1 | | | |
| **SM** | 0.33* | 0.11 | 0.27* | 0.14 | 0.12 | 1 | | |
| **ET** | 0.05 | 0.08 | -0.03 | 0.11 | 0.05 | 0.02 | 1 | |
| **LT** | 0.26* | 0.11 | 0.26* | 0.04 | 0.15* | 0.07 | 0.13 | 1 |
| | **LH** | **SS** | **Age** | **Edu** | **TPS** | **SM** | **ET** | **LT** |

**Legend:** LH = left hemisphere lesion size. SS = initial speaking severity. Age = age at stroke. Edu = education amount. TPS = time post-stroke. SM = semantic memory. ET = early therapy group. LT = later therapy hours. Blue shading = negative value. Red shading = positive value. * = statistically significant at p=<0.05.

## 3.4. Results

### A: The strongest predictors of overall language ability

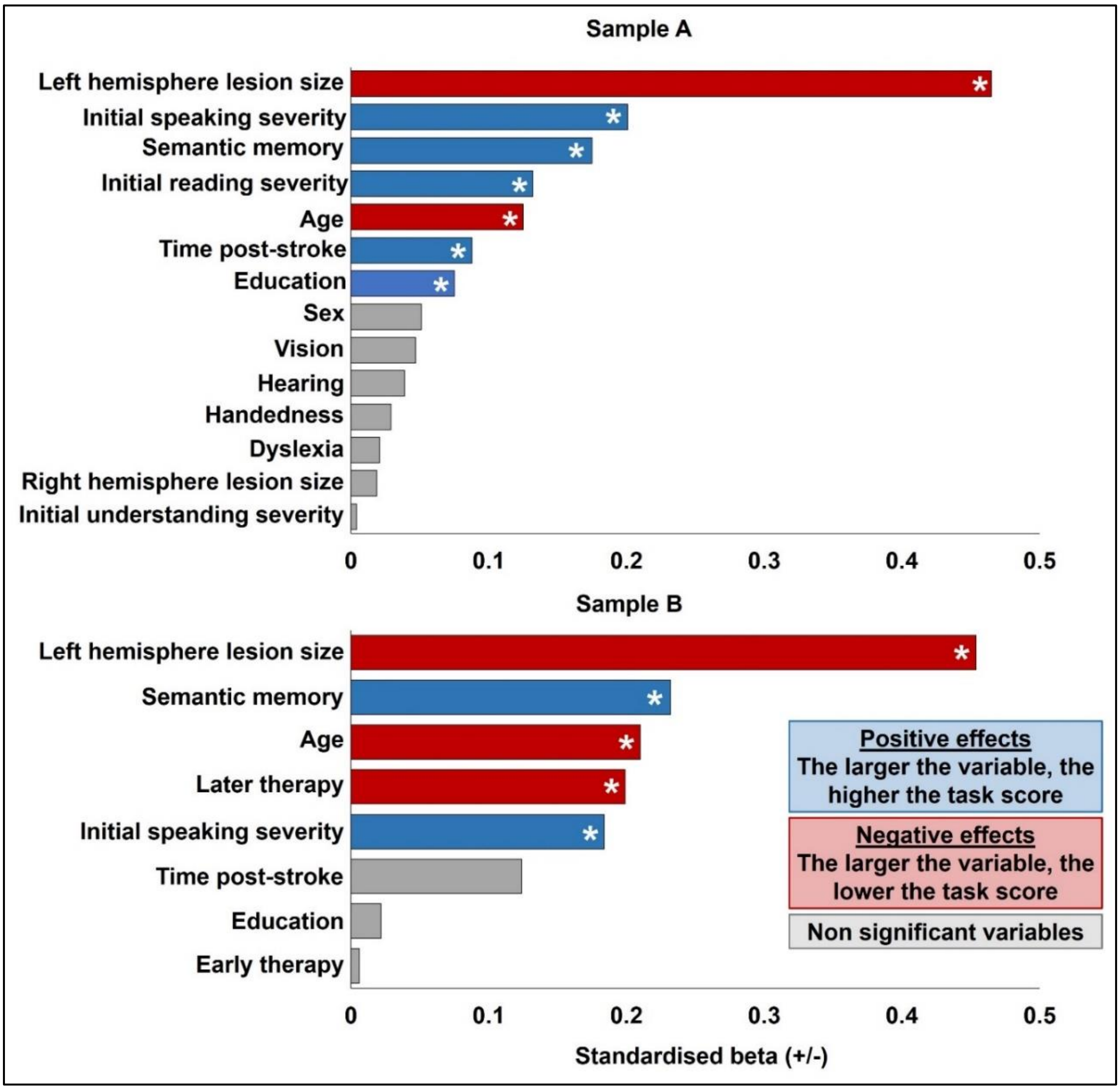
Across all 569 participants, five predictor variables had a highly significant influence on long-term overall language ability, which was worse with: larger left hemisphere lesions, more severe aphasia (speaking and reading) at one week post-stroke, older age at stroke and less pre-stroke education. In addition, long-term language ability was significantly worse earlier than later post-stroke. In total, the model accounted for 63% of variance in overall language ability ($R^2$ = 0.632), see Table 3.4 for the relative effect size of each variable, indicated by β.

**Table 3.4** The effect sizes and significance of 14 variables on overall language ability in Sample A.

| Variable | β | p value |
|---|---|---|
| Left hemisphere lesion size | -0.465 | <0.001 |
| Initial speaking severity | 0.201 | <0.001 |
| Semantic memory | 0.175 | <0.001 |
| Initial reading severity | 0.132 | <0.001 |
| Age at stroke | -0.125 | <0.001 |
| Time post-stroke | 0.088 | 0.002 |
| Education | 0.075 | 0.004 |
| Sex | 0.051 | 0.056 |
| Vision | -0.047 | 0.072 |
| Hearing | -0.039 | 0.144 |
| Handedness | -0.029 | 0.273 |
| Dyslexia | -0.021 | 0.437 |
| Right hemisphere lesion size | -0.019 | 0.469 |
| Initial understanding severity | 0.004 | 0.101 |

The upper part of Figure 3.1 illustrates that the predictor variable with the largest effect size was left hemisphere lesion size, followed by initial severity (speaking and reading), age at stroke, time post-stroke and education. The effect of these variables was significant irrespective of whether the analysis factored out semantic memory scores or not; however, when it was removed, the amount of variance explained dropped to 61%, and significant effects of initial understanding severity and vision status emerged.

**Figure 3.1** The effect size of different variables on overall language ability in Sample A (top) and B (bottom).



Sex, handedness, dyslexia, right hemisphere lesion size and hearing status had no significant effect on overall language ability, irrespective of whether the analyses included or excluded semantic memory scores at the time of the CAT. However, when these five nonsignificant variables were excluded, a significant effect of initial understanding severity emerged.

### B: The effect of early speech and language therapy on overall language ability

There was no significant effect of early therapy on overall language ability (p=0.925, β=0.006). In contrast, and counter-intuitively, more hours of later therapy was significantly associated with worse overall language ability (p=0.004, β=-0.199). This result could occur in the likely scenario that more hours of later therapy are provided to aphasic participants who were not recovering, either with or without early therapy. This explanation justifies the focus of this analysis on participants with moderate or severe aphasia in the first week post-stroke, and factoring out initial severity (moderate or severe). However, it was not possible to control for participants' capacity for recovery as the determinants of recovery are still unknown, and other (uncontrolled) variables may be at play.

The therapy analysis also revealed a change in the variables that had a significant effect on overall language ability score. As with Sample A, there were significant effects of left hemisphere lesion size, initial speaking severity, and age at stroke in Sample B. However, unlike Sample A, the effects of time post-stroke and education were not significant in Sample B. Overall, the total amount of variance explained was 49% ($R^2 = 0.491$) in Sample B, compared to 63% in Sample A. See Table 3.5 and the lower part of Figure 3.1 for the relative effect size of each variable.

**Table 3.5** The effect sizes and significance of 8 variables, including early therapy, on overall language ability in Sample B.

| Variable | β | p value |
|---|---|---|
| Left hemisphere lesion size | -0.454 | <0.001 |
| Semantic memory | 0.232 | 0.001 |
| Age at stroke | -0.210 | 0.003 |
| Later therapy amount | -0.199 | 0.004 |
| Initial speaking severity | 0.184 | 0.006 |
| Time post-stroke | 0.124 | 0.080 |
| Early therapy group | 0.006 | 0.925 |
| Education | 0.022 | 0.734 |

### 3.5. Discussion

Using data from 569 stroke survivors, this experiment found that four predictor variables significantly influenced long-term overall language ability. Larger left hemisphere lesions, more severe initial aphasia, older age at stroke, and less pre-stroke education were all associated with worse language outcomes. The model, which controlled for the influence of other stroke-related, demographic and perceptual/assessment variables, explained 63% of variance in outcomes; an improvement on many prior models (30% to 55% in Godecke et al. (2013), Johnson et al. (2022), Harvey et al. (2022)) – but within the same range as other models (40% to 70% in Ramsey et al. (2017), Osa García et al. (2020)). In Sample B, with therapy data (n=137), there was no benefit of early therapy on overall language ability, but 49% of variance in outcomes was explained by other predictor variables.

Novelties of this study include its very large sample size, the inclusion of many predictor and control variables, which have not all been available in prior studies, and the use of a participant-reported measure of initial aphasia severity. Furthermore, the therapy analysis enabled both a controlled investigation of early therapy within an atypically large sample, and the observation that significant effects varied with the sample tested. Finally, this study could be treated as a pilot analysis, paving the way for further detailed analyses, in subsequent experiments, that (1) focus on the effect of individual predictors and their relationship to other variables, and (2) explore how a range of specific language functions, measured using both objective assessments and participant-reported measures, are influenced by the variables of interest. These more detailed analyses have been absent from prior studies, and may help to explain inconsistencies in the evidence to date.

The discussion that follows considers the results in relation to the study objectives and hypotheses.

### 3.5.1. A: What are the strongest predictors of overall language ability?

In line with expectations from prior studies (Hope et al., 2013, Ramsey et al., 2017, Benghanem et al., 2019), larger left hemisphere lesions and worse initial aphasia symptoms had the greatest impact on overall language ability score, in both samples. As these effects were expected, they are not investigated as predictors of interest in subsequent experiments. That said, the influence of both was controlled in all further analyses to allow the investigation of how other predictors interact with lesion size and initial severity.

With regard to the second hypothesis, younger age at stroke and more education were found to significantly benefit overall language ability. The benefit of younger age prevailed in both analyses, but interestingly for education, its benefit was only observed in Sample A, suggesting that the effect of education varies depending on the characteristics of the sample. If the effect of education was driven by participants with, for example, small lesions or mild initial symptoms, its effects would not be observed in Sample B, which excluded those participants. Detecting age and education effects may therefore be limited by (1) small sample sizes, and (2) limited/no investigation of how an effect depends on other variables. This dependency may partially explain the mixed and inconclusive evidence from prior studies of both education and age in stroke survivors. Furthermore, these effects may be stronger or weaker for certain language tasks – for example, those which require wider vocabulary, or advanced cognitive and linguistic strategies. These issues could all be addressed with the data available in the PLORAS dataset, thus motivating me to investigate age at stroke and pre-stroke education further in Experiments 2 and 3 respectively.

### 3.5.2. B: What is the effect of early speech and language therapy on overall language ability?

No benefit of early therapy was observed on overall language ability. Plausibly, the lack of significance could arise from (i) insufficient therapy amount, frequency or intensity (none of which were manipulated in this experiment), (ii) a mismatch between the therapy content and the outcome measure (which was plausible in this experiment given that therapy is most likely to target speaking abilities but the overall language ability score is not limited to speaking); (iii) individual variation in response to therapy, resulting in an attenuated or absent effect across participants, or (iv) the therapy itself being ineffective. Experiment 4 pursues the second explanation (mismatch between therapy and outcome measure) in further investigations of how early therapy influences specific language functions, such as Naming.

The counter-intuitive and negative effect of later therapy is likely to reflect a limitation of the data available. Specifically, it was not possible to monitor the speed of recovery over the time course of the therapy, or to differentiate the type of therapy received. It is likely that the participants who received more later therapy were those who were making the slowest recovery (i.e. not responding to therapy). This would result in a negative relationship (as observed) between amount of therapy and language scores. It is also possible that the type of therapy changed in those needing more therapy – with later therapy targeting functional and/or compensatory communication strategies rather than language impairments (Doedens and Meteyard, 2022). More hours of later therapy, in this case, would not result in improved assessment scores, but rather improved participation and quality of life (which were not measured for this work). Later therapy was therefore not investigated further because of insufficient details about the type of therapy or the timing of therapy-related recovery.

### 3.5.3. The influence of other variables

Longer time post-stroke was significantly associated with better language ability in Sample A, indicating recovery of language skills over time. A non-significant trend was observed in Sample B – suggesting that this group also experienced recovery, but which was slower perhaps due to the worse initial symptoms in Sample B. The influence of time post-stroke on the effects of age and education was considered thoroughly in the subsequent experimental chapters. For example, an effect of education that was stronger later than earlier post-stroke would be consistent with a benefit of education on recovery. In contrast, Experiment 4 (investigating early therapy) did not investigate how the effect of therapy depended on time post-stroke because the outcome timepoints were strictly controlled (one month and one year post-stroke).

Better semantic memory ability was significantly associated with better overall language ability in both samples, as expected. This suggests that processing required by the semantic memory task is essential for CAT tasks tapping language functions. For example, the ability to visually perceive or recognise an object from a picture is required for most of the CAT comprehension tasks. Semantic memory is also fundamental for tasks involving lexical access and retrieval, and integrating of multiple ideas and concepts during spoken picture description and sentence comprehension. If, as the results suggest, the semantic memory task is accounting for variance related to language processing, then the estimation of variance for individual tasks may be conflated by co-linearity. This concern is mitigated by three points. First, the CAT semantic memory task isolates semantic knowledge from language as much as possible, by using non-verbal stimuli (pictures) and requiring a non-verbal response (pointing). Second, only small-medium correlations were observed between semantic memory task scores and each of the summary task scores as well as overall language ability (Pearson’s r= 0.242 to 0.407); suggesting that whilst they share similarities, the tasks capture discrete abilities. Third, and critically for the experiments reported in this thesis, when semantic memory was removed from the model of analysis A, the effects of left hemisphere lesion size, initial severity, age and education remained significant, and their effect sizes were largely unchanged. This suggests that semantic memory explains some unique variance in language outcomes, which warrant further investigation; ideally using a longer and more detailed assessment of semantic abilities, both during the acute and chronic stages of recovery. Further investigations were not possible in the current work because only 82 participants had impairments on the semantic memory task.

The variables with no significant influence (sex, vision status, hearing status, handedness, dyslexia status, right hemisphere lesion size and initial understanding severity) on overall language ability were expected, given that the prior evidence for their effects are inconsistent or negligible. Initial understanding severity only emerged as significant when semantic memory was removed from the model; even then, its effect was smaller than that of initial speaking and reading severity. As discussed in Chapter 2 (section 2.5.2) it may be less reliable than speaking severity; being more difficult to assess and measure. This may explain its inconsistent and/or context-dependent effect. That said, these variables may still have sample-dependent effects on a small number of participants, and so these factors were (1) controlled for in subsequent experimental analyses, and (2) used to test how they influenced the impact of other predictor variables of interest.

### 3.5.4. Limitations

The study design was shaped by the data held in the PLORAS database. The limitations of these data, with respect to the current study, are discussed below.

First, the retrospective measurement of initial severity relies on participants' memories, which may be inaccurate and/or susceptible to bias depending on how well participants recovered. For example, a good recovery may motivate a more positive retrospective rating of one's initial symptoms, whereas a poor recovery may motivate a more severe retrospective rating (Kress and Aue, 2017). This could be mitigated in future experiments by using initial assessment scores, such as NIHSS (National Institutes of Health Stroke Scale), which are favoured in acute clinical settings because they assess overall aphasia severity and are quick to administer (NICE, 2022). These scores were not collected for the participants included in this study. Furthermore, the timing and type of clinical assessment may vary between patients, whereas the current approach had the advantage of inter-subject consistency because every participant completed the same measure, for the same timepoint.

A second limitation relates to the measurement of education amount. It was recorded as the number of years completed, consistent with prior studies. However, this disregards the possible influence of different education settings, content, and attainment, as well as other factors that may moderate education such as socioeconomic status, intelligence, and learning difficulties. It was therefore gratifying, that despite this limitation, a significant effect of education was still observed in Sample A with 569 participants. The advantages and limitations of this measurement of education will be discussed in more detail in Chapter 5.

The third limitation is that semantic matching abilities were assessed concurrently with the language assessment. Whilst they may explain low scores on language tasks, they cannot be used to prospectively predict language outcomes and recovery in this sample. Future studies should consider measuring these abilities early post-stroke, to understand how acute semantic impairments, and other cognitive and perceptual abilities influence long-term language outcomes and recovery.

A fourth limitation was the omission of the many other variables that may influence language outcomes and recovery, because the data were not available in the PLORAS database. These include the cognitive, health and social factors introduced in Chapter 1, such as attention, brain age, BMI, depression, and social support (Busby et al., 2023, Gilmore et al., 2019a, Johnson et al., 2019). While overfitting is a risk, and covariate selection must be well-motivated (as discussed in Chapter 2, section 2.9), future studies would benefit from consideration of these variables, including their relationship to/dependence on each other.

Finally, details of therapy were participant-reported, and therefore also susceptible to inaccurate memory recall and/or optimism or pessimism bias. Furthermore, because the therapy was received clinically, and naturally varies between participants, it was not possible to control for amount, content, or frequency. However, being clinically provided, it was also more likely to be tailored to the individual’s needs and goals (and therefore more motivating and effective); which is not necessarily the case with experimental trial therapy. The advantages and limitations of this measurement of therapy details are discussed in more detail in Chapter 6.

### 3.5.5. Conclusions and Summary of subsequent experiments

This experiment identified two predictor variables – age at stroke and amount of pre-stroke education - that significantly influenced long-term overall language ability, and for which the prior evidence has been unclear or inconsistent. Further investigation is required to fully understand the influence of these variables in different subgroups of participants, and on different language functions, in order to determine whether they can be used to prospectively predict outcomes and recovery. Experiments 2 and 3 therefore consider: (a) the independent effect of age and education on a range of different language scores; (b) how the effect of each variable depends on other variables; and (c) whether and how each variable influences language recovery. The investigation of early therapy, in Experiment 4, focused on its effect on speech production, whilst controlling for other influential lesion and demographic variables.

# Chapter 4: Experiment 2. The effect of age at stroke on language outcomes and recovery after stroke

### 4.1. Summary of experiment

**Introduction:** The goals of this study were to explain prior inconsistency in reports of how age affects post-stroke language outcomes and recovery by establishing (A) which language task scores are sensitive to the chronological age of stroke survivors at the time of their stroke, (B) how the effect of age on post-stroke language performance depends on other stroke and non-stroke factors, and (C) whether and how age at stroke influences recovery from aphasia.

**Methods:** Post-stroke language abilities were assessed in 749 participants using (i) a range of language tasks from the CAT, and (ii) self-reported measures of speaking, understanding, reading and writing ability at three timepoints within the first year of stroke onset. The predictor variable of interest was chronological age at stroke. Covariates included: left and right hemisphere lesion size, amount of pre-stroke education, sex, handedness, presence of developmental dyslexia, time post-stroke, initial aphasia severity, and hearing and vision status. Multiple regression analyses were conducted to investigate the effect of age at stroke on (1) CAT tasks and (2) the degree of self-reported recovery from one week to one year post-stroke. Bayesian regressions were conducted to assess the strength of the evidence for any observed effects.

**Results:** (A) Older age at stroke was associated with worse overall language ability ($\beta$=-0.125), and specifically worse nonword ($\beta$=-0.258) and word ($\beta$=-0.201) repetition, semantic ($\beta$=-0.234) and letter ($\beta$=-0.119) fluency, spoken and written word ($\beta$=-0.164 and $\beta$=-0.061) and sentence ($\beta$=-0.140 and $\beta$=-0.151) comprehension, and object naming ($\beta$=-0.106). (B) The disadvantage of older age at stroke was observed in all participants irrespective of left hemisphere lesion size and initial aphasia severity, but was worse amongst participants with the largest lesions and most severe initial aphasia. (C) Consistent with many but not all past studies, there was no significant influence of age at stroke on recovery of language, measured using either CAT scores or participant-reported measures.

**Conclusion:** Older age at stroke results in worse performance on language scores compared to younger age at stroke, suggesting that the CAT is sensitive to age irrespective of stroke/aphasia symptoms, and that scores should be interpreted with regards to age to help inform diagnosis and/or further assessment. For participants with large lesions and severe symptoms, both language recovery, and the disadvantage of older age were observed to be worse. An implication of these results is that understanding recovery potential and the success of future therapy could be optimised by considering the impact of age at stroke.

## 4.2. Introduction

This study investigates how the chronological age of stroke survivors, at the time of their stroke, influences post-stroke speech and language outcomes and recovery. Specifically, it investigates (A) which language task scores are influenced by age at stroke, (B) whether an effect of age on language abilities depends on other variables, and (C) whether age influences language recovery. The goals of the study were to understand whether (1) the clinical management of aphasia might be influenced by age at stroke, and (2) prognostic models for aphasia outcomes and recovery could be improved by the inclusion of age at stroke.

Age at stroke is already known to influence a variety of stroke outcomes. Older stroke patients have reduced functional independence and poorer motor outcomes, compared to younger patients; both at the time of stroke and well into the chronic phase of recovery (Meyer et al., 2015, Yoo et al., 2020). This may be for a number of reasons. One possibility is that older age is associated with an increased presence of comorbid conditions (Yousufuddin et al., 2017), and an increased presence of risk factors for stroke itself (e.g. diabetes, hypertension), all of which may inhibit stroke recovery. A second possibility is that neuroplasticity may be compromised with increasing age, as has been observed in animal studies (Hermann and Chopp, 2012), meaning that the mechanisms of recovery are reduced in older compared to younger people. A third possibility is that the effect of age on stroke recovery is a natural consequence of the physical and cognitive changes that occur with normal healthy ageing. Older age is associated with less grey and white matter volume (Harada et al., 2013) which reduces the integrity of brain networks and increases biological ‘brain age’. Critically, brain age does not necessarily match an individual’s chronological age exactly. For example, brain age is accelerated, relative to chronological age, in individuals with neurodegeneration (Biondo et al., 2022), excessive alcohol use (Bøstrand et al., 2022), higher visceral fat levels (Wing et al., 2022), physio-cognitive decline (Kuo et al., 2023) and poorer cognitive function after stroke (Aamodt et al., 2023). These correlations/associations need to be interpreted with caution as they are not necessarily causal (Salthouse, 2010); but the point being made here is that chronological age effects after stroke might be the consequence of brain age.

The effect of age at stroke on aphasia outcomes and recovery is not yet clearly understood, as revealed by the literature review below. Further understanding is a high research priority, because the combination of a decreasing average age at stroke (Evans, 2018), and longer survival as a result of better treatments (Sally et al., 2011), means that the number of people living with aphasia is likely to increase. An understanding of how age affects aphasia outcomes across the lifespan also has the potential to improve prognostic models, thereby offering patients more realistic and accurate information about their likely recovery.

### 4.2.1. The effect of age on speech and language abilities in healthy adults

Prior literature has found that older adults have significantly poorer performance than younger adults on verbal fluency tasks and memory/learning tasks, in a variety of different languages (Santos Nogueira et al., 2016, Lubrini et al., 2022, Shirdel et al., 2022, Kempler et al., 1998, Tombaugh et al., 1999, Jebahi et al., 2022, Messinis et al., 2016). That said, other studies have reported no, or inconsistent effects of age, or have shown that age effects depend on which type of language function is being tested. For example, receptive verbal abilities and semantics are relatively stable; whereas skills such as lexical retrieval may deteriorate with age (Cohen et al., 2019) – even though vocabulary size is typically larger amongst older compared to younger adults (Verhaeghen, 2003, Keuleers et al., 2015). Likewise, age effects were observed on semantic but not phonemic fluency (Aki Ö et al., 2022, Abdel Aziz et al., 2017, Pereira et al., 2018, Mathuranath et al., 2003).

A few studies have also reported non-linear relationships with age; where scores improved (or remained stable) with age during mid-life, then began to decrease during later life (Muayqil et al., 2021, Van der Elst et al., 2006). Such effects could be lost in studies searching for linear effects of age. In other studies, the effect of age was found to depend on the effect of education. For example, in Turkish adults, older age reduced the benefit of more education on phonemic fluency (Aki Ö et al., 2022); and in Spanish adults, the disadvantage of older age was worse in patients with less education (Lubrini et al., 2022).

The cause of ageing effects on language abilities is likely to be due to a combination of factors. As discussed above, ageing increases brain atrophy, reducing white matter connectivity and synaptic activity and changing neurotransmitter and cerebrovascular function (Cohen et al., 2019). In addition, a number of other physical changes occur during normal, healthy ageing, which may affect various aspects of speech and language production and comprehension. For example, muscle atrophy and changes in elasticity in the vocal cords can reduce vocal volume, alter vocal pitch, and result in voice weakness. Hearing loss will affect speech comprehension; visual loss will affect reading ability; reduced motor control could affect speech production and writing, and slower processing speed and reduced attention could affect the ability to follow and participate in a conversation (Cohen et al., 2019).

### 4.2.2. The effect of age on speech and language abilities post-stroke

This section reviews prior studies that investigated the effect of age on (i) aphasia outcomes, and (ii) aphasia recovery. Table 4.1 shows a summary of the studies reviewed, grouped according to their subject and conclusions.

**Table 4.1** Summary of literature investigating the influence of chronological age on aphasia outcomes and recovery.

| Subject | Conclusion | Study |
|---|---|---|
| Age & outcomes | Significantly worse aphasia outcomes in older patients. | Naess et al. (2009)<br>El Hachioui et al. (2013b)<br>Nakagawa et al. (2019)<br>Shin et al. (2022)<br>Dresang et al. (2022) |
| | No, or limited evidence for a relationship between age and aphasia outcomes. | Keenan and Brassell (1974)<br>de Riesthal and Wertz (2004)<br>Oliveira and Damasceno (2009)<br>Basilakos et al. (2019)<br>Gadson et al. (2022)<br>Johnson et al. (2022) |
| Age & recovery | Significant effect, or correlation/trend, demonstrating worse recovery in older patients | Kertesz and McCabe (1977)<br>Holland et al. (1989)<br>Laska et al. (2001)<br>Kastrau et al. (2005)<br>Johnson et al. (2019)<br>Shin et al. (2022)<br>Brady et al. (2022)<br>Kristinsson et al. (2023) |
| | No, or limited, evidence for a relationship between age and aphasia recovery | Pickersgill and Lincoln (1983)<br>Lendrem and Lincoln (1985)<br>Pedersen et al. (1995)<br>Pedersen et al. (2004)<br>Lazar et al. (2008)<br>Inatomi et al. (2008)<br>Seniów et al. (2009)<br>Basilakos et al. (2019)<br>Kristinsson et al. (2022) |

### *(i) Age and post-stroke language outcomes (aphasia severity)*

Eleven studies that investigated the effect of age on language outcomes in participants with post-stroke aphasia were identified. Five of these studies (Naess et al., 2009, El Hachioui et al., 2013b, Nakagawa et al., 2019, Shin et al., 2022, Dresang et al., 2022) reported a benefit of younger age on language outcome. One (Johnson et al., 2022) of the eleven studies showed an effect of age on aphasia severity when demographic factors were controlled and not when health and lesion factors were also controlled. The five remaining studies reported no relationship between age and aphasia outcome (Keenan and Brassell, 1974, de Riesthal and Wertz, 2004, Oliveira and Damasceno, 2009, Basilakos et al., 2019, Gadson et al., 2022).

The absence of age effects in 5/11 studies might be a consequence of insufficient sample sizes because all five of these studies had less than 100 participants (four had less than 40 participants) whereas 5/6 of the studies showing age effects (including Johnson et al., 2022) included more than 100 participants. Larger sample sizes are more likely to control the impact of other factors across age, thereby increasing statistical power. The Johnson et al. (2022) study also shows that sensitivity to age effects depends on how well other sources of variance are controlled. Specifically, in addition to showing that age effects were lost when health and lesion factors were controlled, the authors also reported a significant correlation between age and white matter integrity. It is therefore plausible that the effect of age on aphasia severity, when observed, may be related to the integrity of residual brain networks (i.e. brain age).

A comparison of the effects of chronological versus brain age was reported by Kristinsson et al. (2022). They found that brain age had a greater impact on language outcomes than chronological age; specifically, accelerated brain age relative to chronological age (i.e. an ‘older’ brain) was associated with worse performance than decelerated brain age relative to chronological age (i.e. a ‘younger’ brain). This study also revealed how age effects can be task dependent: they were observed on Naming, Repetition and overall score, but not on auditory comprehension. Both of these findings emphasise the importance of (i) careful and well-motivated selection of covariates, and (ii) fully investigating any reported effects to understand how age influences language in the presence of other variables, and how those variables interact with each other.

Prior investigations of how post-stroke language outcomes are influenced by other factors are limited. Indeed, only one study (Dresang et al., 2022) investigated interactions between age and other variables on aphasia severity; reporting that the disadvantage of older age was stronger in patients who were carriers of one genetic biomarker (ValVal allele) compared to another (ValMet allele). The current study therefore planned to investigate how the effect of age on post-stroke language outcomes was influenced by a range of different variables. For example, as reported in neurotypical populations, age effects may depend on the amount of education undertaken (Lubrini et al., 2022).

In addition, and in contrast to the majority of past studies that used broad language outcomes (e.g. measuring overall aphasia severity; most often with the Western Aphasia Battery Aphasia Quotient, or WAB-AQ), the current study investigated how age effects depend on the specific language process being tested. For example, as suggested with neurotypical populations, Naming may be more vulnerable to ageing, whereas Comprehension is more robust (Shafto and Tyler, 2014, Cohen et al., 2019). This could only be observed if specific language tasks, which test a range of processes, are analysed, (e.g. lexical retrieval, semantics) – which was not the case in prior studies of chronological age effects on post-stroke aphasia severity. As noted above, however, Kristinsson et al. (2022) found that brain age had a greater effect on naming and repetition than auditory comprehension (whereas chronological age had a greater effect on naming than both repetition and auditory comprehension.

### *(ii) Age and aphasia recovery*

Seventeen studies were identified that investigated how age influences aphasia recovery (i.e. the degree of improvement in language scores over time). Eight of the seventeen studies reported either a significant relationship or trend, where older age resulted in worse recovery. The remaining nine studies found no relationship between age and recovery. Factors that might explain why age effects are inconsistent include: (1) sample size, (2) time-frame for recovery, (3) the type of recovery being measured, (4) the use of intervention to facilitate recovery, and (5) how the influence of other variables was factored out. However, there were no notable differences between these factors for studies that did and did not detect an effect of age. For example:

(1) Sample size ranged from very small to very large in studies that found an effect of age (n=14 in Kastrau et al. (2005); n=>4000 in Shin et al. (2022)) as well as studies that did not find a significant effect of age (n=22 in Lazar et al. (2008); n=>800 in Pedersen et al. (1995), Inatomi et al. (2008)).

(2) The timeframe for recovery also ranged widely in studies that found a significant effect of age: first three months in Holland et al. (1989) to many years post-stroke (Kertesz and McCabe, 1977, Johnson et al., 2019, Kristinsson et al., 2023, Shin et al., 2022) as well as studies that did not find a significant effect of age: 10 days in Inatomi et al. (2008) to years post-stroke (Pedersen et al., 2004, Basilakos et al., 2019).

(3) The type of recovery measured was very inconsistent both within and between studies that found, or did not find an effect of age (e.g. aphasia severity measured by the Western Aphasia Battery (Johnson et al., 2019), the Porch Index of Communicative Ability in (de Riesthal and Wertz, 2004, Pickersgill and Lincoln, 1983, Lendrem and Lincoln, 1985); the Boston Diagnostic Aphasia Examination (Lazar et al., 2008, Seniów et al., 2009), and change in Naming ability (Kristinsson et al., 2023).

(4) Treatment to facilitate recovery was provided in some of the studies with significant effects of age (Kristinsson et al., 2023, Brady et al., 2022), as well as some of the studies with no significant effect of age (Seniów et al., 2009, Lazar et al., 2008).

(5) Finally, the variables that were factored out of the analysis also differed within and across studies that found and did not find an effect of age. Studies finding a significant effect of age typically included some, but not all of the following covariates: sex, initial severity, lesion size, stroke type, diabetes status, cognitive abilities, education, lesioned hemisphere, ethnicity, history of previous stroke, time post-stroke.

Surprisingly, lesion size was only included as a covariate in one study (Laska et al., 2011). This provides a strong motivation to look at whether lesion size influences the effect of age in the current study. Furthermore, as with the studies investigating the effect of age on language outcome, very few studies of age effects on language recovery investigated how age interacted with other variables. One exception to this was a study by Johnson et al. (2019) who reported a negative effect of older age on recovery that was worse in patients with diabetes. A central hypothesis of the current study is that the effect of age on recovery will vary across subgroups of participants with different characteristics (such as larger versus smaller lesions).

### 4.2.3. Summary of evidence and Motivation for the current study

There is inconsistent evidence that age influences aphasia outcomes and recovery, with an almost equal split between studies finding and not finding an effect of age. Prior studies have not sufficiently explained this inconsistency, potentially because of (1) a lack of specific outcome measurements, and (2) insufficient investigation of the effects in relation to other stroke and non-stroke variables.

The current study addresses these two problems by drawing participants from the PLORAS database, for whom language scores are available for a variety of language tasks from the CAT, alongside other lesion and demographic data. The study had three goals. Goal A investigated which language tasks were most sensitive to age at stroke. Goal B investigated whether the effect of age depended on other variables, such as initial severity or lesion size. Goal C investigated whether age at stroke influenced recovery, measured as change between initial severity and (i) long-term language scores and (ii) participant reports of one-year ability.

### 4.2.4. Hypotheses:

For goal A, the expectation was that older age at stroke would result in poorer performance on tasks that use word retrieval, such as Naming, Picture description, and Fluency (Kavé and Goral, 2017). Conversely, for language tasks that rely on receptive abilities and semantics, no effect of age was expected (Cohen et al., 2019).

For goal B, the expectation was that the disadvantage of older age would be worse in participants with the most severe aphasia, at one week post-stroke. In this scenario, an interaction of age with both initial severity and also lesion size could be observed (because larger lesions can result in worse initial symptoms). Prior literature also suggests that the effect of age might depend on the effect of education (Aki Ö et al., 2022, Lubrini et al., 2022). Alternatively, if the disadvantage of older age at stroke was the same for all participants, including those with normal initial language ability, it is more likely to be due to normal healthy ageing.

For goal C, despite the inconsistent conclusions from prior studies (see above), older age at stroke was expected to hinder language recovery; after controlling for a wider range of variables than previous studies.

## 4.3. Methods

### 4.3.1. Participant selection

Participants were selected from the PLORAS database (Seghier et al., 2016) if they were native English speakers. They were excluded if they (a) had unknown initial speaking severity (n=212), (b) were medically unwell (e.g. in a coma) one week post-stroke, and therefore unable to attempt speech production nor rate their initial severity (n=145), (c) had an unknown (n=6) or atypical amount of pre-stroke education after the age of 14 (n=5; two with 16 years, two with 17 years, and one with 18 years); or (d) had inconsistent self-reported language ability compared to their objectively assessed language abilities (i.e. those rating themselves as ‘Normal’ but with aphasic speech production according to CAT scores; n=12). These criteria identified 749 eligible participants, details for whom can be found in Table 4.2. Most of these participants had behavioural, demographic and lesion data available. However, participants with missing data for some covariates were excluded from some analyses; see the description of the statistical analyses in section 4.3.6 for more details.

**Table 4.2** Participant characteristics.

| | | |
|---|---|---|
| **Number** | | 749 |
| **Age at stroke (years)** | Mean (SD) | 57 (13) |
| | Range | 17 to 99 |
| **Sex** | N male (%) | 518 (69) |
| | N female (%) | 231 (31) |
| **Initial speaking severity** | Severe (%) | 243 (32) |
| | Moderate (%) | 118 (16) |
| | Mild (%) | 242 (32) |
| | Normal (%) | 146 (19) |
| **Education amount after age 14 (years)** | Mean (SD) | 5 (3) |
| | Range | 0 to 13 |
| **Years post-stroke of CAT** | Mean (SD) | 4 (5) |
| | Range | 0 to 42 |
| **Lesioned hemisphere** | Left (%) | 383 (51) |
| | Right (%) | 201 (27) |
| | Both (%) | 134 (18) |
| | Undetected (%) | 31 (4) |
| **Left hemisphere lesion size (cm$^3$)** | Mean (SD) | 31 (58) |
| | Range | 0 to 428 |
| **Right hemisphere lesion size (cm$^3$)** | Mean (SD) | 17 (46) |
| | Range | 0 to 357 |
| **Pre-stroke handedness** | Right (%) | 648 (87) |
| | Left (%) | 77 (10) |
| | Ambidextrous (%) | 24 (3) |
| **Developmental dyslexia status** | Formal diagnosis (%) | 14 (2) |
| | None (%) | 425 (57) |
| | Difficulties relative to peers [a] (%) | 52 (7) |
| | Unknown (%) | 258 (34) |

**Legend:** [a] 'Difficulties relative to peers' is reported when participants had reading or spelling difficulties compared to their peers that may indicate developmental dyslexia, but did not receive a formal diagnosis (usually because they were not assessed for dyslexia). For statistical analyses, these participants were grouped with those whose dyslexia status was 'unknown'.

### 4.3.2. Age at stroke and other demographic variables

Questionnaires capture each participant's date of birth and date of stroke, from which chronological age at stroke was subsequently calculated. The same questionnaires capture each participant's biological sex, pre-stroke handedness, developmental dyslexia status, and amount of pre-stroke education (measured as the number of years completed).

### 4.3.3. Initial aphasia severity

Each participant was assigned an initial severity rating, derived from the ARTQ (see Chapter 2 section 2.5). The 4 severity categories were: Severe (unable to produce any speech, or only able to produce automatic speech), Moderate (able to produce single words and phrases), Mild (able to produce lexically meaningful short sentences), and Normal. The ratings were repeated for understanding, reading and writing.

### 4.3.4. Lesion data

Separate estimates of left and right hemisphere lesion size were obtained from structural MRI brain scans (see Chapter 2 section 2.7 for details of scan acquisition and lesion identification). 497 participants were imaged on a 3T Trio (53 of which were after Prisma upgrade), 11 were imaged on a 3T Allegra, 176 were imaged on a 1.5T Avanto, and 65 were imaged on a 1.5T Sonata.

### 4.3.5. Language outcomes

The focus of the current study was 19 tasks and 6 ‘summary’ scores from the CAT (Naming, Repetition, Reading, Spoken picture description, Spoken Comprehension and Written Comprehension), and ‘overall language ability’. See Chapter 2 section 2.4 for full details of the CAT and the language tasks.

### 4.3.6. Statistical analyses

Described below are a series of analyses that are designed to answer each question of interest (A, B, and C).

#### A. Which language tasks are most sensitive to age at stroke?

This question was approached using three steps. (1) Is there a main effect of age on overall language ability score? (2) Which summary scores (e.g. naming, repetition, etc.) drive this effect? (3) Which individual tasks drive effects of age in the summary scores? As there is only one statistical analysis in the first step, there is no need to correct for multiple comparisons. The second and third steps involve 6 different analyses (one for each summary score) and 24 different analyses (one for each of the individual tasks and subscores). If the main effect in step one is significant, the second and third steps can be considered as descriptive, and therefore do not need to be corrected for multiple comparisons. However, the Table legends also note which summary and task scores would survive a correction for multiple comparisons if they were considered independently.

Each of the analyses described above was conducted using multiple linear regression. The following covariates were included in all analyses to factor out the effects of: education, time post-stroke of CAT, left hemisphere lesion size, right hemisphere lesion size, sex, pre-stroke handedness and developmental dyslexia status. The addition of covariates measuring initial speaking, reading and understanding severity, vision status, hearing status and semantic memory at the time of the CAT depended on the task scores being considered. Specifically: all covariates were included in the analysis of overall language ability; initial speaking severity was included in the analyses of speaking scores (naming, repetition, spoken picture description); initial understanding severity was considered for the analyses of comprehension scores, initial reading severity was included in the analyses of reading scores, vision status was considered for tasks that depend on intact visual perception (e.g. naming and reading), and hearing status was included for tasks which depend on intact hearing ability (e.g. repetition and spoken comprehension). Semantic memory scores were included for tasks that included visual or conceptual stimuli (naming, spoken picture description). The covariates selected for each task analysis can also be seen in Table 4.3. Each analysis had a different sample size, due to missing data for some covariates. For example, 173 participants had missing data from the self-reported reading and/or understanding severity ratings; 6 more had missing hearing or vision data, and 3 more had missing CAT scores for various tasks. Participants were excluded from the analyses that used the covariates for which they had missing data. The sample size for each analysis is reported in Table 4.5. The impact of excluding these participants is considered in the Discussion.

**Table 4.3** The covariates factored out of each CAT analysis.

| | SS | RS | US | SM | Vis | Hea |
|---|---|---|---|---|---|---|
| **Overall language ability** | + | + | + | + | + | + |
| **Naming:** | + | - | - | + | + | - |
| Semantic fluency | + | - | - | + | - | - |
| Letter fluency | + | - | - | + | - | - |
| Object naming | + | - | - | + | + | - |
| Action naming | + | - | - | + | + | - |
| **Repetition:** | + | - | - | + | - | + |
| Word repetition | + | - | - | + | - | + |
| Complex word repetition | + | - | - | + | - | + |
| Nonword repetition | + | - | - | - | - | + |
| Digit span | + | - | - | - | - | + |
| Sentence repetition | + | - | - | + | - | + |
| **Reading:** | - | + | - | + | + | - |
| Reading words | - | + | - | + | + | - |
| Reading complex words | - | + | - | + | + | - |
| Reading function words | - | + | - | + | + | - |
| Reading nonwords | - | + | - | - | + | - |
| **Spoken picture description:** | + | - | - | + | + | - |
| Total number of words | + | - | - | + | + | - |
| Number of appropriate information-carrying words | + | - | - | + | + | - |
| Number of inappropriate information-carrying words | + | - | - | + | + | - |
| Grammatical well-formedness | + | - | - | + | + | - |
| Syntactic variety | + | - | - | + | + | - |
| Speed | + | - | - | + | + | - |
| **Spoken comprehension:** | - | - | + | + | + | + |
| Spoken word-to-picture | - | - | + | + | + | + |
| Spoken sentence-to-picture | - | - | + | + | + | + |
| Spoken paragraph | - | - | + | + | - | + |
| **Written comprehension:** | - | + | - | + | + | - |
| Written word-to-picture | - | + | - | + | + | - |
| Written sentence-to-picture | - | + | - | + | + | - |

**Legend:** SS = initial speaking severity. RS = initial reading severity. US initial understanding severity. SM = semantic memory. Vis = vision status. Hea = hearing status. **+** indicates the covariate was included in that task analysis. – indicates the covariate was not included in that task analysis. All analyses factored out: sex, handedness, education, time post-stroke of CAT, left hemisphere lesion size, right hemisphere lesion size, and dyslexia status.

Bayesian statistics were used to validate the results further by estimating evidence for or against an effect of age on overall language ability, and each summary skill, task and sub-score. This was done by first conducting a multiple linear regression for each outcome with all the relevant covariates, but not age, to obtain the standardised residuals, and then regressing these residuals with age. The Bayes Factors were then converted into Log BF. A negative log BF indicates evidence for no effect of age, whereas a positive log BF indicates evidence for an effect of age. The strength of evidence can also range from ‘anecdotal’ to ‘decisive’ (Jeffreys, 1961), see Table 4.5 legend for categories of evidence strength.

**B. How does the effect of age at stroke depend on other variables?**

Moderator analyses were conducted to investigate potential interactions between age and five other variables of interest, on overall language ability. The interaction terms were created by multiplying ‘age at stroke’ (the regressor of interest) with each variable: left hemisphere lesion size, right hemisphere lesion size, initial severity, education amount, and time post-stroke. Two-way interaction terms were added into a second block of the model; three-way interaction terms were added into a third block of the model (along with the relevant two-way interaction terms) and four-way interaction terms were added into a fourth block of the model (along with the relevant two-way and three-way interaction terms). The first block of the model contained the same covariates as the main analyses. These analyses were conducted first for overall language ability, then repeated as post-hoc tests for each of the six summary scores.

Post hoc analyses of significant interactions were conducted by testing and illustrating the effect of age in different subgroups of participants. The subgroups were created by systematically categorising participants, first by age, then by initial severity, and finally by left hemisphere lesion size. This process proceeded as follows:

First, two age groups were created by categorising participants as ‘younger’ (≤57.88 years) or ‘older’ (≥57.93 years) according to the full sample median. Next, for initial severity, the four severity categories were used (Normal, Mild, Moderate, and Severe). Finally, for left hemisphere lesion size, two approaches were taken. First, participants were allocated to one of two lesion size subgroups: ‘smaller’ (0-3.33cm$^3$) or ‘larger’ (≥9cm$^3$). These group thresholds were chosen in two steps: first by determining the median lesion size of the full sample (3.34cm$^3$). This resulted in 8 subgroups each within the ‘smaller’ and ‘larger’ lesion size categories (i.e. 4 severity subgroups and 2 age subgroups). Second, participants with the largest and smallest lesions, and those who were assessed very late post-stroke were removed, in order to ensure that each ‘younger’ and ‘older’ subgroup pair were matched for lesion size and time post-stroke, and so any disadvantage of older age could not be attributed to these participants having larger lesions or shorter time post-stroke compared to their younger counterparts. This resulted in the exclusion of 177 participants from the ‘larger’ groups, leaving 374 participants in the ‘smaller’ groups and 198 participants in the ‘larger’ groups (Figure 4.1).

The second lesion size approach allocated all participants from the full sample of 749 into one of three lesion size subgroups: <1cm$^3$, 1-50cm$^3$, and ≥50cm$^3$. This resulted in 24 (smaller) subgroups (2 age, 4 severity, and 3 lesion size) – and so to increase the subgroup sizes, participants with Moderate initial severity were combined with participants with Severe initial severity, reducing the number to 18 subgroups. Then once again, to match the younger and older subgroup pairs for lesion size, 29 participants with the largest and smallest lesions were removed (Table 4.4 and Figure 4.3).

**Table 4.4** Mean and range of Left hemisphere lesion size (cm$^3$) within each lesion and severity group, in 720 participants.

| Lesion group | Initial severity | Younger | | | Older | | |
|---|---|---|---|---|---|---|---|
| | | N | Mean | Range | N | Mean | Range |
| LH: <1cm$^3$ | Normal | 48 | 0.036 | 0 to 0.848 | 51 | 0.035 | 0 to 0.89 |
| | Mild | 75 | 0.054 | 0 to 0.992 | 56 | 0.053 | 0 to 0.85 |
| | Severe | 46 | 0.087 | 0 to 0.960 | 22 | 0.080 | 0 to 0.94 |
| LH: 1 to 50cm$^3$ | Normal | 11 | 7.70 | 1.42 to 46.86 | 32 | 7.68 | 1.02 to 31.22 |
| | Mild | 35 | 9.54 | 1.17 to 48.62 | 49 | 9.28 | 1.06 to 31.50 |
| | Severe | 74 | 16.82 | 1.21 to 44.34 | 81 | 16.70 | 1.10 to 44.43 |
| LH: >50cm$^3$ | Normal | n/a | n/a | n/a | n/a | n/a | n/a |
| | Mild | 7 | 96.39 | 62.16 to 182.52 | 8 | 86.84 | 51.11 to 171.08 |
| | Severe | 66 | 117.70 | 50.08 to 225.34 | 59 | 112.51 | 50.49 to 235.12 |

**Legend:** Lesion size groups were determined by the size of the left hemisphere lesion, irrespective of the size of the right hemisphere lesion. ‘Severe’ includes participants with Moderate initial severity. The age cutoff was determined using the median of the full sample: Younger = up to 57.88 years; Older = 57.93 years and above. 29 participants were removed from the full sample to match the groups for left hemisphere lesion size.

### C. Does age at stroke influence the degree of recovery?

Two methods were used to investigate the relationship between age and recovery.

**CAT scores:** An effect of age on recovery was expected to result in a significant interaction between age and time post-stroke (as investigated in section B above), with post hoc tests showing that, as time post-stroke increased, the disadvantage of older age worsened. However, as the CAT scores were mostly acquired years post-stroke, this analysis will not be sensitive to the influence of age on early recovery (i.e. in the first year post stroke).

**Self-rated scores:** A proportional ‘recovery score’ was calculated for each participant by dividing [actual improvement at one year post-stroke] by [improvement potential]. Actual improvement is the number of ability levels by which a participant increases between one week and one year (Hope et al., 2019, Kim et al., 2019). For example, a participant with Severe initial severity who improved to Mild by one year improved by 2 severity levels (severe to moderate and moderate to mild). Actual recovery (or absolute recovery) does not distinguish participants who were recovering from severe or mild impairments. In contrast, improvement potential is the total number of severity levels available to increase by. This is maximum for participants with Severe initial symptoms who can move up three levels (severe to moderate, moderate to mild and mild to normal). Conversely, those with mild initial symptoms can only move up one level (mild to normal). The proportional recovery score is the proportion of the maximum recovery available (see Chapter 2 section 2.5.3). For example, a participant who improved by 2 out of 3 possible performance increases would receive a recovery score of 0.66. This method of measuring recovery score enables all participants to be analysed together, regardless of their initial severity, but has limitations that are discussed below (and in Chapter 2). The analysis was repeated for each of the four language skills, resulting in each participant having a recovery score for speaking, understanding, reading and writing. The mean of these recovery scores was also calculated for each participant. Multiple linear regression was used to investigate the relationship between age and recovery score, when left and right hemisphere lesion size, education, sex, and handedness were factored out.

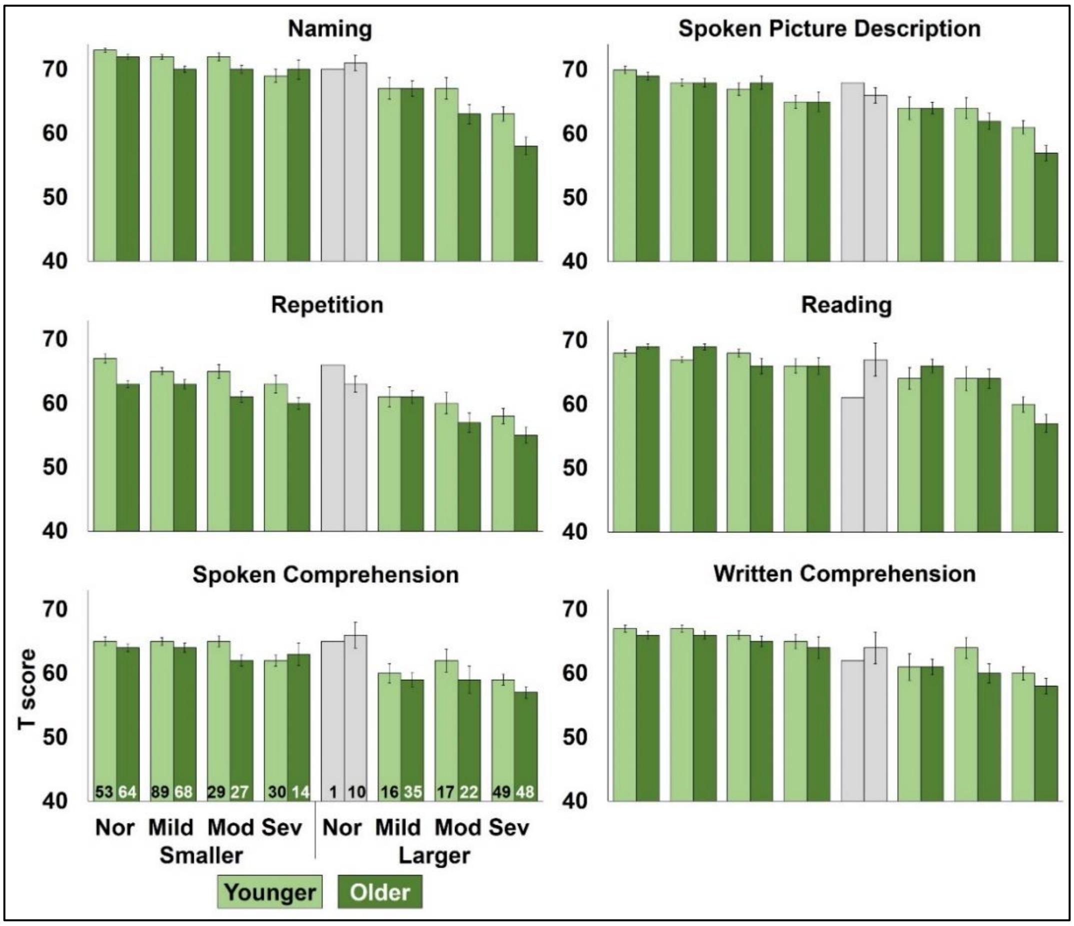
## 4.4. Results

### A. Which language scores are most sensitive to age at stroke?

Across 749 participants, there was a significant, negative effect of age on overall language ability, after factoring out the effect of other variables. As expected, scores were significantly lower for participants who were older than younger at stroke. Figure 4.1 and Table 4.5 and show how this disadvantage of older age was observed for five of the six summary scores: Naming, Spoken Picture Description, Repetition, and Spoken and Written Comprehension (but not Reading). Bayesian evidence was ‘Decisive’ for Naming, Repetition and Spoken Comprehension; and ‘Anecdotal’ for Written Comprehension (Table 4.5).

For each of the tasks that contributed to the summary scores, Table 4.5 provides (i) the $R^2$, β, and p values from separate regression analyses, and (ii) the Log Bayes evidence from the Bayesian regression, when other variables were factored out of the analyses. Figure 4.2 shows the amount of variance explained independently by age for each task – which was relatively small, compared to the amount collectively explained by the other covariates. The effect of age was strongest for: (1) semantic fluency within the naming score, (2) nonword repetition within the repetition score, (3) spoken word-to-picture matching within the spoken comprehension score, (4) written sentence-to-picture matching within the written comprehension score, and (5) the number of inappropriate words produced during spoken picture description (i.e. more inappropriate words were produced by older participants).

**Figure 4.1** The effect of age in 572 participants, grouped according to left hemisphere lesion size and initial symptom severity, on 6 summary scores from the CAT.



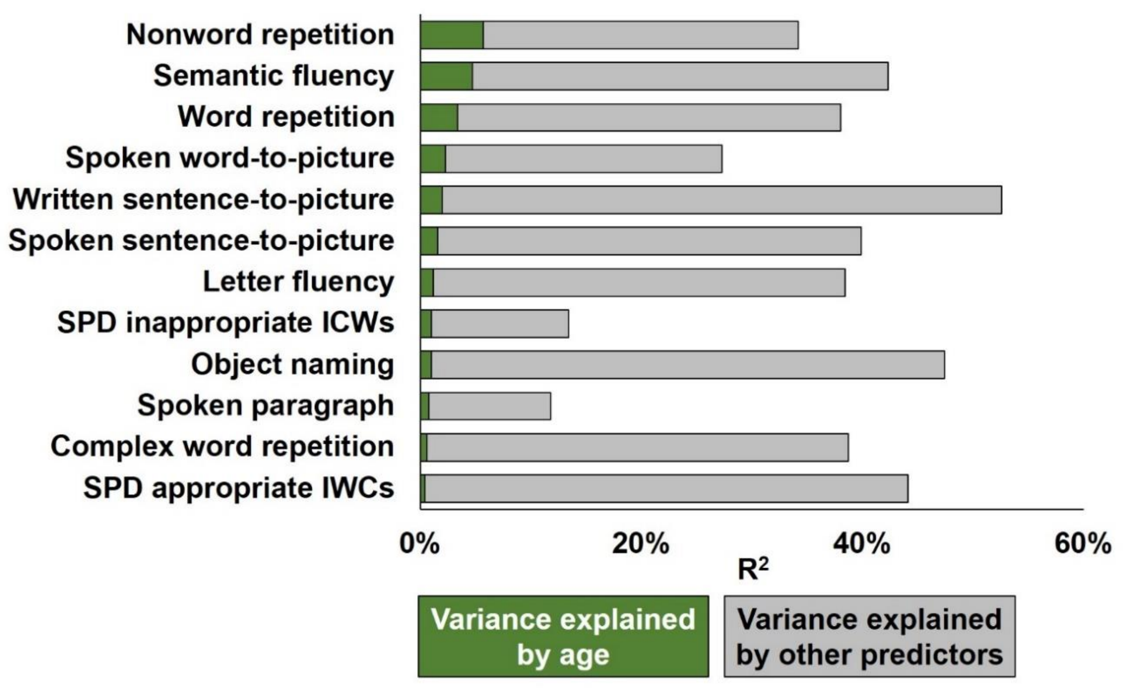
**Legend:** T scores for each of the six CAT summary scores, grouped by age, left hemisphere lesion size, and initial speaking severity. The colours indicate younger (light green; ≤57.88 years) and older (dark green; ≥57.93 years), the threshold for which was set using the full sample median. Grey bars indicate where the group comprised 10 or fewer participants (i.e. not well-sampled). Nor, Mild, Mod and Sev refer to the four severity levels (Normal, mild aphasia, moderate aphasia or severe aphasia. 'Smaller' (first 8 bars) refers to participants with smaller left hemisphere lesions (0 to 3cm$^3$). 'Larger' (last 8 bars) refers to participants with larger left hemisphere lesions (more than 9cm$^3$). The number of participants in each group is shown in the plots for spoken comprehension (bottom left) and is the same for all summary scores. The total number of participants illustrated is 572, after 177 were excluded to match groups for left hemisphere lesion size and time post-stroke. Score thresholds indicating aphasia**:** <63 for Naming, <61 for Spoken picture description <60 for Repetition, <61 for Reading, <57 for Spoken comprehension, <60 for Written comprehension.

**Table 4.5** The effects of age on CAT scores.

| **Analysis type:** | | **Multiple linear regression** | | | **Bayesian regression** | |
|---|---|---|---|---|---|---|
| **Outcome measure** | **N** | **$R^2$** | **p value** | **β** | **Log BF** | **Evidence strength** |
| **Overall language ability** | 569 | 0.632 | <0.001 | -0.125 | 3.04 | Decisive |
| **Naming:** | 741 | 0.580 | <0.001* | -0.149 | 4.47 | Decisive |
| Semantic fluency | 748 | 0.423 | <0.001* | -0.234 | 9.16 | Decisive |
| Letter fluency | 748 | 0.384 | <0.001* | -0.119 | 1.14 | Strong |
| Object naming | 741 | 0.473 | <0.001* | -0.106 | 0.91 | Substantial |
| Action naming | 741 | 0.455 | ns | -0.049 | | |
| **Picture Description:** | 741 | 0.516 | 0.008 | -0.074 | -0.22 | - |
| Total number of words | 741 | 0.415 | ns | -0.055 | | |
| Number of appropriate ICWs | 741 | 0.441 | 0.026 | -0.067 | -0.61 | - |
| Number of inappropriate ICWs | 741 | 0.134 | 0.003 | 0.111 | 0.10 | Anecdotal |
| Grammatical well-formedness | 741 | 0.433 | ns | -0.002 | | |
| Syntactic variety | 741 | 0.418 | ns | -0.005 | | |
| Speed | 741 | 0.438 | ns | -0.055 | | |
| **Repetition:** | 743 | 0.478 | <0.001* | -0.169 | 4.59 | Decisive |
| Word repetition | 743 | 0.380 | <0.001* | -0.201 | 5.68 | Decisive |
| Complex word repetition | 743 | 0.387 | 0.008 | -0.083 | -0.25 | - |
| Nonword repetition | 743 | 0.342 | <0.001* | -0.258 | 9.81 | Decisive |
| Digit span | 744 | 0.391 | ns | -0.043 | | |
| Sentence repetition | 743 | 0.432 | ns | 0.027 | | |
| **Reading:** | 572 | 0.444 | ns | 0.029 | | |
| Reading words | 573 | 0.435 | ns | 0.027 | | |
| Reading complex words | 572 | 0.431 | ns | 0.027 | | |
| Reading function words | 572 | 0.188 | ns | -0.015 | | |
| Reading nonwords | 572 | 0.394 | ns | -0.007 | | |
| **Spoken Comprehension:** | 709 | 0.388 | <0.001* | -0.167 | 3.32 | Decisive |
| Spoken word-to-picture | 711 | 0.272 | <0.001* | -0.164 | 2.39 | Decisive |
| Spoken sentence-to-picture | 710 | 0.400 | <0.001* | -0.140 | 1.97 | Very strong |
| Spoken paragraph | 710 | 0.118 | 0.014 | -0.096 | -0.40 | - |
| **Written Comprehension:** | 573 | 0.525 | 0.001* | -0.099 | 0.49 | Anecdotal |
| Written word-to-picture | 574 | 0.339 | ns | 0.061 | | |
| Written sentence-to-picture | 573 | 0.526 | <0.001* | -0.151 | 3.01 | Decisive |

**Legend:** Effects of age on overall language ability, and summary and individual tasks scores from the CAT, using (i) multiple linear regression and (ii) Bayesian regression, when other variables were factored out. The full sample includes 749 participants, but the number in each analysis (N) varies due to missing data for certain covariates. ICWs = Information-carrying words. ns = not significant at a threshold of p=<0.05. BF = Bayes factor: Positive BF supports an effect of age. Evidence strength: >2 = decisive; 1.5-2 = very strong; 1-1.5 = strong; 0.5-1 = substantial; 0-0.5 = anecdotal (Jeffreys, 1961). Negative BF refutes the effect (empty cells indicate negative BF and non-significant p value). * = significant after correction for multiple comparisons (= 6 summary scores; 24 tasks/subscores).

**Figure 4.2** The variance in CAT task scores explained by (i) age at stroke and (ii) other variables.



**Legend:** The amount of variance explained by age (green portion of bar) and by all other covariates (grey portion of bar) for the 12 CAT tasks/sub-scores which were significantly influenced by age at stroke. SPD = spoken picture description. ICW = information-carrying word. The remaining 12 CAT tasks/sub-scores (on which there was no significant effect of age) are not shown here.

### B. How does the effect of age at stroke depend on other variables?

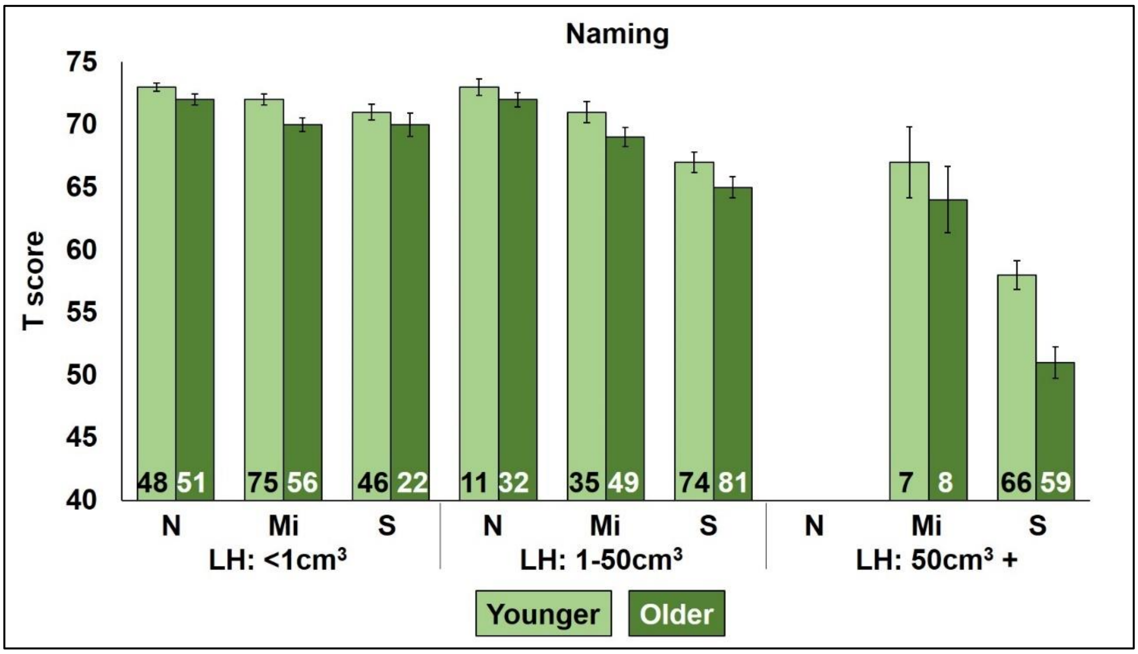
For all 6 summary scores, the effect of age was greater for participants with larger than smaller lesions. This is illustrated in Figure 4.1 and resulted in significant two-way interactions between age and lesion size that additionally interacted with initial severity (i.e. a three-way interaction) for all four speech production scores (Naming, Spoken picture description, Repetition and Reading), but not the comprehension scores. As illustrated in Figure 4.1, the effect of age is greater for larger lesions that were associated with severe or moderate initial symptoms than for larger lesions associated with normal or mild initial symptoms. Figure 4.3 examines the three-way interaction in more detail by (i) grouping the participants according to three different left hemisphere lesion size groups ($<1cm^3$, 1-50 $cm^3$, $>50$ $cm^3$) rather than the two lesion size groups in Figure 4.1 that were based on the mean left hemisphere lesion size for the whole sample (0-3.3 $cm^3$ / $>3.3$ $cm^3$), (ii) summing over participants with moderate and severe initial symptoms and (iii) focusing on naming because the interaction between age and initial severity was strongest for naming (see

Table **4.6**). As can be seen clearly in Figure 4.3, the effect of age was greater in participants who had lesions larger than 50cm$^3$ than lesions that were 1-50 cm$^3$ – and this effect was greater when participants had severe/moderate initial symptoms than normal or mild initial symptoms.

The effect of age also interacted with education (see

Table **4.6**) with a significant four-way interaction between age, education, left hemisphere lesion size, and initial severity for Naming and Picture Description. These four-way interactions arose because, in participants with large lesions (over 50cm$^3$), and severe/moderate initial symptoms, the effect of age was enhanced (see Figure 4.3) and the effect of education was reduced. This result is explored and discussed further in Experiment 3.

**Figure 4.3** The interaction between age, left hemisphere lesion size and initial severity on Naming in 720 participants.



**Legend:** Average T scores for the Naming summary score, for participants who were younger versus older, grouped by left hemisphere lesion size and initial speaking severity. As in Figure 4.1, younger/older are indicated by light/dark green, and were determined using the median of the full sample. Younger= ≤57.88 years; older = ≥57.93 years. N, Mi and S refer to three initial severity groups (Normal, Mild, Severe). Participants with Moderate initial severity were included in the Severe groups because of small sample sizes. LH <1cm$^3$ (first six columns) refers to participants with very small left hemisphere lesions (<1cm$^3$) and/or larger right hemisphere lesions. LH 1-50cm$^3$ (middle six columns) refers to participants with left hemisphere lesions that were between 1 and 50cm$^{3,}$ irrespective of right hemisphere damage. LH >50cm$^3$ (last four columns) refers to participants with left hemisphere lesions that were larger than 50cm$^3$, irrespective of right hemisphere damage. 29 participants were removed from the full sample to match each pair of age groups for left hemisphere lesion size.

**Table 4.6** Interactions between age and other variables on CAT summary scores.

| **CAT summary score:** | Two-way: Age & | | Three-way: Age & | Two-way: Age & | Four-way: Age & |
|---|---|---|---|---|---|
| | LH lesion size | Initial severity | LH lesion size & initial severity | Education | Education, LH lesion size, & initial severity |
| Overall language ability | <0.001 | ns | <0.001 | 0.049 | ns |
| Naming | <0.001* | 0.001* | <0.001* | 0.038 | 0.019 |
| Picture description | <0.001* | ns | <0.001* | ns | 0.030 |
| Repetition | 0.002* | ns | 0.005* | 0.050 | ns |
| Reading | <0.001* | ns | 0.025 | 0.020 | ns |
| Spoken comprehension | <0.001* | ns | ns | ns | ns |
| Written comprehension | <0.017 | ns | ns | ns | ns |

**Legend**: p values for the two- and three-way interactions between age, left hemisphere lesion size and initial severity, the two-way interaction between age and education, and the four-way interaction between age, education, left hemisphere lesion size and initial severity, in 749 participants, for overall language ability, and for each CAT summary score. Due to missing data for some variables, 180 participants were excluded from the overall language ability analysis; 8 were excluded from the Naming and Picture description analyses; 6 were excluded from the Repetition analysis; 177 and 176 were excluded from the Reading and Written comprehension analyses respectively (primarily because of missing data for initial reading severity), and 40 were excluded from the Spoken comprehension analysis. ns = not significant at an uncorrected threshold of p=>0.05. * = significant after correction for multiple comparisons (= 6 summary scores). The interactions with education are also reported and discussed in Experiment 3.

### C. The effect of age at stroke on the degree of recovery

**CAT scores:** Evidence of recovery was observed, as a significant and positive effect of time post-stroke on overall language ability (p=0.002, β=0.088). However, there were no significant interactions between age and time post-stroke, and therefore no evidence that age influenced the degree of recovery when initial severity was factored out.

**Self-rated scores:** No effect of age was observed on recovery score for any of the four self-rated language skills (speaking, understanding, reading and writing), in 559 participants with data for initial aphasia severity, and after controlling for left and right hemisphere lesion size, amount of education, sex and handedness.

## 4.5. Discussion

Using data from 749 stroke survivors, and controlling for a range of stroke and non-stroke variables that influence aphasia outcomes and recovery, the results from this study demonstrate that older age at stroke is associated with worse language scores on a range of tasks. Whilst the effects of age are reasonably small, these results are novel because no prior studies have investigated how the effect of age varies with task, or other predictors of outcome. The discussion that follows is centred around the three study questions: (A) Which CAT tasks are most sensitive to age at stroke, and why? (B) Does the effect of age at stroke depend on other variables? and (C) Does age at stroke affect the degree of recovery? The clinical implications of these findings are then discussed, along with study limitations and future directions.

### 4.5.1. A. Which language tasks are most sensitive to age at stroke, and why?

Within the Naming score, the disadvantage of older age was greatest for semantic and letter fluency, and object naming. During Picture description, the disadvantage of older age was greatest for the number of inappropriate words produced. Within the Repetition score, the disadvantage of older age was greatest for nonword and word repetition. For the Comprehension tasks, the disadvantage of older age was greatest for spoken word- and sentence-to-picture matching and written sentence-to-picture matching. Finally, across tasks, the three strongest effects of ageing were observed, as hypothesized, on expressive language tasks (Nonword repetition, Semantic fluency, and Word repetition), see Figure 4.2. Contrary to the study hypotheses (see the study Introduction), however, the results show (1) surprisingly strong effects of older age on the comprehension tasks, which is inconsistent with the conclusions of Cohen et al. (2019); and (2) the effects of older age were greater on comprehension tasks than Object naming, Spoken picture description and Letter fluency (Figure 4.2); whereas the expectation was that ageing effects would be stronger for word retrieval tasks, such as Object naming, Picture description, and Fluency (Kavé and Goral, 2017) than receptive tasks (Cohen et al., 2019).

The influence of age may vary with task because of its varying impact on the underlying cognitive functions that are required for those tasks. Below, each of these cognitive functions is addressed in turn. First, executive functions (e.g. working memory, attention, cognitive control and flexible thinking) are essential for attending to and remembering task instructions. Reduced executive functioning amongst healthy older people (Idowu and Szameitat, 2023, Cohen et al., 2019) could therefore partly explain the consistent disadvantage of older age observed in most groups of participants in the current study (including those who were never aphasic) – if, for example, their lower scores resulted from taking longer to recall the task demands, process the stimuli, consider their response options, and/or provide a response in the correct modality. Furthermore, some tasks rely on these skills more than others, potentially explaining why the disadvantage of age was stronger. For example, spoken picture description places a particularly high load on executive functions because participants require visual working memory to hold the contents of the picture online, attention to form and organise ideas and to systematically retrieve speech for one minute, and working memory and cognitive control to inhibit repeating parts of the description that they had already produced. Likewise, semantic and letter fluency require flexible thinking to generate a variety of words within a specific category, and working memory and cognitive control to inhibit the retrieval of words that have already been produced.

A second function impacted by age is sensory skills (Chia et al., 2006, World Health Organisation, 2019, World Health Organisation, 2021). Age-related hearing and vision loss may impact performance on tasks involving perception of auditory stimuli (e.g. repetition, spoken word- and sentence-to-picture matching) or visual stimuli (e.g. all comprehension tasks, picture description and naming). In particular, item-specific effects of age may be apparent: for example, when distinguishing between words like *‘faith’* and *‘face’* in word repetition, because perception of high-frequency phonemes like ‘th’ and ‘s’ can deteriorate with older age (Bance, 2007). The current study attempted to control for age-related sensory loss by factoring ‘self-reported’ hearing and vision abilities out of the analyses. However, these self-reports would not be sensitive to high level sensory loss which participants might not have been aware of. Performance on the comprehension tasks used in the current study relies heavily on being able to

(i) hear the spoken speech, and (ii) match the speech to multi-component pictures. Therefore, it is plausible that the surprisingly strong effects of ageing on comprehension tasks, illustrated in Figure 4.2, could be the consequence of age-related sensory loss.

High level hearing loss could also reduce the efficiency of auditory working memory which has been shown to deteriorate with age (Chevalère et al., 2020, Biran et al., 2023). This would partly explain why nonword repetition and spoken sentence comprehension deteriorates with age because both these tasks place high demands on auditory/phonological working memory (Chevalère et al., 2020, Biran et al., 2023).

Third, language skills, over and above other cognitive processing involved in language tasks, are influenced by age. Whilst vocabulary size has been shown to be typically larger among older compared to younger adults (Verhaeghen, 2003, Keuleers et al., 2015), expressive language, in particular word retrieval, is more likely to decline with older age (Shafto and Tyler, 2014, Biran et al., 2023); as was observed for naming in the current study. The effect of age on naming ability may arise from increased word-finding failures and an increased number of ‘tip-of-the-tongue’ states amongst older compared to younger adults (Burke and Shafto, 2004, Facal et al., 2012). This could also explain the greater number of incorrect words that the current study’s older compared to younger participants produced during spoken picture description.

Finally, the slower overall processing speed known to affect older adults (Faroqi-Shah and Gehman, 2021, Lemke and Zimprich, 2005) may result in slower comprehension of the task stimuli and decision times, which could result in lower scores, for speaking and comprehension tasks, when (correct) responses are recorded as ‘delayed’ (longer than 5 seconds for most CAT tasks).

### 4.5.2. B. How the effect of age at stroke depends on other variables

The disadvantage of older age was observed to be robust across a range of left hemisphere lesion sizes and initial severity. Such effects could be attributed to normal healthy ageing. Nonetheless, as hypothesised (see study Introduction), age interacted significantly with left hemisphere lesion size on overall language ability and each of the summary scores; and with initial severity and lesion size on the four speech production scores (but not comprehension). These interactions arose because the disadvantage of older age at stroke was worse in participants with large left hemisphere lesions and severe initial symptoms.

Prior literature, reviewed in the Introduction of this chapter, supports an interpretation of these age effects in terms of accelerated ‘brain age’ (Kristinsson et al., 2023, Brady et al., 2022) which reduces recovery capacity. Larger lesions are more likely to cause substantial damage to the neural networks that have the potential to support language functions, leaving less residual neural capacity available for recovery of lost functions. More severe initial symptoms mean more relearning is needed and therefore more residual capacity is needed. Additionally, severe aphasia can aggravate the existing challenges of older age (e.g. word retrieval skills), and introduce new impairments (e.g. comprehension, semantic and/or syntactic abilities) which are not typically experienced in normal healthy ageing (Cohen et al., 2019, Shafto and Tyler, 2014, Burke and Mackay, 1997).

The disadvantages of older age on other (non-language) stroke and overall health outcomes may also contribute to reduced language recovery. For example, reduced functional independence and poorer motor outcomes may result in older patients being less physically able to participate in stroke support groups compared to younger patients.

Finally, as hypothesized from prior literature (Aki Ö et al., 2022, Lubrini et al., 2022), the effect of age interacted with the effect of education. This emerged in a significant four-way interaction between age, left hemisphere lesion size, initial severity and education on Naming and Picture description. In brief, the disadvantage of older age, overrode the benefit of more education, plausibly because a lack of residual neural networks impairs the use of cognitive strategies that are otherwise more efficient in those with more education. Critically, these results support the findings of (Aki Ö et al., 2022); who found that older age reduced the benefit of more education on phonemic fluency, however, the current study found no evidence to support the finding that the disadvantage of older age was worse in participants with less education (Lubrini et al., 2022).

### 4.5.3. C. The effect of age at stroke on the degree of recovery

Despite hypothesizing that older age at stroke would hinder language recovery when the influence of a wide range of variables was controlled, the current study found no evidence that older age at stroke significantly disadvantages language recovery (measured as the difference between initial severity and (i) long-term CAT scores (years post-stroke) and (ii) self-reported abilities at one year post-stroke). These null effects add to 9/17 prior studies that also failed to find an effect of age on language recovery (see Table 4.1).

Nonetheless, the current study suggests that older participants with large lesions and severe initial symptoms performed worse (particularly on Naming and Picture description) than all other groups. In other words, recovery was limited in these participants. A similar observation was made by Pickersgill and Lincoln (1983), who suggested that recovery is worse in older patients with severe, compared to moderate, aphasia. Irrespective of age, larger lesions are more likely to damage regions that are critical to language learning such as the inferior frontal gyrus (Navarrete-Orejudo et al., 2023). Likewise, irrespective of age, severe initial symptoms exacerbate the loss of learning ability (Dignam et al., 2016) in people with aphasia (Kelly and Armstrong, 2009, Tuomiranta et al., 2011, Tuomiranta et al., 2014a, Peñaloza et al., 2022) and this may impede the response to language therapy (Tuomiranta et al., 2014b, Dignam et al., 2016). On top of the loss of learning capacity that affects younger and older participants, word learning is further reduced in older compared to younger adults, potentially because the encoding and consolidation processes required for word learning are less effective with older age, along with other declines in sensory and processing abilities (Blachstein and Vakil, 2016, Shing et al., 2008, Craik and Rose, 2012, Craik and Bialystok, 2006) that are discussed above.

Furthermore, learning ability relies on a number of cognitive and meta-cognitive processes, which, as discussed in section A, can also be eroded with older age (e.g. working memory, attention, cognitive flexibility, use of learning strategies, ability to plan and structure learning).

A final consideration is that older age is inherently linked to shorter remaining lifespan, as well as being associated with increased risk of comorbid health conditions. People who are older when they have a stroke may simply have less time available to consolidate any new learning (Peñaloza et al., 2022), and/or have more health complications that prevent or interfere with language recovery.

In summary, the impact of the most debilitating stroke and aphasia characteristics, exacerbated by the challenge of older age, appears to severely reduce word-learning ability, and therefore language recovery; potentially explaining some of the heterogeneity in response to therapy often observed in therapy trials (Brady et al., 2022).

### 4.5.4. Clinical implications

These findings could influence how aphasia is diagnosed and managed. The CAT appears to be sensitive to age irrespective of language impairments (because the disadvantage of older age at stroke was observed in participants both with and without aphasia). Interpreting the severity of assessments scores should therefore be made with reference to an individual's age at stroke (as has been suggested for the Boston Diagnostic Aphasia Examination, and the Western Aphasia Battery (Rosselli et al., 1990, Kim and Na, 2004). Understanding how the effects of age vary across different tasks could also be crucial for directing further assessment, and for differential diagnosis of language impairments as distinct from other cognitive and sensory impairments. The efficacy of language therapy is also augmented when it incorporates strategies to support other non-language cognitive and/or sensory impairments.

For participants in whom the effect of age was biggest – those with large left hemisphere lesions and severe initial symptoms - the implications for therapy may be different. Needless to say, older age should never preclude any individual from receiving therapy. However, if recovery of impairments is reduced in participants who were older at stroke, expectations for the speed of recovery could be modified and therapy content and approach could have a more functional and compensatory focus, targeting functional communication strategies, use of alternative and augmentative communication (AAC), and communication partner training. The intention is not to deprioritise these patients, nor reduce the quantity or quality of therapeutic input that they receive. Rather it is to manage expectations and allocate resources most appropriately. For example, a fatiguing and time-intensive impairment-based therapy may be inappropriate for an older participant if it could cause detrimental effects (e.g. fatigue and low motivation if observable improvements are not made); whereas a less intensive therapy may be more impactful. Conversely, if recovery of impairments is more likely in patients who were younger when they had their stroke, a restorative therapy approach may be more beneficial. Such recommendations do not deny or undermine other ways that therapy goalsetting is already adjusted for patient age; for example, older patients who are retired, may want to focus on social communication, hobbies, and activities of daily living etc., whereas younger patients may be more focused on returning to work or supporting their families. Altogether, consideration of age when setting future targets would support a more personalised ‘precision rehabilitation’ approach to aphasia therapy (Brady et al., 2022).

### 4.5.5. Study limitations and Future directions

Despite its large sample size, which allowed for control of many factors known to influence language outcomes and recovery, this study had limitations.

First, the retrospective measurement of initial severity was dependent on participants’ memories, and therefore subject to inaccuracy or bias. For example, a positive recovery could prompt a more positive initial severity rating, whereas a negative recovery could prompt a worse initial severity rating. To mitigate this, the current study excluded participants with vague reports, or whose reports conflicted with either that of a carer, or with their CAT scores (as described in Chapter 2, section 2.5).

A second limitation is the potential impact of missing participant-reported severity scores. 173 participants had missing initial reading or understanding abilities, and were excluded from the analyses that included these covariates. If this data was missing because the participants’ abilities were too severe at one week, then their exclusion could bias the sample towards those with milder abilities. However, it is more likely that the data were missing due to participants not being able to remember their early abilities. Conversely, missing CAT data are more likely to occur if participants were unable to complete a task due to aphasia severity. Only 3 participants in the sample had missing CAT data, and so the impact of their exclusion from the analyses is relatively small, and unlikely to inflate the effect of age.

Third, hearing and vision abilities were measured via self-reports. Future studies could instead use instrumental measurements (e.g. audiometry) to detect subtle age-related sensory changes which might be undetected by an individual, but which could influence language scores in participants both with and without aphasia.

The fourth point relates to the inclusion and investigation of other variables in relation to chronological age. Many variables with a potential influence on aphasia outcomes and recovery could not be controlled because the data were not available. These include: cognitive abilities, medical factors (e.g. diabetes, hypertension), speech and language therapy, and ‘brain age’. Brain age, calculated in Busby et al. (2023) from non-lesioned brain tissue, has been shown to explain more variance in aphasia outcomes than chronological age (Kristinsson et al., 2022, Busby et al., 2023). A measurement of brain age may give valuable insight into the effects of pre-clinical dementia or other age-related degeneration on post-stroke language outcomes.

Even though no formal measure of ‘brain age’ was included in this experiment, there is a risk that a biomarker of older brain age could be misidentified as lesion. For example, atrophied regions or enlarged ventricles could be erroneously detected as ‘abnormal’, and designated as ‘lesioned’ voxels during the lesion identification process – when they in fact reflect changes that are associated with older chronological age (Scahill et al., 2003). If this were the case, further work would need to be conducted to dissociate the effect of chronological age from that of brain age. However, this risk is mitigated during the lesion identification process, because older subjects were included in the 64 neurologically-normal subjects from which the normative structural data were derived. This ensures that normal age-related structural changes such as atrophy are recognised as ‘normal’, and not misidentified as lesion. To further minimise the risk of misidentifying age-related changes as lesion, all scans and lesion identifications were checked by eye by a neurologist, as recommended in Seghier et al. (2008).

Inclusion of all these other variables has two additional benefits: first, correlations between chronological age and other variables (such as brain age or non-verbal cognitive function) can be determined (as was found by (Johnson et al., 2022)). Second, investigations can be conducted to assess how the effect of age depends on other variables effects (as was done in the current study). Future studies are therefore needed to investigate the independent contributions of different measures of ageing and brain health, and how they interact with each other.

A fifth limitation was the measurement of recovery. The two approaches used in this study may not have been sufficiently sensitive to age effects. Specifically, the first approach, which measured the difference between initial severity and CAT scores, may not be sensitive to an effect of age on early recovery, because the CAT scores were typically obtained years post-stroke. The second method, which used participant reports of language ability to measure change during the first year post-stroke, may be too subjective to detect age effects – for example, if participants rate their abilities in relation to their age. In other words, the meaning of 'Normal' may shift for an individual over time, to account for age-related changes. The limits of the self-rated measurement itself, with only 4 ability categories and limited detail about specific language impairments - may also result in insufficient statistical variance to detect changes in abilities. Additionally, by using a proportion to measure actual recovery compared to potential recovery, there could be bias towards participants with Mild initial severity - because less improvement is required for them to recover to normal (compared to those with Severe initial severity, in whom more improvement was required for full recovery to normal). Future studies could also investigate the effect of age on recovery of functional communication abilities, in addition to recovery of impairments.

Finally, the median age at stroke (57 years) in this study's sample was notably younger than the average age at stroke in the UK - which was 71 years in 2016 (Evans, 2018). The 'younger' and 'older' threshold based on this median and used for this study was chosen for illustrative purposes, and not for any clinical labelling of individuals. Indeed there is no definitive age at which a person is considered 'old', and the health and social care system rightly takes more aspects of an individual into account than just their chronological age, that might more accurately reflect their care needs, such as frailty (Skills for Health, 2018). That said, it is striking that the age range within the 'older' group in the current study was reasonably young, and so there is a need to further investigate how age influences outcomes and recovery amongst the 'older old'.

### 4.5.6. Summary and Conclusions

The goals of this study were to understand whether (1) the clinical management of aphasia might be influenced by age at stroke, and (2) prognostic models for aphasia outcomes and recovery could be improved by the inclusion of age at stroke. The results showed that participants who were older when they had a stroke performed worse on a range of language tasks from the Comprehensive Aphasia Test compared to participants who were younger when they had a stroke. This disadvantage of older age was found to prevail independent of left hemisphere lesion size and initial severity, but was noticeably worse amongst participants with large left hemisphere lesions and severe initial aphasia, suggesting these participants experienced slower/limited recovery of language abilities. In support for goal (1), these results may inform further assessment to guide differential diagnosis of language from age-related cognitive and sensory impairments, and may guide therapy planning. Regarding goal (2), the predictive value of age at stroke (independent to that of lesion size and initial severity) motivates its inclusion in prognostic models of aphasia outcomes and recovery.

# Chapter 5: Experiment 3. The effect of amount of pre-stroke formal education on language outcomes and recovery after stroke

## 5.1 Summary

**Introduction:** The goals of this study were to establish (A) which language tasks from a standardised aphasia assessment were most sensitive to the amount of formal education participants received prior to their stroke; and whether any benefit of this education (B) depends on other stroke and non-stroke-related variables, and/or (C) facilitates recovery from post-stroke aphasia.

**Methods:** The participants were 749 stroke survivors, who ranged from one month to 42 years post-stroke. Their post-stroke language abilities were assessed using (i) a wide range of language task scores from the CAT, and (ii) self-reported ratings of language ability at one week and one year post-stroke. The predictor variable of interest was the number of years of pre-stroke education, established in questionnaires that also recorded age at stroke, handedness, sex, and presence of developmental dyslexia. Left and right hemisphere lesion size was established from structural MRI brain scans. Multiple regression analyses investigated the effect of education amount, and its interaction with other variables, on (1) naming, repetition, reading, spoken picture description, spoken comprehension and written comprehension abilities, and (2) the degree of recovery (from one week to one year, and from one week to the time of the CAT assessment). Bayesian regressions were conducted to assess the strength of the evidence for any observed effects.

**Results:** (A) Across participants and tasks, more years of education were associated with greater overall language ability (β=0.075) with the most robust effects of education observed on semantic and letter fluency (β=0.123 and (β=0.166) and the number of words produced (β=0.085) and their grammatical well-formedness (β=0.087) during spoken picture description. (B) The benefit of more pre-stroke education was additive with the effects of other variables including initial severity of aphasia symptoms and left hemisphere lesion size, except that the benefit of education was reduced in older participants who had large lesions with severe initial symptoms. Finally, (C) whilst language ability improved over time, there was no effect of education on the speed of language recovery, measured using CAT scores or participant reports.

**Conclusion:** More pre-stroke education is associated with higher post-stroke language scores on a wide range of tasks, but this advantage is diminished by older age in participants with large lesions that cause severe aphasia. These

results suggest that diagnostic thresholds for aphasia should consider whether borderline aphasic language scores indicate fewer years of education rather than the presence of aphasia.

## 5.2 Introduction

This study investigates how the amount of pre-stroke education a person received influences post-stroke speech and language ability (measured months-to-years after a stroke) and recovery. Specifically, it investigates (A) which language task scores are influenced by amount of education, (B) whether an effect of education on language abilities depends on other variables, such as presence or absence of initial aphasia and the severity of initial symptoms, lesion size and age, and (C) whether education influences language recovery. The goals of the study were to understand whether (1) the clinical diagnosis and/or management of aphasia might be influenced by educational differences, and (2) future aphasia research could benefit from the inclusion of education as a prognostic factor for recovery.

Formal education refers to structured and organised education that is provided by schools, colleges, universities, or other educational institutions. It involves a planned curriculum, specific learning objectives, and a systematic approach to education delivered by trained teachers or instructors. Students in formal education are introduced to new knowledge, subjects, people, cultures, and experiences that may otherwise not be encountered, and which are accompanied by different vocabularies and language opportunities (Romeo, 2019). Further Education (which in the UK refers to optional formal schooling after the end of compulsory secondary education, typically after age 16) and Higher Education (which refers to tertiary education, such as university) involve deeper study with exposure to more specialised vocabularies, independent learning, critical and flexible thinking, as well as enhancing one’s existing cognitive, communication, and problem-solving skills. As described in the literature review below, the long-term influence of a rich language environment can manifest through better performance on tasks such as verbal fluency, object naming, and digit span, which require good lexical knowledge and retrieval, semantic memory, working memory and/or executive control. A wider vocabulary, syntactic variety and correct grammar may also result in higher scores on tasks that involve sentence comprehension and production.

The effect of education on language abilities after stroke is expected to be more complex than that observed in neurotypical people. One possibility is that post-stroke aphasia outcome and/or recovery will be better in people who received more education and have more efficient cognitive strategies, particularly if these individuals also have higher socioeconomic status that may enable greater access to healthcare and therapies, which promote a faster recovery. A second possibility is that the benefit of more education on language abilities is lost when patients have extensive damage to the neural systems that normally support language. In this case, the effect of education on post-stroke language ability will depend on one or more of the multiple lesion- and non-lesion factors that are known to influence aphasia outcome and recovery.

To frame the study's approach and reasoning, prior evidence for and against an impact of education on language ability and change is first evaluated below for: (i) neurotypical populations, (ii) people with dementia, and (iii) stroke populations.

### 5.2.1 The effect of education in neurotypical populations

Borod et al. (1980) conducted one of the earliest studies into education effects on language ability, finding that neurotypical adults who completed more education performed better on a range of tasks from the Boston Diagnostic Aphasia Examination, including verbal agility, word repetition, oral sentence reading, spelling to dictation, and the graded naming test, but not animal naming or narrative writing. Numerous subsequent studies have reported similar findings: adults who received more compared to fewer years of education performed better on tests of naming, verbal fluency, spoken picture description and digit span (Neils et al., 1995, Roberts and Doucet, 2011, Soylu and Cangöz, 2018, Constantinidou et al., 2012, Ratcliff et al., 1998, Brucki and Rocha, 2004, Abdel Aziz et al., 2017, Rodríguez-Lorenzana et al., 2020, Van der Elst et al., 2006, Ardila and Rosselli, 1996, Zimmermann et al., 2015, Choi et al., 2014). In contrast, but as expected, there have been no reports of better scores in people who received fewer compared to more years of education.

The advantage of a greater amount of education on language test performance in the above studies was observed despite variation in how each study was conducted. For example, the way education was defined varied greatly, with 'more education' ranging from 'more than 5 years' in total (Avila et al., 2009) to 'more than 13 years' in total (Neils et al., 1995), and 'less education' ranging from 'illiterate' (Ratcliff et al., 1998, Brucki and Rocha, 2004) to 'fewer than 9 years' (Roberts and Doucet, 2011). There was also variance in the groups' ethnicities, languages spoken, education systems and the covariates considered.

The strength of the evidence for educational benefits on language scores has led to proposals that scores from clinical language assessments such as the Boston Naming Test (Zec et al., 2007) and verbal fluency should be adjusted for amount of education. In clinical practice, this would mean that diagnosis and management of aphasia would still consider an individual's scores, but in the context of the amount of education they completed (along with observations, informal assessments and patient/carer reports).

### 5.2.2 **The effect of education on dementia risk and cognitive decline**

Two systematic reviews, by Valenzuela and Sachdev (2006), which included 33 studies, and Sharp and Gatz (2011), which included 88 studies, found a significant and consistent relationship between more education and reduced dementia risk. Additionally, the Valenzuela and Sachdev review found a protective effect of higher occupational status and participation in mentally stimulating leisure activities. The larger review, by Sharp and Gatz, included many of the studies from the Valenzuela and Sachdev review; and found the relationship between more education and reduced dementia risk to be reported by 58% of studies. In addition, both reviews emphasise that, even when an effect of education amount was reported, it was not clear whether it was mediated by other uncontrolled factors (e.g. familial risk of dementia and socioeconomic status).

Other studies that were not included in either of the reviews above have also investigated the effect of education (amount or level) on dementia risk, again with inconsistent conclusions (Fritsch et al., 2005, Rantalainen et al., 2018, Osler et al., 2017) that may relate to variable sample sizes and the degree to which variance from other moderating factors was controlled (Rodriguez and Lachmann, 2020). The definition and measurement of education also varied across studies.

Regarding the effect of education on cognitive decline, five studies analysed data from the 1921 and 1936 Lothian Birth Cohort (Gow et al., 2008, Ritchie et al., 2016, Staff et al., 2018, Bourne et al., 2007, Gow et al., 2011) and a sixth study analysed the 1946 National Survey of Health and Development (Richards et al., 2004). Only two of these six studies (Staff et al., 2018, Richards et al., 2004) found that less education was associated with faster cognitive decline. However, many other studies have argued that education contributes to cognitive reserve (along with other factors such as social engagement, mentally stimulating pursuits, physical activity and genetics). Cognitive reserve describes a property of the brain that reduces an individual’s susceptibility to, or compensates for, pathology or injury, by utilising behavioural strategies, or engaging alternative neural networks (Stern et al., 2023). Greater cognitive reserve may delay the onset of cognitive decline (Meng and D’Arcy, 2012, Stern, 2009, Stern, 2012, Stern et al., 2020), or the detection of cognitive decline when premorbid abilities are very high.

### 5.2.3 The effect of education on speech and language scores after stroke

Very few studies have demonstrated a relationship between post-stroke language ability and years of education. In a sample of only 39 participants, Connor et al. (2001) reported that initial severity of aphasia was significantly greater for subjects with lower education and occupational status. A larger study with 173 stroke patients (González-Fernández et al. (2011) compared language performance in patients with more than, or less than, 12 years of education (the numbers of patients in each group and their range of educational level was not specified). After controlling for age, sex, lesion volume and socio-economic status, the group with more education had better reading scores (written word comprehension and oral reading) but not better auditory comprehension or naming scores. In contrast, other studies have reported an absence of significant effects of education on aphasia outcome. For example, a recent study of 147 participants by Johnson et al. (2022) reported that years of education was not significantly predictive of chronic aphasia severity when other demographic, lesion, and cognitive variables were factored out.

Given the benefit of cognitive reserve on delaying cognitive decline in patients with dementia (see above), it might also be the case that cognitive reserve protects against cognitive decline after stroke. However, several systematic reviews have shown very inconsistent evidence for this, and highlighted the need for more studies (Contador et al., 2023, Tao et al., 2023, Laura and Roza, 2022, Nunnari et al., 2014, Rosenich et al., 2020).

Three prior studies have investigated the effects of education on post-stroke recovery of speaking ability. One of these studies suggested that education level (measured as the highest stage completed, from none, primary, secondary or Higher education) was significantly associated with the change in score from one week to one year post-stroke in a cohort of 235 patients (Kim et al., 2019) after factoring out variance from age, initial severity and lesion size. Although this demonstrates a benefit of education on aphasia recovery, Kim et al. focused on many predictive factors in addition to education, and so did not explore the education effect in depth, nor determine how the effect of education was influenced by the effect of other variables. Moreover, language ability was measured using the Korean version of the Frenchay Aphasia Screening Test (K-FAST), which combines speaking, comprehension, reading and writing ability into one composite score, hindering any detailed understanding of which language skills are sensitive to educational level during the course of recovery. The other two studies found no significant effect of education on recovery, perhaps because they (a) used smaller sample sizes (39 patients in Connor et al. (2001) and 22 patients in Lazar et al. (2008)) and/or (b) measured change according to the Boston Diagnostic Aphasia Examination aphasia threshold, rather than measuring a change in raw scores for specific language tasks.

In summary, prior literature suggests that performance on a range of speech and language tasks is influenced by amount/level of education in neurotypical participants, participants with dementia and in participants with post-stroke aphasia. However, the influence of education on language performance varied with the participant group. For example, an effect of education on object naming was reported for neurotypical participants (Borod et al., 1980) but not for participants with post-stroke aphasia (González-Fernández et al., 2011). There is also limited evidence that education amount influences post-stroke language recovery (Kim et al., 2019).

### 5.2.4 **The current study**

The current study investigated these contradictory effects in the prior stroke literature by investigating the effect of amount of pre-stroke education on (i) language scores on a range of tasks, and (ii) language recovery, indicated by change in both participant-reported ability between one week and one year post-stroke, and long-term language scores compared to initial severity. It also controlled for the influence of many variables on language performance, including some that have not been considered in the prior studies (e.g. semantic memory, dyslexia status and discrete measurements of lesion size in the left and right hemisphere). A distinction is made between variables that are measured at the time of the stroke (e.g. amount of pre-stroke education), and variables that are measured later, at the time of assessment (e.g. time post-stroke).

This experimental design addressed the three study goals (A-C). Goal A investigated which language tasks were most sensitive to pre-stroke amount of education. Goal B investigated whether the effect of education on language score depends on other stroke and non-stroke factors, such that participants benefitted differently from their education amount (Reis and Petersson, 2003), or whether it was additive with the effect of stroke, such that all participants benefitted similarly from the effect of education regardless of other variables. Finally, goal C investigated how education amount affects language recovery by looking at both change in participant-reported aphasia severity from one week post-stroke to (1) participant-reported severity one year post-stroke and (2) objective language scores.

### 5.2.5 **Hypotheses:**

For Goal A the expectation was that more education would result in better scores for Reading and Written comprehension, but not for Naming, Repetition, or Spoken comprehension, as found by González-Fernández et al. (2011). A benefit of more education on picture description scores was also expected because this task rewards a wider vocabulary, syntactic variety and correct grammar, all of which are associated with education amount. The ‘open’ nature of the task may also enable participants to circumvent any aphasic difficulties. For example, multiple different words could be used to describe the picture, as opposed to Naming (a ‘closed’ task), where usually only one target answer is deemed correct. In this scenario, the cognitive flexibility and wider vocabulary required may also stem from more education. That said, the greater demands of this task (i.e. drawing on a wide range of linguistic and cognitive skills) will make it particularly difficult for participants with aphasia and, in this context, the greatest amount of variance in task performance will be accounted for by stroke factors (e.g. lesion size and location), potentially masking the effect of education.

For Goal B the expectation was that the effect of education would be stronger in participants who had normal post-stroke language abilities compared to those who were aphasic one week post-stroke, because aphasia was expected to interfere with, or mask the effect of education on language ability. Alternatively, if the effect of education is similar in participants both with and without initial aphasia, then either it is (a) additive with other factors (i.e. independent of the effects of lesion size, symptom severity, age, etc.), or (b) dependent on other factors (i.e. greater for some participants, and weaker for other participants) with this variance cancelling out when participants are pooled together.

For Goal C, a greater amount of education was expected to enhance recovery (Kim et al., 2019).

## 5.3 Methods

### 5.3.1 Participant selection

Participants were selected from the PLORAS database if they were native English speakers. They were excluded if they (a) had unknown initial speaking severity (n=212), (b) were medically unwell (e.g. in a coma) one week post-stroke, and therefore unable to attempt speech production nor rate their initial severity (n=145), (c) had inconsistent self-reported language ability compared to their objectively assessed language abilities (i.e. those rating themselves as 'Normal' had aphasic speech production according to CAT scores; n=12), or (d) had an unknown (n=6) or outlying amount (n=5; two with 16 years, two with 17 years, and one with 18 years) of pre-stroke education (see Chapter 2, section 2.6). These criteria identified 749 eligible participants, details for whom can be found in Table 5.1. A complete set of behavioural, demographic and lesion data was available for most of the 749 eligible participants. However, 173 participants had missing severity data from the self-reported reading and/or comprehension rating scales; 6 more had missing hearing or vision data, and 3 more had missing CAT scores for various tasks. Participants were excluded from the analyses that used the covariates for which they had missing data; see the description of statistical analyses in section 5.3.5 for more details. The impact of excluding these participants is considered in the study Discussion.

**Table 5.1** Participant characteristics

| | | **Less education** | **More education** |
|---|---|---|---|
| | Number | 413 | 336 |
| **Years of education after age 14** | Mean (SD) | 2 (1) | 8 (2) |
| | Range | 0 to 4 | 5 to 13 |
| **Initial speaking severity** | Severe (%) | 145 (35) | 98 (29) |
| | Moderate (%) | 59 (14) | 59 (18) |
| | Mild (%) | 131 (32) | 111 (33) |
| | Normal (%) | 78 (19) | 68 (20) |
| **Sex** | Male (%) | 287 (69) | 231 (69) |
| | Female (%) | 126 (31) | 105 (31) |
| **Pre-stroke handedness** | Right (%) | 356 (86) | 292 (87) |
| | Left (%) | 48 (12) | 29 (9) |
| | Ambidextrous (%) | 9 (2) | 15 (4) |
| **Age at stroke (years)** | Mean (SD) | 58 (12.5) | 57 (14) |
| | Range | 17 to 88 | 18 to 98 |
| **Years post-stroke of CAT** | Mean (SD) | 4.2 (5.2) | 3.8 (4.5) |
| | Range | 0.1 to 42.4 | 0.1 to 36.0 |
| **Lesioned hemisphere** | Left (%) | 211 (51) | 172 (51) |
| | Right (%) | 111 (27) | 90 (27) |
| | Both (%) | 73 (18) | 61 (18) |
| | Undetected (%) | 18 (4) | 13 (4) |
| **Left hemisphere lesion size (cm3)** | Mean (SD) | 33 (62) | 28 (53) |
| | Range | 0 to 429 | 0 to 355 |
| **Right hemisphere lesion size (cm3)** | Mean (SD) | 19 (43) | 15 (43) |
| | Range | 0 to 280 | 0 to 357 |
| **Developmental dyslexia status** | Formal diagnosis (%) | 6 (1) | 8 (2) |
| | None (%) | 232 (56) | 193 (57) |
| | Difficulties relative to peers [a] (%) | 31 (8) | 21 (6) |
| | Unknown (%) | 144 (35) | 114 (34) |

**Legend:** [a] 'Difficulties relative to peers' is reported when participants had reading or spelling difficulties compared to their peers that may indicate developmental dyslexia, did not receive a formal diagnosis (usually because they were not assessed for dyslexia). For these analyses, these participants were grouped with those whose dyslexia status was 'unknown'.

### 5.3.2 Pre-stroke education

Education amount was measured as the number of years of formal education completed. This study's definition of formal education includes subject-based qualifications such as GCSEs, AS and A levels (or equivalent qualifications), and undergraduate and post-graduate degrees and diplomas. Apprenticeships, work-related qualifications and continuing professional development courses are not included, as they typically involve different (less traditionally academic) learning content, style and assessment methods compared to school or university-based education, which relies more heavily on written material and examination. The number of years of education was primarily treated as a continuous variable, but was also used to allocate participants into groups (Less versus More) for some analyses and illustrations. The groups were ‘Less’ education, which comprised participants who underwent education up to age 18 (i.e. A Levels or equivalent), and ‘More’ education, which comprised participants who continued their formal education beyond age 18 (i.e. degrees and diplomas).

### 5.3.3 Language outcomes

The focus of the current study was on 19 tasks and 6 ‘summary’ skills (Naming, Repetition, Reading, Spoken picture description, Spoken Comprehension and Written Comprehension) from the CAT, and ‘overall language ability’. See Chapter 2 section 2.4 for full details of the CAT and the language tasks selected.

### 5.3.4. Initial aphasia severity

Each participant was assigned a one week post-stroke initial aphasia severity rating derived from the ARTQ (see Chapter 2 section 2.5). The 4 severity categories were: ‘Severe’ (unable to produce any speech, or only able to produce automatic speech), ‘Moderate’ (able to produce single words and phrases), ‘Mild’ (able to produce lexically meaningful short sentences), and ‘Normal’. The ratings were repeated for comprehension, reading, and writing.

### 5.3.5. Lesion data

Separate estimates of left and right hemisphere lesion size were obtained from structural MRI brain scans (see Chapter 2 section 2.7 for details of scan acquisition and lesion identification). 497 participants were imaged on a 3T Trio (53 of which were after Prisma upgrade), 11 were imaged on a 3T Allegra, 176 were imaged on a 1.5T Avanto, 65 were imaged on a 1.5T Sonata.

### 5.3.6. Demographic and assessment variables

Study questionnaires capture each participant’s biological sex, pre-stroke handedness, developmental dyslexia status, date of birth, and date of stroke (see Table 5.1). Age at stroke was subsequently calculated from date of birth and date of stroke. Object recognition impairments were controlled for using scores from the CAT ‘semantic memory’ task, and hearing and vision status were included to control for perceptual abilities that might impact the ability to perceive the assessment stimuli.

### 5.3.7 Statistical analyses

The results are reported from a series of analyses that are designed to answer each question of interest (A, B and C).

#### A. Which language scores are most sensitive to pre-stroke education amount?

This question is approached in three steps: (1) Is there a main effect of education on overall language ability score? (2) Which summary scores (e.g. naming, repetition, etc.) drive this effect? (3) Which individual tasks drive effects of education in the summary scores? As there is only one statistical analysis in the first step, there is no need to correct for multiple comparisons. The second and third steps involve 6 different analyses (one for each summary score) and 24 different analyses (one for each of the individual tasks and subscores). If the main effect in step one is significant, the second and third steps can be considered as descriptive, and therefore do not need to be corrected for multiple comparisons. However, in the Table legends, it is also noted which summary and task scores would survive a correction for multiple comparisons if they were considered independently.

Each of the analyses described above was conducted using multiple linear regression. The following covariates were included in all analyses to factor out the effects of: age at stroke, time post-stroke of CAT, left hemisphere lesion size, right hemisphere lesion size, sex, pre-stroke handedness and developmental dyslexia status. The addition of covariates measuring initial speaking, reading and understanding severity, vision status, hearing status and semantic memory at the time of the CAT depended on the task scores being considered. Specifically, initial speaking severity was included in the analyses of speaking scores (naming, repetition, spoken picture description); initial understanding severity was considered for the analyses of comprehension scores, initial reading severity was included in the analyses of reading scores, vision status was considered for tasks that depend on intact visual perception (e.g. naming and reading), and hearing status was included for tasks which depend on intact hearing ability (e.g. repetition and spoken comprehension). Semantic memory scores were included for tasks that included visual or conceptual stimuli (naming, spoken picture description). The covariates selected for each task analysis can also be seen in Table 5.2. Each analysis had a different sample size, due to missing data for some covariates. For example, hearing status was unknown for 5 participants, so the sample size in the analyses including this covariate was reduced from 749 to 744. The sample size for each analysis is reported in Table 5.3.

**Table 5.2** The covariates factored out of each CAT analysis.

| | SS | RS | US | SM | Vis | Hea |
|---|---|---|---|---|---|---|
| **Overall language ability** | + | + | + | + | + | + |
| **Naming:** | + | - | - | + | + | - |
| Semantic fluency | + | - | - | + | - | - |
| Letter fluency | + | - | - | + | - | - |
| Object naming | + | - | - | + | + | - |
| Action naming | + | - | - | + | + | - |
| **Repetition:** | + | - | - | + | - | + |
| Word repetition | + | - | - | + | - | + |
| Complex word repetition | + | - | - | + | - | + |
| Nonword repetition | + | - | - | - | - | + |
| Digit span | + | - | - | - | - | + |
| Sentence repetition | + | - | - | + | - | + |
| **Reading:** | - | + | - | + | + | - |
| Reading words | - | + | - | + | + | - |
| Reading complex words | - | + | - | + | + | - |
| Reading function words | - | + | - | + | + | - |
| Reading nonwords | - | + | - | - | + | - |
| **Spoken picture description:** | + | - | - | + | + | - |
| Total number of words | + | - | - | + | + | - |
| Number of appropriate information-carrying words | + | - | - | + | + | - |
| Number of inappropriate information-carrying words | + | - | - | + | + | - |
| Grammatical well-formedness | + | - | - | + | + | - |
| Syntactic variety | + | - | - | + | + | - |
| Speed | + | - | - | + | + | - |
| **Spoken comprehension:** | - | - | + | + | + | + |
| Spoken word-to-picture | - | - | + | + | + | + |
| Spoken sentence-to-picture | - | - | + | + | + | + |
| Spoken paragraph | - | - | + | + | - | + |
| **Written comprehension:** | - | + | - | + | + | - |
| Written word-to-picture | - | + | - | + | + | - |
| Written sentence-to-picture | - | + | - | + | + | - |

**Legend**: SS = initial speaking severity. RS = initial reading severity. US initial understanding severity. SM = semantic memory. Vis = vision status. Hea = hearing status. **+** indicates the covariate was included in that task analysis. – indicates the covariate was not included in that task analysis. All analyses factored out: sex, handedness, age at stroke, time post-stroke of CAT, left hemisphere lesion size, right hemisphere lesion size, and dyslexia status.

Bayesian statistics were used to validate the results further by estimating evidence for or against an effect of education on overall language ability, and each summary skill, task and sub-score. This was done by first conducting a multiple linear regression for each outcome with all the relevant covariates, but not education, to obtain the standardised residuals, and then regressing these residuals with education. The Bayes Factors were then converted into a log BF. A negative log BF indicates evidence for no effect of education, whereas a positive log BF indicates evidence for an effect of education. The strength of evidence can also range from ‘anecdotal’ to ‘decisive’ (Jeffreys, 1961), see Table 5.3 legend for categories of evidence strength.

**B: How does the effect of education depend on other variables?**

Moderator analyses were used to investigate potential interactions between education and six other variables of interest. The interaction terms were created by multiplying ‘years of education’ (the regressor of interest) with each variable: age at stroke, time post-stroke, initial severity, semantic memory, right hemisphere lesion size and left hemisphere lesion size. Two-way interaction terms were added into a second block of the regression; three-way interaction terms were added into a third block of the regression (along with the relevant two-way interaction terms) and four-way interaction terms were added into a fourth block of the regression (along with the relevant two-way and three-way interaction terms). The first block of the model contained the same covariates as the main analyses. These moderator analyses were conducted first for overall language ability, then repeated as post hoc tests for each of the six summary scores.

Post hoc analyses of significant interactions were conducted by testing whether there was an effect of education in different subgroups of participants. For age at stroke, the subgroups were created using the median age: the ‘younger’ group were 17 to 57.88 years (n=375) and the ‘older’ group were 57.93 to 99 years (n=374). For initial severity, the four categories were used (Normal, n=146; Mild, n=242; Moderate, n=118, and Severe, n=243) – although to increase group sizes, participants with Moderate initial severity were combined with participants with Severe initial severity. For left hemisphere lesion size, two approaches were taken. First, to illustrate the effect of education in participants with smaller versus larger lesions, subgroups were created using the median lesion size (3.3cm$^3$); 177 participants in the ‘larger’ subgroups with the largest and smallest lesions, and who were assessed very late post-stroke were subsequently removed to match them for left hemisphere lesion size and time post-stroke (resulting in 374 participants in the ‘smaller’ subgroup, and 198 participants in the ‘larger’ subgroup) (Figure 5.1). Second, to illustrate the interactive effects between education, lesion size, age and initial severity, subgroups were created for participants with left hemisphere lesions of less than 1cm$^3$, and greater than 1cm$^3$ (Figure 5.3).

**C: Does amount of education influence the degree of recovery?**

Two methods were used to investigate the relationship between education and recovery.

**CAT scores:** An effect of education on recovery was expected to result in a significant interaction between education and time post-stroke (as investigated in section B above), with post hoc tests showing that, as time post-stroke increased, the advantage of more education also increased. However, as the CAT scores were mostly acquired years post-stroke (see Table 5.1), this analysis will not be sensitive to the influence of education on early recovery (i.e. in the first year post stroke).

**Self-rated scores**: As discussed in Chapter 2 section 2.5.3, a proportional 'recovery score' was calculated for each participant by dividing [actual improvement at one year post-stroke] by [improvement potential]. Actual improvement is the number of ability levels by which a participant increases between one week and one year (Hope et al., 2019, Kim et al., 2019). For example, a participant with Severe initial severity who improved to Mild by one year improved by 2 severity levels (Severe to Moderate and Moderate to Mild). Improvement potential is the total number of severity levels available to increase by. This is maximum for participants with Severe initial symptoms who can move up three levels (Severe to Moderate, Moderate to Mild and Mild to Normal). Conversely, those with mild initial symptoms can only move up one level (Mild to Normal). The recovery score is a proportion; for example, a participant who improved by 2 out of 3 possible performance increases would receive a recovery score of 0.66. This method of measuring recovery score enables all participants to be analysed together, regardless of their initial severity. This was repeated for each of the four language skills, resulting in each participant having a recovery score for speaking, understanding, reading and writing. The mean of these recovery scores was also calculated for each participant. Multiple linear regression was used to investigate the relationship between education and recovery score, when left and right hemisphere lesion size, age, sex, and handedness were factored out.

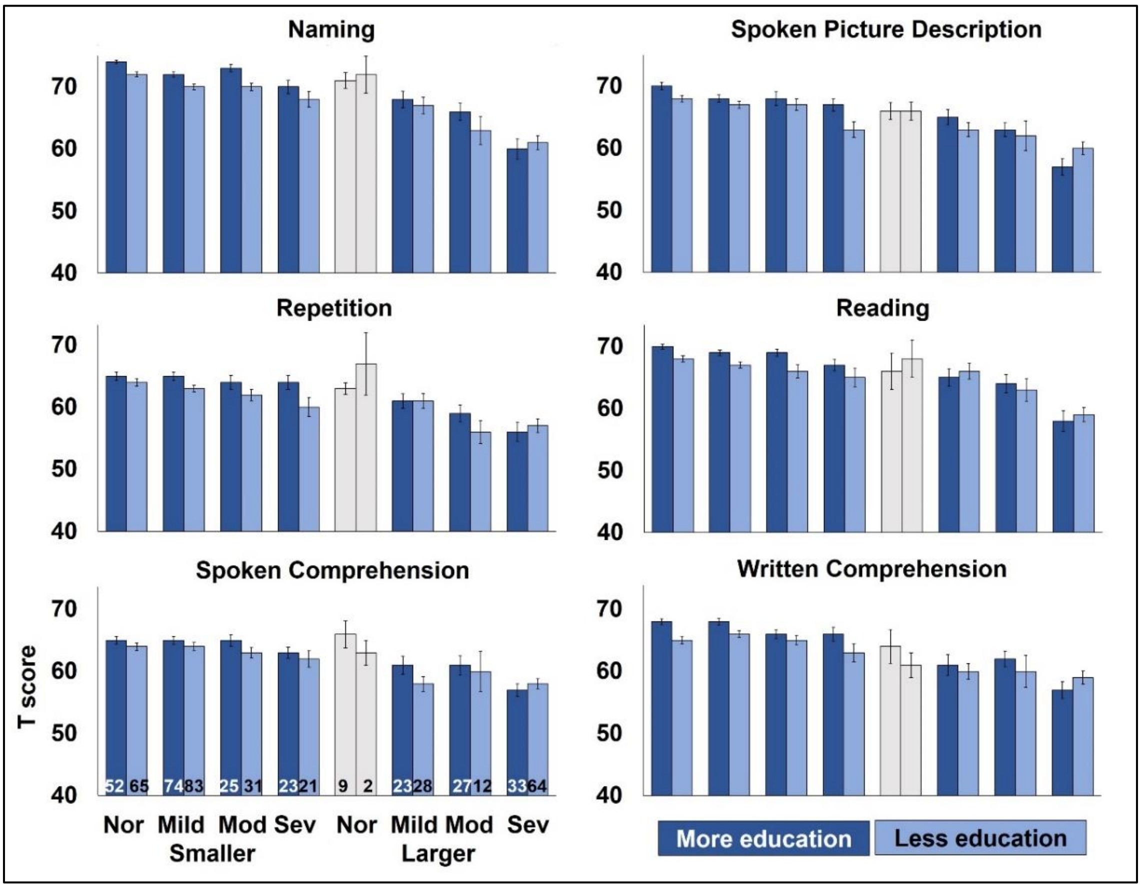
## 5.4 Results

### A. Which language scores are most sensitive to pre-stroke education amount?

Across all participants, the multiple regression analysis revealed that years of education had a significant and positive, albeit small, effect on overall language ability (β=0.075, p=0.004), after factoring out the effect of other variables. Figure 5.1 illustrates that this advantage of more education was also observed, independently, on all six summary scores.

For each of the tasks that contributed to the summary scores, Table 5.3 provides (i) the $R^2$, β, and p values from separate regression analyses and (ii) the Log Bayes evidence from the Bayesian regressions, when other variables were factored out of the analyses. The effects of education were strongest for: (1) letter fluency and semantic fluency within the naming score, (2) nonword repetition and digit span within the repetition score, (3) nonword reading within the reading score, (4) the total number of words produced during spoken picture description and their grammatical well-formedness within the spoken picture description score, and (5) sentence comprehension within both auditory and written comprehension scores. Figure 5.2 shows the amount of variance explained independently by education for each task, compared to the amount collectively explained by the other covariates.

**Figure 5.1** The effect of education on performance on six summary scores from the CAT.



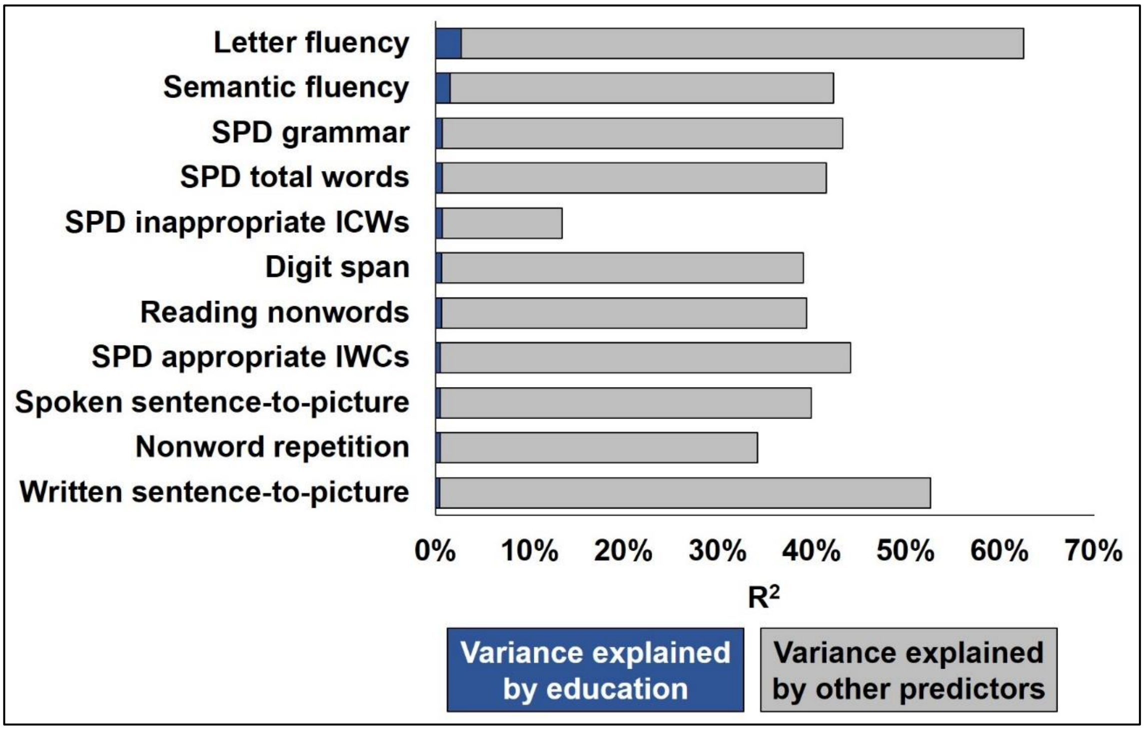
**Legend:** CAT T scores for each of the six summary scores, grouped by education amount, initial speaking severity and lesion size. The colours indicate more (dark blue) or less (light blue) education, with grey indicating where the group comprised fewer than 10 participants (i.e. not well-sampled). Nor, Mild, Mod, Sev refer to the four levels of initial severity (Normal, Mild aphasia, Moderate aphasia or Severe aphasia). Smaller (first 8 bars) refers to participants with smaller left hemisphere lesions (0-3cm$^3$). Larger (last 8 bars) refers to participants with larger left hemisphere lesions (more than 9cm$^3$). The number of participants in each group is shown in the plots for spoken comprehension (lower left) and is the same for all summary scores. The total number of participants illustrated is 572, after 177 were excluded to match groups, for left hemisphere lesion size and time post-stroke. Scores are higher for More compared to Less education for all 6 tasks except for (i) Large lesions with severe initial symptoms, where scores for all 6 tasks are consistently low, irrespective of education; (ii) Large lesions and mild initial symptoms for 2 tasks (repetition and reading); and (iii) Larger lesions with normal initial language, that were not well-sampled. Score thresholds indicating aphasia: <63 for Naming, <61 for Spoken picture description <60 for Repetition, <61 for Reading, <57 for Spoken comprehension, <60 for Written comprehension.

**Table 5.3** The effect of education on post-stroke language assessment scores.

| | | **Multiple linear regression** | | | **Bayesian regression** | |
|---|---|---|---|---|---|---|
| **Outcome measure** | **N** | **$R^2$** | **β** | **p** | **Log BF** | **Evidence strength** |
| **Overall language ability** | 569 | 0.632 | 0.075 | 0.004 | 0.29 | Anecdotal |
| **Naming:** | 741 | 0.580 | 0.079 | 0.001* | 0.74 | Strong |
| Semantic fluency | 748 | 0.423 | 0.123 | <0.001* | 2.48 | Decisive |
| Letter fluency | 748 | 0.384 | 0.166 | <0.001* | 5.27 | Decisive |
| Object naming | 741 | 0.473 | 0.014 | ns | | |
| Action naming | 741 | 0.455 | 0.024 | ns | | |
| **Repetition:** | 743 | 0.478 | 0.054 | 0.047 | | |
| Word repetition | 743 | 0.380 | -0.017 | ns | | |
| Complex word repetition | 743 | 0.387 | -0.021 | ns | | |
| Nonword repetition | 743 | 0.342 | 0.072 | 0.019 | | |
| Digit span | 744 | 0.391 | 0.076 | 0.010 | | |
| Sentence repetition | 743 | 0.432 | 0.012 | ns | | |
| **Reading:** | 572 | 0.444 | 0.058 | ns | | |
| Reading words | 573 | 0.435 | 0.043 | ns | | |
| Reading complex words | 572 | 0.431 | 0.027 | ns | | |
| Reading function words | 572 | 0.188 | 0.044 | ns | | |
| Reading nonwords | 572 | 0.394 | 0.076 | 0.022 | | |
| **Spoken picture description:** | 741 | 0.516 | 0.059 | 0.024 | | |
| Total number of words | 741 | 0.415 | 0.085 | 0.003 | 0.36 | Anecdotal |
| Number of appropriate ICWs | 741 | 0.441 | 0.073 | 0.009 | | |
| Number of inappropriate ICWs | 741 | 0.134 | 0.087 | 0.013 | | |
| Grammatical well-formedness | 741 | 0.433 | 0.087 | 0.002 | 0.48 | Anecdotal |
| Syntactic variety | 741 | 0.418 | 0.052 | ns | | |
| Speed | 741 | 0.438 | 0.025 | ns | | |
| **Spoken comprehension:** | 709 | 0.388 | 0.066 | 0.029 | | |
| Spoken word-to-picture | 711 | 0.272 | 0.039 | ns | | |
| Spoken sentence-to-picture | 710 | 0.400 | 0.073 | 0.014 | | |
| Spoken paragraph | 710 | 0.118 | 0.016 | ns | | |
| **Written comprehension:** | 573 | 0.525 | 0.076 | 0.010 | | |
| Written word-to-picture | 574 | 0.339 | 0.054 | ns | | |
| Written sentence-to-picture | 573 | 0.526 | 0.067 | 0.022 | | |

**Legend:** Effects of education on overall language ability, and summary and individual task scores from the CAT, using (i) multiple linear regression and (ii) Bayesian regression, when other variables were factored out. The full sample includes 749 participants, but the number in each analysis varies due to missing data for certain covariates. ns = not significant at a threshold of p=<0.05. BF = Bayes Factor: Positive BF supports an effect of education; Evidence strength: >2=decisive; 1.5-2=very strong; 1-1.5=strong; 0.5-1=substantial; 0-0.5=anecdotal (Jeffreys, 1961). Negative BF refutes the effect (empty cells indicate negative BF). * = significant after correction for multiple comparisons (= 6 summary scores; 20 tasks). ICW = information carrying word.

**Figure 5.2** The variance in CAT task scores explained by (i) education amount and (ii) other predictors.



**Legend:** See Table 5.2 for details of the 'other predictors' factored out of each task analysis.

### B. How does the effect of education depend on other variables?

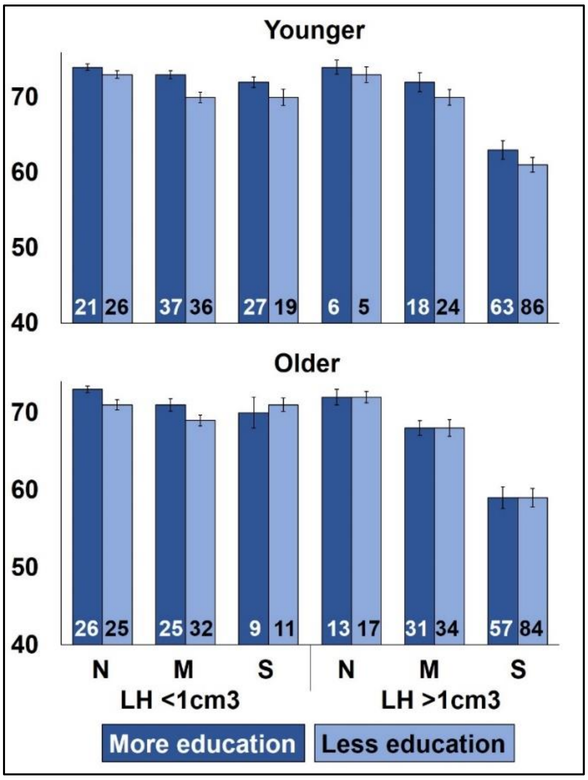
The moderator analyses indicated that the benefit of more education was dependent on age for most outcome measures (except for picture description, spoken comprehension and written comprehension), with post-hoc tests indicating that the benefit of education was primarily driven by younger participants (Table 5.4). For naming and spoken picture description, there was a four-way interaction between education, age, left hemisphere lesion size and initial speaking severity (see Table 5.4). This four-way interaction arose because a benefit of more education was observed for younger participants irrespective of lesion size or initial severity (see upper part of Figure 5.3); whereas in older participants, a benefit of education was only observed when left hemisphere damage was less than 1cm$^3$ and initial speaking severity was normal or mildly impaired (see lower part of Figure 5.3). In contrast, no benefit of education was observed on naming scores for older participants whose lesions were larger than 1cm$^3$, irrespective of initial severity.

**Table 5.4** Interactions between Education and Age.

| | 2-way interaction: Ed & Age | Effect of education on: | | 4-way interaction: Ed, Age, LH & initial severity |
|---|---|---|---|---|
| | | Younger | Older | |
| Overall language ability | 0.049 | 0.004 | ns | ns |
| Naming | 0.038 | 0.001* | ns | 0.019 |
| Repetition | 0.050 | 0.042 | ns | ns |
| Reading | 0.020 | 0.020 | ns | ns |
| Spoken Picture Description | ns | ns | ns | 0.030 |
| Spoken Comprehension | ns | 0.048 | ns | ns |
| Written Comprehension | ns | 0.013 | ns | ns |

**Legend:** p values for the two-way interaction between education (Ed) and age (n=749), the main effect of education in younger (n=375) and older (n=374) participants, and the four-way interaction between education, age, left hemisphere lesion size (LH) and initial severity: for overall language ability, and for each summary score. ns = not significant (p=>0.05 uncorrected). * = significant after correction for multiple comparisons (= 6 summary scores; 20 tasks). Due to missing data for some variables, 8 participants were excluded from the Naming analysis; 6 were excluded from the Repetition analysis; 1 was excluded from the fluency analyses and 177 were excluded from the Reading analysis (primarily because of missing data for initial reading severity).

**Figure 5.3** Interaction between education, age, lesion size and initial speaking severity on naming scores.



**Legend**: Average CAT T scores for the summary naming score, for participants who were younger (n=368, upper part) or older (n=364, lower part), grouped by education amount, initial speaking severity and left hemisphere damage. As in Figure 5.1, more/less education is indicated by dark/light blue. N, M, S refer to three levels of initial speaking severity (Normal, Mild aphasia, Severe aphasia). Participants with Moderate initial severity are included in the Severe aphasia groups because of very small sample sizes. LH <1cm$^3$ (first six columns) refers to participants with very small left hemisphere lesions (<1 cm$^3$) and/or larger right hemisphere lesions. LH >1cm$^3$ (last six columns) refers to participants with left hemisphere lesions that were larger than 1cm$^3$, irrespective of right hemisphere damage. The effect of education is larger in younger participants than older participants (see Table 5.4). For older participants, there is no apparent effect of education in any of the groups with LH lesions >1cm$^3$ or for the participants with less LH damage but severe initial severity.

There were no other significant interactions between education and other variables on overall language ability or any of the summary scores or individual task scores, after correction for multiple comparisons. Indeed, the benefit of education was remarkably robust across groups of participants with different lesion sizes and different initial severities, with the exception that, for participants with large lesions and severe initial aphasia, all scores were low, irrespective of education amount (Figure 5.1).

### C. The effect of education on the degree of recovery

**CAT scores:** Evidence of recovery was observed as a significant positive effect of time-post stroke on overall language ability (p=0.002, β=0.088), after factoring out the influence of all other variables. However, there were no significant interactions between education and time post-stroke and therefore no evidence that more education increased the rate or degree of recovery.

**Self-rated scores:** Recovery score (measuring the degree of improvement between one week and one year) was not significantly associated with the amount of education for any of the 4 language skills (speaking, understanding, reading and writing), their mean, or any of the initial severity groups (severe, moderate, mild).

## 5.5 Discussion

Using data from 749 stroke survivors, and controlling for a range of stroke-related, demographic and perceptual factors that were expected to influence language outcomes after stroke, this study demonstrates that more education is associated with better language ability, in participants with, as well as without, aphasia. The discussion below addresses the three questions: (A) which CAT tasks are most sensitive to education amount? (B) how does the effect of education depend on other variables? and (C) does the degree of recovery depend on amount of pre-stroke education? I than discuss the clinical implications of the findings for the diagnosis and management of aphasia post-stroke, study limitations and future directions.

### 5.5.1 A. Which CAT tasks are most sensitive to education amount, and why?

Prior studies have demonstrated that aphasia assessments are sensitive to amount of education in neurotypical people. The results of the current study are the first to show that the benefit of more education on language performance prevails across a range of language production and comprehension tasks, regardless of whether participants had aphasia or not after their stroke. This latter finding is especially crucial, because it demonstrates a sensitivity of the aphasia assessment itself to the amount of education, rather than an assistive benefit of education on aphasia outcome. For naming tasks, the effect of education was strongest on semantic and letter fluency scores, according to both the size of the effects and Bayesian evidence. During spoken picture description, the effect of education was significant on the number of words produced and their grammatical well-formedness. During the repetition tasks, the effect of education was significant for nonword repetition and digit span, and during the reading tasks, it was significant for reading nonwords. Finally, during the comprehension tasks, the effect of education was significant for sentence comprehension but not for word comprehension.

It is worth recognising that the effects of education were reasonably small (see Figure 5.1). The clinical implications of these small effects will be discussed in further detail later on. However, this observation is relevant in the context of the prior studies of education effects in aphasia discussed in this chapter's introduction. These prior studies typically involved small sample sizes. If, as identified in the current study, education effects on language ability are relatively small, a large sample size is necessary to reveal these effects. This may partly explain why an effect of education was observed on recovery in Kim et al. (2019), which involved 235 participants, but not in the studies by Connor et al. (2001) and Lazar et al. (2008) which involved fewer participants (39 and 22 respectively).

The generic benefits of education on verbal skills are well recognised (as discussed in the Introduction, section 5.2). For example, education enriches vocabulary and exposes people to a wider range of linguistic structures for organising, embellishing, and understanding speech. In addition to increasing proficiency in language use, more education and examination experience may increase one's ability to focus attention on verbal details during post-stroke language assessment. More education may also support and enhance the underlying cognitive processes that support language production and comprehension. For example, during picture description, visual perception and working memory are required to perceive the stimuli and hold the contents online; semantic knowledge and processing skills are required to process the information presented and to retrieve semantic information about the individual components and how they relate to each other (i.e. to distinguish the fact that the cat has knocked the books off the shelf; rather than the books simply falling). Greater linguistic cohesion and coherence, which are also developed during education, may result in a more complex, richer, and/or more meaningful narrative (Zanichelli et al., 2020). Additionally, for participants with mild aphasic impairments, more education may confer additional skills in self-monitoring and self-repair of errors, and of circumventing word-finding difficulties by retrieving another appropriate word.

Semantic and letter fluency may benefit most from more education because they are facilitated by a range of language and cognitive skills that are enhanced during education, such as a wide-ranging vocabulary, proficient semantic and phonological retrieval strategies, verbal short-term memory and executive control. For both picture description and fluency, cognitive control is required to suppress incorrect responses, and attention/working memory to recall what has already been said and avoid repeating responses. Like in this study, González-Fernández et al. (2011) found no advantage of education on word repetition or object naming. However, unlike this study, they did find an advantage of education on reading, plausibly because their reading vocabulary tasks were more extensive and/or demanding.

Crucially, the effects of education reported in this study were found to prevail regardless of whether participants had aphasia or not after their stroke. The effects of education were mostly additive, i.e. independent of the effects of lesion size and initial severity. The only exception was that the benefit of education was reduced in older participants with larger lesions and more severe initial symptoms. These findings indicate that there was no protective effect of education on aphasia outcome; contrasting with the wealth of evidence supporting the cognitive reserve theory in dementia (see the study Introduction, section 5.2.2).

### 5.5.2 B. How the effect of education on task performance depends on other variables

The only variable that showed a significant interaction with education was age. This was observed on overall language ability, and specifically on the naming, repetition and reading summary scores, but not on spoken picture description or any of the comprehension tasks. In each case, the interaction indicated a sub-additive relationship between age and education, with a reduced benefit of more education in older compared to younger participants. As discussed in Experiment 2, this might be a consequence of older participants having reduced cognitive and learning abilities (e.g. tip-of-the-tongue states) that could not be overcome by educational experience (Blachstein and Vakil, 2016). Moreover, the effect of education did not interact significantly with left hemisphere lesion size, initial severity, time post-stroke, handedness, sex, semantic memory or developmental dyslexia. In accordance with the hypotheses outlined in the study Introduction, it is clear that the benefit of more education is additive with the multiple other factors that influence post-stroke language performance. The one exception was the benefit of education was eroded in older participants with large left hemisphere lesions and severe initial symptoms. These participants may have sustained substantial damage to the brain regions that can support language after stroke making it much harder to improve or recover their abilities irrespective of education. Indeed, lesion size (whole brain and left hemisphere) was significantly larger in participants with Severe compared to Moderate and Mild initial severity (p=<0.001 for both) and in participants with Moderate compared to Mild initial severity (p=0.001).

### 5.5.3 C. The effect of education on recovery

The current study did not find any evidence that more education facilitated the degree of recovery from aphasia, using either long-term CAT scores, or self-reports of language ability between one week and one year. This contrasts with a prior study by Kim et al. (2019), which is the only one to demonstrate an effect of education on aphasia recovery. They reported significant effects of education on change in overall language function from one week to one year post-stroke, in a cohort of 235 patients, after factoring out variance from age, initial severity and lesion size. The language score used by Kim et al. was a standardised behavioural measure that integrated scores from speaking, comprehension, reading and writing tasks; albeit with no further exploration of which of those language skills were driving the effect on recovery. As in the current study, their participants were categorised according to severity of aphasia symptoms one week post-stroke (according to the composite K-FAST score), and recovery measured as change in score from one week to one year. However, differences from the current study include the language spoken (Korean versus English), sample size (235 versus 749), and likely cross-national and cultural differences in education systems and attitudes to formal schooling.

Unlike the study by Kim et al., the measurement of initial severity in the current study was based on retrospective participant reports because the CAT was not typically administered early post-stroke in the sample of participants available.

Retrospective self-reported scores will undoubtedly result in measurement error, which could de-sensitise the study to the effect of education on recovery.

Further work is therefore needed to understand exactly when and how a greater amount of education facilitates recovery, importantly by adopting both self-rated measures, and objective measures. Although this approach was taken in the current study, the objective (CAT) measures were acquired very late (typically years) post-stroke and compared to the participant-reported assessments of initial abilities. Future studies could measure the effect of education on language recovery by monitoring, assessing and comparing language ability longitudinally, using the same standardised language tasks, at regular time-points post-stroke, starting within the first days or weeks post-onset. CAT assessment at regular ongoing timepoints would allow us to detect subtle score changes between shorter periods of time, increasing sensitivity to more subtle effects on recovery. In parallel, the PLORAS study plans to improve participant-reported outcome measures to more accurately capture participants' own perspectives of their recovery, and to identify more subtle improvements that cannot be detected by objective language assessments.

### 5.5.4 Clinical and research implications for interpreting post-stroke language impairments

The sensitivity of the CAT to educational differences in participants both with and without aphasia has implications for clinical practice and for research. Specifically, it raises the possibility that patients with normal language abilities but less pre-stroke education may be identified as aphasic (according to the assessment thresholds), and that patients with mild language impairments but more pre-stroke education may be considered non-aphasic. In clinical settings, this could potentially influence decisions about subsequent assessments (e.g. to categorise the aphasia subtype, or identify higher-level language impairments) or intervention plans. That said, the effects of education were small, with a maximum 3-point T-score difference in naming, reading, and spoken and written comprehension, and a 4-point T-score difference in spoken picture description, and repetition (Figure 5.1). Therefore, this concern only applies to patients whose scores are within 2 to 4 points below or above the aphasia score threshold. For these patients, an understanding of their educational background may aid interpretation of their test results, and inform clinical management. On the other hand, the likelihood of misdiagnosis is low because clinicians draw information from a range of assessments and observations, including patients' functional abilities and self-reports, as well as their impairments. This also highlights the value of using a participant-reported outcome measure to supplement objective assessment scores that are sensitive to the amount of education. For example, a patient with less education may self-report normal language when the CAT indicates mild aphasia. Conversely a patient with more education may consider themselves to have aphasia due to subtle post-stroke language changes which go undetected on the CAT. In the latter case, the patient would likely not be prioritised to receive therapy from the UK's national health service, but might still benefit from recommendations for self-directed therapy strategies, such as techniques to retrieve personally relevant words.

To address the sensitivity of assessments to influential variables, a 'normative modelling' approach has been adopted in prior studies. This approach provides a calculation to adjust any individual's score for the influence of variables such as education, as well as age and sex, based on the performance of neurotypical controls. The resulting adjusted score indicates how that individual performed with respect to others with similar demographics to them (Shirk et al., 2011, Casaletto et al., 2015). No such approach has been implemented for the CAT thus far, but it could be transformative; enabling a personalised understanding of an individual's language profile. That said, care must be taken to ensure the normative data are derived from sufficiently large and representative samples of participants; to avoid biasing the prediction model to a demographically and socioeconomically homogeneous group. Of note, the studies cited above included over 3000 participants to derive their norms.

There are also implications for research when task scores are used to categorise participants either as having aphasia or not, or into different levels of aphasia severity (as in the current study). Plausibly, if those categories are used when comparing outcomes or recovery, then participants miscategorised - because of their education amount - could be erroneously treated as 'recovered to normal' when they in fact have mild impairments – or erroneously treated as mildly aphasic when their abilities were in fact back to normal. In this study, 10-45% of participants fell within the score range that, according to the results, could be miscategorised as 'aphasic' or 'normal', with the percentage range depending on which language skill is used. For example, for Naming, 122/749 participants received a T score from 60 to 65, which could be miscategorised if education amount is not taken into account. Ultimately, this leads to inaccuracies in analyses of outcomes and recovery.

Given these issues, the results of this study advocate that borderline CAT task scores should be interpreted in light of the amount of education a patient completed before their stroke; as has been proposed for the Boston Naming Test and verbal fluency (Zec et al., 2007, Tombaugh et al., 1999). By considering how individual factors contribute to variation in outcomes, this study also supports a personalised approach to assessment, management and prediction of outcome and recovery, advocated by Brady et al. (2022).

### 5.5.5 Future directions

The measurement and definition of amount of education has varied across prior studies, and there is currently no agreement as to what constitutes ‘more’ or ‘less’ education. In the current study, for illustrative purposes, ‘more education’ was defined as that typically done after age 18 in the UK – i.e. university or college; and ‘less education’ was defined as that typically done before age 18, including GCSEs and A Levels. The alternative is to use years of formal education as a continuous variable (as was also done in the current and other studies). The problem with both of these measurements is that they disregard the possible effect of different education settings or content, which may easily arise within, or across, countries, cultures, and generations (for example, academic versus vocational subjects). A more rigorous approach is therefore needed which incorporates more detail about education setting, content, attainment, and considers other factors that may moderate education such as socioeconomic status, intelligence, and learning difficulties, as well as traditionally non-academic learning experiences, such as job-specific training or learning for professional development.

A second and related limitation is that the current study was unable to control for all of the many factors that influence aphasia outcome and recovery, particularly when analysing the effect of education on recovery. In addition to the factors that may moderate education, mentioned in the previous paragraph, other factors that contribute to language recovery were not considered. For example, the amount of speech and language therapy received may have influenced the speed of recovery, and/or interacted with the effect of education on recovery (Brady et al., 2016, Roberts et al., 2021). Real-world language practice, such as occupation, and the other daily activities and interactions a participant engages in, may influence recovery, and/or correlate with education, if they draw on linguistic strategies (e.g. completing crosswords or reading for pleasure). Similarly, non-linguistic cognitive ability, measured in Johnson et al. (2022) using the WAIS Matrices subtest as an indicator of non-verbal problem solving and reasoning, has been correlated with education amount in prior studies. This may overlap with, or explain the effect of education, if for example non-linguistic skills are associated with more education.

A third limitation is the potential impact of missing CAT or participant-reported severity scores. 173 participants had missing initial reading or comprehension abilities, and were excluded from the analyses that included these covariates. If this data was missing because the participants' abilities were severely impaired at one week, then their exclusion could bias the sample towards those with milder impairments. However, it is more likely that the data were missing due to participants not being able to remember their early abilities. Conversely, missing CAT data are more likely to occur if participants were unable to complete a task due to aphasia severity. Only 3 participants in the sample had missing CAT data, and so the impact of their exclusion from the analyses is relatively small, and unlikely to inflate the effect of education.

Finally, the finding that the CAT is sensitive to education amount in stroke participants who never had aphasia suggests that future studies should investigate the effect of other non-stroke factors on language performance, such as cultural background and/or the ability to speak English among non-native English speakers (who were excluded from this study).

### 5.5.6 Summary and Conclusions

Using a large sample of stroke survivors, this experiment found that participants with more pre-stroke education had better performance on a range of language tasks from the CAT when assessed in the chronic stage of recovery. Moreover, the benefit of more education on language outcome was found to be largely independent of (i.e. additive with) the influence of other variables known to impact language performance after stroke. No effect of education was observed on recovery of language abilities. In practice, these results suggest that diagnostic thresholds for aphasia should be interpreted with care for patients with borderline scores. In addition, future studies should consider assessing whether education influences language abilities measured at earlier and more regular timepoints after a stroke, to determine whether a more subtle effect of education on recovery exists.

# Chapter 6: Experiment 4. The effect of early clinical speech and language therapy on language outcomes and recovery

## 6.1 Summary of study

**Introduction**: Establishing whether speech and language therapy after stroke benefits speaking ability is challenging because of the need to control for multiple non-therapy factors known to influence recovery. The aim of this study was to investigate how speaking ability at three time points post-stroke differed in participants who received varying amounts of clinical therapy in the first month post-stroke.

**Methods:** The participants were 143 stroke survivors with left or bilateral lesions, Severe or Moderate initial speaking severity, and for whom details of early speech and language therapy (within the first four weeks of stroke onset) were known. They were allocated into one of two groups: Therapy (received at least 1 hour of early therapy) and No Therapy (received no early therapy). Language abilities were assessed using (1) participant-reported outcomes at one month and one year post-stroke, and (2) Naming, Repetition and Spoken picture description scores from the CAT. The effect of early therapy on language abilities was assessed using binary logistic regression and multiple linear regression analyses, whilst controlling for variance from: initial speaking severity, amount of later therapy, left and right hemisphere lesion size, handedness, semantic memory and auditory comprehension abilities, and time post-stroke.

**Results:** The results showed that participant-reported speaking ability at one month post-stroke (OR=2.256) and long-term Naming scores (β=0.158) were significantly better in participants who received early therapy (n=79), versus those who did not (n=64), and the number of hours of early therapy was positively related to recovery at one year post-stroke (OR=1.117).

**Conclusion:** Two non-mutually exclusive interpretations of these results are presented: (1) participants may have benefitted from the early provision of therapy and self-management strategies; and (2) therapy is more likely to be provided to patients who have a better chance of recovery, possibly due to better physical and/or mental health. Both interpretations have implications for future studies aiming to predict individual patients' speech and language outcomes after stroke, and their response to therapy.

This study was published in Neuropsychological Rehabilitation (Roberts et al. (2021).

## 6.2. Introduction

The goal of this experiment was to investigate whether recovery of speaking ability was related to the provision, and/or amount, of clinical speech and language therapy received in the first month after stroke, after controlling for initial speaking severity, amount of later therapy received, handedness, auditory comprehension ability, time post-stroke and lesion size.

Prior studies have yielded inconsistent conclusions about the benefit of early speech and language therapy for post-stroke aphasia. Even when positive effects of therapy are reported, there are concerns that these effects are weak, not clinically meaningful, and without generalisation or maintenance (Brady et al., 2016). This may be because establishing true therapy effects requires strictly controlled trials where patients who did and did not receive therapy are matched for other influential predictors of outcome such as lesion site and size and initial aphasia severity, (Watila and Balarabe, 2015). Controlling for lesion factors has typically not been possible in prior therapy studies because neuroimaging data were not available.

To investigate therapy effects reported by prior studies, I conducted a review of the literature focusing on the effect of timing and amount of speech and language therapy. As a starting point, I examined Bhogal et al.'s (2003) highly influential review, which investigated the relationship between therapy intensity and change in language score for ten different therapy trials. They observed that the 5 'positive' trials that demonstrated a significant impact of therapy provided an average of 8.8 hours per week for an average of 11.2 weeks (averaging 98 hours in total per patient), whereas the 5 'negative' trials that failed to demonstrate significant therapy effects provided an average of 2 hours per week for an average of 22.9 weeks (averaging 44 hours in total per patient). From this, they concluded that high-intensity therapy is more beneficial than low-intensity therapy. Critically, however, the 'positive' versus 'negative' trials were not matched for lesion site, initial severity, time post-stroke, or therapy type. Regarding timing of therapy, the literature search did not reveal any individual study that directly compared the effect of early- versus late-initiated therapy on speaking outcomes, although a between-study meta-analysis of 55 therapy studies by Robey (1998) concluded that therapy outcome was better in studies that provided earlier than later therapy. Nonetheless, there were, again, multiple uncontrolled factors across the meta-analysed studies, including initial symptom severity.

The primary source of therapy trials for this review was the 2016 Cochrane systematic review of speech and language therapy for aphasia following stroke (Brady et al., 2016). Guided by this, other prior reviews of aphasia therapy studies (Robey, 1998, Bhogal et al., 2003, Nouwens et al., 2015, Coleman et al., 2017), and a search of PubMed using search terms such as aphasia/anomia, therapy/rehabilitation/treatment, intensity/duration/amount, and timing/early/acute/subacute, trials were selected and reviewed if they met the following criteria: i) compared speaking recovery in at least two groups of stroke survivors who primarily differed in whether or not they had received speech and language therapy, or in the intensity of therapy; (ii) sufficient numbers of participants to allow the significance of therapy effect sizes to be evaluated (i.e. all single case studies were excluded), (iii) aphasia was caused by stroke, not neurodegeneration (e.g. primary progressive aphasia); and (iv) changes in speaking ability were part of the outcomes measured.

I found 15 studies that met these inclusion criteria, including seven published since the 2016 Cochrane review. These are listed in Table 6.1, with more details given in Table 6.2. No trials were found that directly compared the effect of early- versus late-initiated therapy on speaking outcomes. Unlike the Cochrane systematic review, I categorised studies according to whether the therapy was administered early (within the first 4 weeks of stroke onset) or later (more than 4 weeks post-onset), allowing an exploration of therapy intensity effects at different time-points, and to extend the discussion offered by the Cochrane systematic review. Although the studies differed from one another in terms of the specific type of speech and language therapy administered, speaking ability was consistently defined as one of the outcome measures.

**Table 6.1** Therapy trials reviewed.

| Study reference | Study |
|---|---|
| **Therapy commenced within the first month post-stroke** | |
| 1 | Godecke et al. (2012) |
| 2 | Godecke et al. (2014) |
| 3 | Mattioli et al. (2014) |
| 4 | Nouwens et al. (2017) |
| 5 | Laska et al. (2011) |
| 6 | Bowen et al. (2012b) |
| 7 | Woldag et al. (2017) |
| 8 | Godecke et al. (2020) |
| **Therapy commenced after the first month post-stroke** | |
| 9 | Breitenstein et al. (2017) |
| 10 | Dignam et al. (2015) |
| 11 | Kesav et al. (2017) |
| 12 | Denes et al. (1996) |
| 13 | Martins et al. (2013) |
| 14 | Stahl et al. (2017) |
| 15 | Bakheit et al. (2007) |

**Table 6.2** Results of therapy trials that compared different amounts of therapy on speaking outcomes.

| Study | Therapy onset | Therapy duration | Low-intensity | | High-intensity | | Best therapy intensity |
|---|---|---|---|---|---|---|---|
| | | | Hours per week | n | Hours per week | n | |
| | Days since stroke | Weeks | Acute period (therapy onset within one month of stroke). | | | | |
| 1 [a] | 3 | 4 | <0.25 | 27 | 2.5 | 32 | High |
| 2 [a] | 6 | Low=3<br>High=4.5 | <0.25 | 27 | 4.25 | 20 | High |
| 3 | 2 | 2 | 0 | 6 | 5 | 6 | High [b] |
| 4 | 12 | 4 | 0 | 72 | 7 | 23 | High [c] |
| 5 | 3 | 4 | 0 | 61 | 3.75 | 62 | Neither |
| 6 | 15 | 16 | 0 | 72 | 1.3 | 81 | Neither |
| 7 | 19 [d] | 2 | 7 | 20 [e] | 15 | 20 | Neither |
| 8 | 8-10 | 4 | 2.3 | 70 | 5 | 147 | Neither |
| | Months since stroke | Weeks | Sub-acute/chronic period (therapy onset > 4 weeks post-stroke) | | | | |
| 9 | 27-43 | Low=4<br>High=4.8 | 1.5 | 78 | 10 | 78 | High [f] |
| 10 | 31–47 | Low=8<br>High=3 | 6 | 18 | 16 | 16 | Low |
| 11 | 1 [d] | 4 | 3 | 9 | 6 | 11 | Low |
| 12 | 3 | 24 | 1.8 to 2.5 | 9 | 3.75 to 5 | 8 | Neither |
| 13 | 2 [d] | Low=50<br>High=10 | 2 | 9 | 10 | 9 | Neither |
| 14 | 65 | 4 | 6 | 15 | 12 | 15 | Neither |
| 15 | 1 [d] | 12 | 1.6 | 46 | 4.3 | 51 | Neither |

**Legend:** [a] Studies 1 and 2 are not independent of one another: they used the same control group. [b] Therapy benefit was maintained at 6-month follow-up. [c] A further 57 participants were included in the High-intensity group who did not complete the full 7 hours per week. There was no significant benefit of either therapy intensity in this analysis. [d] Did not control for spontaneous recovery. [e] A third group [n=20] received 15 hours per week of conventional communication therapy (to compare with the same intensity of Constraint Induced Aphasia Therapy received by the first intervention group); CIAT resulted in significantly better outcomes than conventional therapy (same intensity). [f] The control group made similar improvements after receiving the same intensity of therapy as the intervention group after a 3-week deferral period.

### 6.2.1 Effect of early therapy

I identified a total of 8 studies that compared the effect of early speech and language therapy (received in the first month post-stroke) to no therapy or a lower amount of therapy (studies 1-8 in Table 6.1). Four of these studies showed a significant advantage of therapy compared to no or very little therapy (see Table 6.2). However, their conclusions are weakened by a number of factors. The first two studies, both by Godecke et al. (2012, 2014), report that initial severity was worse in the low-intensity therapy group than the high-intensity groups. More critically, a longitudinal follow-up of outcome at 6 months in patients from Godecke et al. (2012) revealed non-significant results suggesting that better outcomes in the high-intensity therapy group were lost over time. The third study, by Mattioli et al. (2014) used a very small sample size (only 6 patients per group), and did not control for all the non-therapy factors that influence recovery. The fourth study, by Nouwens et al. (2017), only observed significant effects of therapy (versus no therapy) after removing 71% of the participants who were unable to complete the intervention. The authors therefore noted that the therapy benefit should be interpreted with great caution, because patients who were unable to tolerate intensive therapy were only excluded from the therapy group. When the groups were better matched, speaking outcome did not differ for the therapy versus no therapy groups.

No significant advantage of early high- versus low-intensity therapy was found in the remaining four studies. The fifth study, by Laska et al. (2011), compared no therapy to 3.75 hours per week of therapy that predominantly targeted language comprehension. Unsurprisingly, this resulted in no significant benefits on speech production, the primary outcome. The sixth study, by Bowen et al. (2012b), provided standard clinical therapy at a lower intensity (1.3 hours per week) than those studies by Godecke et al. (2012), Mattioli et al. (2014), and Nouwens et al. (2017), which found significant therapy effects (2.5, 5 or 7 hours per week). Moreover, because this was typical clinical therapy delivered through the UK National Health Service, the 1.3 hours of speech and language therapy per week included broader contact time (including, for example, multidisciplinary team planning) rather than just intervention time. It may therefore be possible that the amount of therapy delivered to their patients was not enough to result in a detectable improvement. The seventh study, by Woldag et al. (2017), found no significant advantage of 15 hours per week of Constraint-Induced Aphasia

Therapy (CIAT) compared to 7 hours per week of conventional therapy. This outcome led the authors to conclude that 7 hours of conventional communication therapy was as effective as 15 hours of CIAT but there was no demonstration of the effectiveness of 7 hours of conventional communication therapy compared to no therapy. The eighth study, by Godecke et al. (2020), was a randomised controlled trial borne from the positive significant effect of early therapy found in their pilot trials described above (Godecke et al. 2012 and 2014). Contrary to their expectations, the authors found no significant advantage of early intensive therapy (5 hours per week) compared to usual care (2.3 hours per week). Whilst both patient groups improved over time, the lack of a no-therapy control makes it impossible to infer that the improvement was the consequence of therapy.

It is also relevant to note that sample size was at least twice as large in studies showing non-significant intensity effects as studies showing significant intensity effects. For example, the studies with non-significant therapy intensity effects by Laska et al. (2011), Bowen et al. (2012b), and Godecke et al. (2020) included 60-147 participants per group, whereas only 6-30 participants per group were included in the studies with significant therapy intensity effects by Godecke et al. (2012), Godecke et al. (2014), Mattioli et al. (2014), and Nouwens et al. (2017). This raises two possibilities: first, that group differences in therapy outcomes might be confounded by differences in the patients' potential for spontaneous recovery, and second, that the influence of outliers is greater where sample sizes are smaller. Both of these can be better controlled with larger sample sizes.

In summary, the evidence for the benefit of early therapy is currently weak. Inconsistent conclusions across studies might reflect discrepancies in (i) what constitutes a high versus low intensity therapy; (ii) participant numbers, (iii) how well groups are matched for non-therapy factors that are known to influence recovery (e.g. initial severity and lesion site), (iv) the criteria used to determine therapy time and (v) the type of therapy administered.

### 6.2.2 The effect of later therapy

I identified another 7 studies that compared high- versus low-intensity speech and language therapy at a later stage post-stroke, i.e. more than one month after stroke onset (studies 9-15 in Table 6.1). Only one of these studies, by Breitenstein et al. (2017), showed a significant benefit of high-intensity therapy.

This was a well-powered study, which reported a medium effect size (Cohen's d = 0.58), with 78 participants per group, conducted 2 to 3 years after stroke for a period of 4-5 weeks, with 10 hours or more per week of therapy in the high-intensity group and 1.5 hours per week of therapy in the low-intensity group (see Table 6.2).

Five of the other 6 studies (that did not find a significant therapy effect) had (i) fewer participants (9-18 per group) and (ii) a higher amount of therapy in the low intensity therapy group (1.8-2.5, 2, 3, 6, 6 hours per week). These were reported by Dignam et al. (2015)., Kesav et al. (2017), Denes et al. (1996), Martins et al. (2013), and Stahl et al. (2018). The remaining study, by Bakheit et al. (2007), included more participants per group (approximately 50 in each) but the high therapy intensity (4.3 hours per week) was less than half that of the Breitenstein study (2017) showing a significant effect (10 hours or more per week).

Contrary to expectation, two studies of later therapy by Kesav et al. (2017) and Dignam et al. (2015) found that the low-intensity intervention was significantly more beneficial than the high-intensity intervention. Specifically, better speaking outcomes were reported after 3 hours per week than 6 hours per week of therapy in Kesav et al. (2017) and 6 hours per week than 16 hours per week of therapy in Dignam et al. (2015). These findings suggest that too much therapy is at best unnecessary, and at worst, disadvantageous in sub-acute or chronic stroke patients, for example, when the benefits of high-intensity therapy are outweighed by the higher dropout rate (Brady et al., 2016, Bakheit et al., 2007). Furthermore, there may be a "therapy ceiling effect", which, if reached, may result in patients ceasing to improve, and potentially even declining (Stahl et al., 2018).

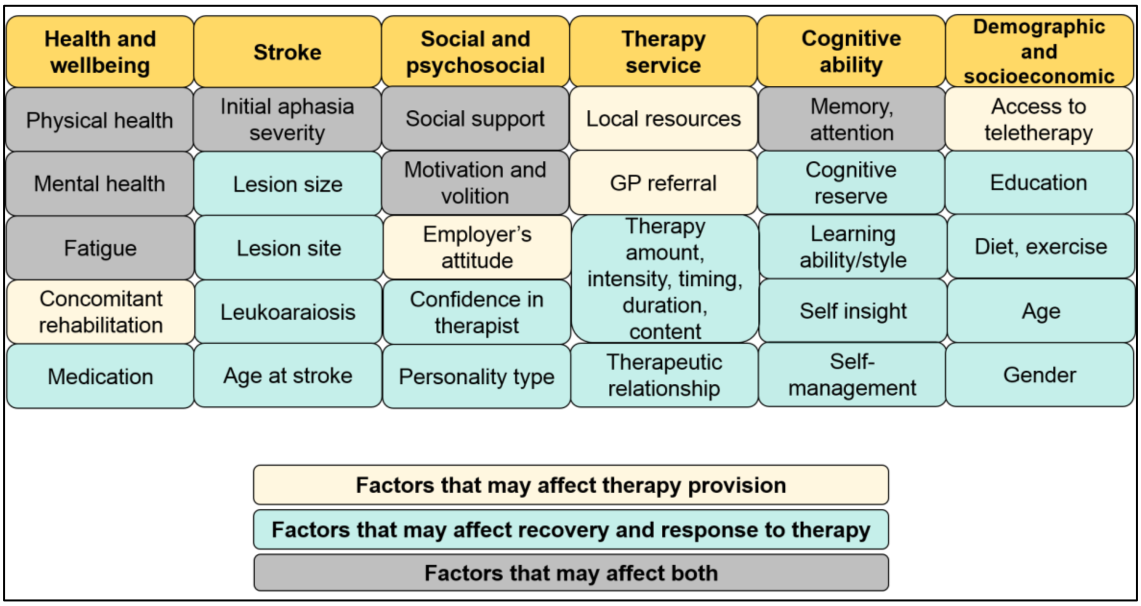
### 6.2.3 Summary of prior evidence & motivation for the current experiment

Overall, the review of studies presented above, that compared high- versus low-intensity speech and language therapy in chronic stroke patients, only found inconclusive evidence that high-intensity therapy is beneficial. The strongest results come from Breitenstein et al. (2017) who administered their therapy 2-3 years after stroke. The results of early therapy studies were more consistent but did not provide evidence for long-term maintenance of the early therapy benefit. No studies that directly compared the effect of early- versus late-initiated therapy on speaking outcomes were identified, although a review by Robey (1998) suggested that early therapy might be more beneficial.

This experiment investigated how speaking outcomes in stroke survivors with aphasia differed in those who did and did not receive clinical speech and language therapy, administered by health services, in the first month after stroke; and whether any benefit of therapy was enhanced by the number of hours received. In contrast to previous studies, I investigated (1) clinical therapy rather than experimental therapy, and (2) long-term maintenance of the therapy gains. I controlled for (i) initial severity in the first week after stroke; (ii) intervening therapy between one month and one year and (iii) left and right hemisphere lesion size, all whilst maintaining a large sample size (n=143). Motivated by the absence of a therapy effect on overall language ability in Experiment 1, this study focused specifically on objectively-measured speech production abilities (measured late post-stroke), and on participant-reported speaking ability (measured early post-stroke).

By controlling for multiple known predictors, the study was sensitised as closely as possible to the effects of early clinical therapy. If recovery is better in participants who (A) did, versus did not, receive early therapy and (B) had greater versus fewer hours of early therapy, it could be inferred that the early therapeutic activities had a beneficial effect. If only (B) is true, it could be inferred that early therapy had a beneficial effect when the hours of therapy were sufficiently high. If only (A) is true, it could be inferred that either the participants benefitted from the early provision of self-managed relearning strategies, or that therapy was only available or suitable for participants who had the most chance of recovery (e.g. those with better physical and/or mental health); see Figure 6.1. Both interpretations have important implications for understanding how the effect of speech and language recovery can be predicted and optimised in the future.

**Figure 6.1** Factors that may affect the provision of speech therapy, and recovery and response to therapy.



| **Health and wellbeing** | **Stroke** | **Social and psychosocial** | **Therapy service** | **Cognitive ability** | **Demographic and socioeconomic** |
|---|---|---|---|---|---|
| Physical health | Initial aphasia severity | Social support | Local resources | Memory, attention | Access to teletherapy |
| Mental health | Lesion size | Motivation and volition | GP referral | Cognitive reserve | Education |
| Fatigue | Lesion site | Employer's attitude | Therapy amount, intensity, timing, duration, content | Learning ability/style | Diet, exercise |
| Concomitant rehabilitation | Leukoaraiosis | Confidence in therapist | | Self insight | Age |
| Medication | Age at stroke | Personality type | Therapeutic relationship | Self-management | Gender |

**Factors that may affect therapy provision**

**Factors that may affect recovery and response to therapy**

**Factors that may affect both**

The effect of early clinical therapy on speaking ability was assessed using scores from both the objective CAT assessment and the participant-reported ARTQ, see Chapter 2 sections 2.4 and 2.5 for more details. The two major advantages of using the participant-reported data are: (i) the impact of therapy was assessed from the participant's perspective, and (ii) the time-points at which the effect of therapy was assessed (i.e. one month and one year post-stroke) were controlled across participants. The disadvantages of the ARTQ are discussed in detail in Chapter 2 section 2.5.2, but predominantly lie in their reliance on memory. Although participants were excluded if their memory of therapy time was vague or inconsistent, a false negative result (i.e. the absence of a therapy effect) could still occur if inter-participant variability in memory accuracy was greater than a true therapy effect. In contrast, false positives could result if participants were more likely to perceive an improvement over time when they received more, rather than less, therapy. To avoid such bias, speaking abilities were also measured using 3 summary scores from the CAT: Naming, Repetition, and Spoken Picture Description (see Chapter 2, section 2.4). The disadvantages of the objective scores are: (i) they were obtained at different time points post-stroke (months to years), depending on when the participant entered the study and, (ii) performance on any one task depends on many interrelated perceptual, cognitive and motor functions (the task impurity problem), therefore isolating speaking outcomes requires us to dissociate variance that is related or unrelated to speech production: e.g. auditory word repetition from auditory word comprehension and object naming from object recognition; dissociations I was able to address in this investigation.

A robust outcome would be one where an effect of early therapy on the ability to speak was observed in both the participant-reported and objective scores, after controlling for time post-stroke of test, and other cognitive abilities. In addition, unlike prior studies, left and right hemisphere lesion size were included as an estimate of the brain's capacity to recover (Thye and Mirman, 2018, Meier et al., 2019, Benghanem et al., 2019). If participants who did not receive early speech and language therapy had larger lesions, worse speaking outcomes may have been the consequence of less capacity for recovery (because of the larger lesions) rather than the absence of therapy. In addition, participants with larger lesions may have had less time and cognitive resources for speech and language therapy because of the presence of concomitant impairments that needed a variety of interventions, such as physiotherapy and occupational therapy.

## 6.3 Methods

### 6.3.1 Participant selection

This experiment included 143 stroke survivors. All were native English speakers who (i) had a left hemisphere lesion that was larger than 1cm$^3$, and (ii) were categorised as Severely or Moderately aphasic one week after their stroke (see Chapter 2, section 2.5). Critically, the selected participants had reported, via questionnaires, how many hours of therapy they had received and when this therapy took place. See Table 6.3 for participant characteristics.

**Table 6.3** Participant characteristics.

| | | **Severe** | | **Moderate** | |
|---|---|---|---|---|---|
| | | **No Therapy** | **Therapy** | **No Therapy** | **Therapy** |
| Group size | | 49 | 59 | 15 | 20 |
| Early therapy amount [a] | Mean (SD) | 0 | 8 (6) | 0 | 9 (8) |
| | Range | - | 1-20 | - | 2-30 |
| Later therapy amount [a] | Mean (SD) | 33 (29) | 37 (35) | 10.9 (12) | 38 (41) |
| | Range | 0-100 | 0-220 | 0-32 | 0-191 |
| Sex | Male (%) | 36 (73) | 43 (73) | 13 (87) | 15 (75) |
| | Female (%) | 13 (27) | 16 (27) | 2 (13) | 5 (25) |
| Pre-stroke handedness | Right (%) | 47 (96) | 49 (83) | 15 (100) | 14 (70) |
| | Left (%) | 2 (4) | 6 (10) | 0 (0) | 2 (10) |
| | Ambidextrous (%) | 0 (0) | 4 (7) | 0 (0) | 4 (20) |
| Age at stroke (years) | Mean (SD) | 56 (14) | 57 (13) | 61 (9) | 57 (11) |
| | Range | 33-85 | 30-82 | 45-73 | 37-70 |
| Years post-stroke of CAT | Mean (SD) | 4 (4) | 4 (4) | 4 (3) | 4 (4) |
| | Range | 0-23 | 1-17 | 1-9 | 1-17 |
| Lesioned hemisphere | Left (%) | 42 (86) | 50 (85) | 10 (66) | 18 (90) |
| | Both (%) | 7 (14) | 9 (15) | 5 (33) | 2 (10) |
| Left hemisphere lesion size ($cm^3$) | Mean (SD) | 70 (70) | 73 (59) | 32 (37) | 43 (55) |
| | Range | 3-355 | 1-235 | 2-119 | 1-194 |
| Initial understanding severity [b] | Severe (%) | 14 (29) | 19 (32) | 1 (7) | 0 (0) |
| | Moderate (%) | 14 (29) | 10 (17) | 4 (27) | 8 (40) |
| | Mild (%) | 5 (10) | 13 (22) | 7 (47) | 7 (35) |
| | Normal (%) | 15 (31) | 16 (27) | 3 (20) | 5 (25) |

**Legend:** [a] Therapy amount was measured in hours. [b] Initial understanding severity scores are missing for: 1 in the No Therapy-Severe group and 1 in the Therapy-Severe group.

Participants with Mild initial difficulties were excluded, because, of the 63 participants with therapy data and Mild initial severity: (i) only 3 (5%) received more than 4 hours of early therapy (within the first month) and (ii) only 12 (19%) of those who did not receive any early therapy failed to recover to normal within a year post-stroke. Therefore, there was insufficient data to assess the effect of therapy in these participants. As with Experiments 1-3, participants were excluded if they reported being medically unwell in the first week after their stroke - because the severity of their aphasia could not be assessed, and they would not have been well enough to engage in, or benefit from, therapy.

### 6.3.2 Speech and language therapy

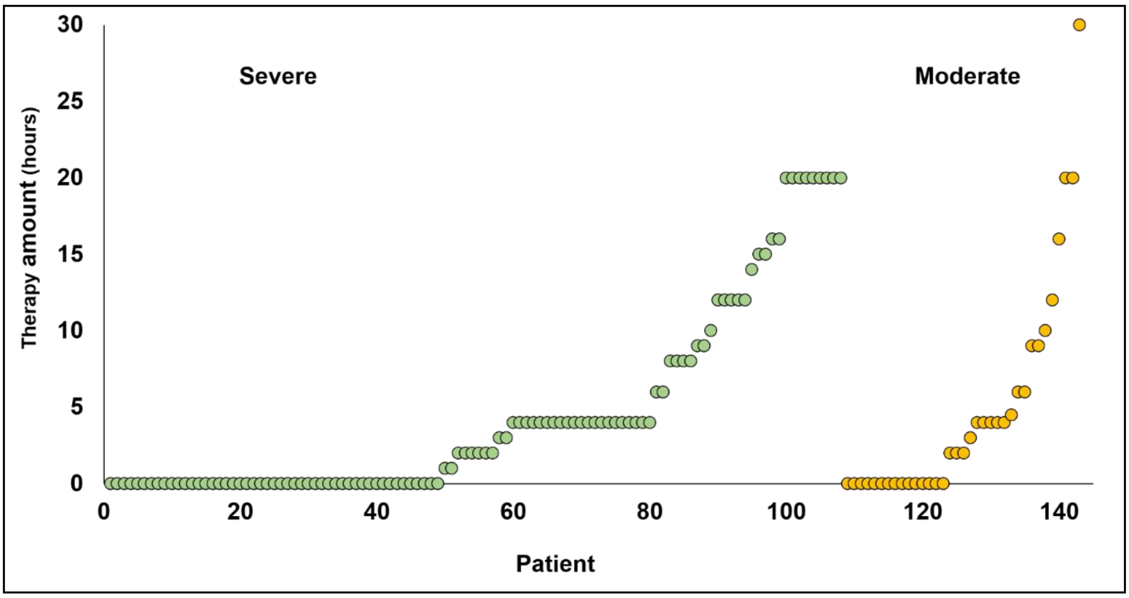
Each participant provided details of: (i) how many hours of speech and language therapy they had received, (ii) when speech and language therapy was received (start and end points, and frequency) and (iii) therapy activities (see Chapter 2, section 2.5). Participants are supported in providing this information by a research speech and language therapist, who facilitates both the participants' understanding of the questions, and their responses, and helps them to differentiate direct language therapy from other typical therapist input delivered during the acute stage post-stroke; e.g. assessment, dysphagia management, or information and monitoring.

The speech and language therapy received can be described as 'clinical' (i.e. part of routine care) rather than 'experimental' (i.e. delivered as part of a research project). Speech and language therapy provision in the UK is guided by NICE (National Institute for Health and Care Excellence) clinical guidelines, which recommend screening for communication impairments within 72 hours of stroke (NICE, 2023). The current guidelines do not stipulate a set amount of speech therapy; but rather recommend 3 hours of multidisciplinary input per day, for 5 days per week, which includes speech, physio and occupational therapies. At the time these participants were recruited to the PLORAS study, the guidelines recommended 3-4 hours of speech therapy per week (NICE, 2013). Further rehabilitation can then be offered for those who can participate and continue to make functional gains. There are no additional specific recommendations for the timing, amount, intensity and duration of therapy. If communication difficulties persist at 6-month or annual review, patients can be re-referred to speech therapy services, and offered further treatment if they have the potential for functional improvement. Clinical aphasia therapy typically comprises two broad approaches: 1) functional, and 2) impairment-based, targeting expressive and/or receptive language as required. All therapy aims to reduce impairment severity, increase communicative ability, and/or implement alternative methods of communicating, as well as to provide information and support to both the patient and the carer or communication partner. Impairment-based therapy may target speech sounds, single words, or sentences, and therapists provide support and guidance through strategies such as modelling and cues.

As these participants were recruited from a broad range of geographical locations and clinical settings across England, their experience of rehabilitation differs due to regional variation in service provision. The benefit of this is that the data accurately reflects typical clinical speech and language therapy provision. The disadvantage is that it was impossible to control for therapy content or approach in the sample. There are many ways a speech therapist could have contact with a patient which do not include language therapy; e.g. assessment time or dysphagia management. Nonetheless, because all these participants had difficulties producing speech in the first week after their stroke, I could assume that the clinical therapy they received included some therapy or strategies aimed to improve their speaking difficulties. In summary, it is highly likely that therapy included activities to improve speaking ability, although the actual amount of therapeutic input may be over-estimated.

The amount of therapy received by each of the participants included in this experiment can be seen in Figure 6.2, and it can be observed that the majority of the participants received less than the recommended 3-4 hours per week Nevertheless, it might be close to the actual amount of therapy typically received, given that recent national audit data shows that the target amount of therapy is frequently unmet due to time spent on administration (therapy planning and documenting), and patient factors (medical instability and fatigue) (Royal College of Physicians, 2017, Clarke et al., 2018). There may also be a regional disparity in the amount of therapy provided by the National Health Service in the UK due to differences in service structure (SSNAP). This experiment capitalised on this regional variability to reveal the influence of therapy.

**Figure 6.2** Amount of early therapy (hours) reported by each participant according to whether the initial severity of symptoms was Severe or Moderate.



An advantage of clinical therapy, over experimental therapy, is that clinical therapy is tailored around the patient's personal needs, goals and learning style. A natural disadvantage follows, namely that therapy differs for everybody, even if the overall goal of improving speaking remains the same. Another disadvantage is that the amount of clinical therapy received is typically lower than that reported in experimental trials. Although fewer hours of therapy could have desensitised the experiment and resulted in false negative results, (i) this was found not to be the case and (ii) by analysing effects of what is typically received, rather than ideally received, I reduce the gap between what is often researched, and what is actually being delivered. Focusing retrospectively on clinical therapy also allows us to assess how well participants recovered after receiving no speech and language therapy, without facing ethical issues related to prospectively withholding therapy.

### 6.3.3 Behavioural assessment

This study selected three CAT summary scores on which to evaluate therapy effects: (1) Repetition; (2) Naming, and (3) Spoken Picture Description (see Chapter 2, section 2.4). Additionally, I controlled for the impact of perceptual and semantic impairments on the ability to perform these speaking tasks by factoring out performance on two other CAT tasks: (4) spoken word comprehension (because poor auditory word comprehension will affect the ability to repeat words); and (5) semantic memory (because poor performance on this task will affect the ability to name objects and describe pictures).

### 6.3.4 Participant-reported measures of speaking ability

108 participants were categorised as Severely aphasic one week post-stroke, and 35 were categorised as Moderately aphasic. None of the 143 participants had inconsistent reports of ability between carer and participant (15 had already been excluded for this reason).

Unlike Experiments 2 and 3, which reported proportional recovery scores, this experiment reported actual recovery (see Chapter 2, section 2.5.3). This was because initial speaking severity was reasonably controlled (all had either Severe or Moderate initial symptoms), such that participants could be analysed together. Additionally, raw recovery scores represent a clinically meaningful improvement, being both functionally important, and indicating a reduction in impairment, which will be essential when considering the impact of therapy. ‘Recovery’ was therefore calculated as any change in participant-reported score between two timepoints. In statistical analyses, this was treated as a binary outcome (see below in section 6.3.5 for more details). For illustrations, the raw recovery is shown. As described in Chapter 2, the maximum raw recovery score is 3 for Severe participants (Severe to Moderate, Moderate to Mild, and Mild Normal) and 2 for Moderate participants (Moderate to Mild and Mild to Normal).

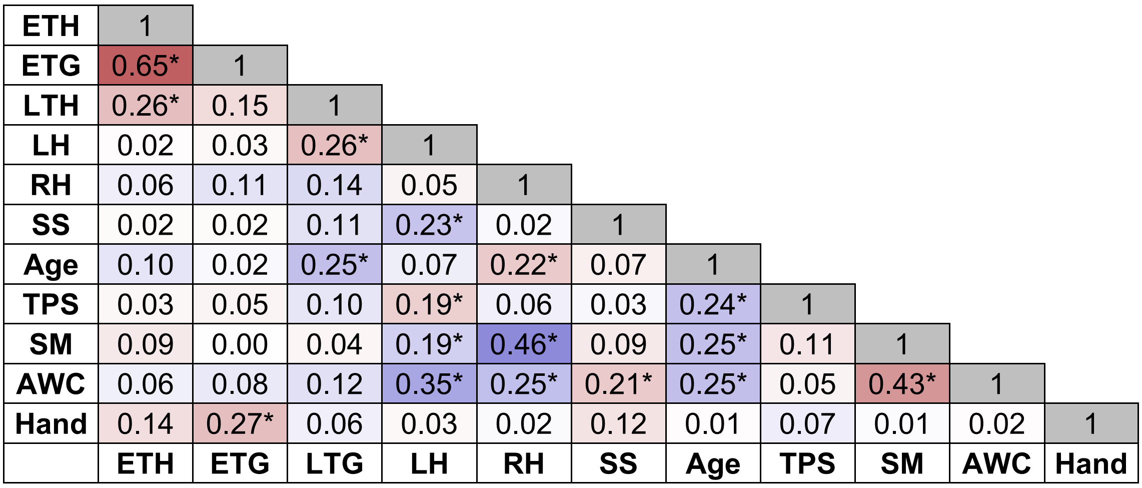
### 6.3.5 Statistical analysis of therapy effects on behaviour

Two different types of regression analysis were used for the two different outcome measures. Binary logistic regression was used to assess the effect of therapy on participant-reported improvement from one week to one month post-stroke, and from one month to one year post-stroke (i.e. a binary outcome of improvement or no improvement). Multiple linear regression was used to assess the effect of therapy on the linear CAT scores. Each of these 5 analyses were performed twice (10 analyses in total) with therapy either treated as a binary variable (presence versus absence of therapy in the first month after stroke) or a continuous variable (number of therapy hours in the first month after stroke).

All 10 analyses factored out variance of no interest by including the following covariates: initial severity (Severe or Moderate), age at stroke, handedness, left hemisphere lesion size and right hemisphere lesion size. In addition, amount of later therapy (between one month and one year) was factored out of all analyses except improvement between one week and one month. For the analyses using CAT scores, I also factored out the time post-stroke that the CAT was administered, semantic memory scores (to control for object recognition and semantic impairments) and auditory word comprehension scores (to control for auditory perception and comprehension).

There was minimal collinearity in the data because none of the analyses yielded Tolerance values of less than 0.1, nor Variance Inflation Factor (VIF) values that were greater than 10. Moreover, none of the independent variables have correlation values greater than 0.7 (see Table 6.4).

**Table 6.4** Correlation matrix for the 11 covariates included.



| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **ETH** | 1 | | | | | | | | | | |
| **ETG** | 0.65* | 1 | | | | | | | | | |
| **LTH** | 0.26* | 0.15 | 1 | | | | | | | | |
| **LH** | 0.02 | 0.03 | 0.26* | 1 | | | | | | | |
| **RH** | 0.06 | 0.11 | 0.14 | 0.05 | 1 | | | | | | |
| **SS** | 0.02 | 0.02 | 0.11 | 0.23* | 0.02 | 1 | | | | | |
| **Age** | 0.10 | 0.02 | 0.25* | 0.07 | 0.22* | 0.07 | 1 | | | | |
| **TPS** | 0.03 | 0.05 | 0.10 | 0.19* | 0.06 | 0.03 | 0.24* | 1 | | | |
| **SM** | 0.09 | 0.00 | 0.04 | 0.19* | 0.46* | 0.09 | 0.25* | 0.11 | 1 | | |
| **AWC** | 0.06 | 0.08 | 0.12 | 0.35* | 0.25* | 0.21* | 0.25* | 0.05 | 0.43* | 1 | |
| **Hand** | 0.14 | 0.27* | 0.06 | 0.03 | 0.02 | 0.12 | 0.01 | 0.07 | 0.01 | 0.02 | 1 |
| | **ETH** | **ETG** | **LTG** | **LH** | **RH** | **SS** | **Age** | **TPS** | **SM** | **AWC** | **Hand** |

**Legend:** ETH = Early therapy hours. ETG = Early therapy group. LTH = Later therapy hours. LH = left hemisphere lesion size. RH = right hemisphere lesion size. SS = initial speaking severity. Age = age at stroke. TPS = time post-stroke of CAT. SM = Semantic memory. AWC = auditory word comprehension. Hand = handedness. Red shading indicates a positive correlation; blue shading indicates a negative correlation. * = correlation is significant at p=<0.05.

### 6.3.6 Lesion analyses

40 participants were imaged on a 3T Trio scanner, 42 on a 3T Prisma scanner, 23 on a 1.5T Sonata scanner, 37 on a 1.5T Avanto scanner, and 1 on a 3T Allegra scanner.

To investigate whether lesion site differed in participants who did and did not receive early clinical therapy, the whole brain fuzzy images were entered into a voxel-based morphometry, or VBM (Ashburner and Friston, 2000) analysis performed in SPM12, using the general linear model (Ashburner and Friston, 2000). Like other voxel-based lesion symptom mapping (VLSM) methods, VBM searches the whole brain for voxels where local brain structure varies with a symptom or other variable of interest (here the presence or absence of therapy or the amount of therapy). Many previous studies have demonstrated the sensitivity of VBM/VSLM but there are also limitations to the approach (Gajardo-Vidal et al., 2018, Zhang et al., 2014, Ivanova et al., 2020). For example, the effect of damage to one area may depend on the presence or absence of damage to another area. Understanding such combinatorics requires multivariate lesion analyses on large samples of participants but as these methods are still in their infancy and have numerous interpretation problems (Ivanova et al., 2020) they were not used in this experiment.

The fuzzy images used in the VBM analysis index the degree to which each voxel differs from the normal range and therefore avoids reliance on binary cut-offs (Gajardo-Vidal et al., 2018). The statistical model for the VBM was an ANOVA with four different independent groups of fuzzy images, with unequal variance: (1) Severe initial severity and early clinical therapy (n=59); (2) Severe initial severity and no early therapy (n=49); (3) Moderate initial severity and early clinical therapy (n=20); and (4) Moderate initial severity and no early therapy (n=15). The statistical contrasts computed the main effects of Therapy versus No Therapy (across initial severity), Severe versus Moderate initial severity (across therapy groups), and the interaction between therapy and initial severity.

To investigate whether lesion site differed in participants who received more versus less therapy, I repeated the four-group analysis again, this time including two covariates (i.e. ANCOVA): (1) number of hours of early therapy; and (2) number of hours of later therapy. The results from both analyses (ANOVA and

ANCOVA) are reported at a significance level of p=<0.05 corrected for multiple comparisons in extent or in height within the left hemisphere search volume, defined as all the voxels (36,037 = 288.3 $cm^3$) that were damaged in at least 10 of the 143 participants in the sample.

## 6.4 Results

### 6.4.1 The effect of early therapy on language outcomes

When early clinical therapy was treated as a binary measure, there were two significant effects: compared to no therapy, participants who received early clinical therapy had better (i) participant-reported improvement at one month and (ii) Naming scores (Table 6.5). A significant effect of hours of later therapy was also seen on Naming (p=0.036) at the time of the CAT assessment.

When early clinical therapy was treated as a continuous measure, participant-reported improvement at one year was positively related to the number of hours of early therapy. This was observed across the whole sample (n=143, see Table 6.5). These results remained significant when the one participant who received an outlying amount of early therapy (30 hours, see Figure 6.2 in the Moderate initial severity group) was removed from the analyses. Furthermore, the results did not change when participants with outlying scores on Naming (n=6) and Repetition (n=4) were also removed from the relevant analyses (see Table 6.6). There were no significant improvements related to later therapy in the analyses using continuous measures of early therapy.

**Table 6.5** Outcomes in 143 participants who reported (i) whether they did versus did not receive early therapy and (ii) varying amounts of early therapy.

| Outcome measure | N | $R^2$ | β | p value |
|---|---|---|---|---|
| Binary analyses: Early therapy versus No early therapy | | | | |
| Participant-reported improvement at one month | 143 | 0.203 | 2.256 | 0.045 |
| Participant-reported improvement at one year | 142 | 0.186 | 1.779 | ns |
| Naming | 142 | 0.557 | 0.158 | 0.013 |
| Repetition | 143 | 0.462 | 0.109 | ns |
| Spoken picture description | 142 | 0.499 | 0.043 | ns |
| Continuous analyses: hours of early therapy | | | | |
| Participant-reported improvement at one month | 143 | 0.172 | 1.018 | ns |
| Participant-reported improvement at one year | 142 | 0.222 | 1.117 | 0.045 |
| Naming | 142 | 0.540 | 0.074 | ns |
| Repetition | 143 | 0.460 | 0.099 | ns |
| Spoken picture description | 142 | 0.499 | 0.044 | ns |

**Legend**: Covariates: initial severity (Severe or Moderate), handedness, age at stroke, left hemisphere lesion size, right hemisphere lesion size (all models), amount of later therapy (all models except ‘improvement at one month’), semantic memory CAT scores, auditory word comprehension CAT scores, and time post-stroke of CAT (for analyses of Naming, Repetition and Spoken Picture Description). β = standardised beta. ns = not significant at p=<0.05.

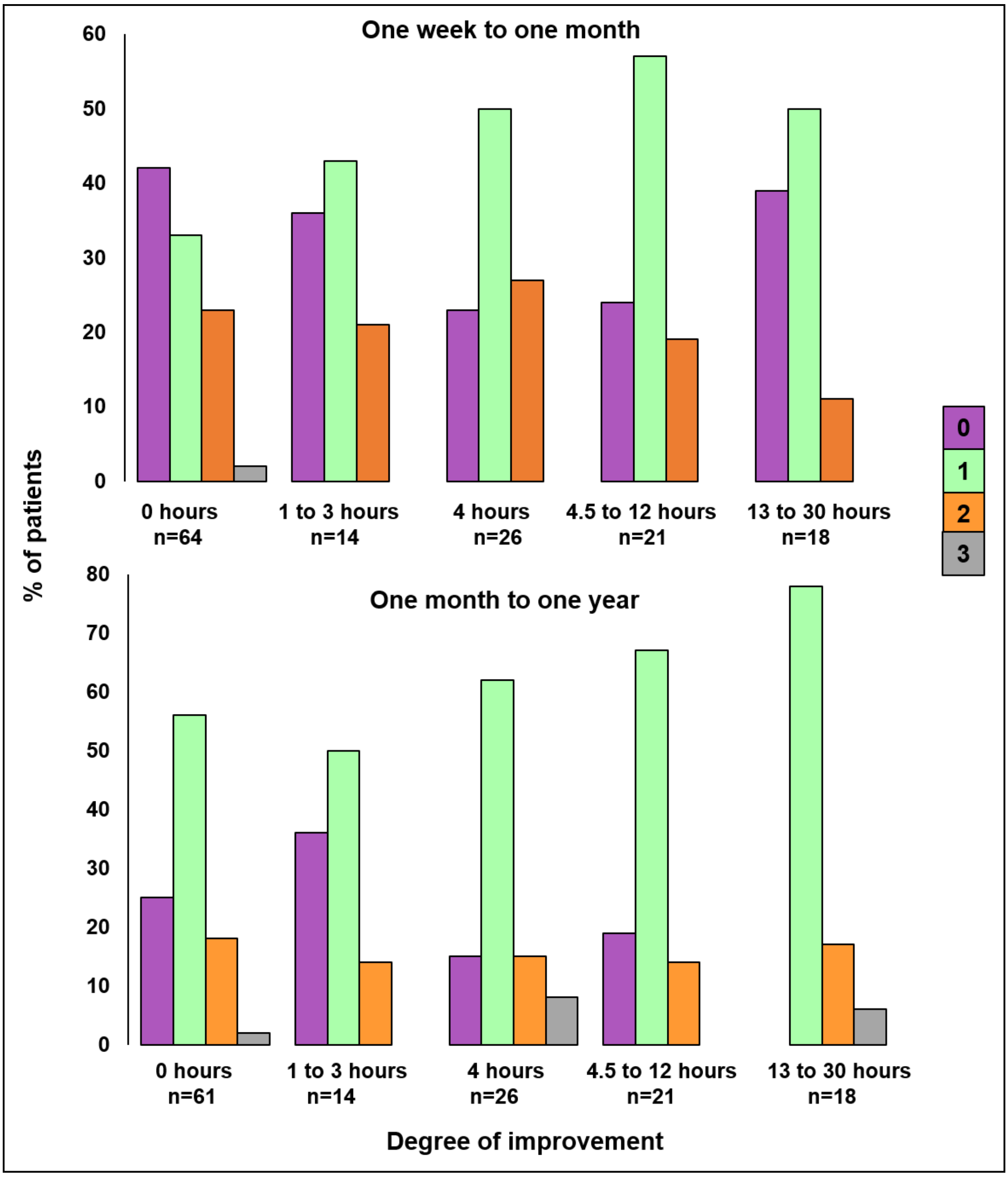
**Table 6.6** Outcomes in participants who reported (i) whether they did versus did not receive early therapy and (ii) varying amounts of early therapy, with outliers removed from analyses.

| Outcome measure | N | $R^2$ | β | p value |
|---|---|---|---|---|
| Binary analyses: Early therapy versus No early therapy | | | | |
| Participant-reported improvement at one month | 142 | 0.216 | 2.455 | 0.030 |
| Participant-reported improvement at one year | 141 | 0.183 | 1.765 | ns |
| Naming | 135 | 0.446 | 0.189 | 0.009 |
| Repetition | 138 | 0.414 | 0.120 | ns |
| Spoken picture description | 141 | 0.502 | 0.048 | ns |
| Continuous analyses: hours of early therapy | | | | |
| Participant-reported improvement at one month | 142 | 0.187 | 1.039 | ns |
| Participant-reported improvement at one year | 141 | 0.218 | 1.117 | 0.046 |
| Naming | 135 | 0.419 | 0.074 | ns |
| Repetition | 138 | 0.415 | 0.125 | ns |
| Spoken picture description | 141 | 0.504 | 0.068 | ns |

**Legend:** Covariates were the same as in Table 6.5. One outlier who received 30 hours of therapy was removed from all models. Six participants with outlying Naming scores were removed from Naming analyses. Four participants with outlying Repetition scores were removed from Repetition analyses. β = standardised beta. ns = not significant at p=<0.05.

The differing one month and one year outcomes after varying therapy amounts are illustrated in Figure 6.3. At one month (upper part of Figure 6.3), the proportion of participants who did not improve is greater for those who received 13-30 hours of early compared to those who received 4 hours. In contrast, at one year (lower part of Figure 6.3), the proportion of participants who did improve is higher after 13-30 hours of early therapy compared to 4-12 hours of early therapy.

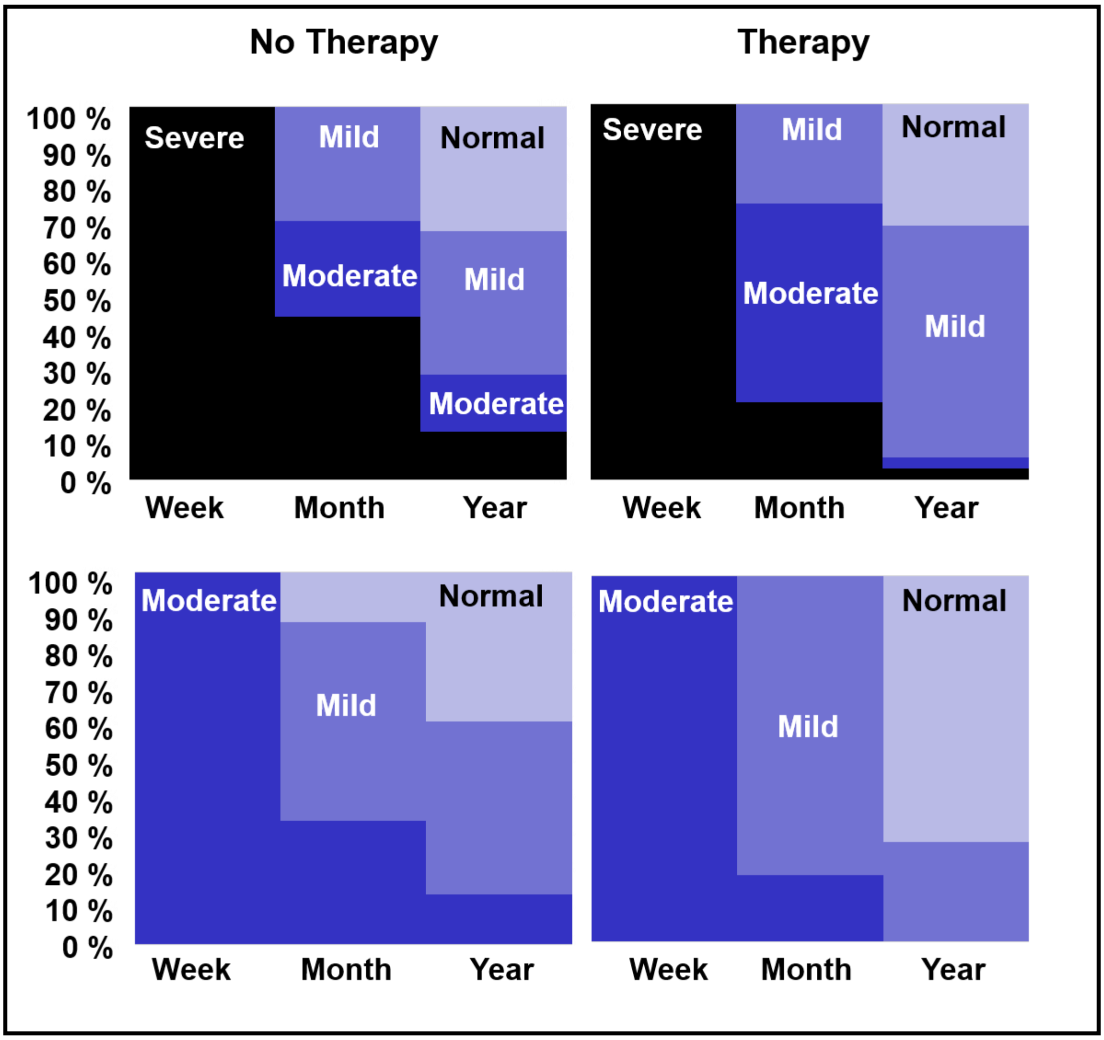
**Figure 6.3** Raw improvement in participant-reported speaking ability from one week to one month post-stroke, and one month to one year post-stroke, according to five different participant-reported therapy amounts.



**Legend:** The colour of the bar indicates the degree of improvement: 0 (purple) = no change in score, 1 (green) = a change of one (e.g. Severe to Moderate, Moderate to Mild), 2 (orange) = a change of two (Severe to Mild, Moderate to Normal) and 3 (grey) = a change of three (Severe to Normal).

Figure 6.4 illustrates participant-reported outcomes from one week post-stroke through to one year post-stroke in four different groups of participants who either had: (1) Severe initial severity and early clinical therapy (n=39), (2) Severe initial severity and no early clinical therapy (n=34), (3) Moderate initial severity and early clinical therapy (n=15), or (4) Moderate initial severity and no early clinical therapy (n=11). To maximise between-group differences in the amount of early therapy, 14 participants who received 1 to 3 hours of early therapy were excluded. To focus on the influence of early therapy, 30 participants who received 50 hours or more of later therapy (between one month and one year) were also excluded. After these exclusions, the average number of hours of early therapy was: 9 (Severe group, with therapy) and 7 (Moderate group with therapy), compared to zero in both the groups with No Therapy. The average amount of later therapy was 21.7 (Severe group with early therapy), 20.5 (Severe group with no early therapy), 22.5 (Moderate group with early therapy) and 10.9 hours (Moderate group with no early therapy. It was not possible to match later therapy in the Moderate groups more closely because of the smaller group size.

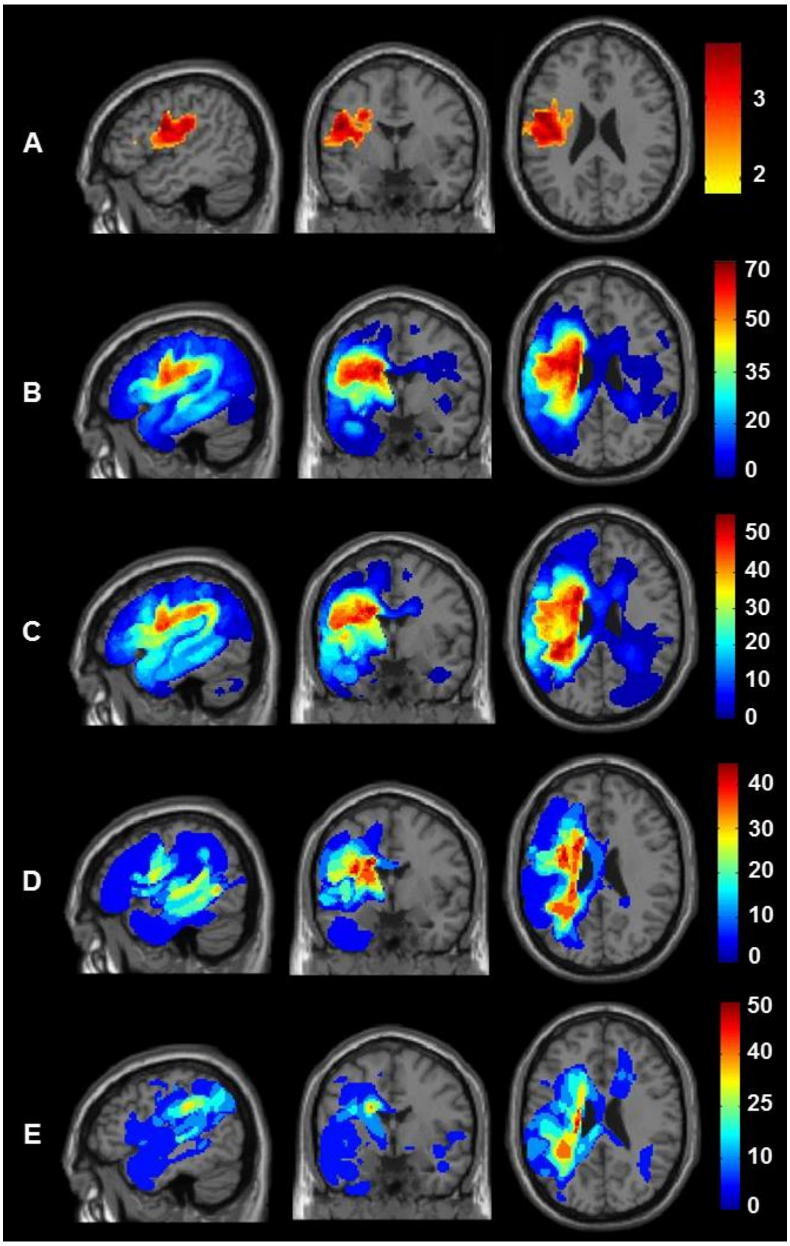
**Figure 6.4** Degree of improvement for four participant groups with either Severe (top) or Moderate (bottom) initial severity, and who reported either no early therapy (left) or more than 4 hours of early therapy (right).



### 6.4.2 Lesion analyses

The lesion analyses showed that poorer outcomes in participants who did not receive early therapy could not be explained by (i) larger lesion sites or (ii) more damage to regions required for recovery because global and local brain structure did not vary between therapy groups; nor did it depend on the number of hours of early therapy received. The same analyses were, nevertheless, highly sensitive to two other effects. First, in both analyses (with and without the therapy covariates), participants who had Severe initial severity had more extensive damage to the left premotor cortex and underlying white matter than participants with Moderate initial severity (see row A of Figure 6.5). This was significant, after correction for multiple comparisons within the left hemisphere search volume in both height and extent (peak Z score = 4.5 at MNI co-ordinates [-48, -2, +24], with 3679 voxels at p=<0.001 uncorrected). Second, in the analysis with early and later therapy added as covariates, there was a significant effect of later therapy: Participants who received more hours of later therapy had more extensive lesions in the white matter beneath the left superior temporal gyrus. This was significant, after correction for multiple comparisons within the left hemisphere search volume in both height and extent (peak Z score = 4.34 at MNI co-ordinates [-42, -40, +8], with 384 voxels at p=<0.001 uncorrected). The latter result suggests that participants with damage involving this part of the left temporal lobe required, or were offered, more therapy.

**Figure 6.5** Comparison of lesion sites across the different groups.



**Legend:** Lesion site results after assigning the 143 participants to four groups according to initial severity (Severe versus Moderate) and the provision of early therapy (Therapy versus None). Top row (A) shows the brain regions where damage was significantly greater in those with Severe versus Moderate initial severity. The sagittal, coronal and axial slices are at the co-ordinates of the most significant group difference (-48, -2, +24). Rows B to E show the lesion overlap maps (showing the frequency of damage) for the four groups at the same co-ordinates. There were no significant differences in lesion site between the groups. From top to bottom the groups are: (B) Participants with Severe initial severity who received Therapy (n=59); (C) Participants with Severe initial severity who did not receive therapy (n=49); (D) Participants with Moderate initial severity who received Therapy (n=20); (E) Participants with Moderate initial severity who did not receive therapy (n=15). The colour scale indicates percentage of participants for each group.

## 6.5 Discussion

In a review of the literature comparing the effect of speech and language therapy on speaking outcomes, I found no evidence that therapy in the first month after stroke resulted in long-term behavioural gains. The majority of these studies investigated experimental therapy, in varying amounts – but even the study by Bowen et al. (2012b), which closely resembled the current experiment by investigating a relatively little amount of early clinical therapy, did not find a significant benefit of therapy. I hypothesized that these studies might not have been able to detect the beneficial effect of therapy because of their small sample sizes, or because they did not control for the multiple non-therapy factors that are known to influence recovery, including initial severity, left hemisphere lesion size, age at stroke, therapy received between the intervention period and the follow-up time point, and other neuropsychological impairments. When all these factors were controlled, I found evidence for both short- and long-term speaking improvements in those who received early therapy.

In the introduction to this chapter, I distinguished between three different scenarios. The first is when better recovery is observed in participants who (A) did versus did not receive early clinical speech and language therapy (i.e. binary therapy analyses), and (B) had greater versus fewer hours of early therapy (i.e. continuous therapy analyses). I did not find an effect of both (A) and (B) on the same outcome measure, therefore this scenario is not relevant here.

The second scenario is when only (B) is true (better recovery in participants who had greater versus fewer hours of early therapy). The positive relationship I found between the number of hours of early therapy and participant-reported outcome at one year post-stroke may be explained by therapy being more beneficial when the amount delivered is sufficiently high. At one month post-stroke, proportionally more participants recovered if they received 4-12 hours of therapy compared to 0-3 hours of therapy (see Figure 6.3) but this did not reach significance.

The third scenario is when only (A) is true (i.e. better recovery in participants who did, versus did not, receive early clinical speech and language therapy). This was the case for one month outcomes, and long-term Naming ability.

Below, I discuss alternative interpretations of the findings in relation to these two scenarios, along with the limitations of the study, and new directions for future studies.

### 6.5.1 Interpreting the long-term benefits in those who received early therapy

The following discussion offers two different speculative explanations that might individually, or in combination with one another, explain these findings. First, a few hours of early therapy in the first month after stroke may be sufficient to (i) impart learning strategies that can be used by the patient with their families to facilitate recovery, (ii) motivate participants to self-manage extensive practice and training of functional skills, and/or (iii) encourage and facilitate participants to join communication support groups that positively influence their recovery. Such activities could explain why a significantly greater improvement was found for participant-reported outcomes from one month to one year (i.e. after the early therapy period) when participants received more early therapy. In other words, the suggestion is that the participants who received therapy may have continued to practise and benefit from the self-management strategies that they were taught early post-stroke, compared to those who did not receive therapy and who had little/no guidance for recovery. Examples of self-management strategies that these participants described include (i) reading aloud from newspapers or magazine articles, (ii) identifying personally-relevant key words around the home, and increasing their salience in daily conversation, and (iii) using school literacy books. If this interpretation is correct, it motivates future research to understand more about these therapeutic activities, both those done in a clinical setting with a therapist, and those done outside of the clinic (including activities not traditionally seen as ‘therapy’, such as attending a stroke club).

Second, better outcomes in participants who received early clinical therapy were not the consequence of the therapy itself but instead reflect other factors which influence early therapy provision (or lack thereof). Patients may be prioritised over others for therapy if they either have greater potential to benefit, or better resources and/or social support in place to support their therapy (see Figure 6.1). For example, if a patient has better general health, greater motivation and attention, early signs of language improvement, or extra family support, therapists may use these assets to the patient’s advantage, capitalising on their recovery potential. If those patients who received early therapy also enjoyed a positive therapeutic relationship, it may have further encouraged and enhanced the two-way engagement needed for successful rehabilitation (Lawton et al., 2018). On the other hand, patients who have poorer health, concomitant impairments, less motivation, or who are socially and/or technologically isolated may not be as able to engage in therapy as quickly, or may need rehabilitation from other professionals first – delaying speech and language therapy. These patients may also require further support to enhance their volition to engage with the therapy (Hart et al., 2019).

If this interpretation is correct, it highlights the need to understand in much greater detail the reasons why therapy may be prioritised or delayed – be they patient factors (e.g. cognitive, health, social etc.) or service delivery factors (e.g. limited therapy resources, long waiting lists, logistics of reaching patients in the community etc.), see Figure 6.1. Whilst we already have some understanding of why patients do not receive the recommended amounts of therapy (e.g. medical instability, fatigue, administration, and planning (Clarke et al., 2018)), a holistic and multidisciplinary understanding of speech therapy prioritisation and resource allocation will provide new insights into the factors that need to be entered into predictive models of outcome after stroke (Godecke et al., 2013, Hope et al., 2013). This will enable better treatment planning, which keeps the patient at the core of the rehabilitation model and takes into account all of the factors which may influence their recovery.

Both of these speculative interpretations warrant further investigation, and highlight important directions for future studies. Regardless of which interpretation is correct, the early interaction between the patient and the therapeutic activity may harness and boost the neuroplastic changes that take place in the early weeks after stroke, thereby improving language outcomes beyond any spontaneous recovery (Nouwens et al., 2015) and leading to greater behavioural changes that are maintained over subsequent months.

### 6.5.2 Limitations and future directions

As with Experiments 1-3, it was impossible to control these retrospective analyses for the many variables known to affect language outcomes at the point of entry into the study. This was tackled by factoring out the influence of many variables that are not usually controlled (e.g. initial severity and lesion size). However, other sources of variance may have biased the results, as discussed below.

The reliability of the retrospective ARTQ has been discussed extensively thus far. With specific reference to the current experiment, there are three potential sources of inaccuracy: (i) the participant's memory for the therapy received and the timing of their recovery; (ii) insight into own ability that might be more accurate for those who make a better recovery; and (iii) optimism or pessimism bias, such that participants who reported receiving more therapy may have reported better language outcomes due to recalling a positive experience with the therapist and/or input, rather than reflecting genuine language improvement. These challenges were mitigated by supplementing the participant-reported outcomes with the objective measures of spoken language.

A second limitation is that variability was measured in the chronic rather than the acute stage post-stroke. Future studies would benefit from a detailed understanding of how the provision and benefits of early therapy depend on variables measured in the first week or two after stroke, including perceptual and cognitive impairments, initial severity of language symptoms, medical history, mental health, fatigue, attention, motivation, education, socioeconomic status and social support (see Figure 6.1). This would help to distinguish whether better outcomes in those who received therapy were the consequence of the therapy itself or a better potential to recover, or a combination of the two.

A third limitation is that, although there was no evidence that the Therapy and No Therapy groups differed in lesion size or damage to any part of the brain, it remains possible that those who made a poor recovery in the No Therapy group had a critical combination of damage that precluded recovery. This could be investigated in future by multivariate lesion analyses that are informed by prior knowledge of the critical combination of damage that impedes recovery.

A fourth limitation is the lack of specific detail about therapy content, duration and frequency, done both within and outside of the clinic. Participants' estimates of their therapy may not reflect the amount of actual therapeutic input received. For example, a therapy session reported to last one hour may involve very few therapeutic activities targeting speaking impairments, e.g. if the therapist spent more time supporting psychological well-being. Further studies should (i) focus on the proportion of time given to language-related therapeutic activities, rather than the less-exact ‘total number of hours’ measure, (ii) obtain more detail on the diverse contexts in which therapy may occur and (iii) ask patients to formally record their ‘therapy homework’ and other language practice activities. By quantifying the number of specific therapeutic inputs, a measure of cumulative intervention intensity can be calculated (Brogan et al., 2020), and evaluated alongside all the other factors found to influence inter-patient variability in recovery (e.g. physical and mental health, cognitive abilities, level of engagement, demographics, social support). Ultimately, this will improve predictions of outcomes and the response to therapy at the individual patient level (Aguilar et al., 2018). Such studies would also offer a deeper insight into the therapeutic ingredients driving positive behavioural change and lead to more precise guidelines in the type, content and duration of the therapy recommended to stroke survivors with severe aphasia.

Finally, although this study demonstrated that participants who received more hours of early therapy had better outcomes at one year post-stroke than those who received fewer hours, it is not possible to infer what the optimal number of hours would be because no participants in the sample reported receiving more than 30 hours of therapy in the first month.

### 6.5.3 Conclusions

In a relatively large sample of stroke survivors with moderate to severe aphasia, those who reported receiving clinical speech and language therapy in the first month after their stroke achieved better long-term speaking abilities than those who reported not receiving therapy. This was observed after controlling for: time post-stroke at test, left and right hemisphere lesion size, initial severity of speaking impairment, amount of later therapy and speech comprehension ability. These data suggest two, non-mutually exclusive, conclusions. The first is that early therapy may help patients develop self-management strategies that have a long-term benefit on their recovery. The second is that patients who receive more early clinical therapy have a better potential for recovery due to other influential factors. If this conclusion is correct, we need to first understand the reasons why patients do and do not receive early therapy, and then devise a framework for measuring and entering these reasons into predictive models of individual-level outcomes and response to therapy. This retrospective study also has implications for future therapy studies, and may inform directions for future trials: first by emphasising the factors that need to be controlled, second by identifying additional variables which may influence outcomes, and third by highlighting the benefits of combining objective behavioural assessments with participant-reported outcome measures of language function.

# Chapter 7: General Discussion

The overarching goal of this thesis was to further our current understanding of individual variation in aphasia outcomes and recovery, by: investigating (i) which variables have the biggest impact on a range of language outcomes and recovery in stroke survivors, and whether the influence of these variables is (ii) specific to the type of language function, (iii) dependent on other variables, and (iv) also observed on the degree of recovery from aphasia. These questions were answered in the four reported experiments, which, in addition, collectively demonstrate that multivariate analyses are a powerful tool for exploring how many different variables influence language outcomes. By investigating interactions between variables, these findings have deepened our knowledge, and help us to understand how the tapestry of characteristics that make up an individual weave together and collectively influence language recovery.

The contribution that this thesis makes to aphasia research will be emphasised in this final discussion chapter. I will first summarise the most relevant and interesting findings from Experiments 1-4 (section 7.1), and then discuss the scientific novelty of the findings and methods used (section 7.2), the clinical implications of the findings (section 7.3), how they may inform future research (section 7.4) and my conclusions (section 7.5).

## 7.1. Summary of experimental findings

In **Experiment 1**, I demonstrated that left hemisphere lesion size and initial severity had the largest effect on long-term overall language ability in stroke survivors. This was expected and in line with prior literature – however, what is particularly striking is that the size of the effect of left hemisphere lesion size was over twice that of initial severity, and over three times the size of other variables. Nonetheless, even when the enormous influences of lesion size and initial severity were controlled, there were still significant, albeit very small, effects of both age at stroke and amount of pre-stroke education on overall language ability. These effects are noteworthy because the evidence for their effect in smaller prior studies has been inconclusive – prompting me to question whether the effect size is dependent on language task, and how well other variables (e.g. lesion size, initial severity, etc.) are controlled. These questions motivated the investigations presented in subsequent experiments.

A second important outcome from Experiment 1 was the absence of a significant influence of early therapy on long-term overall language ability. As with age and education, it is plausible that an effect of early therapy may depend on language task or how well other variables were controlled – again motivating further investigation.

A third finding of note from Experiment 1 was the demonstration that the effect of education was only observed in a large sample that included participants with a wide range of lesion sizes and language abilities but not in a second sample that predominantly included participants with large lesions and severe symptoms. This context-dependency may explain the inconsistent prior evidence for education effects, and emphasises the value of heterogeneous participant samples, that can be studied both as a whole, and grouped according to certain characteristics.

In **Experiments 2 and 3**, I demonstrated that the disadvantage of older age, and the benefit of more pre-stroke education that were observed for overall language ability in Experiment 1, prevailed and were surprisingly robust across a range of different language tasks, irrespective of lesion size and initial aphasia severity. This included demonstrating that age and education effects were observed in participants who did and did not have aphasia; a novel finding that was critical for emphasising that the language tasks themselves are sensitive to age and education. By probing in more detail, I was able to identify which tasks were most sensitive to age or education effects; as well as a very specific context in which the effect of age was exacerbated and the benefit of education was diminished.

Regarding which tasks were most sensitive to age or education effects, I found that the effects of both age and education were remarkably strong during verbal fluency tasks and notably weaker during reading tasks (with one exception detailed below). In addition, the effect of age was particularly strong for auditory tasks like repetition and auditory comprehension. In contrast, the effect of education was particularly strong for describing pictures and reading nonwords. The task-dependency of the effects are likely due to innate task properties. For example, age effects may be stronger during auditory tasks because they require good hearing ability, which is disadvantaged by older age. In contrast, education effects may be stronger during picture description which requires a wide vocabulary, and accurate grammar, both of which benefit from more education.

Only by conducting these task-level analyses are we able to understand the driving force behind the variables influencing post-stroke language performance.

Regarding the specific context when the effects of age and education were influenced by other variables, I observed that the disadvantage of older age is worsened in participants who have both large left hemisphere lesions and severe initial symptoms. Conversely, the benefit of more education was eroded in these older participants with large left hemisphere lesions and severe symptoms. Identifying this context dependency was only possible because the analysis included a very large sample of heterogeneous participants, which afforded a range of variance across multiple factors, and allowed their interactions to be tested.

In accordance with the findings reported in Experiment 1, we can recognise that left hemisphere lesion size not only has a big independent influence on language; it also moderates the influence of other variables.

In **Experiment 4**, I showed that participants who received early clinical therapy had better short- and long-term speech production compared to participants who received no early therapy. This is inconsistent with Experiment 1, where no benefit of early therapy was observed. One reason that Experiment 4 might have been more sensitive to therapy effects is that the speech production tasks used corresponded better with the content of the early therapy (compared with the broad measure of language ability used in Experiment 1). A second possibility was that effects of early therapy were strongest in the period immediately following therapy provision, and diminish with time, meaning they were more detectible on the short-term language measurements included in Experiment 4. A third possibility was that the presence of therapy effects depended on how other variables were controlled.

The benefits of early therapy are especially interesting because (i) the amount of therapy provided was relatively small – far smaller than that often provided by experimental therapy trials, and, (ii) these participants were deemed less likely to recover in Experiments 2 and 3 partly because of their severe initial symptoms. The short- and long-term benefits of even a small amount of clinically-provided therapy are therefore enormously encouraging. The results motivate a deeper understanding of exactly what elements of this early therapy stimulated the improvement; specifically, whether it was the content of the therapy itself, or that its provision encouraged participants to seek further learning and practice opportunities – or that its provision reflected greater cognitive, social or health factors that assisted the recovery.

## 7.2. Scientific novelty

The findings from all four experiments have advanced our knowledge of the variables that influence performance on a range of language tasks. The small but consistent effects of age and education have not been reported in any prior studies. Furthermore, they were observed irrespective of the presence or absence of aphasia which encourages more consideration of the many other potential sources of 'normal' variance – be they related to early, mid or later life experiences.

Despite the broad consistency of the effects of age and education, I identified one group of participants in whom these effects were atypical. Participants with large lesions and severe initial aphasia were (i) more vulnerable to the challenges of older age, (ii) less able to capitalise on an educational advantage, and (iii) less likely to recover language impairments. Encouragingly however, language improvements were still observed after therapy, suggesting that the early provision of therapy may be a critical factor for maximising recovery in participants with the most adverse stroke outcomes. These participants are, unfortunately, more likely to be excluded from aphasia research, due to the challenges of obtaining informed consent, and carrying out research procedures in the face of co-morbid impairments. Nonetheless, an uncontrolled or unidentified variable (such as lesion location, or non-linguistic cognitive abilities) may be a strong determinant of recovery in these participants, providing a compelling rationale to continue investigating the determinants of recovery specifically within this group, and to maximise their inclusion in future research.

These novel findings were able to be revealed because of the extremely large sample of participants, which enabled (i) high statistical power, and thus detection of small effects, (ii) consideration of many covariates and their interactions, and (iii) systematic subgroup analyses that illustrated the striking consistency of these effects irrespective of a range of clinical characteristics.

## 7.3. Clinical implications

There is immense value in understanding the independent influence of both age and education on specific language tasks. In support of a personalised approach to assessment and rehabilitation, individual patients' CAT scores should be interpreted with respect to these demographic characteristics. This 'normative modelling' approach has been demonstrated by prior studies (Shirk et al., 2011, Casaletto et al., 2015), but has not yet been replicated in the CAT. It also paves the way for investigation of many other variables that may influence assessment performance – such as multilingualism. Adopting this approach would force direct emphasis on the individual, and on recognition that a multitude of personal characteristics and life experiences should be recognised when interpreting assessment performance.

Furthermore, therapy planning may be enhanced when there is prior evidence that a patient has less recovery capacity for language impairments. This knowledge would guide expectation management, and help to provide realistic expectations of recovery that can be facilitated by a therapy approach that focuses on compensation for impairments and functional communication.

Finally, the predictive value of age at stroke and education compel their inclusion in future prediction models of aphasia recovery, and advances us towards the broader goal of providing individualised recovery prognoses.

## 7.4 Ethical implications for prediction modelling

Despite the potential for prediction modelling to drive individualised approaches to assessment and rehabilitation, consideration must be given to the fact that sampling bias may limit the generalisability of the results to the wider stroke population, and force the question of whether it is appropriate for prediction models to inform clinical practice and future prognostic modelling. The concern is that if participants who are excluded from research are also those who are disadvantaged by existing health inequalities, we may be driving and exacerbating those health inequalities. Specifically, in these experiments, it is likely that patients who had more severe strokes and were less well-recovered were likely to be underrepresented. It is also plausible that there is limited geographical diversity, given the study's requirement to travel to London. Finally, data pertaining to demographic characteristics that drive health inequalities (e.g. ethnicity and socioeconomic status) were not obtained, meaning it was not possible to (a) assess the diversity of the sample with respect to these characteristics, nor (b) extend the findings to all participants irrespective of these variables. The result is that a prediction model could over- or underestimate a score for a new patient, resulting in a potentially incorrect interpretation of their language abilities. However, this risk is likely to be minimal because therapists use a number of sources to inform their assessment and management plans. The addition of model-based prognosis should be seen as another "tool in the toolkit", which supports their clinical management for a given individual. Another risk is that patients who are deemed less likely to recover by prognostic modelling could receive less therapy, or are deprioritised over those deemed more likely to recover. It is paramount that this does not occur. The aim of incorporating prognosis into rehabilitation is not to redirect resources away from patients who need them; it is to guide and optimise the therapy that is already being delivered and to ensure that it is maximally effective.

## 7.5 Future directions

The amount of variance that was collectively explained by the set of lesion and demographic variables ranged from as little as 12% (for understanding spoken paragraphs) to as much as 63% for overall language ability. The large amount of unexplained variance must be acknowledged, and suggestions for future studies to tackle this are addressed below. There are three angles to consider: (1) analyses that would develop and progress the experimental findings reported in this thesis, (2) investigation of new variables, and (3) improving the measurement of known variables.

The first point, relating to future analyses, is motivated by the collective findings from Experiments 2-4. When participants have large left hemisphere lesions and severe initial symptoms, the disadvantage of older age is worsened, and, in those older participants, the benefit of more education is eroded. Nonetheless, early therapy was found to benefit both short- and long-term language abilities amongst participants with severe initial symptoms. It would therefore be interesting to examine whether therapy also interacts with age and/or education. Specifically, is the effect of therapy different for younger compared to older participants, or for participants who received less versus more education. These analyses were not possible in this thesis due to insufficient therapy data.

The second point relates to the variables whose influence on language recovery is speculated but not yet fully understood. One notable example discussed in this thesis is brain heath, which is thought to worsen with advancing age, and also to be associated with other health variables. Brain health could be measured by assessing white matter hyperintensities (e.g. using the Fazekas rating scale), cholinergic pathways (e.g. using the Cholinergic Pathways Hyperintensities Scale) (Zhong et al., 2021), or by using the undamaged hemisphere to calculate a measure of 'brain age', as opposed to chronological age (Kristinsson et al., 2022). Other variables whose influence on aphasia recovery warrants more investigation include demographic, cognitive, social and health variables that were not captured by the PLORAS study, and remain under-researched despite being considered in three recent studies (Harvey et al., 2022, Shin et al., 2022, Jacobs et al., 2023). Once again, it would be critical to understand how the potential effects of these variables depend on each other - adopting an intersectional approach, as advocated by Evans et al. (2023).

The third point relates to the improvement of how known variables are measured. Specifically, the measurement of therapy in these experiments, and in many others, is typically reported as number of hours received over a set period of time. Whilst simpler to estimate, this measurement disregards the multidimensional nature of therapy (Hayward et al., 2021), and could be enriched by also considering timing, content, patient engagement, family support, and everyday language practice. Examples of the latter include work, volunteering, attendance at stroke clubs or other social groups, and speaking with family and friends. Additionally, it would be valuable to understand how therapy is prioritised and allocated, and any reasons its delivery may be delayed or stopped – such as physical and mental health, additional therapy priorities, logistical barriers, or limited therapy resources. These more detailed data about speech and language therapy, supplemented by novel insights into real-world language practice, may help to identify the key strategies or activities that drive improvement or motivate further practice; sometimes referred to as the 'active ingredients' of therapy (Brogan et al., 2020), and of the social and health factors that may explain variation in therapy response. Altogether, understanding the independent and collective influence of many clinical and non-clinical variables, and how they interact with each other, will enrich our knowledge of recovery. With an understanding of natural, individual recovery trajectories, it will be possible to provide individualised prognoses, and fairly evaluate therapy effects such that therapy can be optimised for each patient.

Future work also needs to address some of the challenges related to multivariate analyses which require (i) extensive datasets encompassing many variables, from (ii) large numbers of participants. One solution to this challenge could involve standardisation of data collection across multiple trials and research groups. For example, the DESCRIBE checklist (Wallace et al., 2022) identifies 14 participant characteristics (at a minimum) that should be recorded by all aphasia trials, and a standard measurement for that variable (for example, number of years of education completed). Whilst this endeavour for standardisation should be encouraged, there are several glaring omissions from the DESCRIBE checklist. For example, there is no mention of the importance of measuring and controlling for lesion size or location. Furthermore, recommendations that ‘education’ is measured in ‘number of years completed’ disregards the potential influence of other aspects of education discussed in Experiment 3 (e.g. level of attainment). That said, whilst research into these other variables is ongoing, a standardised checklist remains a pragmatic solution – and one which can be adapted as we grow our knowledge of the predictors of recovery. In particular, a standardised approach to data collection between trials encourages data-sharing and collaboration between research groups. It will dramatically increase the number of participants included in analyses and meta-analyses, as illustrated by the meta-analyses conducted and published by the RELEASE team within the Collaboration of Aphasia Trialists (Brady et al., 2022).

## 7.6 Conclusions

In this thesis, I have demonstrated how age at stroke and amount of pre-stroke education affect a variety of language task scores from the Comprehensive Aphasia Test (CAT). In other words, the CAT is sensitive to both the challenge of normal healthy ageing, and the benefit of more education. The disadvantage of older age, and benefit of more education, were observed across all participants regardless of their initial language abilities and impairments, and were found to be additive with the influence of highly predictive variables such as lesion size and initial aphasia severity. The exception is participants with the largest left hemisphere lesions, and the most severe initial language symptoms, in whom the challenge of older age was worsened, and any benefit of more education was eroded. This motivates further study of participants with these clinical characteristics, specifically with regard to therapy effects.

The second major finding is that language outcomes are better when early therapy is received, compared to when no early therapy is received – and critically, in participants with Severe and Moderate initial symptoms. If this benefit of early therapy is observed regardless of age, symptom severity and lesion size, the disadvantage of age may still be surmounted. On the other hand, if therapy response varies with age, symptom severity or lesion size, this might help to explain variation in therapy response found by prior trials. It would also provide a rationale for adjusting the content and goal of therapy for age.

Overall, the findings reported in this thesis (1) advance our knowledge of the influence of age, education and early therapy on language outcomes, (2) motivate the inclusion of these variables in predictive models of language outcomes after stroke, (3) encourage investigation of other, under-studied variables, and (4) contribute to the wider goal of equipping future patients with knowledge of their likely individual recovery.

## References

AAMODT, E. B., ALNÆS, D., DE LANGE, A. G., AAM, S., SCHELLHORN, T., SALTVEDT, I., BEYER, M. K. & WESTLYE, L. T. 2023. Longitudinal brain age prediction and cognitive function after stroke. *Neurobiol Aging,* 122**,** 55-64.

ABDEL AZIZ, K., KHATER, M. S., EMARA, T., TAWFIK, H. M., RASHEEDY, D., MOHAMMEDIN, A. S., TOLBA, M. F., EL-GABRY, D. A. & QASSEM, T. 2017. Effects of age, education, and gender on verbal fluency in healthy adult Arabic-speakers in Egypt. *Appl Neuropsychol Adult,* 24**,** 331-341.

AGUILAR, O. M., KERRY, S. J., ONG, Y.-H., CALLAGHAN, M. F., CRINION, J., WOODHEAD, Z. V. J., PRICE, C. J., LEFF, A. P. & HOPE, T. M. H. 2018. Lesion-site-dependent responses to therapy after aphasic stroke. *Journal of Neurology, Neurosurgery &amp;amp; Psychiatry,* 89**,** 1352.

AHALT, C., WALTER, L. C., YOURMAN, L., ENG, C., PÉREZ-STABLE, E. J. & SMITH, A. K. 2012. "Knowing is better": preferences of diverse older adults for discussing prognosis. *J Gen Intern Med,* 27**,** 568-75.

AKI Ö, E., ALKAN, B., DEMIRSÖZ, T., VELIBAŞOĞLU, B., TAŞDEMIR, T., ERBAŞ, S. P., SELVI, K., ERGENÇ, İ., BARIŞKIN, E., ÖZDEMIR, P. & DEMIR, B. 2022. Effects of Age, Gender and Education on Phonemic and Semantic Verbal Fluency. *Turk Psikiyatri Derg,* 33**,** 53-64.

ALEXANDER, M. P., NAESER, M. A. & PALUMBO, C. 1990. Broca's area aphasias. *Neurology,* 40**,** 353-353.

ALIA, C., SPALLETTI, C., LAI, S., PANARESE, A., LAMOLA, G., BERTOLUCCI, F., VALLONE, F., DI GARBO, A., CHISARI, C., MICERA, S. & CALEO, M. 2017. Neuroplastic Changes Following Brain Ischemia and their Contribution to Stroke Recovery: Novel Approaches in Neurorehabilitation. *Frontiers in Cellular Neuroscience,* 11.

ARDILA, A. & ROSSELLI, M. 1996. Spontaneous Language Production and Aging: Sex and Educational Effects. *International Journal of Neuroscience,* 87**,** 71-78.

ASHBURNER, J. & FRISTON, K. J. 2000. Voxel-based morphometry--the methods. *Neuroimage,* 11**,** 805-21.

AVILA, R., MOSCOSO, M. A. A., RIBEIZ, S., ARRAIS, J., JALUUL, O. & BOTTINO, C. M. C. 2009. Influence of education and depressive symptoms on cognitive function in the elderly. *International Psychogeriatrics,* 21**,** 560-567.

BAKHEIT, A. M., SHAW, S., BARRETT, L., WOOD, J., CARRINGTON, S., GRIFFITHS, S., SEARLE, K. & KOUTSI, F. 2007. A prospective, randomized, parallel group, controlled study of the effect of intensity of speech and language therapy on early recovery from poststroke aphasia. *Clin Rehabil,* 21**,** 885-94.

BANCE, M. 2007. Hearing and aging. *Cmaj,* 176**,** 925-7.

BASILAKOS, A., STARK, B. C., JOHNSON, L., RORDEN, C., YOURGANOV, G., BONILHA, L. & FRIDRIKSSON, J. 2019. Leukoaraiosis Is Associated With a Decline in Language Abilities in Chronic Aphasia. *Neurorehabilitation and Neural Repair,* 33**,** 718-729.

BENGHANEM, S., ROSSO, C., ARBIZU, C., MOULTON, E., DORMONT, D., LEGER, A., PIRES, C. & SAMSON, Y. 2019. Aphasia outcome: the interactions between initial severity, lesion size and location. *J Neurol,* 266**,** 1303-1309.

BHOGAL, S. K., TEASELL, R. & SPEECHLEY, M. 2003. Intensity of aphasia therapy, impact on recovery. *Stroke,* 34**,** 987-93.

BIONDO, F., JEWELL, A., PRITCHARD, M., AARSLAND, D., STEVES, C. J., MUELLER, C. & COLE, J. H. 2022. Brain-age is associated with progression to dementia in memory clinic patients. *Neuroimage Clin,* 36**,** 103175.

BIRAN, M., GVION, A. & SHMUELY-SAMUEL, S. 2023. Language in Healthy Ageing: A Comparison across Language Domains. *Folia Phoniatrica et Logopaedica,* 75**,** 90-103.

BLACHSTEIN, H. & VAKIL, E. 2016. Verbal learning across the lifespan: an analysis of the components of the learning curve. *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn,* 23**,** 133-53.

BONKHOFF, A. K., HOPE, T., BZDOK, D., GUGGISBERG, A. G., HAWE, R. L., DUKELOW, S. P., REHME, A. K., FINK, G. R., GREFKES, C. & BOWMAN, H. 2020. Bringing proportional recovery into proportion: Bayesian modelling of post-stroke motor impairment. *Brain,* 143**,** 2189-2206.

BOROD, J. C., GOODGLASS, H. & KAPLAN, E. 1980. Normative data on the boston diagnostic aphasia examination, parietal lobe battery, and the boston naming Test. *Journal of Clinical Neuropsychology,* 2**,** 209-215.

BØSTRAND, S. M. K., VAHER, K., DE NOOIJ, L., HARRIS, M. A., COLE, J. H., COX, S. R., MARIONI, R. E., MCCARTNEY, D. L., WALKER, R. M., MCINTOSH, A. M., EVANS, K. L., WHALLEY, H. C., WOOTTON, R. E. & CLARKE, T. K. 2022. Associations between alcohol use and accelerated biological ageing. *Addict Biol,* 27**,** e13100.

BOURNE, V. J., FOX, H. C., DEARY, I. J. & WHALLEY, L. J. 2007. Does childhood intelligence predict variation in cognitive change in later life? *Personality and Individual Differences,* 42**,** 1551-1559.

BOWEN, A., HESKETH, A., PATCHICK, E., YOUNG, A., DAVIES, L., VAIL, A., LONG, A., WATKINS, C., WILKINSON, M., PEARL, G., LAMBON RALPH, M. & TYRRELL, P. 2012a. Clinical effectiveness, cost-effectiveness and service users' perceptions of early, well-resourced communication therapy following a stroke: a randomised controlled trial (the ACT NoW Study). *Health Technol Assess,* 16**,** 1-160.

BOWEN, A., HESKETH, A., PATCHICK, E., YOUNG, A., DAVIES, L., VAIL, A., LONG, A. F., WATKINS, C., WILKINSON, M., PEARL, G., RALPH, M. A. & TYRRELL, P. 2012b. Effectiveness of enhanced communication therapy in the first four months after stroke for aphasia and dysarthria: a randomised controlled trial. *BMJ,* 345**,** e4407.

BOWMAN, H., BONKHOFF, A., HOPE, T., GREFKES, C. & PRICE, C. 2021. Inflated Estimates of Proportional Recovery From Stroke. *Stroke,* 52**,** 1915-1920.

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, T. B., PIETRO-BACHMANN, M. D., ENDERBY, P., FILLINGHAM, J., LUCIA GALLI, F., GANDOLFI, M., GLIZE, B., GODECKE, E., HAWKINS, N., HILARI, K., HINCKLEY, J., HORTON, S., HOWARD, D., JAECKS, P., JEFFERIES, E., JESUS, L. M., KAMBANAROS, M., KYOUNG KANG, E., KHEDR, E. M., PAK-HIN KONG, A., KUKKONEN, T., LAGANARO, M., LAMBON RALPH, M. A., CHARLOTTE LASKA, A., LEEMANN, B., LEFF, A. P., LIMA, R. R., LORENZ, A., MACWHINNEY, B., SHISLER

MARSHALL, R., MATTIOLI, F., MAVIŞ, İ., MEINZER, M., NILIPOUR, R., NOÉ, E., PAIK, N. J., PALMER, R., PAPATHANASIOU, I., PATRICIO, B., PAVÃO MARTINS, I., PRICE, C., PRIZL JAKOVAC, T., ROCHON, E., ROSE, M. L., ROSSO, C., RUBI-FESSEN, I., RUITER, M. B., SNELL, C., STAHL, B., SZAFLARSKI, J. P., THOMAS, S. A., VAN DE SANDT-KOENDERMAN, M., VAN DER MEULEN, I., VISCH-BRINK, E., WORRALL, L. & HARRIS WRIGHT, H. 2022. Precision rehabilitation for aphasia by patient age, sex, aphasia severity, and time since stroke? A prespecified, systematic review-based, individual participant data, network, subgroup meta-analysis. *Int J Stroke,* 17**,** 1067-1077.

BRADY, M. C., KELLY, H., GODWIN, J., ENDERBY, P. & CAMPBELL, P. 2016. Speech and language therapy for aphasia following stroke. *Cochrane Database Syst Rev***,** CD000425.

BRAIS, B. 1992. The third left frontal convolution plays no role in language. *Neurology,* 42**,** 690-690.

BREITENSTEIN, C., GREWE, T., FLÖEL, A., ZIEGLER, W., SPRINGER, L., MARTUS, P., HUBER, W., WILLMES, K., RINGELSTEIN, E. B., HAEUSLER, K. G., ABEL, S., GLINDEMANN, R., DOMAHS, F., REGENBRECHT, F., SCHLENCK, K.-J., THOMAS, M., OBRIG, H., DE LANGEN, E., ROCKER, R., WIGBERS, F., RÜHMKORF, C., HEMPEN, I., LIST, J., BAUMGAERTNER, A., VILLRINGER, A., BLEY, M., JÖBGES, M., HALM, K., SCHULZ, J., WERNER, C., GOLDENBERG, G., KLINGENBERG, G., KÖNIG, E., MÜLLER, F., GRÖNE, B., KNECHT, S., BAAKE, R., KNAUSS, J., MIETHE, S., STELLER, U., SUDHOFF, R., SCHILLIKOWSKI, E., PFEIFFER, G., BILLO, K., HOFFMANN, H., FERNEDING, F. J., RUNGE, S., KECK, T., MIDDELDORF, V., KRÜGER, S., WILDE, B., KRAKOW, K., BERGHOFF, C., REINHUBER, F., MASER, I., HOFMANN, W., SOUS-KULKE, C., SCHUPP, W., OERTEL, A., BÄTZ, D., HAMZEI, F., SCHULZ, K., MEYER, A., KARTMANN, A., SOM, O. N., SCHIPKE, S. B. & BAMBORSCHKE, S. 2017. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. *The Lancet,* 389**,** 1528-1538.

BROCA, P. 1861. Remarques sur le siège de la faculté du langage articulé, suivies d'une observation d'aphémie (perte de la parole). *Bulletin et mémoires de la Société Anatomique de Paris,* 6**,** 330-357.

BROGAN, E., CICCONE, N. & GODECKE, E. 2020. An exploration of aphasia therapy dosage in the first six months of stroke recovery. *Neuropsychological Rehabilitation***,** 1-35.

BROTT, T., ADAMS, H. P., OLINGER, C. P., MARLER, J. R., BARSAN, W. G., BILLER, J., SPILKER, J., HOLLERAN, R., EBERLE, R. & HERTZBERG, V. 1989. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke,* 20**,** 864-870.

BRUCKI, S. M. & ROCHA, M. S. 2004. Category fluency test: effects of age, gender and education on total scores, clustering and switching in Brazilian Portuguese-speaking subjects. *Braz J Med Biol Res,* 37**,** 1771-7.

BULLIER, B., CASSOUDESALLE, H., VILLAIN, M., COGNÉ, M., MOLLO, C., DE GABORY, I., DEHAIL, P., JOSEPH, P.-A., SIBON, I. & GLIZE, B. 2020. New factors that affect quality of life in patients with aphasia. *Annals of Physical and Rehabilitation Medicine,* 63**,** 33-37.

BUNKER, L. D. & HILLIS, A. E. 2022. Chapter 3 - Vascular syndromes: Revisiting classification of poststroke aphasia. *In:* HILLIS, A. E. & FRIDRIKSSON, J. (eds.) *Handbook of Clinical Neurology.* Elsevier.

BURKE, D. M. & MACKAY, D. G. 1997. Memory, language, and ageing. *Philos Trans R Soc Lond B Biol Sci,* 352**,** 1845-56.

BURKE, D. M. & SHAFTO, M. A. 2004. Aging and Language Production. *Curr Dir Psychol Sci,* 13**,** 21-24.

BUSBY, N., WILMSKOETTER, J., GLEICHGERRCHT, E., RORDEN, C., ROTH, R., NEWMAN-NORLUND, R., HILLIS, A. E., KELLER, S. S., DE BEZENAC, C., KRISTINSSON, S., FRIDRIKSSON, J. & BONILHA, L. 2023. Advanced Brain Age and Chronic Poststroke Aphasia Severity. *Neurology,* 100**,** e1166-e1176.

CARAMAZZA, A. 1984. The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language,* 21**,** 9-20.

CASALETTO, K. B., UMLAUF, A., BEAUMONT, J., GERSHON, R., SLOTKIN, J., AKSHOOMOFF, N. & HEATON, R. K. 2015. Demographically Corrected Normative Standards for the English Version of the NIH Toolbox Cognition Battery. *J Int Neuropsychol Soc,* 21**,** 378-91.

CASTRO, N., HULA, W. D. & ASHAIE, S. A. 2023. Defining aphasia: Content analysis of six aphasia diagnostic batteries. *Cortex,* 166**,** 19-32.

CHEVALÈRE, J., LEMAIRE, P. & CAMOS, V. 2020. Age-Related Changes in Verbal Working Memory Strategies. *Experimental Aging Research,* 46**,** 93-127.

CHIA, E. M., MITCHELL, P., ROCHTCHINA, E., FORAN, S., GOLDING, M. & WANG, J. J. 2006. Association between vision and hearing impairments and their combined effects on quality of life. *Arch Ophthalmol,* 124**,** 1465-70.

CHOI, H. J., LEE, D. Y., SEO, E. H., JO, M. K., SOHN, B. K., CHOE, Y. M., BYUN, M. S., KIM, J. W., KIM, S. G., YOON, J. C., JHOO, J. H., KIM, K. W. & WOO, J. I. 2014. A normative study of the digit span in an educationally diverse elderly population. *Psychiatry Investig,* 11**,** 39-43.

CHURRUCA, K., POMARE, C., ELLIS, L. A., LONG, J. C., HENDERSON, S. B., MURPHY, L. E. D., LEAHY, C. J. & BRAITHWAITE, J. 2021. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. *Health Expect,* 24**,** 1015-1024.

CLARKE, D. J., BURTON, L. J., TYSON, S. F., RODGERS, H., DRUMMOND, A., PALMER, R., HOFFMAN, A., PRESCOTT, M., TYRRELL, P., BRKIC, L., GRENFELL, K. & FORSTER, A. 2018. Why do stroke survivors not receive recommended amounts of active therapy? Findings from the ReAcT study, a mixed-methods case-study evaluation in eight stroke units. *Clin Rehabil,* 32**,** 1119-1132.

CODE, C. & HERRMANN, M. 2003. The relevance of emotional and psychosocial factors in aphasia to rehabilitation. *Neuropsychol Rehabil,* 13**,** 109-32.

CODE, C., PAPATHANASIOU, I., RUBIO-BRUNO, S., CABANA MDE, L., VILLANUEVA, M. M., HAALAND-JOHANSEN, L., PRIZL-JAKOVAC, T., LEKO, A., ZEMVA, N., PATTERSON, R., BERRY, R., ROCHON, E., LEONARD, C. & ROBERT, A. 2016. International patterns of the public awareness of aphasia. *Int J Lang Commun Disord,* 51**,** 276-84.

CODE, C., TORNEY, A., GILDEA-HOWARDINE, E. & WILLMES, K. 2010. Outcome of a one-month therapy intensive for chronic aphasia: variable individual responses. *Semin Speech Lang,* 31**,** 21-33.

COHEN, R. A., MARSISKE, M. M. & SMITH, G. E. 2019. Chapter 10 - Neuropsychology of aging. *In:* DEKOSKY, S. T. & ASTHANA, S. (eds.) *Handbook of Clinical Neurology.* Elsevier.

COLEMAN, E. R., MOUDGAL, R., LANG, K., HYACINTH, H. I., AWOSIKA, O. O., KISSELA, B. M. & FENG, W. 2017. Early Rehabilitation After Stroke: a Narrative Review. *Curr Atheroscler Rep,* 19**,** 59.

CONNOR, L. T., OBLER, L. K., TOCCO, M., FITZPATRICK, P. M. & ALBERT, M. L. 2001. Effect of socioeconomic status on aphasia severity and recovery. *Brain Lang,* 78**,** 254-7.

CONSTANTINIDOU, F., CHRISTODOULOU, M. & PROKOPIOU, J. 2012. The effects of age and education on executive functioning and oral naming performance in greek cypriot adults: the neurocognitive study for the aging. *Folia Phoniatr Logop,* 64**,** 187-98.

CONTADOR, I., ALZOLA, P., STERN, Y., DE LA TORRE-LUQUE, A., BERMEJO-PAREJA, F. & FERNÁNDEZ-CALVO, B. 2023. Is cognitive reserve associated with the prevention of cognitive decline after stroke? A Systematic review and meta-analysis. *Ageing Res Rev,* 84**,** 101814.

CRAIK, F. I. M. & BIALYSTOK, E. 2006. Cognition through the lifespan: mechanisms of change. *Trends in Cognitive Sciences,* 10**,** 131-138.

CRAIK, F. I. M. & ROSE, N. S. 2012. Memory encoding and aging: A neurocognitive perspective. *Neuroscience & Biobehavioral Reviews,* 36**,** 1729-1739.

CURRIE, S., HOGGARD, N., CRAVEN, I. J., HADJIVASSILIOU, M. & WILKINSON, I. D. 2013. Understanding MRI: basic MR physics for physicians. *Postgrad Med J,* 89**,** 209-23.

CURTIN, F. & SCHULZ, P. 1998. Multiple correlations and bonferroni's correction. *Biological Psychiatry,* 44**,** 775-777.

DE RIESTHAL, M. & WERTZ, R. 2004. Prognosis for aphasia: Relationship between selected biographical and behavioural variables and outcome and improvement. *Aphasiology,* 18**,** 899-915.

DENES, G., PERAZZOLO, C., PIANI, A. & PICCIONE, F. 1996. Intensive versus regular speech therapy in global aphasia: A controlled study. *Aphasiology,* 10**,** 385-394.

DIEDRICHS, V. A., JEWELL, C. & HARNISH, S. M. 2022. A SCOPING REVIEW OF THE RELATIONSHIP BETWEEN NONLINGUISTIC COGNITIVE FACTORS AND APHASIA TREATMENT RESPONSE. *Top Lang Disord,* 42**,** 212-235.

DIGNAM, J., COPLAND, D., MCKINNON, E., BURFEIN, P., O'BRIEN, K., FARRELL, A. & RODRIGUEZ, A. D. 2015. Intensive Versus Distributed Aphasia Therapy: A Nonrandomized, Parallel-Group, Dosage-Controlled Study. *Stroke,* 46**,** 2206-11.

DIGNAM, J., COPLAND, D., RAWLINGS, A., O'BRIEN, K., BURFEIN, P. & RODRIGUEZ, A. D. 2016. The relationship between novel word learning and anomia treatment success in adults with chronic aphasia. *Neuropsychologia,* 81**,** 186-197.

DOEDENS, W. J. & METEYARD, L. 2020. Measures of functional, real-world communication for aphasia: a critical review. *Aphasiology,* 34**,** 492-514.

DOEDENS, W. J. & METEYARD, L. 2022. What is Functional Communication? A Theoretical Framework for Real-World Communication Applied to Aphasia Rehabilitation. *Neuropsychol Rev,* 32**,** 937-973.

DRESANG, H. C., HARVEY, D. Y., XIE, S. X., SHAH-BASAK, P. P., DELORETTA, L., WURZMAN, R., PARCHURE, S. Y., SACCHETTI, D.,

FASEYITAN, O., LOHOFF, F. W. & HAMILTON, R. H. 2022. Genetic and Neurophysiological Biomarkers of Neuroplasticity Inform Post-Stroke Language Recovery. *Neurorehabil Neural Repair,* 36, 371-380.

EL HACHIOUI, H., LINGSMA, H. F., VAN DE SANDT-KOENDERMAN, M. E., DIPPEL, D. W. J., KOUDSTAAL, P. J. & VISCH-BRINK, E. G. 2013a. Recovery of aphasia after stroke: a 1-year follow-up study. *Journal of Neurology,* 260, 166-171.

EL HACHIOUI, H., LINGSMA, H. F., VAN DE SANDT-KOENDERMAN, M. W., DIPPEL, D. W., KOUDSTAAL, P. J. & VISCH-BRINK, E. G. 2013b. Long-term prognosis of aphasia after stroke. *J Neurol Neurosurg Psychiatry,* 84, 310-5.

ENDERBY, P. & CROW, E. 1996. Frenchay Aphasia Screening Test: validity and comparability. *Disability and Rehabilitation,* 18, 238-240.

ERBELI, F., RICE, M. & PARACCHINI, S. 2022. Insights into Dyslexia Genetics Research from the Last Two Decades. *Brain Sciences* [Online], 12.

EVANS, E., JACOBS, M. & ELLIS, C. 2023. The Intersection of Social Determinants of Health and Post-stroke Aphasia Outcomes: A Need for Intersectional Analysis. *Semin Speech Lang,* 45, 084-098.

EVANS, T. 2018. Briefing document: First incidence of stroke. Estimates for England 2007 to 2016. *In:* ENGLAND, P. H. (ed.). PHE Publications.

FACAL, D., JUNCOS-RABADÁN, O., RODRÍGUEZ, M. S. & PEREIRO, A. X. 2012. Tip-of-the-tongue in aging: influence of vocabulary, working memory and processing speed. *Aging Clin Exp Res,* 24, 647-56.

FAROQI-SHAH, Y. & GEHMAN, M. 2021. The Role of Processing Speed and Cognitive Control on Word Retrieval in Aging and Aphasia. *Journal of Speech, Language, and Hearing Research,* 64, 949-964.

FEIGIN, V. L., BRAININ, M., NORRVING, B., MARTINS, S., SACCO, R. L., HACKE, W., FISHER, M., PANDIAN, J. & LINDSAY, P. 2022. World Stroke Organization (WSO): Global Stroke Fact Sheet 2022. *Int J Stroke,* 17, 18-29.

FERRO, J. M., MARIANO, G. & MADUREIRA, S. 1999. Recovery from aphasia and neglect. *Cerebrovasc Dis,* 9 Suppl 5, 6-22.

FLEMING, V., BROWNSETT, S., KRASON, A., MAEGLI, M. A., COLEY-FISHER, H., ONG, Y. H., NARDO, D., LEACH, R., HOWARD, D., ROBSON, H., WARBURTON, E., ASHBURNER, J., PRICE, C. J., CRINION, J. T. & LEFF, A. P. 2020. Efficacy of spoken word comprehension therapy in patients with chronic aphasia: a cross-over randomised controlled trial with structural imaging. *J Neurol Neurosurg Psychiatry*.

FONSECA, J., RAPOSO, A. & MARTINS, I. P. 2019. Cognitive functioning in chronic post-stroke aphasia. *Appl Neuropsychol Adult,* 26, 355-364.

FRIDRIKSSON, J., BONILHA, L. & RORDEN, C. 2007. Severe Broca’s Aphasia without Broca’s Area Damage. *Behavioural Neurology,* 18, 785280.

FRIDRIKSSON, J., FILLMORE, P., GUO, D. & RORDEN, C. 2015. Chronic Broca's Aphasia Is Caused by Damage to Broca's and Wernicke's Areas. *Cereb Cortex,* 25, 4689-96.

FRIER, A., BARNETT, F. & DEVINE, S. 2017. The relationship between social determinants of health, and rehabilitation of neurological conditions: a systematic literature review. *Disabil Rehabil,* 39, 941-948.

FRITSCH, T., SMYTH, K. A., MCCLENDON, M. J., OGROCKI, P. K., SANTILLAN, C., LARSEN, J. D. & STRAUSS, M. E. 2005. Associations between dementia/mild cognitive impairment and cognitive performance and activity levels in youth. *J Am Geriatr Soc,* 53**,** 1191-6.

FUJII, M., MAESAWA, S., ISHIAI, S., IWAMI, K., FUTAMURA, M. & SAITO, K. 2016. Neural Basis of Language: An Overview of An Evolving Model. *Neurol Med Chir (Tokyo),* 56**,** 379-86.

GADSON, D. S., WESLEY, D. B., VAN DER STELT, C. M., LACEY, E., DEMARCO, A. T., SNIDER, S. F. & TURKELTAUB, P. E. 2022. Aphasia severity is modulated by race and lesion size in chronic survivors: A retrospective study. *Journal of Communication Disorders,* 100**,** 106270.

GAJARDO-VIDAL, A., LORCA-PULS, D. L., CRINION, J. T., WHITE, J., SEGHIER, M. L., LEFF, A. P., HOPE, T. M. H., LUDERSDORFER, P., GREEN, D. W., BOWMAN, H. & PRICE, C. J. 2018. How distributed processing produces false negatives in voxel-based lesion-deficit analyses. *Neuropsychologia,* 115**,** 124-133.

GERSTENECKER, A. & LAZAR, R. M. 2019. Language recovery following stroke. *Clin Neuropsychol,* 33**,** 928-947.

GESCHWIND, N. 1965. DISCONNEXION SYNDROMES IN ANIMALS AND MAN1. *Brain,* 88**,** 237-237.

GILMORE, N., DWYER, M. & KIRAN, S. 2019a. Benchmarks of Significant Change After Aphasia Rehabilitation. *Archives of physical medicine and rehabilitation,* 100**,** 1131-1139.e87.

GILMORE, N., MEIER, E. L., JOHNSON, J. P. & KIRAN, S. 2019b. Nonlinguistic Cognitive Factors Predict Treatment-Induced Recovery in Chronic Poststroke Aphasia. *Archives of Physical Medicine and Rehabilitation,* 100**,** 1251-1258.

GODECKE, E., ARMSTRONG, E., RAI, T., CICCONE, N., ROSE, M. L., MIDDLETON, S., WHITWORTH, A., HOLLAND, A., ELLERY, F., HANKEY, G. J., CADILHAC, D. A. & BERNHARDT, J. 2020. A randomized control trial of intensive aphasia therapy after acute stroke: The Very Early Rehabilitation for SpEech (VERSE) study. *International Journal of Stroke***,** 1747493020961926.

GODECKE, E., CICCONE, N. A., GRANGER, A. S., RAI, T., WEST, D., CREAM, A., CARTWRIGHT, J. & HANKEY, G. J. 2014. A comparison of aphasia therapy outcomes before and after a Very Early Rehabilitation programme following stroke. *Int J Lang Commun Disord,* 49**,** 149-61.

GODECKE, E., HIRD, K., LALOR, E. E., RAI, T. & PHILLIPS, M. R. 2012. Very early poststroke aphasia therapy: a pilot randomized controlled efficacy trial. *Int J Stroke,* 7**,** 635-44.

GODECKE, E., RAI, T., CICCONE, N., ARMSTRONG, E., GRANGER, A. & HANKEY, G. J. 2013. Amount of therapy matters in very early aphasia rehabilitation after stroke: a clinical prognostic model. *Semin Speech Lang,* 34**,** 129-41.

GOLDENBERG, G. & SPATT, J. 1994. Influence of size and site of cerebral lesions on spontaneous recovery of aphasia and on success of language therapy. *Brain and language,* 47**,** 684-698.

GONZÁLEZ-FERNÁNDEZ, M., DAVIS, C., MOLITORIS, J. J., NEWHART, M., LEIGH, R. & HILLIS, A. E. 2011. Formal education, socioeconomic status, and the severity of aphasia after stroke. *Arch Phys Med Rehabil,* 92**,** 1809-13.

GOODGLASS, H. 1993. *Understanding aphasia,* San Diego, CA, US, Academic Press.

GOODGLASS, H., KAPLAN, E. & WEINTRAUB, S. 2001. *BDAE: The Boston diagnostic aphasia examination*, Lippincott Williams & Wilkins Philadelphia, PA.

GOW, A. J., JOHNSON, W., PATTIE, A., BRETT, C. E., ROBERTS, B., STARR, J. M. & DEARY, I. J. 2011. Stability and change in intelligence from age 11 to ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936. *Psychol Aging,* 26**,** 232-40.

GOW, A. J., JOHNSON, W., PATTIE, A., WHITEMAN, M. C., STARR, J. & DEARY, I. J. 2008. Mental ability in childhood and cognitive aging. *Gerontology,* 54**,** 177-86.

GRAUGAARD, P. K., ROGG, L., EIDE, H., UHLIG, T. & LOGE, J. H. 2011. Ways of providing the patient with a prognosis: A terminology of employed strategies based on qualitative data. *Patient Education and Counseling,* 83**,** 80-86.

GU, Y., ZHOU, C., PIAO, Z., YUAN, H., JIANG, H., WEI, H., ZHOU, Y., NAN, G. & JI, X. 2022. Cerebral edema after ischemic stroke: Pathophysiology and underlying mechanisms. *Front Neurosci,* 16**,** 988283.

HACKENBERGER, B. K. 2019. Bayes or not Bayes, is this the question? *Croat Med J,* 60**,** 50-52.

HANNAN, J., WILMSKOETTER, J., FRIDRIKSSON, J., HILLIS, A. E., BONILHA, L. & BUSBY, N. 2023. Brain health imaging markers, post-stroke aphasia and Cognition: A scoping review. *NeuroImage: Clinical,* 39**,** 103480.

HARADA, C. N., NATELSON LOVE, M. C. & TRIEBEL, K. L. 2013. Normal cognitive aging. *Clin Geriatr Med,* 29**,** 737-52.

HARRELL, J. F. E. 2015. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis,* Netherlands, Springer Nature.

HART, T., DIJKERS, M. P., WHYTE, J., TURKSTRA, L. S., ZANCA, J. M., PACKEL, A., VAN STAN, J. H., FERRARO, M. & CHEN, C. 2019. A Theory-Driven System for the Specification of Rehabilitation Treatments. *Arch Phys Med Rehabil,* 100**,** 172-180.

HARVEY, D. Y. & HAMILTON, R. 2022. Chapter 15 - Noninvasive brain stimulation to augment language therapy for poststroke aphasia. *In:* HILLIS, A. E. & FRIDRIKSSON, J. (eds.) *Handbook of Clinical Neurology.* Elsevier.

HARVEY, D. Y., PARCHURE, S. & HAMILTON, R. H. 2022. Factors predicting long-term recovery from post-stroke aphasia. *Aphasiology,* 36**,** 1351-1372.

HAYWARD, K. S., CHURILOV, L., DALTON, E. J., BRODTMANN, A., CAMPBELL, B. C. V., COPLAND, D., DANCAUSE, N., GODECKE, E., HOFFMANN, T. C., LANNIN, N. A., MCDONALD, M. W., CORBETT, D. & BERNHARDT, J. 2021. Advancing Stroke Recovery Through Improved Articulation of Nonpharmacological Intervention Dose. *Stroke,* 52**,** 761-769.

HERMANN, D. M. & CHOPP, M. 2012. Promoting brain remodelling and plasticity for stroke recovery: therapeutic promise and potential pitfalls of clinical translation. *The Lancet Neurology,* 11**,** 369-380.

HICKOK, G. 2009. The functional neuroanatomy of language. *Phys Life Rev,* 6**,** 121-43.

HILARI, K. 2011. The impact of stroke: are people with aphasia different to those without? *Disability and Rehabilitation,* 33**,** 211-218.

HILL, A., BLEVINS, R. & CODE, C. 2019. Revisiting the public awareness of aphasia in Exeter: 16 years on. *International Journal of Speech-Language Pathology,* 21**,** 504-512.

HILLIS, A. E. & HEIDLER, J. 2002. Mechanisms of early aphasia recovery. *Aphasiology,* 16**,** 885-895.

HOLLAND, A., FROMM, D., FORBES, M. & MACWHINNEY, B. 2017. Long-term Recovery in Stroke Accompanied by Aphasia: A Reconsideration. *Aphasiology,* 31**,** 152-165.

HOLLAND, A. L., GREENHOUSE, J. B., FROMM, D. & SWINDELL, C. S. 1989. Predictors of language restitution following stroke: a multivariate analysis. *J Speech Hear Res,* 32**,** 232-8.

HOLLAND, R. & CRINION, J. 2012. Can tDCS enhance treatment of aphasia after stroke? *Aphasiology,* 26**,** 1169-1191.

HOPE, T. M. H., FRISTON, K., PRICE, C. J., LEFF, A. P., ROTSHTEIN, P. & BOWMAN, H. 2019. Recovery after stroke: not so proportional after all? *Brain,* 142**,** 15-22.

HOPE, T. M. H., LEFF, A. P., PREJAWA, S., BRUCE, R., HAIGH, Z., LIM, L., RAMSDEN, S., OBERHUBER, M., LUDERSDORFER, P., CRINION, J., SEGHIER, M. L. & PRICE, C. J. 2017. Right hemisphere structural adaptation and changing language skills years after left hemisphere stroke. *Brain,* 140**,** 1718-1728.

HOPE, T. M. H., NARDO, D., HOLLAND, R., ONDOBAKA, S., AKKAD, H., PRICE, C. J., LEFF, A. P. & CRINION, J. 2021. Lesion site and therapy time predict responses to a therapy for anomia after stroke: a prognostic model development study. *bioRxiv***,** 2021.02.05.429894.

HOPE, T. M. H., SEGHIER, M. L., LEFF, A. P. & PRICE, C. J. 2013. Predicting outcome and recovery after stroke with lesions extracted from MRI images. *NeuroImage: Clinical,* 2**,** 424-433.

HUDSON, R., HIGH, L. & AL OTAIBA, S. 2007. Dyslexia and the Brain: What Does Current Research Tell Us? *Reading Teacher - READ TEACH,* 60**,** 506-515.

IDOWU, M. I. & SZAMEITAT, A. J. 2023. Executive function abilities in cognitively healthy young and older adults—A cross-sectional study. *Frontiers in Aging Neuroscience,* 15.

INATOMI, Y., YONEHARA, T., OMIYA, S., HASHIMOTO, Y., HIRANO, T. & UCHINO, M. 2008. Aphasia during the Acute Phase in Ischemic Stroke. *Cerebrovascular Diseases,* 25**,** 316-323.

IVANOVA, M. V., HERRON, T. J., DRONKERS, N. F. & BALDO, J. V. 2020. An empirical comparison of univariate versus multivariate methods for the analysis of brain-behavior mapping. *bioRxiv***,** 2020.04.13.039958.

JACOBS, M., EVANS, E. & ELLIS, C. 2023. Intersectional sociodemographic and neurological relationships in the naming ability of persons with post-stroke aphasia. *J Commun Disord,* 105**,** 106352.

JEBAHI, F., ABOU JAOUDE, R. & ELLIS, C. 2022. Semantic verbal fluency task: The effects of age, educational level, and sex in Lebanese-speaking adults. *Appl Neuropsychol Adult,* 29**,** 936-940.

JEFFREYS, H. 1961. *Theory of Probability,* Oxford, Oxford University Press.

JOHNSON, L., BASILAKOS, A., YOURGANOV, G., CAI, B., BONILHA, L., RORDEN, C. & FRIDRIKSSON, J. 2019. Progression of Aphasia Severity in the Chronic Stages of Stroke. *Am J Speech Lang Pathol,* 28**,** 639-649.

JOHNSON, L., NEMATI, S., BONILHA, L., RORDEN, C., BUSBY, N., BASILAKOS, A., NEWMAN-NORLUND, R., HILLIS, A. E., HICKOK, G. &

FRIDRIKSSON, J. 2022. Predictors beyond the lesion: Health and demographic factors associated with aphasia severity. *Cortex,* 154**,** 375-389.

KASTRAU, F., WOLTER, M., HUBER, W. & BLOCK, F. 2005. Recovery from aphasia after hemicraniectomy for infarction of the speech-dominant hemisphere. *Stroke,* 36**,** 825-9.

KAVÉ, G. & GORAL, M. 2017. Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Neuropsychol Dev Cogn B Aging Neuropsychol Cogn,* 24**,** 508-527.

KEENAN, J. S. & BRASSELL, E. G. 1974. A study of factors related to prognosis for individual aphasic patients. *J Speech Hear Disord,* 39**,** 257-69.

KELLY, H. & ARMSTRONG, L. 2009. New word learning in people with aphasia. *Aphasiology,* 23**,** 1398-1417.

KEMPLER, D., TENG, E. L., DICK, M., TAUSSIG, I. M. & DAVIS, D. S. 1998. The effects of age, education, and ethnicity on verbal fluency. *J Int Neuropsychol Soc,* 4**,** 531-8.

KERTESZ, A. 2007. Western Aphasia Battery--Revised.

KERTESZ, A. & MCCABE, P. 1977. RECOVERY PATTERNS AND PROGNOSIS IN APHASIA. *Brain,* 100**,** 1-18.

KESAV, P., VRINDA, S. L., SUKUMARAN, S., SARMA, P. S. & SYLAJA, P. N. 2017. Effectiveness of speech language therapy either alone or with add-on computer-based language therapy software (Malayalam version) for early post stroke aphasia: A feasibility study. *J Neurol Sci,* 380**,** 137-141.

KEULEERS, E., STEVENS, M., MANDERA, P. & BRYSBAERT, M. 2015. Word knowledge in the crowd: Measuring vocabulary size and word prevalence in a massive online experiment. *Quarterly Journal of Experimental Psychology,* 68**,** 1665-1692.

KIM, H. & NA, D. L. 2004. Normative Data on the Korean Version of the Western Aphasia Battery. *Journal of Clinical and Experimental Neuropsychology,* 26**,** 1011-1020.

KIM, K. A., LEE, J. S., CHANG, W. H., KIM, D. Y., SHIN, Y. I., KIM, S. Y., KIM, Y. T., KANG, S. H., CHOI, J. Y. & KIM, Y. H. 2019. Changes in Language Function and Recovery-Related Prognostic Factors in First-Ever Left Hemispheric Ischemic Stroke. *Ann Rehabil Med,* 43**,** 625-634.

KIRAN, S. 2012. What is the nature of poststroke language recovery and reorganization? *ISRN Neurol,* 2012**,** 786872.

KNECHT, S., DRÄGER, B., DEPPE, M., BOBE, L., LOHMANN, H., FLÖEL, A., RINGELSTEIN, E. B. & HENNINGSEN, H. 2000. Handedness and hemispheric language dominance in healthy humans. *Brain,* 123 Pt 12**,** 2512-8.

KRAKAUER, J. W. & MARSHALL, R. S. 2015. The proportional recovery rule for stroke revisited. *Ann Neurol,* 78**,** 845-7.

KRESS, L. & AUE, T. 2017. The link between optimism bias and attention bias: A neurocognitive perspective. *Neuroscience & Biobehavioral Reviews,* 80**,** 688-702.

KRISTINSSON, S., BASILAKOS, A., DEN OUDEN DIRK, B., CASSARLY, C., SPELL LEIGH, A., BONILHA, L., RORDEN, C., HILLIS ARGYE, E., HICKOK, G., JOHNSON, L., BUSBY, N., WALKER GRANT, M., MCLAIN, A. & FRIDRIKSSON, J. 2023. Predicting Outcomes of Language Rehabilitation: Prognostic Factors for Immediate and Long-Term Outcomes After Aphasia Therapy. *Journal of Speech, Language, and Hearing Research,* 66**,** 1068-1084.

KRISTINSSON, S., BUSBY, N., RORDEN, C., NEWMAN-NORLUND, R., DEN OUDEN, D. B., MAGNUSDOTTIR, S., HJALTASON, H., THORS, H., HILLIS, A. E., KJARTANSSON, O., BONILHA, L. & FRIDRIKSSON, J. 2022. Brain age predicts long-term recovery in post-stroke aphasia. *Brain Commun,* 4**,** fcac252.

KUO, C. Y., LEE, P. L., PENG, L. N., LEE, W. J., WANG, P. N., CHEN, L. K., CHOU, K. H., CHUNG, C. P. & LIN, C. P. 2023. Advanced brain age in community-dwelling population with combined physical and cognitive impairments. *Neurobiol Aging,* 130**,** 114-123.

LASKA, A. C., HELLBLOM, A., MURRAY, V., KAHAN, T. & VON ARBIN, M. 2001. Aphasia in acute stroke and relation to outcome. *J Intern Med,* 249**,** 413-22.

LASKA, A. C., KAHAN, T., HELLBLOM, A., MURRAY, V. & VON ARBIN, M. 2011. A randomized controlled trial on very early speech and language therapy in acute stroke patients with aphasia. *Cerebrovasc Dis Extra,* 1**,** 66-74.

LAURA, G. & ROZA, M. U. 2022. Association of cognitive reserve with stroke outcome: a protocol for a systematic review. *BMJ Open,* 12**,** e059378.

LAWTON, M., HADDOCK, G., CONROY, P., SERRANT, L. & SAGE, K. 2018. People with aphasia's perception of the therapeutic alliance in aphasia rehabilitation post stroke: a thematic analysis. *Aphasiology,* 32**,** 1397-1417.

LAZAR, R. M. & ANTONIELLO, D. 2008. Variability in recovery from aphasia. *Curr Neurol Neurosci Rep,* 8**,** 497-502.

LAZAR, R. M., MINZER, B., ANTONIELLO, D., FESTA, J. R., KRAKAUER, J. W. & MARSHALL, R. S. 2010. Improvement in aphasia scores after stroke is well predicted by initial severity. *Stroke,* 41**,** 1485-8.

LAZAR, R. M., SPEIZER, A. E., FESTA, J. R., KRAKAUER, J. W. & MARSHALL, R. S. 2008. Variability in language recovery after first-time stroke. *J Neurol Neurosurg Psychiatry,* 79**,** 530-4.

LEMKE, U. & ZIMPRICH, D. 2005. Longitudinal Changes in Memory Performance and Processing Speed in Old Age abstract Keywords. *Aging, Neuropsychology, and Cognition,* 12**,** 57-77.

LENDREM, W. & LINCOLN, N. B. 1985. Spontaneous recovery of language in patients with aphasia between 4 and 34 weeks after stroke. *Journal of neurology, neurosurgery, and psychiatry,* 48**,** 743-748.

LICHTHEIM, L. 1885. On Aphasia. *Brain,* 7**,** 433-484.

LUBRINI, G., PERIÁÑEZ, J. A., LASECA-ZABALLA, G., BERNABÉU-BROTONS, E. & RÍOS-LAGO, M. 2022. Verbal Fluency Tasks: Influence of Age, Gender, and Education and Normative Data for the Spanish Native Adult Population. *Arch Clin Neuropsychol,* 37**,** 365-375.

MARCHI, N. A., PTAK, R., DI PIETRO, M., SCHNIDER, A. & GUGGISBERG, A. G. 2017. Principles of proportional recovery after stroke generalize to neglect and aphasia. *Eur J Neurol,* 24**,** 1084-1087.

MARSH, E. B. & HILLIS, A. E. 2006. Recovery from aphasia following brain injury: the role of reorganization. *In:* MØLLER, A. R. (ed.) *Progress in Brain Research.* Elsevier.

MARSHALL, J. 2010. Classification of aphasia: Are there benefits for practice? *Aphasiology,* 24**,** 408-412.

MARTINS, I. P., FONSECA, J., MORGADO, J., LEAL, G., FARRAJOTA, L., FONSECA, A. C. & MELO, T. P. 2017. Language improvement one week after thrombolysis in acute stroke. *Acta Neurol Scand,* 135**,** 339-345.

MARTINS, I. P., LEAL, G., FONSECA, I., FARRAJOTA, L., AGUIAR, M., FONSECA, J., LAUTERBACH, M., GONÇALVES, L., CARY, M. C., FERREIRA, J. J. & FERRO, J. M. 2013. A randomized, rater-blinded, parallel trial of intensive speech therapy in sub-acute post-stroke aphasia: the SP-I-R-IT study. *International journal of language & communication disorders,* 48**,** 421-431.

MATHURANATH, P. S., GEORGE, A., CHERIAN, P. J., ALEXANDER, A., SARMA, S. G. & SARMA, P. S. 2003. Effects of age, education and gender on verbal fluency. *J Clin Exp Neuropsychol,* 25**,** 1057-64.

MATTIOLI, F., AMBROSI, C., MASCARO, L., SCARPAZZA, C., PASQUALI, P., FRUGONI, M., MAGONI, M., BIAGI, L. & GASPAROTTI, R. 2014. Early aphasia rehabilitation is associated with functional reactivation of the left inferior frontal gyrus: a pilot study. *Stroke,* 45**,** 545-52.

MAZZONI, M., VISTA, M., PARDOSSI, L., AVILA, L., BIANCHI, F. & MORETTI, P. 1992. Spontaneous evolution of aphasia after ischaemic stroke. *Aphasiology,* 6**,** 387-396.

MEIER, E. L., JOHNSON, J. P., PAN, Y. & KIRAN, S. 2019. The utility of lesion classification in predicting language and treatment outcomes in chronic stroke-induced aphasia. *Brain Imaging Behav,* 13**,** 1510-1525.

MENG, X. & D'ARCY, C. 2012. Education and Dementia in the Context of the Cognitive Reserve Hypothesis: A Systematic Review with Meta-Analyses and Qualitative Analyses. *PLOS ONE,* 7**,** e38268.

MESSINIS, L., NASIOS, G., MOUGIAS, A., POLITIS, A., ZAMPAKIS, P., TSIAMAKI, E., MALEFAKI, S., GOURZIS, P. & PAPATHANASOPOULOS, P. 2016. Age and education adjusted normative data and discriminative validity for Rey's Auditory Verbal Learning Test in the elderly Greek population. *J Clin Exp Neuropsychol,* 38**,** 23-39.

MEYER, S., VERHEYDEN, G., BRINKMANN, N., DEJAEGER, E., DE WEERDT, W., FEYS, H., GANTENBEIN, A. R., JENNI, W., LAENEN, A., LINCOLN, N., PUTMAN, K., SCHUBACK, B., SCHUPP, W., THIJS, V. & DE WIT, L. 2015. Functional and motor outcome 5 years after stroke is equivalent to outcome at 2 months: follow-up of the collaborative evaluation of rehabilitation in stroke across Europe. *Stroke,* 46**,** 1613-9.

MITCHELL, C., GITTINS, M., TYSON, S., VAIL, A., CONROY, P., PALEY, L. & BOWEN, A. 2021. Prevalence of aphasia and dysarthria among inpatient stroke survivors: describing the population, therapy provision and outcomes on discharge. *Aphasiology,* 35**,** 950-960.

MOHR, J. P., PESSIN, M. S., FINKELSTEIN, S., FUNKENSTEIN, H. H., DUNCAN, G. W. & DAVIS, K. R. 1978. Broca aphasia: pathologic and clinical. *Neurology,* 28**,** 311-24.

MUAYQIL, T. A., ALSHATHRI, R. M., ALNASSER, H. A., ALFAWZAN, L. A., ALQARNI, A. A., ALKERIDY, W. A., NAFISAH, I., ALJAFEN, B. N. & ALANAZY, M. H. 2021. Effects of age, sex and education on verbal fluency tasks in a normal Saudi sample. *J Clin Exp Neuropsychol,* 43**,** 879-889.

MURPHY, T. H. & CORBETT, D. 2009. Plasticity during stroke recovery: from synapse to behaviour. *Nature Reviews Neuroscience,* 10**,** 861-872.

NAESS, H., HAMMERSVIK, L. & SKEIE, G. O. 2009. Aphasia among young patients with ischemic stroke on long-term follow-up. *J Stroke Cerebrovasc Dis,* 18**,** 247-50.

NAKAGAWA, Y., SANO, Y., FUNAYAMA, M. & KATO, M. 2019. Prognostic factors for long-term improvement from stroke-related aphasia with adequate linguistic rehabilitation. *Neurol Sci,* 40**,** 2141-2146.

NAVARRETE-OREJUDO, L., CERDA-COMPANY, X., OLIVÉ, G., MARTIN, N., LAINE, M., RODRÍGUEZ-FORNELLS, A. & PEÑALOZA, C. 2023. Expressive recall and recognition as complementary measures to assess novel word learning ability in aphasia. *Brain and Language,* 243**,** 105303.

NEILS, J., BARIS, J. M., CARTER, C., DELL'AIRA, A. L., NORDLOH, S. J., WEILER, E. & WEISIGER, B. 1995. Effects of age, education, and living environment on Boston Naming Test performance. *J Speech Hear Res,* 38**,** 1143-9.

NICE, N. I. F. H. A. C. E. 2013. *Stroke rehabilitation in adults* [Online]. Available: https://www.nice.org.uk/guidance/cg162/resources/stroke-rehabilitation-in-adults-pdf-35109688408261 [Accessed 9 March 2020].

NICE, N. I. F. H. A. C. E. 2022. Stroke and transient ischaemic attack in over 16s: diagnosis and initial management. National Institute for Health and Care Excellence.

NICE, N. I. F. H. A. C. E. 2023. Stroke rehabilitation in adults. National Institute for Health and Care Excellence.

NOUWENS, F., DE LAU, L. M., VISCH-BRINK, E. G., VAN DE SANDT-KOENDERMAN, W. M., LINGSMA, H. F., GOOSEN, S., BLOM, D. M., KOUDSTAAL, P. J. & DIPPEL, D. W. 2017. Efficacy of early cognitive-linguistic treatment for aphasia due to stroke: A randomised controlled trial (Rotterdam Aphasia Therapy Study-3). *Eur Stroke J,* 2**,** 126-136.

NOUWENS, F., VISCH-BRINK, E. G., VAN DE SANDT-KOENDERMAN, M. M., DIPPEL, D. W., KOUDSTAAL, P. J. & DE LAU, L. M. 2015. Optimal timing of speech and language therapy for aphasia after stroke: more evidence needed. *Expert Rev Neurother,* 15**,** 885-93.

NUNNARI, D., BRAMANTI, P. & MARINO, S. 2014. Cognitive reserve in stroke and traumatic brain injury patients. *Neurol Sci,* 35**,** 1513-8.

O'HALLORAN, R., RENTON, J., HARVEY, S., MCSWEEN, M. P. & WALLACE, S. J. 2023. Do social determinants influence post-stroke aphasia outcomes? A scoping review. *Disabil Rehabil***,** 1-14.

OCKLENBURG, S., BESTE, C., ARNING, L., PETERBURS, J. & GÜNTÜRKÜN, O. 2014. The ontogenesis of language lateralization and its relation to handedness. *Neuroscience & Biobehavioral Reviews,* 43**,** 191-198.

OLIVEIRA, F. F. & DAMASCENO, B. P. 2009. Short-term prognosis for speech and language in first stroke patients. *Arq Neuropsiquiatr,* 67**,** 849-55.

OSA GARCÍA, A., BRAMBATI, S. M., BRISEBOIS, A., DÉSILETS-BARNABÉ, M., HOUZÉ, B., BEDETTI, C., ROCHON, E., LEONARD, C., DESAUTELS, A. & MARCOTTE, K. 2020. Predicting Early Post-stroke Aphasia Outcome From Initial Aphasia Severity. *Front Neurol,* 11**,** 120.

OSLER, M., CHRISTENSEN, G. T., GARDE, E., MORTENSEN, E. L. & CHRISTENSEN, K. 2017. Cognitive ability in young adulthood and risk of dementia in a cohort of Danish men, brothers, and twins. *Alzheimers Dement,* 13**,** 1355-1363.

PAPADATOU-PASTOU, M., NTOLKA, E., SCHMITZ, J., MARTIN, M., MUNAFÒ, M. R., OCKLENBURG, S. & PARACCHINI, S. 2020. Human handedness: A meta-analysis. *Psychol Bull,* 146**,** 481-524.

PARR, S. 2007. Living with severe aphasia: Tracking social exclusion. *Aphasiology,* 21**,** 98-123.

PEDERSEN, P. M., JØRGENSEN, H. S., NAKAYAMA, H., RAASCHOU, H. O. & OLSEN, T. S. 1995. Aphasia in acute stroke: incidence, determinants, and recovery. *Ann Neurol,* 38**,** 659-66.

PEDERSEN, P. M., VINTER, K. & OLSEN, T. S. 2004. Aphasia after stroke: type, severity and prognosis. The Copenhagen aphasia study. *Cerebrovasc Dis,* 17**,** 35-43.

PEÑALOZA, C., MARTIN, N., LAINE, M. & RODRÍGUEZ-FORNELLS, A. 2022. Language learning in aphasia: A narrative review and critical analysis of the literature with implications for language therapy. *Neurosci Biobehav Rev,* 141**,** 104825.

PEREIRA, A. H., GONÇALVES, A. B., HOLZ, M., GONÇALVES, H. A., KOCHHANN, R., JOANETTE, Y., ZIMMERMANN, N. & FONSECA, R. P. 2018. Influence of age and education on the processing of clustering and switching in verbal fluency tasks. *Dement Neuropsychol,* 12**,** 360-367.

PICKERSGILL, M. J. & LINCOLN, N. B. 1983. Prognostic indicators and the pattern of recovery of communication in aphasic stroke patients. *J Neurol Neurosurg Psychiatry,* 46**,** 130-9.

PISANO, F., MANFREDINI, A., CASTELLANO, A., CALTAGIRONE, C. & MARANGOLO, P. 2022. Does Executive Function Training Impact on Communication? A Randomized Controlled tDCS Study on Post-Stroke Aphasia. *Brain Sciences* [Online], 12.

PLOWMAN, E., HENTZ, B. & ELLIS, C., JR. 2012. Post-stroke aphasia prognosis: a review of patient-related and stroke-related factors. *J Eval Clin Pract,* 18**,** 689-94.

POEPPEL, D., EMMOREY, K., HICKOK, G. & PYLKKÄNEN, L. 2012. Towards a new neurobiology of language. *J Neurosci,* 32**,** 14125-31.

PRICE, C. J. 2000. The anatomy of language: contributions from functional neuroimaging. *J Anat,* 197 Pt 3**,** 335-59.

PRICE, C. J. 2018. The evolution of cognitive models: From neuropsychology to neuroimaging and back. *Cortex,* 107**,** 37-49.

PRICE, C. J., HOPE, T. M. & SEGHIER, M. L. 2017. Ten problems and solutions when predicting individual outcome from lesion site after stroke. *Neuroimage,* 145**,** 200-208.

PRICE, C. J., SEGHIER, M. L. & LEFF, A. P. 2010. Predicting language outcome and recovery after stroke: the PLORAS system. *Nat Rev Neurol,* 6**,** 202-10.

RAMSEY, L. E., SIEGEL, J. S., LANG, C. E., STRUBE, M., SHULMAN, G. L. & CORBETTA, M. 2017. Behavioural clusters and predictors of performance during recovery from stroke. *Nat Hum Behav,* 1.

RANTALAINEN, V., LAHTI, J., HENRIKSSON, M., KAJANTIE, E., ERIKSSON, J. G. & RÄIKKÖNEN, K. 2018. Cognitive ability in young adulthood predicts risk of early-onset dementia in Finnish men. *Neurology,* 91**,** e171-e179.

RAO, P. R. 1994. The aphasia syndromes: Localization and classification. *Top Stroke Rehabil,* 1**,** 1-13.

RATCLIFF, G., GANGULI, M., CHANDRA, V., SHARMA, S., BELLE, S., SEABERG, E. & PANDAV, R. 1998. Effects of literacy and education on measures of word fluency. *Brain Lang,* 61**,** 115-22.

REF2014. 2014. Transforming clinical practice in aphasia: the Comprehensive Aphasia Test (CAT). Available: https://ref2014impact.azurewebsites.net/casestudies2/refservice.svc/GetCaseStudyPDF/21714.

REIS, A. & PETERSSON, K. M. 2003. Educational level, socioeconomic status and aphasia research: A comment on Connor et al. (2001)—Effect of socioeconomic status on aphasia severity and recovery. *Brain and Language,* 87**,** 449-452.

RESHETNYAK, E., NTAMATUNGIRO, M., PINHEIRO, L. C., HOWARD, V. J., CARSON, A. P., MARTIN, K. D. & SAFFORD, M. M. 2020. Impact of Multiple Social Determinants of Health on Incident Stroke. *Stroke,* 51**,** 2445-2453.

RICHARDS, M., SHIPLEY, B., FUHRER, R. & WADSWORTH, M. E. J. 2004. Cognitive ability in childhood and cognitive decline in mid-life: longitudinal birth cohort study. *BMJ (Clinical research ed.),* 328**,** 552-552.

RITCHIE, S. J., TUCKER-DROB, E. M., COX, S. R., CORLEY, J., DYKIERT, D., REDMOND, P., PATTIE, A., TAYLOR, A. M., SIBBETT, R., STARR, J. M. & DEARY, I. J. 2016. Predictors of ageing-related decline across multiple cognitive functions. *Intelligence,* 59**,** 115-126.

ROBERTS, P. M. & DOUCET, N. 2011. Performance of French-speaking Quebec adults on the Boston Naming Test. *Canadian Journal of Speech-Language Pathology and Audiology,* 35**,** 254-267.

ROBERTS, S., BRUCE, R. M., LIM, L., WOODGATE, H., LEDINGHAM, K., ANDERSON, S., LORCA-PULS, D. L., GAJARDO-VIDAL, A., LEFF, A. P., HOPE, T. M. H., GREEN, D. W., CRINION, J. T. & PRICE, C. J. 2021. Better long-term speech outcomes in stroke survivors who received early clinical speech and language therapy: What's driving recovery? *Neuropsychol Rehabil***,** 1-23.

ROBEY, R. R. 1998. A meta-analysis of clinical outcomes in the treatment of aphasia. *J Speech Lang Hear Res,* 41**,** 172-87.

RODRÍGUEZ-LORENZANA, A., BENITO-SÁNCHEZ, I., ADANA-DÍAZ, L., PAZ, C. P., YACELGA PONCE, T., RIVERA, D. & ARANGO-LASPRILLA, J. C. 2020. Normative Data for Test of Verbal Fluency and Naming on Ecuadorian Adult Population. *Front Psychol,* 11**,** 830.

RODRIGUEZ, F. S. & LACHMANN, T. 2020. Systematic Review on the Impact of Intelligence on Cognitive Decline and Dementia Risk. *Frontiers in Psychiatry,* 11.

ROHRER, J. M. 2018. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. *Advances in Methods and Practices in Psychological Science,* 1**,** 27-42.

ROMEO, R. R. 2019. Socioeconomic and experiential influences on the neurobiology of language development. *Perspect ASHA Spec Interest Groups,* 4**,** 1229-1238.

ROSE, S. J. 2009. Identifying and Teaching Children and Young People with Dyslexia and Literacy Difficulties. The Dyslexia-SpLD trust

ROSENICH, E., HORDACRE, B., PAQUET, C., KOBLAR, S. A. & HILLIER, S. L. 2020. Cognitive Reserve as an Emerging Concept in Stroke Recovery. *Neurorehabilitation and Neural Repair,* 34**,** 187-199.

ROSSELLI, M., ARDILA, A., FLOREZ, A. & CASTRO, C. 1990. Normative data on the Boston Diagnostic Aphasia Examination in a Spanish-speaking population. *J Clin Exp Neuropsychol,* 12**,** 313-22.

ROYAL COLLEGE OF PHYSICIANS, C. E. A. E. U. O. B. O. T. I. S. W. P. 2017. Changes over Time: 4 years of data April 2013-March 2017. Sentinel Stroke National Audit Programme (SSNAP).

SALLY, L., ANNA, C. E. S. & MARTIN, R. C. 2011. UK stroke incidence, mortality and cardiovascular risk management 1999–2008: time-trend analysis from the General Practice Research Database. *BMJ Open,* 1**,** e000269.

SALTHOUSE, T. A. 2010. Selective review of cognitive aging. *J Int Neuropsychol Soc,* 16**,** 754-60.

SANTOS NOGUEIRA, D., AZEVEDO REIS, E. & VIEIRA, A. 2016. Verbal Fluency Tasks: Effects of Age, Gender, and Education. *Folia Phoniatr Logop,* 68**,** 124-133.

SAUR, D., LANGE, R., BAUMGAERTNER, A., SCHRAKNEPPER, V., WILLMES, K., RIJNTJES, M. & WEILLER, C. 2006. Dynamics of language reorganization after stroke. *Brain,* 129**,** 1371-84.

SCAHILL, R. I., FROST, C., JENKINS, R., WHITWELL, J. L., ROSSOR, M. N. & FOX, N. C. 2003. A Longitudinal Study of Brain Volume Changes in Normal Aging Using Serial Registered Magnetic Resonance Imaging. *Archives of Neurology,* 60**,** 989-994.

SEGHIER, M. L., PATEL, E., PREJAWA, S., RAMSDEN, S., SELMER, A., LIM, L., BROWNE, R., RAE, J., HAIGH, Z., EZEKIEL, D., HOPE, T. M. H., LEFF, A. P. & PRICE, C. J. 2016. The PLORAS Database: A data repository for Predicting Language Outcome and Recovery After Stroke. *Neuroimage,* 124**,** 1208-1212.

SEGHIER, M. L., RAMLACKHANSINGH, A., CRINION, J., LEFF, A. P. & PRICE, C. J. 2008. Lesion identification using unified segmentation-normalisation models and fuzzy clustering. *NeuroImage,* 41**,** 1253-1266.

SENIÓW, J., LITWIN, M. & LEŚNIAK, M. 2009. The relationship between non-linguistic cognitive deficits and language recovery in patients with aphasia. *J Neurol Sci,* 283**,** 91-4.

SHAFTO, M. A. & TYLER, L. K. 2014. Language in the aging brain: the network dynamics of cognitive decline and preservation. *Science,* 346**,** 583-7.

SHARP, E. S. & GATZ, M. 2011. Relationship between education and dementia: an updated systematic review. *Alzheimer Dis Assoc Disord,* 25**,** 289-304.

SHAYWITZ, B. A., SHAYWLTZ, S. E., PUGH, K. R., CONSTABLE, R. T., SKUDLARSKI, P., FULBRIGHT, R. K., BRONEN, R. A., FLETCHER, J. M., SHANKWEILER, D. P., KATZ, L. & GORE, J. C. 1995. Sex differences in the functional organization of the brain for language. *Nature,* 373**,** 607-609.

SHEEHY LAURIE, M. 2006. Crossed Aphasia: A Review of the Syndrome. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders,* 16**,** 11-16.

SHEPPARD, S. M. & SEBASTIAN, R. 2021. Diagnosing and managing post-stroke aphasia. *Expert Review of Neurotherapeutics,* 21**,** 221-234.

SHIN, S., LEE, Y., CHANG, W. H., SOHN, M. K., LEE, J., KIM, D. Y., SHIN, Y. I., OH, G. J., LEE, Y. S., JOO, M. C., LEE, S. Y., SONG, M. K., HAN, J., AHN, J. & KIM, Y. H. 2022. Multifaceted Assessment of Functional Outcomes in Survivors of First-time Stroke. *JAMA Netw Open,* 5**,** e2233094.

SHING, Y. L., WERKLE-BERGNER, M., LI, S.-C. & LINDENBERGER, U. 2008. Associative and strategic components of episodic memory: A life-span dissociation. *Journal of Experimental Psychology: General,* 137**,** 495-513.

SHIRDEL, S., ESMAEELI, S., ALAVI, K., GHAEMMAGHAMI, P. & SHARIAT, S. V. 2022. Verbal Fluency Performance in Normal Adult Population in Iran: Norms and Effects of Age, Education, and Gender. *Basic Clin Neurosci,* 13**,** 129-138.

SHIRK, S. D., MITCHELL, M. B., SHAUGHNESSY, L. W., SHERMAN, J. C., LOCASCIO, J. J., WEINTRAUB, S. & ATRI, A. 2011. A web-based normative calculator for the uniform data set (UDS) neuropsychological test battery. *Alzheimer's Research & Therapy,* 3**,** 32.

SIMMONS-MACKIE, N. & KAGAN, A. 2007. Application of the ICF in aphasia. *Semin Speech Lang,* 28**,** 244-53.

SIMON, R. 2010. Clinical trial designs for evaluating the medical utility of prognostic and predictive biomarkers in oncology. *Per Med,* 7**,** 33-47.

SKILLS FOR HEALTH. 2018. Frailty: a framework of core capabilities. Available: https://www.skillsforhealth.org.uk/wp-content/uploads/2021/01/Frailty-framework.pdf.

SMANIA, N., GANDOLFI, M., AGLIOTI, S. M., GIRARDI, P., FIASCHI, A. & GIRARDI, F. 2010. How Long Is the Recovery of Global Aphasia? Twenty-Five Years of Follow-up in a Patient With Left Hemisphere Stroke. *Neurorehabilitation and Neural Repair,* 24**,** 871-875.

SMITH, M. A., LISABETH, L. D., BONIKOWSKI, F. & MORGENSTERN, L. B. 2010. The role of ethnicity, sex, and language on delay to hospital arrival for acute ischemic stroke. *Stroke,* 41**,** 905-9.

SOYLU, A. E. & CANGÖZ, B. 2018. Adaptation and Norm Determination Study of the Boston Naming Test for Healthy Turkish Elderly. *Noro Psikiyatr Ars,* 55**,** 341-348.

SSNAP. 2020. *Clinical Audit: Regional Results* [Online]. Available: https://www.strokeaudit.org/results/Clinical-audit/Regional-Results.aspx [Accessed 12th November 2020].

STAFF, R. T., HOGAN, M. J. & WHALLEY, L. J. 2018. The influence of childhood intelligence, social class, education and social mobility on memory and memory decline in late life. *Age Ageing,* 47**,** 847-852.

STAHL, B., MOHR, B., BÜSCHER, V., DREYER, F. R., LUCCHESE, G. & PULVERMÜLLER, F. 2018. Efficacy of intensive aphasia therapy in patients with chronic stroke: a randomised controlled trial. *Journal of neurology, neurosurgery, and psychiatry,* 89**,** 586-592.

STAHL, B., MOHR, B., DREYER, F. R., LUCCHESE, G. & PULVERMÜLLER, F. 2017. Communicative-Pragmatic Assessment Is Sensitive and Time-Effective in Measuring the Outcome of Aphasia Therapy. *Frontiers in human neuroscience,* 11**,** 223-223.

STARK, J. A. 2010. Long-term analysis of chronic Broca's aphasia: an illustrative single case. *Semin Speech Lang,* 31**,** 5-20.

STERN, Y. 2009. Cognitive reserve. *Neuropsychologia,* 47**,** 2015-28.

STERN, Y. 2012. Cognitive reserve in ageing and Alzheimer's disease. *Lancet Neurol,* 11**,** 1006-12.

STERN, Y., ALBERT, M., BARNES, C. A., CABEZA, R., PASCUAL-LEONE, A. & RAPP, P. R. 2023. A framework for concepts of reserve and resilience in aging. *Neurobiology of Aging,* 124**,** 100-103.

STERN, Y., ARENAZA-URQUIJO, E. M., BARTRÉS-FAZ, D., BELLEVILLE, S., CANTILON, M., CHETELAT, G., EWERS, M., FRANZMEIER, N., KEMPERMANN, G., KREMEN, W. S., OKONKWO, O., SCARMEAS, N., SOLDAN, A., UDEH-MOMOH, C., VALENZUELA, M., VEMURI, P. & VUOKSIMAA, E. 2020. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimers Dement,* 16**,** 1305-1311.

STROKE ASSOCIATION 2012. Short-changed by stroke: the financial impact of stroke on people of working age. London, UK: Stroke Association.

SULLIVAN, G. M. & FEINN, R. 2012. Using Effect Size-or Why the P Value Is Not Enough. *J Grad Med Educ,* 4**,** 279-82.

SWINBURN, K., HOWARD, D. & PORTER, G. 2004. Comprehensive Aphasia Test. Hove: Psychology Press.

TAO, C., YUAN, Y., XU, Y., ZHANG, S., WANG, Z., WANG, S., LIANG, J. & WANG, Y. 2023. Role of cognitive reserve in ischemic stroke prognosis: A systematic review. *Frontiers in Neurology,* 14.

THOMPSON, C. K. 2012. Northwestern Assessment of Verbs and Sentences (NAVS). Evanston, IL: Northwestern University.

THYE, M. & MIRMAN, D. 2018. Relative contributions of lesion location and lesion size to predictions of varied language deficits in post-stroke aphasia. *Neuroimage Clin,* 20**,** 1129-1138.

TOMBAUGH, T. N., KOZAK, J. & REES, L. 1999. Normative data stratified by age and education for two measures of verbal fluency: FAS and animal naming. *Arch Clin Neuropsychol,* 14**,** 167-77.

TREMBLAY, P. & DICK, A. S. 2016. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. *Brain and Language,* 162**,** 60-71.

TUOMIRANTA, L., GRÖNHOLM-NYMAN, P., KOHEN, F., RAUTAKOSKI, P., LAINE, M. & MARTIN, N. 2011. Learning and maintaining new vocabulary in persons with aphasia: Two controlled case studies. *Aphasiology,* 25**,** 1030-1052.

TUOMIRANTA, L., GRÖNROOS, A.-M., MARTIN, N. & LAINE, M. 2014a. Vocabulary acquisition in aphasia: Modality can matter. *Journal of Neurolinguistics,* 32**,** 42-58.

TUOMIRANTA, L. M., CÀMARA, E., FROUDIST WALSH, S., RIPOLLÉS, P., SAUNAVAARA, J. P., PARKKOLA, R., MARTIN, N., RODRÍGUEZ-FORNELLS, A. & LAINE, M. 2014b. Hidden word learning capacity through orthography in aphasia. *Cortex,* 50**,** 174-191.

TURKELTAUB, P. E., MESSING, S., NORISE, C. & HAMILTON, R. H. 2011. Are networks for residual language function and recovery consistent across aphasic patients? *Neurology,* 76**,** 1726-34.

VALDERAS, J. M., KOTZEVA, A., ESPALLARGUES, M., GUYATT, G., FERRANS, C. E., HALYARD, M. Y., REVICKI, D. A., SYMONDS, T., PARADA, A. & ALONSO, J. 2008. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Quality of Life Research,* 17**,** 179-193.

VALENZUELA, M. J. & SACHDEV, P. 2006. Brain reserve and dementia: a systematic review. *Psychological Medicine,* 36**,** 441-454.

VAN DER ELST, W., VAN BOXTEL, M. P., VAN BREUKELEN, G. J. & JOLLES, J. 2006. Normative data for the Animal, Profession and Letter M Naming verbal fluency tests for Dutch speaking participants and the effects of age, education, and sex. *J Int Neuropsychol Soc,* 12**,** 80-9.

VARKANITSA, M., GODECKE, E. & KIRAN, S. 2023. How Much Attention Do We Pay to Attention Deficits in Poststroke Aphasia? *Stroke,* 54**,** 55-66.

VARKANITSA, M. & KIRAN, S. 2022. Understanding, facilitating and predicting aphasia recovery after rehabilitation. *Int J Speech Lang Pathol,* 24**,** 248-259.

VARKANITSA, M. & KIRAN, S. 2024. Insights gained over 60 years on factors shaping post-stroke aphasia recovery: A commentary on Vignolo (1964). *Cortex,* 170**,** 90-100.

VERHAEGHEN, P. 2003. Aging and Vocabulary Scores: A Meta-Analysis. *Psychology and aging,* 18**,** 332-9.

WALLACE, S. J., ISAACS, M., ALI, M. & BRADY, M. C. 2022. Establishing reporting standards for participant characteristics in post-stroke aphasia research: An international e-Delphi exercise and consensus meeting. *Clinical Rehabilitation,* 37**,** 199-214.

WATILA, M. M. & BALARABE, S. A. 2015. Factors predicting post-stroke aphasia recovery. *J Neurol Sci,* 352**,** 12-8.

WAWRZYNIAK, M., SCHNEIDER, H. R., KLINGBEIL, J., STOCKERT, A., HARTWIGSEN, G., WEILLER, C. & SAUR, D. 2022. Resolution of diaschisis contributes to early recovery from post-stroke aphasia. *NeuroImage,* 251**,** 119001.

WERNICKE, C. 1874. Der aphasiche Symptomenkomplex: eine psychologische studie auf anatomischer basis. *Breslau: Max Cohn & Weigert;*.

WHATMOUGH, C. & CHERTKOW, H. 2002. Neuroanatomical aspects of naming. *In:* HILLIS, A. (ed.) *Handbook of Adult Language Disorders.* New York: Psychology Press.

WINANS-MITRIK RONDA, L., HULA WILLIAM, D., DICKEY MICHAEL, W., SCHUMACHER JAMES, G., SWOYER, B. & DOYLE PATRICK, J. 2014. Description of an Intensive Residential Aphasia Treatment Program: Rationale, Clinical Processes, and Outcomes. *American Journal of Speech-Language Pathology,* 23**,** S330-S342.

WING, D., EYLER, L. T., LENZE, E. J., WETHERELL, J. L., NICHOLS, J. F., MEEUSEN, R., GODINO, J. G., SHIMONY, J. S., SNYDER, A. Z., NISHINO, T., NICOL, G. E., NAGELS, G. & ROELANDS, B. 2022. Fatness, fitness and the aging brain: A cross sectional study of the associations between a physiological estimate of brain age and physical fitness, activity, sleep, and body composition. *Neuroimage Rep,* 2.

WOLDAG, H., VOIGT, N., BLEY, M. & HUMMELSHEIM, H. 2017. Constraint-Induced Aphasia Therapy in the Acute Stage:What Is the Key Factor for Efficacy? A Randomized Controlled Study. *Neurorehabilitation and Neural Repair,* 31**,** 72-80.

WORLD HEALTH ORGANISATION 2001. World Health Organisation International Classification of Functioning, Disability and Health. Geneva, Switzerland.

WORLD HEALTH ORGANISATION 2008. Closing the gap in a generation: health equity through action on the social determinants of health. Final report of the commission on social determinants of health. *Commission on Social Determinants of Health.* Geneva, Switzerland: World Health Organisation.

WORLD HEALTH ORGANISATION 2019. World report on vision.

WORLD HEALTH ORGANISATION. 2021. World report on hearing. Available: https://www.who.int/publications/i/item/9789240020481.

WORRALL, L., SHERRATT, S., ROGERS, P., HOWE, T., HERSH, D., FERGUSON, A. & DAVIDSON, B. 2011. What people with aphasia want: Their goals according to the ICF. *Aphasiology,* 25**,** 309-322.

YAO, J., LIU, X., LIU, Q., WANG, J., YE, N., LU, X., ZHAO, Y., CHEN, H., HAN, Z., YU, M., WANG, Y., LIU, G. & ZHANG, Y. 2020. Characteristics of Non-linguistic Cognitive Impairment in Post-stroke Aphasia Patients. *Front Neurol,* 11**,** 1038.

YOO, J. W., HONG, B. Y., JO, L., KIM, J. S., PARK, J. G., SHIN, B. K. & LIM, S. H. 2020. Effects of Age on Long-Term Functional Recovery in Patients with Stroke. *Medicina (Kaunas),* 56.

YOUSUFUDDIN, M., BARTLEY, A. C., ALSAWAS, M., SHEELY, H. L., SHULTZ, J., TAKAHASHI, P. Y., YOUNG, N. P. & MURAD, M. H. 2017. Impact of Multiple Chronic Conditions in Patients Hospitalized with Stroke and Transient Ischemic Attack. *Journal of Stroke and Cerebrovascular Diseases,* 26**,** 1239-1248.

ZANICHELLI, L., FONSECA, R. P. & ORTIZ, K. Z. 2020. Influence of age and schooling in written discourse of healthy adults. *Psicol Reflex Crit,* 33**,** 10.

ZEC, R. F., BURKETT, N. R., MARKWELL, S. J. & LARSEN, D. L. 2007. Normative Data Stratified for Age, Education, and Gender on the Boston Naming Test. *The Clinical Neuropsychologist,* 21**,** 617-637.

ZHANG, Y., KIMBERG, D. Y., COSLETT, H. B., SCHWARTZ, M. F. & WANG, Z. 2014. Multivariate lesion-symptom mapping using support vector regression. *Hum Brain Mapp,* 35**,** 5861-76.

ZHONG, H. H., QU, J. F., XIAO, W. M., CHEN, Y. K., LIU, Y. L., WU, Z. Q., QIU, D. H. & LIANG, W. C. 2021. Severity of Lesions Involving the Cortical Cholinergic Pathways May Be Associated With Cognitive Impairment in Subacute Ischemic Stroke. *Front Neurol,* 12**,** 606897.

ZIMMERMANN, N., CARDOSO, C. O., TRENTINI, C. M., GRASSI-OLIVEIRA, R. & FONSECA, R. P. 2015. Brazilian preliminary norms and investigation of age and education effects on the Modified Wisconsin Card Sorting Test, Stroop Color and Word test and Digit Span test in adults. *Dement Neuropsychol,* 9**,** 120-127.