# The Phenomenology of Post-Stroke Depression

Joshua Blake

Registration: 100299484

Thesis submitted in partial fulfilment of the degree of Doctorate in Clinical Psychology

August 2020

Thesis portfolio word count: 36,779

## Thesis Portfolio Abstract

**Aims:** This portfolio reviews the concept of Post-Stroke Depression (PSD) by exploring differences in the symptoms of depression between stroke survivors and members of the general population. The portfolio focuses on concerns that other stroke symptoms, such as fatigue, might be misinterpreted as depression, or vice versa.

**Structure:** This portfolio includes a systematic review of existing studies that investigated symptom differences between PSD and general population depression. An empirical paper follows, which examined a commonly used depression questionnaire, the Patient Health Questionnaire-9 (PHQ-9) and how it might unintentionally measure stroke symptoms other than depression. This involved Confirmatory Factor Analysis (CFA), a statistical method. The portfolio finishes by reflecting on this research.

**Results:** The systematic review found many similarities in the symptoms of depression between stroke and the general population. It found that loss of interest or pleasure in activity was less common in PSD. It found that problems with changeable emotions and disruption to working were more common in PSD. The empirical paper found that the PHQ-9 measures depression in stroke without problematic effects of other stroke symptoms, but that post-stroke fatigue might raise people's scores on the question relating to tiredness. Researchers should therefore not compare PHQ-9 scores between stroke and non-stroke groups without accounting for this.

**Conclusions:** We found evidence for similarities and differences in experiences of depression between stroke and non-stroke. Depression measurement was robust to other physical consequences of stroke in both papers, but questionnaire items concerning tiredness and fatigue may be more prone to measuring these problems as well as depression. PSD may therefore be treated similarly to depression in other populations, but with consideration of a person's experiences of loss, changes to personal roles and work, their mood and well-being, and changes to how quickly their emotions change.

## Table of Contents

## List of Figures

### Systematic Review

## List of Tables

**Systematic Review**



**Empirical Paper**

**Acknowledgements**

First and foremost, I would like to thank my research supervisors, Niall Broomfield, and Fergus Gracey, for their continued and kind support throughout this journey. I would also like to offer my utmost gratitude to Theresa Munyombwe and Felix Fischer for their careful statistical guidance; I could not have navigated the complex world of factor analytic research without your generous support. I would like to thank Shiona Whitmore, Assistant Psychologist, for her role as a second reviewer for my systematic review and the theoretical contributions she made to this work. I wish to acknowledge the help and support offered by the DEPRESSD team and their collaborators for inviting me to be part of their influential and exciting research project. Special thanks to Professor Brett Thombs, Dr Brooke Levis, and Dr Andrea Benedetti of the DEPRESSD steering committee, and to Sheryl Sun for her kind and patient support in managing the data transfer of eleven data clusters from across the world. Finally, I would like to thank all the DEPRESSD collaborators who contributed data to this project; I could not have accomplished this research without you.

## Chapter 1. Introduction to the Thesis Portfolio

This thesis portfolio presents research that critically examines the construct of Post-Stroke Depression (PSD) and contemporary approaches to its measurement (Burton and Tyson, 2015). The systematic review appraised existing literature relevant to this debate and presents specific hypotheses for future research. Factor analysis and measurement invariance testing were applied in the subsequent empirical paper, with the aim of providing novel insights into phenomenological and measurement differences between PSD and depression in the general population.

### Background and Rationale

Significant advancements to neuroimaging in the 1980s and 1990s supported improved understanding of the interactions between neurovascular diseases and depression, leading to an explosion of research interest in the phenomenology of PSD (Alexopoulos et al., 1997; Gainotti et al., 1997; Lipsey et al., 1986; Robinson et al., 1983). It was through these seminal research papers that the complexity of PSD became clear. Many predictors of elevated distress after a stroke were identified: lesion location (Lipsey et al., 1983), stroke severity (Lipsey et al., 1983; Pohjasvaara et al., 1998; Robinson et al., 1983), neurochemical 'endogenous' factors (Grasso et al., 1994; Starkstein and Robinson, 1989), cognitive dysfunction (Kauhanen et al., 1999), physical impairment (Kauhanen et al., 2000), residential setting and available support (Burvill et al., 1997), post-stroke emotionalism (House et al., 1989), and psychological adjustment to loss (Gainotti et al., 1997).

These observations generated competing viewpoints regarding the potential mechanisms for both the formation and maintenance of PSD: (1) that PSD is primarily endogenous, as indicated by the 'vascular depression hypothesis' (Alexopoulos et al., 1997), (2) that PSD should be considered similar in phenomenology and aetiology to depression in the general population (Lipsey et al., 1986), and/or (3) that cognitive and physical impairments of stroke contribute to the maintenance of depressed mood and negative cognitions in a biopsychosocial model (Broomfield et al., 2011). Notably, these perspectives span several philosophical stances, from the assumed *Realism* and implications of a fundamental context-independent biological reality implied by pure alignment to the vascular depression hypothesis, to *Critical Realist* and contextualist standpoints, which indicate that depression

exists but that its expression depends on individual context, implicit in the arguments of Broomfield et al. (Pilgrim, 2013).

This lack of clarity in the literature about the underlying ontology and epistemology of PSD, as well as previous limitations to methodological capabilities, means that many previous attempts to understand the phenomenology are subject to criticism. For example, some researchers may have assumed the applicability of depression measurement to stroke populations (Gainotti et al., 1997), prompting concerns that the adopted methods were biased by possible measurement interference of post-stroke emotionalism and psychological adjustment to loss, which have been posited as distinct processes that nonetheless load onto measures of depressed mood (Calvert et al., 1998; Taylor et al., 2011). Others have conceptualised elevated distress relating to, for example, emotion dysregulation or emotionalism, to exist within the construct of post-stroke depression, and designed new measures to incorporate this (Gainotti et al., 1997). Furthermore, there have been frequent concerns among clinicians and researchers that the numerous physical and cognitive sequelae may be mistaken for the somatic symptoms of depression (Cumming et al., 2010; De Coster et al., 2005), obfuscating our ability to differentiate depression-related experiences from extraneous sources. Such comorbid difficulties are, however, also likely to covary with depression because they can be formulated as part of the psychological maintenance of depression (Broomfield et al., 2011).

Since this early flurry of interest in the phenomenology of PSD, there have been considerable advancements in methodological tools that support greater potential to differentiate depression phenomenology from measurement bias. These include Factor Analysis (Picardi et al., 2008), Differential Item Functioning (DIF) and Item Response Theory (IRT) methods (Katzan et al., 2021), and latent class sub-group analysis (Ulbricht et al., 2018). This thesis portfolio, therefore, aimed to provide an updated examination of the PSD construct through phenomenological comparisons with depression in the general population.

**Philosophical Position**

As outlined above, there exists substantial diversity of opinion about the ontology and epistemology of depression after stroke, which carries implications for methodological choice and interpretation of

findings (Alexopoulos et al., 1997; Lipsey et al., 1986; Broomfield et al. 2011). Because of the contextual influences on depression measurement between populations (Maul et al., 2016) and evidence that the development and maintenance of depression after stroke is complex, multifactorial, and person-specific (Broomfield et al., 2011), we moved away from *Positivist* philosophical orientations in this portfolio.

In pursuit of the aim of understanding the phenomenology of PSD, an acknowledgement of some real construct of symptoms that tend to cluster together following a stroke has been adopted (Dong et al., 2022; Katzan et al., 2021). As such, the ontological position of this portfolio was *Realist*. This reality, though, is conceptualised in a more nuanced way than typical psychiatric definitions, which tend to view depression as a mental disorder, whose existence in an individual is often binary and dependent on the categorical presence of presupposed symptoms (Pilgrim, 2013). In this thesis, depression has been conceptualised as the presence of relevant affective, behavioural, cognitive, interpersonal and social experiences (symptoms) whose expression in any given individual is determined by an immeasurable set of predictive factors, including reciprocal or circular relationships, spanning biological, psychological, and social levels. Under this view, the superficial 'truth' of depression to those observing it would vary between individual, social, and cultural contexts because of inevitable differences in the immeasurable list of factors that influence its expression between such contexts. Concretely, individuals across nations may conceptualise the symptom profile of depression differently based on how depression is locally expressed (Juhasz et al., 2012), but these local truths can nonetheless be conceptualised as different expressions of the same underlying reality, simply with different inputs. Thus, the underlying reality of depression is seen as a metraphorical blueprint of how symptoms will be expressed depending on the context.

Because of concerns about the fundamental limitation of measurement tools as indicators of an inaccessible internal reality, the dependency of the expression and perception of post-stroke depression on individual, social, and cultural context (i.e. Broomfield et al., 2011), and concerns about a taken-for-granted equivalence of the depression construct between populations, a *Contextualist* epistemological position was adopted (Annis, 1978). Questionnaire items were interpreted only as

*indicators* of an unobservable reality and the interface between the person and the measure was seen as inevitably influenced by individual, group-level, and cultural characteristics. This combination of ontological and epistemological positions is captured by the *Critical Realist* philosophical stance (Pilgrim, 2013).

**Thesis Outline**

The systematic review is presented first, followed by an extended methodology chapter that outlines the numerous methodological considerations employed in the review. A brief bridging chapter summarises the limitations with existing research and outlines the rationale for the subsequent empirical paper. The thesis portfolio concludes with a critical discussion chapter, which integrates the findings of the two papers, outlines methodological strengths and weaknesses, engages with wider issues in PSD phenomenology research, and outlines avenues for future research.

# Chapter 2. A Phenomenological Comparison of Post-Stroke Depression with Depression in the General Population: A Systematic Review

Blake, J.[1], Gracey, F. [1], Whitmore, S.[1], and Broomfield, N. [1]

[1]Department of Clinical Psychology and Psychological Therapies, University of East Anglia, UK

**Corresponding author**

Joshua Blake

Department of Clinical Psychology and Psychological Therapies

Norwich Medical School

University of East Anglia

Norwich Research Park

Norwich, NR4 7TJ

Email: Joshua.blake@nhs.net

Abstract word count: 191

Manuscript word count (not including title page, tables, figures abstract, or references): 7,457

Disclaimer: this manuscript has been formatted for submission to *Neuropsychological Rehabilitation.* In accordance with this guidance, American spellings are used throughout this manuscript.

**Keywords:** Post-stroke Depression, Phenomenology, Symptomatology, Stroke, Systematic Review

**Abstract**

Previous research into the differences in the phenomenology of Post-Stroke Depression (PSD) has typically focused on comparisons of symptom profiles between stroke groups and general population controls. This systematic review aimed to synthesize these findings with results from other methodologies that contributed comparisons of PSD phenomenology. Articles were identified via a search of seven databases and a manual search of references from relevant articles. Twelve articles comparing the symptomatology of PSD between stroke and non-stroke healthy controls were included. Three distinct methodological approaches, relevant to the aim, were identified: comparisons of profiles among groups with similar overall depression severity, comparisons of the strengths of correlations between a symptom and depression, and comparisons of latent symptom severity. A narrative synthesis approach was adopted because of high methodological heterogeneity. The symptomatology of depression was found to be broadly similar between groups. The stroke groups presented with comparatively less severe/prevalent anhedonia and more severe/prevalent emotionalism and disruption to work. The prevalence and severity of somatic symptoms appeared similar between groups, despite the hypothesised interference of comorbid physical stroke effects. Possible mechanisms for observed differences and similarities are explored, and areas for future research outlined.

## Introduction

Depression is a common consequence of stroke, occurring in approximately one-third of survivors (Mitchell et al., 2017). Depression after stroke is associated with poorer functional outcomes, reduced social engagement, and higher rates of mortality. Accordingly, Post-Stroke Depression (PSD) must be assessed accurately so that effective and targeted intervention is made available (Deng et al., 2017; Robinson and Jorge, 2016).

The provision of accurate assessment and support for PSD arguably requires a clear and grounded conceptualization of how depression manifests in this population. However, attempts to understand the phenomenology and etiology of PSD have been complicated by the wide range of morbidities after strokes, such as physical and cognitive disability, functional impairment, fatigue, personality changes, and neurovascular alterations (Duncan, 1994; Hu et al., 2017; Teasdale and Engberg, 2010). Broomfield et al. (2011) outlined four examples of how these factors can interact to heighten the complexity of PSD assessment and conceptualization: 1) the impact of physical impairment on activity engagement and social participation, 2) the depressogenic effect of medical comorbidities and neurobiological alterations, 3) the presence of stroke-specific negative attributions, and 4) the impact of cognitive dysfunction in biasing information processing in favor of depression-reinforcing appraisals. PSD must, therefore, be understood as complex and multi-faceted, with unique interactions between the biological, psychological, and social levels (Dowswell et al., 2000; Mitchell et al., 2017; Newberg et al., 2009; Shi et al., 2017).

Despite the presence of these additional complexities, significant debate in the literature as to whether symptomatology and phenomenology of PSD differ significantly from depression in non-stroke populations has remained. For example, proponents of the vascular depression hypothesis have argued that the etiology of depression in populations with neurovascular diseases may be distinct from major depression in the general population (Aizenstein et al., 2016; Alexopoulos et al., 1997), stemming from findings that neurovascular changes are independent predictors of depressive experiences (Pan et al., 2012; Thomas et al., 2004) and associated with poorer response to treatment (Aizenstein et al., 2014). Furthermore, qualitative studies of depression in stroke populations have

outlined experiences and narratives that appear to be unique to this group; such studies have highlighted themes of identity loss, aloneness in post-stroke experience, self-blame, guilt and burden-related beliefs, when reflecting on life before and after their stroke (Crowe et al., 2016; Taule and Råheim, 2014).

The view that PSD should be considered clinically distinct from major depressions is, however, not without criticism. While qualitative studies have indicated the presence of narratives and meanings that are unique and specific to the experience of stroke recovery (Crowe et al., 2016; Taule and Råheim, 2014), such studies are unable to indicate population-level differences. Differences in narrative or cognitive accounts of, for example, guilt between stroke and non-stroke depression might not be indicative of differences in the frequency, severity, and functional impact of guilt-related cognitions more broadly, and several studies have found evidence of similarities in depression profiles (de Man-van Ginkel et al., 2015; Gainotti et al., 1999; Lipsey et al., 1986). Furthermore, despite concerns that responses to somatic depression items are primarily caused by bias from extraneous sources, such as post-stroke fatigue (Acciarresi et al., 2014), comparisons of profiles between depressed and non-depressed stroke groups indicate that somatic items capture substantial variance attributable to depression, suggesting that somatic depression symptoms also exist in PSD (de Man-van Ginkel et al., 2015; Robinson, 2006).

Studies that compare symptom profiles in this way are, however, only one of many possible methodological approaches to the comparison in symptomatology between depressed and non-depressed groups. With a high diversity of methodologies that could provide insights into phenomenological differences associated with PSD, a systematic review affords a unique opportunity to synthesize the respective insights provided by each methodology and develop a clearer picture of whether PSD should be considered a distinct clinical phenomenon.

As far as we know, no systematic review or meta-analysis has so far investigated the comparative phenomenology of PSD and depressed mood in the general population. This review was, therefore, conducted to answer the following research question: are there population-level differences in symptomatology between PSD and depression in the general population?

## Methods

A scoping search indicated significant heterogeneity in methodology, because of variation in the stroke measures used, the time elapsed since the index stroke event, methods of comparison, nationality, residential setting, and other factors. Accordingly, a narrative synthesis approach, following guidance provided by Popay et al. (2006), was adopted.

The search was conducted in September 2021, followed by an update search in January 2022. The review was registered to Prospero on 18th August 2021 (ID: CRD42021272862). At the time of registration, no similar reviews were registered on Prospero or the Cochrane database.

### Eligibility Criteria

Eligibility criteria were generated using the PICOS (Population, Intervention, Comparison group, Outcomes, and Study design) framework for systematic review design (Methley et al., 2014; Pollock and Berge, 2018). This review did not focus on clinical intervention, so this criterion was removed.

#### *Population*

**Inclusion Criteria.** Articles examining adults with a current diagnosis or history of stroke or strokes, of both ischemic and hemorrhagic origin.

**Exclusion Criteria.** Studies examining people with Transient Ischemic Attacks (TIAs) with or without the presence of a stroke. Articles examining populations with separate or additional acquired or progressive neurological conditions, such as hemorrhages secondary to traumatic brain injury, small vessel disease, or vascular dementia. Articles that were not written in English were excluded because translation services were not available.

#### *Study Outcome*

**Inclusion Criteria.** Articles that used validated quantitative measures. Decisions about sufficient measure validity were based on the quality of initial validation studies and whether the measure has specific stroke or ABI validity evidence.

**Exclusion Criteria.** Studies that did not contain quantitative data on depressive symptoms. Social factors, such as social isolation, were not of interest to the current study because these factors

have been adequately investigated by previous reviews and meta-analyses of predictors of PSD (e.g. Hackett and Anderson, 2005).

### *Comparison Group*

**Inclusion Criteria.** Studies that used a comparison group of healthy individuals without neurological impairment.

**Exclusion Criteria.** Studies that focused on specific health conditions, such as heart disease or orthopedic injury.

### *Study Design and Analysis*

**Inclusion Criteria.** Any studies that utilized quantitative analyses that could provide valid insight into phenomenological between-group differences.

**Exclusion Criteria.** Studies that statistically compared overall depression scores or compared depressive symptoms without accounting for differences in overall depression severity; secondary sources, such as book chapters, systematic reviews or meta-analyses; studies with unextractable data.

## Search Strategy

The search was completed on EBSCOhost, using the following databases: Academic Search Complete, AMED (The Allied and Complementary Medicine Database), APA PsycArticles, APA PsycInfo, CINAHL Complete (Cumulative Index of Nursing and Allied Health Literature), MEDLINE Complete, and OpenDissertations. The search was conducted on each database separately, and the terms were applied to article subject, keyword, title and abstract (see Appendix B for search strategy). The search terms, which included a mixture of keywords and MESH terms, were as follows:

> *("stroke" or "cerebrovascular accident*" or "post-stroke" or "subarachnoid hemorrhage" or "cerebral infarct*" or "lacunar infarct*" or "lacunar stroke" or "cerebral hemorrhage" or "Hypoxia-ischemia, Brain" or "brain infarction") AND ("low mood" or "depress*" or "mood" or "wellbeing" or "distress*" or "affect" or "psychological distress" or "Stress, psychological" or "psychological distress" or "mental depression") AND ("phq-9" or "phq-2" or "phq9" or "patient health questionnaire-9" or "patient health questionnaire" or*

*"patient health questionnaire-2" or "Geriatric Depression Scale" or "GDS" or "GDS-15" or "hospital anxiety and depression scale" or "HADS" or "Center for Epidemiologic Studies Depression Scale" or "CES-D" or "Beck Depression Inventory" or "Beck Depression Inventory-II" or "BDI-II" or "BDI" or "Structured Clinical Interview for DSM-IV" or "SCID" or "SCID-II" or "The Structured Clinical Interview for DSM-5" or "Composite International Diagnostic Interview" or "CIDI" or "Diagnostic Interview Schedule" or "Mini-International Neuropsychiatric Interview" or "MINI" or "M.I.N.I" or "Aphasia Depression Rating Scale" or "ADRS" or "Brief Assessment Schedule Depression Cards" or "BASDEC" or "Montgomery–Asberg Depression Rating Scale" or "MADRS" or "Psychiatric Assessment System" or "Schedule for Affective Disorders and Schizophrenia" or "SADS" or "Schedules for Clinical Assessment in Neuropsychiatry" or "Signs of Depression Scale" or "SODS" or "Visual Analogue Mood Scale" or "VAMS" or "Hamilton Depression Rating Scale" or "HAM-D").*

A manual search was completed by screening reference lists of included articles, reviews, or book chapters that were relevant to the review question (Robinson, 2006; Robinson and Spalletta, 2010). Relevant papers were combined with articles from the above search. Duplicated articles were removed using Mendeley's removal function (Reiswig, 2010) and exported to Rayyan for screening (Ouzzani et al., 2016).

**Screening and Selection**

Articles were first screened for relevance to the criteria by title, followed by abstract and full text. A second reviewer was allocated 10% of the total and followed the same iterative process, blind to the ratings of the primary author. After each stage, the primary and second authors reviewed and resolved incidences of conflict. In cases where the primary author had not considered potentially relevant constructs or methods, the primary author re-screened the excluded articles under the refined criteria.

**Quality Rating**

Quality assessment was employed primarily to assess the risk of bias. There is considerable variation of opinion about the most appropriate quality assessment tool for observational studies (Ma et al.,

2020; Protogerou and Hagger, 2018). The National Heart, Lung, and Blood Institute (NHLBI) Quality Assessment Tool for the Quality Assessment of Observational Cohort and Cross-sectional Studies was selected for the present study (National Heart Lung and Blood Institute, 2013) because of its focus on methodological quality and bias, consideration of the validity of outcome measures, and flexibility to variation in methodology.

The assessment tool consists of fourteen items, with nominal responses of 'Yes', 'No', or 'Other (cannot determine, not reported, not applicable)'. An overall quality rating of 'Good', 'Fair', or 'Poor' is determined based on the judgment of the reviewers, rather than by computation, which supports flexibility in weighting items that are important for the specific methodology of the study. Each shortlisted paper, after full-text screening, was rated for quality using the same criteria.

**Data Extraction and Analysis**

Studies meeting all criteria were extracted for participant and sample characteristics, study design, stroke characteristics (e.g. time since index stroke event, type of stroke), outcome measures used, method of analysis, and key findings. For each study, the findings for each symptom were coded in the direction of significance; studies that found greater prevalence, severity, or correlation of a symptom with depression in the stroke group were coded as 'more', studies that found the reverse were coded as 'less', and non-significant findings were coded as 'no difference'. Significance was based on the method of significance testing reported by the study, or similar studies if significance testing was not conducted; in this latter case, recommendations of significance criteria from other papers (e.g. de Man-van Ginkel et al., 2015) were applied by the reviewers of the current paper to determine effect direction.

For studies that investigated multiple groups or used multiple measures, comparisons of symptom differences were extracted separately for each group and each measure used. For example, a study that sampled stroke groups at three different timepoints post-stroke and compared each to the non-stroke group would have three sets of extracted data. Each study could, therefore, contribute multiple findings.

Categorization of the time since the index stroke event of the stroke group(s) was made for each comparison. The four categories were: '<6 weeks', '6 – 12 weeks', '12 weeks to 1 year', and '>1 year'. These categories were based on approximate thresholds for recovery stages reported in the literature. Most stroke recovery is observed before 12 weeks (Wade et al., 1985) and approaches a flattening of trajectory beyond one year (Kotila et al., 1999).

Many symptoms were extracted because of the variability of measures used and different symptoms assessed by each measure (Cumming et al., 2010; de Man-van Ginkel et al., 2015; House et al., 1991). Dimension reduction was, therefore, performed on the extracted symptoms to consolidate them into a manageable set of broader symptom domains and to support comparisons of similar or overlapping symptoms between measures. In cases where findings for multiple symptoms loaded onto the same domain, a scoring method was used to determine the overall significance category of that new higher-order domain; all symptoms within the domain were scored +1 for a 'more' finding, -1 for 'less', and 0 for 'no difference'. The summed score was divided by the number of symptoms in that category with reported findings. Combined scores between -.5 and +.5 were assigned 'no difference' and scores greater than ±.5 were assigned the category 'more' or 'less'. This ensured that the presence of only one 'more' or one 'less' finding amidst multiple 'no difference' findings did not overstate the level of overall difference within that domain. This approach is consistent with the methods outlined by Thomson and Thomas (2013).

The analysis involved explorations of patterns of 'more', 'less', and 'no difference' findings relating to explanatory variables, such as time since stroke. The main analysis of phenomenological differences involved analysing the proportions of 'more', 'less', and 'no difference' findings for each symptom domain.

Because of many possible methodological approaches capable of answering the research question, the feasibility of aggregating 'more', 'less', and 'no difference' findings across these approaches was considered. For example, whether a finding of greater prevalence of a symptom in stroke could be aggregated with a study reporting a stronger correlation of a symptom with depression. In both cases, a 'more' rating would be ascribed but judgement about the compatibility of

combining these findings was required for analyses that examined patterns in the proportion of 'more', 'less', and 'no difference' findings. In cases of incompatibility, proportional differences in effect direction between groups were analysed separately for each methodology. Such judgments were made with the review team and are reported in the following section.

## Results

### Study Inclusion

The identification and screening process is summarised in Figure 1. From 4462 original articles identified, 58 articles were selected for full-text screening. Most articles were ineligible, most frequently because of the non-reporting of statistics that allowed a valid comparison of depressive symptomology between groups. References for excluded articles at the full-text screening stage are outlined in Appendix C. Twelve eligible studies were included in this review.

### Quality Assessment and Risk of Bias

Results of the quality assessment of each study are summarised in Table 1. Details of item-level responses are provided in Appendix D. Two studies were rated as 'good' in quality, and 9 studies as 'fair'. Two studies were allocated a 'fair to poor' rating, primarily because they featured a newly designed depression measure, the Post-Stroke Depression Rating Scale (PDRS) with limited validation evidence (Gainotti et al., 1999; Gainotti et al., 1997). Despite these weaknesses, these studies were included in the analysis because the methodology was otherwise of high relevance to the research question.

### Study Details

Characteristics of each of the final studies are provided in Table 1.

#### *Design and Methodology*

Three distinct methodologies for indicating symptomatologic differences in depressed mood between stroke and non-stroke participants were identified: (1) comparisons of depression symptom profiles, where depression severity is approximately controlled between groups, (2) comparisons of correlation strengths between a depression symptom and general depression, and (3) Differential Item Functioning (DIF) analysis using Item Response Theory (IRT). Profile comparison studies

investigated either between-groups differences in percentage prevalence of positive endorsement of a depression symptom, or differences in mean symptom score (symptom severity). Comparative correlation studies investigated the correlation between a symptom measure, such as a self-esteem questionnaire, and scores on a depression measure (Vickery et al., 2008). IRT DIF studies offer different insights into phenomenology to profile comparisons or differences in association strengths by comparing differences in the latent symptom severity of items/symptoms, the underlying depression severity to which they are sensitive. In other words, positive responses to some items occur, on average, with only minimal depression, and items are endorsed only in severe depression. IRT arranges items in order of the range of the average underlying depression severity that they are sensitive to, and DIF assesses differences in these latent severities between groups.

Most studies were cross-sectional, except for House et al. (1991), who explored longitudinal changes in depression profiles. Because several studies examined multiple groups (Gainotti et al., 1999; House et al., 1991; Schramke et al., 1998), and therefore contributed multiple comparisons, there were twenty between-group comparisons extracted from twelve studies; Vickery et al. (2008) contributed two findings, Gainotti et al. (1999) three findings, House et al. (1991) three, and Schramke et al. (1998) contributed four. All other studies contributed one finding each.

***Participants***

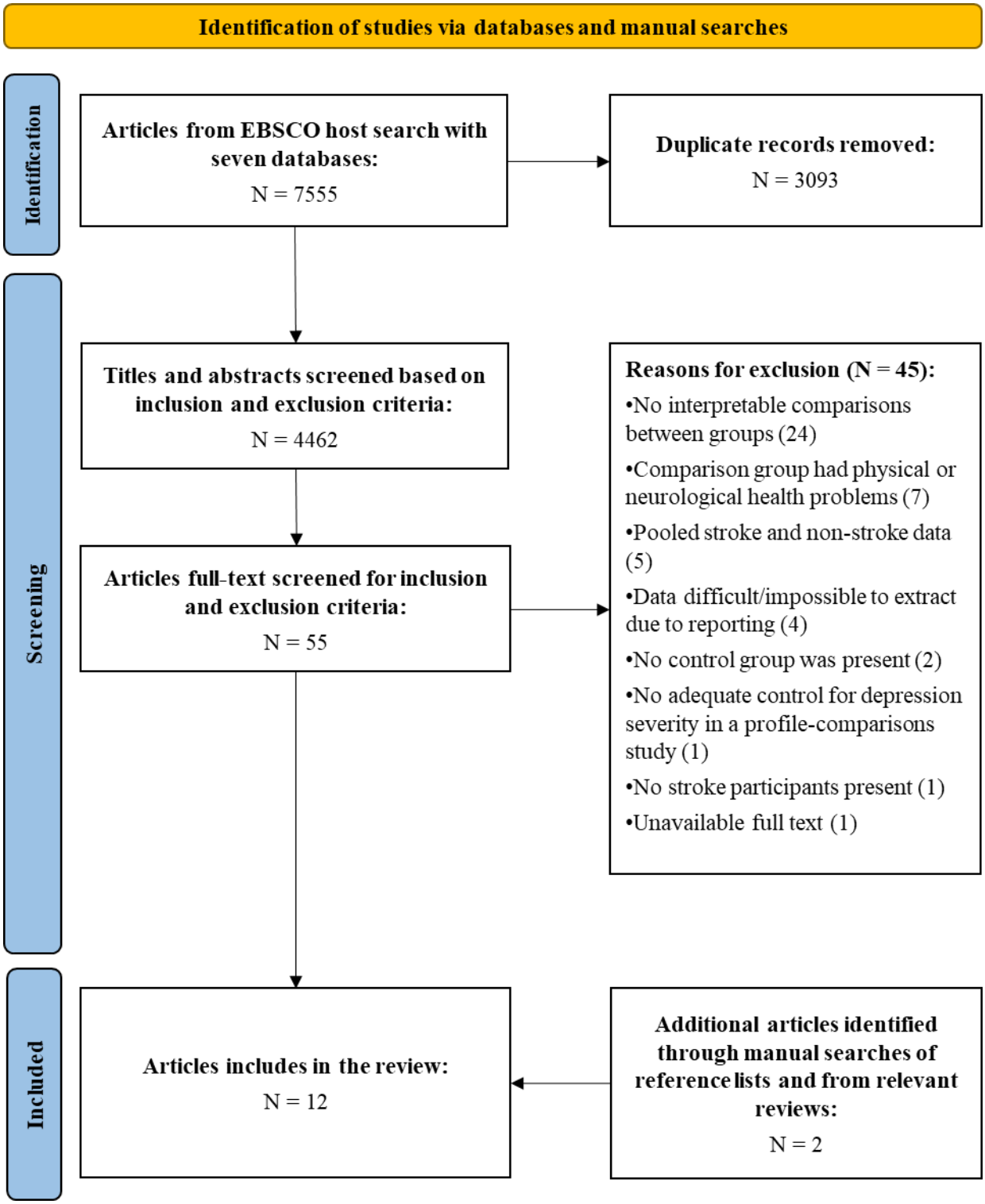
The combined studies featured 1024 stroke-group and 1741 comparison-group participants, with a total sample of 2765. Participants were sampled from seven countries, all Western developed nations. Ethnicity was inconsistently reported and, therefore, could not be analysed.

The time elapsed since the index stroke event varied considerably between studies, from two weeks to many years. Stroke participants were sampled from inpatient settings, which were generally associated with earlier recovery time points, and the community. Most studies did not investigate lateralisation, except for Schramke et al. (1998). Five studies included only participants who experienced a first stroke. Stroke severity was rarely reported; three studies reported scores on Barthel Index for Activities of Daily Living (de Man-van Ginkel et al., 2015; House et al., 1991; Stokes et al., 2011), but none used a specific indicator of stroke severity, such as the Stroke Impact Scale (Duncan

et al., 1999). Stroke sample sizes ranged from 22 (Schramke et al., 1998) to 149 (Cumming et al., 2010); sample size justification was infrequently reported (Pickard et al., 2006).

Study comparison groups were mostly community-based (9/12). The remaining three papers, all profile comparison studies, sampled depressed psychiatric inpatients (Gainotti et al., 1999; Gainotti et al., 1997; Lipsey et al., 1986). Substantial between-groups demographic differences were reported in two studies: de Man-van Ginkel et al. (2015) reported significant differences in several demographic categories, including age, gender, and education level, and Pickard et al. (2006) also reported substantial differences in age, gender, and nationality of included participants. Demographic comparisons were not reported in two studies (Bennett et al., 2006; Gainotti et al., 1999). Control group sample sizes ranged from 24 (Schramke et al., 1998) to 745 (Cumming et al., 2010).

**Figure 1** A flowchart of the article identification, screening and selection process, adapted from guidelines and templates published by Page et al. (2021)




From: Page,et al. (2020). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

For more information, visit: http://www.prisma-statement.org/

**Table 1** Study characteristics of included articles

| Authors (year) | Methodological category and design | Quality rating | Participant characteristics | Setting | Outcome measures | How was significance decided? | Main findings (reported as questionnaire items/depression symptoms before dimension reduction) |
|---|---|---|---|---|---|---|---|
| Gainotti et al. (1999) | **Category:** profile comparisons (symptom severity)<br><br>**Design:** cross-sectional between-groups | Fair to poor | **Stroke group:** Three groups with single first-time monohemispheric stroke. Three groups based on time since stroke: <2 months (N = 58), 2-4 months (N = 52), > 4 months (N = 43). Only depressed people included, selected from the above samples.<br><br>**Control group:** 30 mental health inpatients for endogenous depression. Demographic comparisons between the control and stroke groups were not reported. Only depressed people included. | **Stroke group:** Italian citizens. Unclear residence.<br><br>**Control group:** Italian inpatients | Post-Stroke Depression Rating Scale (PDRS) | No significance test in source paper. The source paper authors used > 1 PDRS score point difference in a symptom between groups as the criterion for difference | Findings depend on each timepoint since the index stroke. Data entered separately during synthesis.<br><br>**Symptoms consistently (≥2/3 stroke groups) more severe in stroke:** catastrophic reactions, hyper-emotionalism, diurnal variations<br><br>**Symptoms consistently (≥2/3 stroke groups) more severe in controls:** mood, suicidal ideation, anhedonia<br><br>**Symptoms with no consistent difference:** vegetative disorders, apathy, anxiety, guilt |
| Gainotti et al. (1997) | **Category:** profile comparisons (symptom severity)<br><br>**Design:** cross-sectional between-groups | Fair to poor | **Stroke group:** Three groups with single first-time monohemispheric stroke (N=124). The time elapsed since the onset of stroke was between two weeks and six months. Stroke patients were sorted into post-stroke major | **Stroke group:** Italian residents. Unclear residence.<br><br>**Control group:** | Post-Stroke Depression Rating Scale (PDRS) | Duncan Test for symptom-level comparisons in source paper.<br><br>Only comparisons made between post-stroke major depression and the | **Symptoms more severe in stroke:** anxiety, catastrophic reactions, hyper-emotionalism, diurnal variations<br><br>**Symptoms more severe in controls:** mood, suicidal ideation, anhedonia |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | depression, minor depression, and no depression.<br><br>**Control group:** 17 mental health inpatients for endogenous depression, matched to the stroke group by age and education. Only depressed people included. | Italian psychiatric inpatients | | control group were extracted, not the 'post-stroke minor depression' or 'no depression' samples, as only this had broadly similar overall depression severity | **Symptoms with no difference:** vegetative disorders, apathy |
| House et al. (1991) | **Category:** profile comparisons (symptom prevalence)<br><br>**Design:** longitudinal between-groups | Good | **Stroke group:** 128 first-stroke patients assessed at two-to-three time points post-stroke: 1 month (n=78), 6 months (n = 107), and 12 months (n =88). Additional participants were recruited between 1 and 6 months. 45% had less than maximum Barthel ADLs immediately post-stroke. Severity difficult to determine.<br><br>**Control group:** 111 participants randomly sampled from general practice, using stratified random sampling approximately matched for age and sex. | **Stroke group:** English people living in the community.<br><br>**Control group:** English people living in the community | Beck Depression Inventory (BDI); positive ratings defined a symptom as present. | No significance test or other criteria for difference was used in the source paper; a prevalence ≥10% was used by the authors of the current paper as a criterion for significance, based on the methodology of de Man-van Ginkel et al. (2015), to maximise consistency of significance appraisal | Findings depend on each timepoint since the index stroke. Data entered separately during synthesis.<br><br>**Symptoms consistently (≥2/3 stroke groups) more prevalent in stroke:** work inhibition/inability to work<br><br>**Symptoms consistently (≥2/3 stroke groups) more prevalent in controls:** sleep problems, loss of libido<br><br>**Symptoms with no consistent difference:** guilt, depressed mood, loss of interest, tiredness, suicidal ideation, sense of failure, self-hate, irritability, social withdrawal, indecisiveness, self-accusation, crying, body image |
| Lipsey et al. (1986) | **Category:** profile comparisons | Fair | **Stroke group:** 43 ischaemic or haemorrhagic stroke patients assessed less than six | **Stroke group:** US citizens in | Present State Examination (PSE) | Chi-square test of significance, used in the source paper | **Symptoms more prevalent in stroke:** slowed down |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (symptom prevalence)<br><br>**Design:** cross-sectional between-groups | | months following their stroke. All stroke patients were depressed accordingly to DSM-III criteria for major depression<br><br>**Control group:** 43 'functionally depressed' participants. Control participants did not differ in any key demographic. Significantly fewer poststroke patients reported previous psychiatric disorder | the community or inpatient wards<br><br>**Control group:** US psychiatric inpatients | | | **Symptoms more prevalent in controls:** loss of interest and concentration<br><br>**Symptoms with no difference:** simple depression, general anxiety, affective flattening, hypomania, overactivity, special features of depression, agitation, self-neglect, ideas of reference, tension, lack of energy, worrying, irritability, social unease, other symptoms of depression |
| Cumming et al. (2010) | **Category:** profile comparisons (symptom prevalence) and factor score comparisons<br><br>**Design:** cross-sectional between-groups | Good | **Stroke group:** 149 ischaemic or haemorrhagic stroke patients assessed eighteen months post-stroke, diagnosed with post-stroke depression<br><br>**Control group:** 745 age- and sex-matched general population controls. Recruited from previous studies. Because they were general population samples, a small percentage (9%) had a history of stroke. Controls had a diagnosis of a major depressive episode | **Stroke group:** Swedish citizens in the community<br><br>**Control group:** Swedish citizens in the community | Montgomery -Asberg Depression Rating Scale (MADRS) | Mann Whitney U significance test, used in the source paper | **Symptoms more prevalent in stroke:** none<br><br>**Symptoms more prevalent in controls:** inability to feel, disturbed sleep<br><br>**Symptoms with no difference:** sadness, suicidal thoughts, observed sadness, inner tension, disturbed appetite, concentration difficulties, loss of energy, pessimistic thoughts |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| de Man-van Ginkel et al. (2015) | **Category:** profile comparisons (symptom prevalence)<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 54 depressed ischaemic or haemorrhagic stroke patients assessed six to eight weeks post-stroke, selected from a sample of 382 stroke patients. Median Barthel score was 97.5, indicating mild stroke severity<br><br>**Control group:** 150 depressed general practice patients, selected from a wider pool of 1160 general practice patients. Uncontrolled demographic differences reported. | **Stroke group:** Dutch citizens in the community<br><br>**Control group:** Dutch citizens in the community | Patient Health Questionnaire-9 (PHQ-9) | No significance test used in the source paper; the source paper authors used a prevalence of ≥10% an indicator of significance | Data presented for non-depressed groups too, but data were only extracted for depressed groups.<br><br>**Symptoms more prevalent in stroke:** concentration, motor retardation, suicidal ideation<br><br>**Symptoms more prevalent in controls:** anhedonia, appetite disruption, guilt<br><br>**Symptoms with no difference:** depressed mood, sleep problems, tiredness/low energy |
| Bennett et al. (2006) | **Category:** comparative correlation strengths<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 79 stroke inpatients assessed two to four weeks post-stroke.<br><br>**Control group:** 49 healthy older adults, recruited via convenience sampling | **Stroke group:** UK subacute stroke inpatients<br><br>**Control group:** UK older adults in the community | **Depression measure:** Hospital Anxiety and Depression Scale depression subscale (HADS-D)<br><br>**Associated symptom and measure:** Self-esteem using the Visual | Correlation coefficient (Spearman rho), used in the source paper.<br><br>A Fisher r to z transformation was performed by the current reviewers to test differences in correlation strengths | Correlation coefficients between depression (HADS-D) and symptom measure (VASES). Fisher r to z supports a significance test between two correlation strengths.<br><br>**Stroke correlation:** r = -.52, p<.001<br><br>**Control correlation:** r= -.54, p<.001<br><br>**Fisher r to z:** non-significant, z = -.15 p =. 88 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | Analog Self-esteem Scale (VASES) | | |
| Fleming et al. (2021) | **Category:** comparative correlation strengths<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 69 stroke participants, recruited via convenience sampling. Participants were >3 months post-stroke<br><br>**Control group:** 63 healthy older adults, recruited via convenience sampling. Controls were similar in age and sex. | **Stroke group:** UK residents in the community<br><br>**Control group:** UK residents in the community | **Depression measure:** Hospital Anxiety and Depression Scale depression subscale (HADS-D)<br><br>**Associated symptom and measure:** Sleep quality/insomnia using the Sleep Condition Indicator (SCI) | A regression of SCI to HADS-D scores was reported in the source paper. If the variable of stroke status contributed to the model r squared, this would be interpreted as an indicator of different association strength | The non-stroke control group independent variable did not contribute significantly to the model, indicating a non-specific difference between groups in the relationship between insomnia and depression. The authors concluded that the relationship between insomnia and depression was not specific to stroke. |
| Schramke et al. (1998) | **Category:** comparative correlation strengths<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 22 participants with a single Right Hemisphere (RH) stroke and 22 participants with a single Left Hemisphere stroke (LH), recruited via convenience sampling. The mean time since stroke for the RH group | **Stroke group:** US residents in the community<br><br>**Control group:** US residents in | **Depression measures:** Centre for Epidemiologic Studies-Depression Scale (CES-D) and the Hamilton | Correlation coefficient with Fisher r to z transformation, reported by the source paper. | **RH stroke correlation of the BAI with CES-D:** $r^2 = .57$, p=.13<br>**LH stroke correlation of the BAI with CES-D:** $r^2 = .72$, p<.01<br>**Control correlation of the BAI with CES-D:** $r^2 = .85$, p<.01<br><br>**R to z comparison RH to control for the BAI and CES-D:** |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | was 3.6 (SD: 3.41) and LH was 3.5 (SD: 2.98).<br><br>**Control group:** 24 controls, recruited via friends and family of stroke participants. Controls were similar in age, sex, and education level. | the community | Depression Rating Scale (HDRS)<br><br>**Associated symptom and measure:** Anxiety using the Beck Anxiety Inventory (BAI) | | Significant difference (less associated in stroke)<br>**R to z comparison LH to control for the BAI and CES-D:** Non-significant<br><br>**RH stroke correlation of the BAI with HDRS:** $r^2 = .47$, p=.05<br>**LH stroke correlation of the BAI with HDRS:** $r^2 = .44$, p=.05<br>**Control correlation of the BAI with HDRS:** $r^2 = .80$, p<.01<br><br>**R to z comparison RH to control for the BAI and HDRS:** Significant difference (less associated in stroke )<br>**R to z comparison LH to control for the BAI and HDRS:** Significant difference (less associated in stroke) |
| Stokes et al. (2011) | **Category:** comparative correlation strengths<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 69 first stroke participants, recruited via convenience sampling. Participants were <3 years post-stroke. Barthel index mean score: 86 (moderate impairment)<br><br>**Control group:** 63 age- and gender-matched controls. | **Stroke group:** Irish residents in the community<br><br>**Control group:** Irish residents in the community | **Depression measure:** Geriatric Depression Scale (GDS)<br><br>**Associated symptom and measure:** Fatigue using the Multidimens | A comparison of differences in the average effect of a one-point increase in GDS score on MFI general scores, reported by the source paper. Higher relative increases in MFI scores caused by higher GDS scores are indicative of a | **Stroke group:** a one-point increase in GDS score corresponds with a 0.1 increase in MFI general score. This is a non-significant relationship (p =.5)<br><br>**Control group:** a one-point increase in GDS score corresponds with a 0.4 increase in MFI general score, reaching significance (p <.05)<br><br>**Comparison:** the difference in effect of a one-point increase in |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | ional Fatigue Inventory (MFI) general domain | stronger correlation. | GDS between the two groups reached statistical significance (p <.05). Fatigue is less correlated with depression in the stroke group. |
| Vickery et al. (2008) | **Category:** comparative correlation strengths<br><br>**Design:** cross-sectional between-groups | Fair | **Stroke group:** 80 stroke inpatients, recruited via convenience sampling. Patients were assessed approximately two weeks post-stroke<br><br>**Control group: 80** volunteers recruited via convenience sampling, matched for age and education. | **Stroke group:** US inpatients<br><br>**Control group:** US citizens in the community | **Depression measure:** Geriatric Depression Scale (GDS)<br><br>**Associated symptom and measure:** Self-esteem, using the Visual Analogue Self-Esteem Scale (VASES), and the Rosenberg Self-Esteem Scale (RSES) | Correlation coefficient with Fisher r to z transformation, reported by the source paper. | **Stroke RSES/GDS correlation:** r = -.75<br>**Control RSES/GDS correlation:** r = -.51<br>**R to z comparison RSES/GDS:** Significant difference (more associated in stroke)<br><br>**Stroke VASES/GDS correlation:** r = -.77<br>**Control VASES/GDS correlation:** r = -.65<br>**R to z comparison RSES/GDS:** No significant difference p = .064 |
| Pickard et al. (2006) | **Category:** Item Response Theory (IRT) Differential Item | Fair | **Stroke group:** 32 depressed stroke inpatients, recruited from secondary sources. Patients were assessed approximately three months post-stroke. Health Utilities | **Stroke group:** Canadian inpatients | **Depression measure:** Centre for Epidemiologic Studies-Depression | DIF t-test statistic of latent item severity between groups, reported by the source paper. T-statistics | Broadly similar hierarchies in latent symptom severity of the items as indicators of depression (r = .75). Significant item misfit was found in the stroke group for the ‘unfriendly’, ‘crying’, and ‘restless’ items. |

| | | | | | | |
|---|---|---|---|---|---|---|
| | Functioning (DIF) analysis<br><br>**Design:** cross-sectional between-groups | Index Mark 3 mean score is .45, indicating 'severe disability' (Feng, 2009).<br><br>**Control group:** 366 primary care depressed adults living in the community. No evidence of demographic matching; substantial age differences reported | **Control group:** US residents living in the community | Scale (CES-D) | relate to a significance test of logit differences between groups. P values not provided, but items with p <.05 are highlighted.<br><br>Items with an infit Mean Squares (MNSQ) >1.4 were deemed to have poor fit to the Rasch model, as specified by the source paper authors. | **Higher latent symptom severity in stroke:** 'I felt disliked by others' (logit diff = .77), restlessness (logit diff = .61).<br><br>**Higher latent symptom severity in primary care:** Crying (logit diff = .48), appetite disruption (logit diff = .65)<br><br>**Items with no significant difference:** Unfriendly, failure, fearful, blues, effort, talked less, sad, as good as, concentration, depressed, get going, bothered, hopeful, happy, lonely, enjoy life. |

***Measures and Symptoms***

Symptom-level data were extracted from five depression measures: the PHQ-9, MADRS, BDI, PSDS, and PSE, resulting in 38 symptoms. Some symptoms were combined before dimension reduction because of overlaps in questionnaire wording: (1) depressed mood and sadness; and (2) fatigue, tiredness, and low energy. Clustering decisions were made by judgment of the reviewers and informed by evidence and theory. The results of the dimension reduction are summarised in Table 2.

The PSE items 'hypomania' and 'overactivity' were excluded from the selection because these are symptoms not typically included in diagnostic criteria of unipolar depression (Bell, 1994; UK National Collaborating Centre for Mental Health, 2010), and because they exhibited low prevalence in both groups in the study that used this measure (Lipsey et al., 1986).

**Table 2** A summary of the dimensions analysed and the constituent symptoms

| Dimension | Composite symptoms and measures | Rationale for clustering |
|---|---|---|
| Negative affect | PHQ-9 (down/depressed)<br>MADRS (observed and reported sadness, inner tension)<br>BDI (sadness)<br>PSDS (depressed mood)<br>PSE (simple depression, agitation, irritability, tension) | Negative affect is seen as a core symptom of depression (Bell, 1994). It is formulated separately to cognitions in Cognitive-Behavioural Therapy (CBT; Fenn and Byrne, 2013).<br><br>Factor analysis studies have generally found that these symptoms cluster together (Clara et al., 2001; González-Blanch et al., 2018; Steer et al., 1999; Storch et al., 2004). |
| Anhedonia and apathy | PHQ-9 (loss of interest in doing things)<br>MADRS (inability to feel)<br>BDI (lack of satisfaction, loss of interest in others)<br>PSDS (anhedonia, apathy/abulia/indifference)<br>PSE (affective flattening, loss of interest and concentration) | Emotional flatness is understood as a core symptom of depression (Bell, 1994).<br><br>Clara et al. (2001) found that anhedonia loaded onto a separate factor to negative affect.<br><br>Anhedonia and apathy appear correlated and often causally linked (Ang et al., 2017). Apathy was therefore added to this dimension |
| Negative cognitions | PHQ-9 (feeling bad about yourself)<br>MADRS (pessimistic thoughts)<br>BDI (guilt, pessimistic thoughts, sense of failure, self-hate, self-blame, punishment, body image)<br>PSDS (guilt feelings)<br>PSE (special features of depression, ideas of reference) | Negative cognitions are identified as a core component in CBT (Beck, 1979).<br><br>Negative cognitions have been found to form a latent factor in factor analytic studies of the BDI (Steer et al., 1999) |

| | | |
|---|---|---|
| Somatic features | PHQ-9 (sleep, tiredness, appetite, slowed down)<br>MADRS (sleep, reduced appetite, lassitude)<br>BDI (sleep, tiredness, appetite, weight, libido. somatic preoccupation)<br>PSDS (vegetative disorders)<br>PSE (other symptoms of depression, slowness, energy) | Somatic features of depression are documented in common depression criteria (Bell, 1994).<br><br>Somatic symptoms have consistently formed a latent factor across multiple depression measures and covary highly (Boothroyd et al., 2019; Cumming et al., 2010; González-Blanch et al., 2018) |
| Behavioural features of depression | BDI (work inhibition)<br>PSE (self-neglect) | Behavioural responses to emotional experiences are understood in CBT to be a primary factor in the maintenance of depressive symptoms and a moderating factor of outcome (Ludman et al., 2003; Moorey, 2010). |
| Cognitive features of depression | PHQ-9 (concentration)<br>MADRS (concentration)<br>BDI (indecisiveness) | Cognitive impairment is a commonly reported symptom of depression (Bell, 1994), and is associated with structural brain changes in neuroimaging studies (Marazziti et al., 2010)<br><br>Symptoms of cognitive impairment have been found to cluster as a latent factor (Adams et al., 2004) |
| Emotional dysregulation | PSDS (catastrophic reactions, hyper-emotionalism, diurnal variations) | Emotion dysregulation, defined here as significant and rapid changes to emotional state, are not typically included in criteria for depression diagnosis (Bell, 1994; UK National Collaborating Centre for Mental Health, 2010). However, the authors hypothesized that elevated emotional variation is part of post-stroke major depression experience (Gainotti et al., 1997) |
| Anxiety | BAI total score<br>PSDS (anxiety)<br>PSD (social unease, worrying) | Worrying, anxiety, and social phobia have been found to load onto a common factor on multiple measures (Lovibond and Lovibond, 1995; Smith et al., 2002) |
| Suicidal ideation | PHQ-9 (thoughts about harming yourself)<br>MADRS (suicidal thoughts)<br>BDI (suicidal thoughts/intent)<br>PSDS (suicidal thoughts/intent) | Maintained as a separate factor because of the importance of understanding differences in this experience as part of informing safe clinical practice (Simon et al., 2013) |

## Main Findings

For analysis of moderating variables, such as time since stroke, and their potential link to results, comparative correlation and profile comparison studies were combined. The results of these two methodologies were expected to broadly correspond because a trait with a higher degree of correlation with depression in one group might also be expected to have greater prevalence and severity in a depressed sample. The single DIF study (Pickard et al., 2006) was excluded from this analysis because differences in latent symptom severity were judged to be conceptually distinct from the association of symptoms with depression or their prevalence.

Analyses of phenomenological differences were, by contrast, completed separately for each methodology because, despite broad epistemological similarities, methodological differences might obscure more nuanced and detailed relationships. See Chapter 3 for more detailed explanations of these decisions.

### *Moderation of Study Characteristics*

Nineteen comparisons were extracted from the 11 profile comparisons and comparative correlation studies. The proportions of 'more' (severe, prevalent, or associated), 'less', and 'no difference' findings were stratified across each level of the predictor variables of interest, for example, for each quality rating category. These results are summarised in Figure 2.

***Methodology.*** Comparative correlation studies reported broadly consistent results to severity-based profile comparison studies. By contrast, prevalence-based profile studies found 'no differences' between groups more often and 'less' findings (i.e., less prevalent, severe, or associated in the stroke group) less often.

Comparative association studies contributed fewer results, providing only one symptom domain per study group comparison, versus an average of seven symptom domains for profile studies. Comparative association studies also reported a narrower spectrum of symptoms, with findings only extracted for anxiety, somatic features and negative cognitions.

***Study Quality.*** Higher rated studies were less likely to report significant differences in either direction. Amongst poorer quality studies, findings of 'less' increased substantially, from 10% to 46.4%. Poorer quality studies, therefore, potentially underestimated symptom severity, prevalence, or association with depression in the stroke groups. This could be explained by the fact that all four 'fair to poor' comparisons sampled psychiatric inpatients for their comparison group.

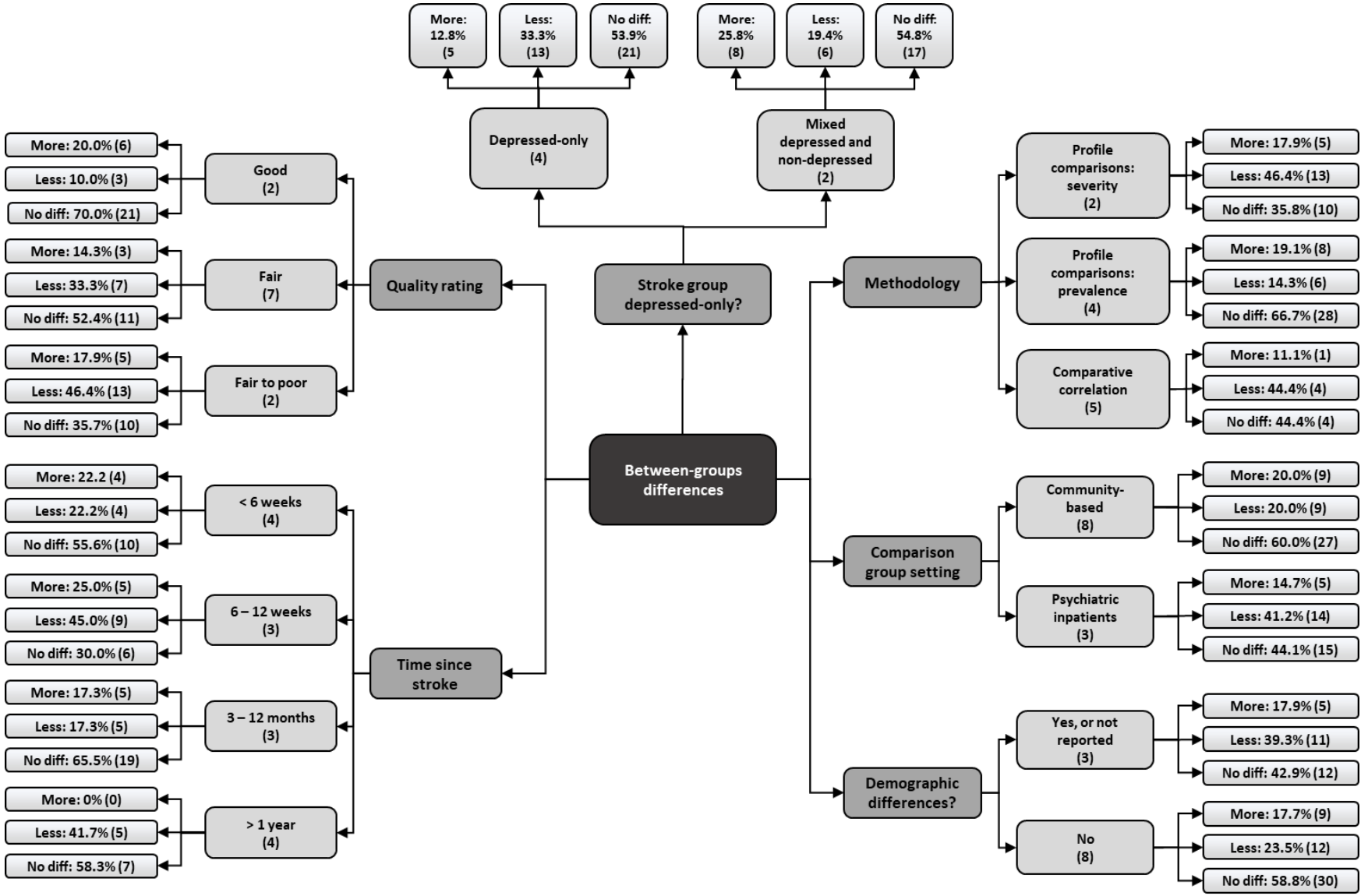
The possible effects of uncontrolled or non-reported demographic differences, a risk factor for bias, was explored. Like findings for overall quality, studies with uncontrolled demographic differences reported more 'less' findings, 39.3% versus 23.5%, but a similar proportion of 'more' findings, 17.9% versus 17.7%.

***Time Since Stroke.*** The proportion of 'more' findings progressively decreased with increased time since stroke. There was no discernible pattern for 'less' or 'no difference' findings.

***Comparison Group Setting.*** Studies sampling psychiatric inpatients in their non-stroke comparison groups were twice as likely to report findings of less prevalence, severity, or association with depression in the stroke group than studies with community-based samples (41.2% versus 20%). All comparisons of suicidal ideation in the inpatient samples were categorised as 'less', compared with 0% in the community-based samples. Studies featuring inpatient comparison groups may have reflected profile differences at the more severe end of post-stroke ***and*** non-stroke depression. The residential setting of the stroke group was not analysed because this had high correspondence with time since stroke.

**Depressed-only Versus Mixed Groups.** Four profile comparison studies featured groups of only depressed participants, and the remaining two had samples with a mixture of depressed and non-depressed participants (Cumming et al., 2010; de Man-van Ginkel et al., 2015; Gainotti et al., 1999; Lipsey et al., 1986). Depressed-only samples provide greater specificity at identifying depression symptomatologic differences, rather than differences in general population characteristics. The frequency of 'no difference' findings between groups was similar, but studies with depressed-only samples were 14% more likely to find 'less' results and reported 13% fewer 'more' results.

**Figure 2** Between-group differences in 'more', 'less', and 'no difference findings, according to each moderating factor, for profile comparisons and comparative associations.




*The number of studies at each level of comparison (e.g. 'good', 'fair', and 'fair to poor' for quality) is summarised in brackets. The number of comparisons for each 'more', 'less', and 'no difference' finding is also provided in brackets.*

***Profile Comparison Studies***

Inferences relating to phenomenological differences were drawn by determining the percentage of 'more', 'less', and 'no difference' findings for each of the nine domains (see Figure 3). Ten profile comparisons were extracted from the six profile comparison studies.

**Negative Affect.** There were mixed findings for negative affect, with 50% of comparisons yielding non-significant differences. Forty percent of comparisons found that negative affect was less severe in the stroke group, but these findings came from two papers that used psychiatric inpatients as a comparison group and compared severity, not prevalence (Gainotti et al., 1999; Gainotti et al., 1997). Therefore, it appeared that the prevalence of problems with negative affect was broadly similar between post-stroke and non-stroke depression groups.

**Anhedonia and Apathy.** Seventy percent of comparisons indicated less prevalence/severity of anhedonia and apathy in the stroke group, and no studies indicated greater prevalence or severity. The three 'no difference' findings were extracted from one of the two papers with mixed depressed and non-depressed groups. This indicated that, amongst those meeting criteria for depression, stroke participants experience less anhedonia and apathy.

**Somatic Features.** In most (80%) cases, no difference between groups was found. When exploring individual symptom differences, 'less' findings in the stroke group were generally due to less prevalent/severe sleep disruption and lost libido. Sleep disruption was less prevalent in stroke in three of the five comparisons that featured a sleep item, and no different in the remaining two. The single 'more' finding was due to a finding of more prevalent appetite disruption and somatic preoccupation in the 1-month post-stroke group (House et al., 1991).

**Negative Cognitions.** The groups did not significantly differ in the presence/severity of indicators of negative cognitions in most cases. Both 'less' findings were explained by a greater prevalence/severity of guilt-related cognitions in the comparison group. Pessimism was more prevalent in two of the three House et al. (1991) stroke groups, but overall negative cognition

prevalence/severity was balanced out by ‘no difference’ findings of other symptoms, and a ‘less’ finding for self-blame in one group.

**Cognitive Difficulties.** Cognitive complaints were more frequently reported in two of the five comparisons, and no different in the remaining three. This indicated a slight trend towards a greater prevalence of difficulties with concentration and indecisiveness post-stroke.

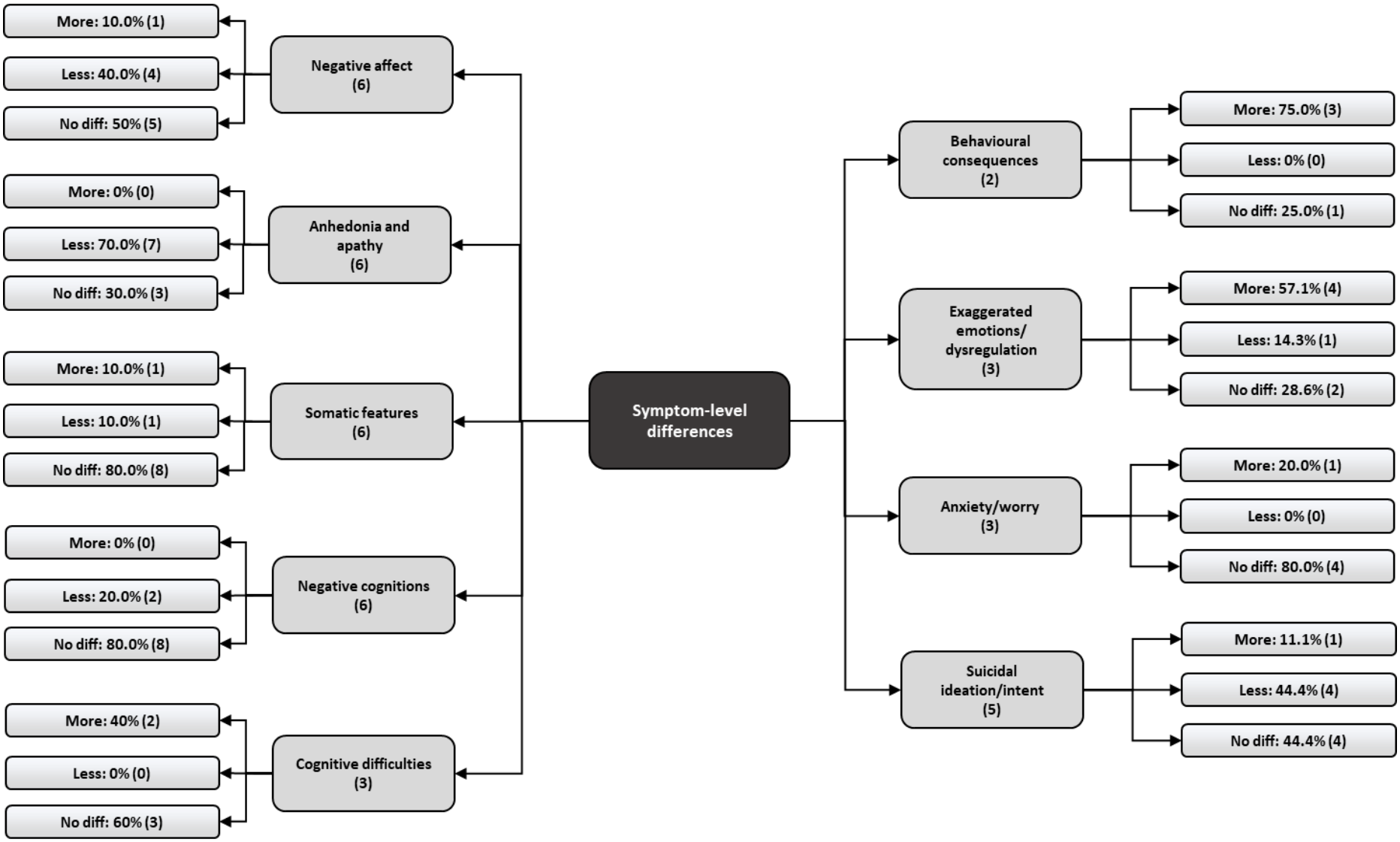
**Behavioural Features.** Three-quarters of comparisons of behavioural symptoms indicated that negative behavioural features of depression are more common after stroke than in comparison groups. All three ‘more’ findings were due to the ‘work inhibition’ item in the BDI (House et al., 1991). In the one study investigating self-neglect, no difference was observed (Lipsey et al., 1986).

**Exaggerated Emotions/Emotional Dysregulation.** Greater severity and prevalence of symptoms of emotional dysregulation were found in the stroke group in most comparisons. All ‘more’ findings were attributable to the PSDS measure (Gainotti et al., 1999; Gainotti et al., 1997). Surprisingly, House et al. (1991) found a lower prevalence of crying in the 1-month post-stroke group, compared to healthy controls, despite suspected loading of post-stroke emotionalism and adjustment processes onto this item.

**Suicidal Ideation/Intent.** Suicidal ideation was less prevalent in the stroke group in 44% of cases, and no difference was found in the same proportion. All four ‘less’ findings were attributable to studies that used psychiatric inpatients as comparison groups (Gainotti et al., 1999; Gainotti et al., 1997). Given that significant risk of suicide is often a criterion for referral to inpatient mental health care, the generalisability of these findings is limited. Excluding these findings, the data indicated no clear difference in the prevalence of suicide between populations.

**Anxiety/Worry.** Anxiety was found to be more severe in one comparison (Gainotti et al., 1997), but there was no evidence for a significant difference in severity or prevalence in the remaining four comparisons. It appeared, therefore, that experiences of comorbid worry/anxiety were broadly similar between groups, according to profile comparison methodology.

**Figure 3** The proportion of 'more', 'less', and 'no difference' findings (absolute number in brackets) for each symptom domain. The number of studies contributing data to each domain is also listed in bracket



### *Comparative Correlation Strength Studies*

Comparative correlation studies reported findings for insomnia (Fleming et al., 2021), fatigue (Stokes et al., 2011), self-esteem (Bennett et al., 2006; Vickery, Sepehri, et al., 2008), and anxiety (Schramke et al., 1998). A stronger degree of correlation was interpreted to indicate that a symptom is a greater predictor of depressive experience, and therefore more central to the phenomenology of depression in that population.

**Somatic Features.** The single study investigating comparative correlation strengths for insomnia found no evidence of a group effect, indicating no differences in association with depression (Fleming et al., 2021). This provided tentative evidence for similarities in the centrality of insomnia to depressive experience between healthy older adults and people > 1-year post-stroke.

A single study found a weaker association between depression and domain-general fatigue in the stroke group (Stokes et al., 2011), by comparing differences in the effect of a one-point increase in fatigue scores on depression. The stroke group were > 1-year post-stroke.

**Negative Cognitions (Self-esteem).** Two studies investigated comparative correlation strengths of self-esteem with depression (Bennett et al., 2006; Vickery, et al., 2008). Vickery et al. contributed two separate comparisons using two separate self-esteem measures, the RSES and the VASES. Vickery found a greater relationship of self-esteem with depression in the stroke group, using the RSES, but no difference in association when using the VASES. Bennett et al. reported no difference in correlation strength, also using the VASES. The findings, therefore, suggested measure dependence.

**Anxiety.** Schramke et al. (1998) contributed four comparisons for anxiety, based on two depression measures, the CES-D and the HDRS, and two stroke groups, a right-hemisphere and left-hemisphere stroke group. In three of the four comparisons, anxiety was less related to depression in the stroke groups than in the control group. There was a non-significant difference in association in the comparison featuring left-hemisphere stroke patients and the CES-D as a measure. Overall, it appeared that there was a trend towards a weaker association between anxiety and depression in

stroke, which indicated that anxiety featured less centrally in post-stroke depressive experience, > 1-year post-stroke, than non-stroke depressive experience.

### *Item Response Theory*

Pickard et al. (2006) used Rasch modelling, an IRT technique, to provide phenomenological insights into differences in post-stroke depressive experience. Pickard et al. found that feeling disliked by others, and feelings of restlessness were indicative of more severe depression in the stroke group, and the presence of crying and appetite disruption were indicative of more severe depression in the primary care group. No differences were found in the remaining CES-D items. Poor model fit was found for 'unfriendly', 'crying', and 'restless' items in the stroke group only, which suggested that these symptoms might have been less specific to depressive experience in this group.

## Discussion

This narrative review aimed to identify similarities and differences in depression phenomenology between stroke survivors and people in the general population. Three distinct methodologies, capable of contributing to this aim, were identified by this review: profile comparisons, comparisons of correlation strengths of depression symptoms with depression, and Differential Item Functioning (DIF). Broad similarities between groups were found, with specific areas of difference that require careful analysis to determine possible causes.

Each methodology contributed unique strengths; profile comparisons contributed rich information for multiple domains but are vulnerable to the effects of extraneous factors. For example, the finding of greater behavioural impairment in the stroke group was affected by the BDI 'work inhibition' item, which might be explained by factors other than depression. Comparisons of correlation strengths had the advantage of identifying the degree of 'closeness' of a symptom to depressive experience and were more robust to the confounds of profile comparisons, despite contributing less information per study. Finally, DIF offered powerful insights into differences in the relative severity of depression symptoms, a different perspective to that offered by the other methodologies.

**Which Characteristics Influence Symptomatology Findings?**

Study quality was found to be a significant influencing factor of symptomatologic differences, with lower-quality studies generally finding lower severity, prevalence, and symptom-depression association in the stroke group in these studies. This finding might be partially explained by the presence of uncontrolled demographic differences in lower-rated studies.

We found tentative evidence to support a trend of fewer 'more' findings with increased time since stroke. This trend may be explained by methodological factors, such as the absence of profile comparison studies in the > 1-year post-stroke range and the observation that comparative correlation strength studies found fewer 'more' results than prevalence-based studies. Stroke-related factors, such as elevated emotion during early adjustment (Taylor et al., 2011), stroke recovery (Wade et al., 1985), or recovery of post-stroke emotionalism (Morris et al., 1993), are also possible. The results of the single longitudinal study (House et al., 1991) suggest that somatic symptoms are more prevalent in early recovery and become less prevalent by one-year post-stroke, which might be indicative of recovery or adjustment-based processes. Unexpectedly, House et al. (1991) found that crying was *less* prevalent in their early recovery group, contrary to theories of the natural course of emotionalism (Fitzgerald, 2021; Morris et al., 1993).

The residential setting of the comparison group appeared relevant; studies investigating psychiatric inpatients in their control groups generally reported reduced symptom severity of negative affect and suicidal ideation in stroke. This difference was somewhat expected, given that inpatient admissions are commonly related to managing clinical risk (Hunt et al., 2012).

**Is the Phenomenology of PSD Different?**

Clear interpretation of patterns was obscured by the high degree of heterogeneity between studies but apparent relationships are summarised here. Profile comparison studies indicated broad similarities between depressed stroke and non-stroke participants in the prevalence/severity of somatic symptoms, negative cognitions, and anxiety. Findings of significantly less severe suicidal ideation and negative affect were reported in a sizeable minority of cases (40% and 44.4%, respectively). When accounting for the observation that these findings all used psychiatric inpatients as comparison groups and that

some had unmatched demographics, the severity and prevalence of negative affect and suicidal ideation appeared to be generally equal between groups.

The absence of consistent differences in prevalence or severity of somatic items between groups is intriguing, given the often-assumed interference of physical health consequences, such as post-stroke fatigue and physical disability, on the reliable measurement of the somatic features of depression (Cumming et al., 2010). This finding supports previous evidence that depression contributes unique variance to these items in both groups (de Man-van Ginkel et al., 2015; Robinson, 2006). One possible explanation for this observed similarity is that somatic depression items are less biased by physical disability than often assumed, and instead primarily capture depression variance. Another possibility is that there is a weaker correlation between depression and somatic problems in stroke, but that this reduced correlation is offset by the presence of elevated baseline somatic symptoms, leading to a cancelling out between groups. Cumming et al., (2010) suggest that stroke patients may, on average, experience less sleep impairment because their fatigue leads to improved sleep, which would also lead to an offsetting effect of overall somatic impairment and post-stroke insomnia (Tang et al., 2015).

A clearer picture emerged for anhedonia and apathy, with 70% of profile comparisons indicating lower severity and prevalence in PSD groups. This finding is surprising, given the evidence in support of apathy as a stroke sequela (Jorge et al., 2010). By contrast, the stroke groups presented with greater severity of problems with emotional dysregulation in 57% of comparisons. Combined, these findings suggest that depression in the general population is more associated with dulled affect and low motivation, and PSD is associated more with greater emotional dysregulation. This picture is, of course, complicated by the distinct phenomena of post-stroke emotionalism (Calvert et al., 1998; Fitzgerald, 2021) and processes of emotional adjustment to loss (Taylor et al., 2011). It is possible that the presence of elevated emotions from adjustment and emotionalism negatively load onto items of anhedonia because strong or changeable emotions might counteract the perception or experience of emotional flatness. Alternatively, it could be that the focus on physical recovery and return to ‘normal

life' after stroke protects against loss of interest or reduced sense of accomplishment (Townend, 2005).

The findings of more prevalent cognitive complaints in the stroke groups in two out of five studies might be confounded by the loading of neuropsychological cognitive deficits post-stroke (Vataja et al., 2003). Care must be taken when interpreting causality, given the observed association between depression and executive dysfunction post-stroke (Jaywant et al., 2022). Studies that compared depressed and non-depressed stroke patients have evinced an overlay of depression onto cognitive items, such as impairment to concentration, but the interaction between these two sources of impairment requires future investigation (de Man-van Ginkel et al., 2015).

Greater prevalence of self-reported negative behavioural features was reported in the stroke group in 75% of comparisons. In all three cases where a 'more' finding was reported, it was due to the 'work inhibition' item, which may have been confounded by post-stroke cognitive and physical impairment, rather than directly because of depression.

Reliable inferences were difficult to draw from correlation comparison studies because of the relatively scarce number of comparisons contributed by this methodology. Findings were mixed for the relative association of self-esteem with depression between groups, meaning it is unclear whether self-esteem can be understood as more intertwined with depressive experience after a stroke. No difference in the association between sleep disruption and depression was found, but this finding should not be overly generalized because this result is derived from only one regression-based study. We found a preliminary indication of a reduced connection between fatigue and depression in the stroke group, but further evidence is required before potential causes are speculated.

The findings of Schramke et al. (1998) suggest that, across left hemisphere and right hemisphere stroke populations, anxiety is generally less associated with depression in stroke. The sample sizes used in this study were small, however. If this finding were to be replicated, anxiety would be less *associated* with depression in stroke but no more or less prevalent or severe between groups, as indicated by profile comparison studies. A possible explanation for this contrast in findings

might be that depression and anxiety may be more likely to form a maintenance cycle, such as the mechanism proposed by Fennell (2005), for people in the wider population, but that the sources of anxiety and depression might be more independent in stroke.

Despite the presence of only one DIF study, the preliminary findings of Pickard et al. (2006) are intriguing. The observed lower latent severity of crying in the stroke group might be explained by a greater likelihood of reported crying, even in milder cases of low mood, because of the common presence of post-stroke emotionalism (Calvert et al., 1998). It is unclear why feeling disliked was indicative of more severe depression in the stroke group. Misfit of the unfriendliness item might be explained by interference from experiences of hospital care (Kitson et al., 2013), crying from post-stroke emotionalism (Calvert et al., 1998), and restlessness from cognitive difficulties and being in hospital (Douiri et al., 2013). However, the sample size was small for IRT modelling (Jiang et al., 2016), and there appeared to be no control of demographic differences between groups, meaning that these findings must be interpreted with caution.

**Clinical Implications**

The findings of this review suggest that depression phenomenology is similar in stroke and the general population, except for possible reduced severity and prevalence of anhedonia and/or apathy. Based on these findings, we recommend an augmented approach to therapeutic support for depression, with consideration of additional stroke-specific factors and mechanisms, such as the influence of adjustment and emotionalism (Broomfield et al., 2011). For example, in their seminal review outlining the need for an augmented approach to CBT in stroke, Broomfield et al. (2011) recommended consideration of additional grief work and motivational interviewing in the context of personal loss and psychological adjustment. Our findings of increased difficulty with high self-reported emotion dysregulation after stroke support the rationale for this approach. Similarly, findings of specific differences in work dysfunction might be of significant psychological relevance, if withdrawal from work was previously a source of value and/or coping (Laidlaw et al., 2008). It may, therefore, be beneficial to augment CBT support for depressed stroke patients, where such processes exist, with Selective Optimisation with Compensation, which aims to support the selection of value-

driven goals that consider the reduced resources associated with stroke-related disability (Baltes, 1997; Broomfield et al., 2011; Grove et al., 2009).

Furthermore, we recommend that a curious approach is taken to somatic and cognitive symptoms, given that the interaction of the physical and cognitive consequences of stroke with depression remains unclear. Similarly, there should be due consideration of the impact of cognitive impairment post-stroke, and its possible role in the maintenance of PSD (Broomfield et al., 2011).

This review also touches on the ongoing debate about whether emotionalism after stroke should be considered part of PSD. Gainotti et al. (1997), who designed the PSDS, hypothesised that emotionalism and adjustment were significant components of 'major' and 'minor' PSD diagnoses. However, this contrasts with evidence for post-stroke emotionalism as a distinct and separate clinical phenomenon to PSD with specific neurological correlates (Andersen et al., 1995; Fitzgerald, 2021), and evidence for 'catastrophic' emotional reactions as an expected process in adjusting to potentially long-term disability (Taylor et al., 2011). Despite the observed relationship between emotionalism and depression (Andersen et al., 1995) and reported narratives of distress associated with emotion dysregulation after stroke (McAleese et al., 2021), emotional dysregulation is, at best, a non-specific indicator of PSD and clinicians should interpret any presentation involving strong or changeable emotions with cautious clinical curiosity.

**Strengths and Limitations of this Review**

A strength of this review is that it is the first paper to synthesise multiple distinct methodologies to identify phenomenological differences in PSD, while considering the myriad of extraneous factors that might load onto commonly used indicators of depression experience, such as post-stroke emotionalism (Calvert et al., 1998). Another strength was the openness of our search strategy, which enabled the identification of many relevant methodological approaches. Previous reviews often focus only on profile comparison studies (e.g. Llorca et al., 2015).

Despite these strengths, several limitations should be highlighted. First, we acknowledge that our specific strategy for coding symptoms, such as during symptom reduction and effect direction

scoring, was one of many possible alternatives. Second, though judgments were agreed across the review team, we must acknowledge that interpretation of symptom similarities and differences and the significance of moderating factors were based on qualitative judgments, and alternative conclusions from the same findings are possible. Furthermore, confidence in our conclusions was evidently confounded by the quality of the studies reviewed, as well as the significant heterogeneity in design and methodology. We included studies rated 'fair to poor' for risk of bias, several studies did not control for demographics, source paper determinants of significance varied substantially, and large between-study variation in factors such as patient setting, time since stroke, and specificity of depressed status, obfuscated possible inferences. The observational, not experimental, nature of the included articles means that any cause of differences in symptoms, such as fatigue or concentration problems, is likely to be a multi-directional consequence of biological, psychological, and social factors, not a linear pathway of depression to symptoms. These abovementioned factors mean that confidence in our conclusions is compromised and therefore interpreted cautiously.

An additional limitation stems from an acknowledgement of the social construction of depression and its developmental context within western medicine. Those adopting constructionist perspectives might argue that there are limitations to comparing broad symptoms between groups if the fundamental way in which these symptoms are understood and constructed is different between populations. For example, self-blame in the context of requiring physical health care from family members may be seen by some as fundamentally incomparable to self-blame or guilt experienced in the context of adult mental health and complex attachment histories. Furthermore, we endeavoured in this review to compare groups with 'like for like' depression severity, but if there is loading from extraneous non-depression-related factors, such as post-stroke emotionalism, this might also mean that the underlying depression severity is not similar between groups, limiting comparability. If the stroke group were, on average, less depressed than observed from the data, because of such interference, then this could explain the findings of less prevalent/severe anhedonia/apathy.

**Future research**

As outlined above, the findings of this review have generated new research questions about the nature of some of the observed between-groups differences and similarities: are findings of similar somatic profiles related to the robustness of these items to extraneous sources of variance, or because of other mechanisms? How might findings of less association of anxiety with depression in stroke be integrated with broad similarities in severity/prevalence of anxiety between groups?

As demonstrated by the presence of only one such study (Pickard et al., 2006), the use of IRT and DIF methodology as a means of deriving phenomenological insights into depression has been under-utilised. Clinicians could use such insights to identify whether the presence of certain symptoms should be interpreted as more concerning than others and if this varies between populations.

Finally, the difficulties associated with determining 'like for like' depression, robust to the loading of distinct phenomena onto items, could be overcome in future research using Structural Equation Modelling (SEM) techniques. For example, Confirmatory Factor Analysis could be used to derive depression factor scores, which would be more robust to loading caused by extraneous sources and observed as differences in intercepts (Kim and Yoon, 2011; Meredith and Teresi, 2006). Comparisons of factor loadings could also be used as an alternative to comparing association strengths of a symptom with depression.

**Conclusions**

Here, we have presented the first synthesis of phenomenological comparisons of depression between stroke and the general population. We identified three unique methodologies that can contribute to this research question. This indicates that phenomenological comparisons cannot be understood from comparisons of profiles alone and that we must consider differences in symptom prevalence, severity, 'closeness' to the construct of depression, and differences in the latent severity of symptoms as indicators of depressive experience. Broad similarities were found, but a more detailed understanding of observed differences, and of mechanisms that help to integrate findings between each methodology, requires future research.

**Declaration of interest statement.** The authors report no conflict of interest.

## References


Acciarresi, M., Bogousslavsky, J., & Paciaroni, M. (2014). Post-Stroke Fatigue: Epidemiology, Clinical Characteristics and Treatment. *European Neurology, 72(5–6),* 255–261. https://doi.org/10.1159/000363763

Adams, K. B., Matto, H. C., & Sanders, S. (2004). Confirmatory factor analysis of the geriatric depression scale. *Gerontologist, 44(6),* 818–826. https://doi.org/10.1093/geront/44.6.818

Aizenstein, H. J., Baskys, A., Boldrini, M., Butters, M. A., Diniz, B. S., Jaiswal, M. K., Jellinger, K. A., Kruglov, L. S., Meshandin, I. A., Mijajlovic, M. D., Niklewski, G., Pospos, S., Raju, K., Richter, K., Steffens, D. C., Taylor, W. D., & Tene, O. (2016). Vascular depression consensus report – a critical update. *BMC Medicine 2016, 14(1),* 1–16. https://doi.org/10.1186/S12916-016-0720-5

Aizenstein, H. J., Khalaf, A., Walker, S. E., & Andreescu, C. (2014). Magnetic resonance imaging predictors of treatment response in late-life depression. *Journal of Geriatric Psychiatry and Neurology, 27(1),* 24–32. https://doi.org/10.1177/0891988713516541

Alexopoulos, G. S., Meyers, B. S., Young, R. C., Campbell, S., Silbersweig, D., & Charlson, M. (1997). "Vascular depression" hypothesis. In *Archives of General Psychiatry, 54(10),* pp. 915–922). https://doi.org/10.1001/archpsyc.1997.01830220033006

Andersen, G., Vestergaard, K., & Ingeman-Nielsen, M. (1995). Post-stroke pathological crying: frequency and correlation to depression. *European Journal of Neurology, 2(1), 45–50.* https://doi.org/10.1111/J.1468-1331.1995.TB00092.X

Ang, Y. S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE, 12(1).* https://doi.org/10.1371/JOURNAL.PONE.0169938

Baltes, P. B. (1997). On the Incomplete Architecture of Human Ontogeny: Selection, Optimization, and Compensation as Foundation of Developmental Theory. *American Psychologist, 52(4),* 366–380. https://doi.org/10.1037/0003-066X.52.4.366

Beck, A. T. (1979). *Cognitive Therapy of Depression.* Guildford Press, New York. https://books.google.com.sg/books?id=L09cRS0xWj0C&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An Inventory for Measuring Clinical Anxiety: Psychometric Properties. *Journal of Consulting and Clinical Psychology, 56(6),* 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Bell, C. C. (1994). DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. JAMA: *The Journal of the American Medical Association, 272(10), 828.* https://doi.org/10.1001/jama.1994.03520100096046

Bennett, H. E., Thomas, S. A., Austen, R., Morris, A. M. S., & Lincoln, N. B. (2006). Validation of screening measures for assessing mood in stroke patients. *British Journal of Clinical Psychology, 45(3),* 367–376. https://doi.org/10.1348/014466505X58277

Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research, 271,* 532–534. https://doi.org/10.1016/j.psychres.2018.12.048

Broomfield, N. M., Laidlaw, K., Hickabottom, E., Murray, M. F., Pendrey, R., Whittick, J. E., & Gillespie, D. C. (2011). Post-stroke depression: the case for augmented, individually tailored cognitive behavioural therapy. *Clinical Psychology & Psychotherapy, 18(3),* 202–217. https://doi.org/10.1002/cpp.711

Brumfitt, S. M., & Sheeran, P. (1999). The development and validation of the visual analogue self-esteem scale (VASES). *British Journal of Clinical Psychology, 38(4),* 387–400. https://doi.org/10.1348/014466599162980

Burton, L-J. & Tyson, S. (2015). Screening for mood disorders after stroke: A systematic review of psychometric properties and clinical utility. *Psychological Medicine, 45(1),* 29–49. https://doi.org/10.1017/S0033291714000336

Calvert, T., Knapp, P., & House, A. (1998). Psychological associations with emotionalism after stroke. *Journal of Neurology Neurosurgery and Psychiatry, 65(6),* 928–929. https://doi.org/10.1136/jnnp.65.6.928

Christensen, D., Johnsen, S. P., Watt, T., Harder, I., Kirkevold, M., & Andersen, G. (2008). Dimensions of post-stroke fatigue: a two-year follow-up study. *Cerebrovascular Diseases, 26(2),* 134–141. https://doi.org/10.1159/000139660

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory Factor Analysis of the Depression-Anxiety-Stress Scales in Depressed and Anxious Patients. *Journal of Psychopathology and Behavioral Assessment, 23(1),* 61–67. https://doi.org/10.1023/A:1011095624717

Cooper, J. E. (1985). The Development of the Present State Examination (P.S.E.). *Clinical Psychopathology Nomenclature and Classification, 133–139.* https://doi.org/10.1007/978-1-4899-5049-9_20

Crowe, C., Coen, R. F., Kidd, N., Hevey, D., Cooney, J., & Harbison, J. (2016). A qualitative study of the experience of psychological distress post-stroke. *Journal of Health Psychology, 21(11),* 2572–2579. https://doi.org/10.1177/1359105315581067

Cumming, T. B., Churilov, L., Skoog, I., Blomstrand, C., & Linden, T. (2010). Little evidence for different phenomenology in poststroke depression. *Acta Psychiatrica Scandinavica, 121(6),* 424–430. https://doi.org/10.1111/j.1600-0447.2010.01558.x

de Man-van Ginkel, J. M., Hafsteinsdóttir, T. B., Lindeman, E., Geerlings, M. I., Grobbee, D. E., & Schuurmans, M. J. (2015). Clinical manifestation of depression after stroke: Is it different from depression in other patient populations? *PLoS ONE, 10(12).* https://doi.org/10.1371/journal.pone.0144450

Deng, L., Sun, X., Qiu, S., Xiong, Y., Li, Y., Wang, L., Wei, Q., Wang, D., & Liu, M. (2017). Interventions for management of post-stroke depression: A Bayesian network meta-analysis of 23 randomized controlled trials. *Scientific Reports 2017, 7(1), 1–12.* https://doi.org/10.1038/s41598-017-16663-0

Desmond, D. W., Remien, R. H., Moroney, J. T., Stern, Y., Sano, M., & Williams, J. B. W. (2003). Ischemic stroke and depression. *Journal of the International Neuropsychological Society: JINS, 9(3),* 429–439. https://doi.org/10.1017/S1355617703930086

Douiri, A., Rudd, A. G., & Wolfe, C. D. A. (2013). Prevalence of poststroke cognitive impairment: South London stroke register 1995-2010. *Stroke, 44(1),* 138–145. https://doi.org/10.1161/STROKEAHA.112.670844

Dowswell, G., Lawler, J., Dowswell, T., Young, J., Forster, A., & Hearn, J. (2000). Investigating recovery from stroke: A qualitative study. *Journal of Clinical Nursing, 9(4),* 507-515. https://doi.org/10.1046/j.1365-2702.2000.00411.x

Duncan, P. W. (1994). Stroke Disability. *Physical Therapy, 74(5),* 399–407. https://doi.org/10.1093/PTJ/74.5.399

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale Version 20: Evaluation of reliability, validity, and sensitivity to change. *Stroke, 30(10),* 2131–2140. https://doi.org/10.1161/01.STR.30.10.2131

Ellis, C. J. (2014). Measurement issues in poststroke depression. *Journal of Neuroscience Nursing, 46(4),* 198. https://doi.org/10.1097/JNN.0000000000000070

Espárrago Llorca, G., Castilla-Guerra, L., Fernández Moreno, M. C., Ruiz Doblado, S., & Jiménez Hernández, M. D. (2015). Post-stroke depression: an update. *Neurologia, 30(1),* 23–31. https://doi.org/10.1016/j.nrl.2012.06.008

Espie, C. A., Kyle, S. D., Hames, P., Gardani, M., Fleming, L., & Cape, J. (2014). The Sleep Condition Indicator: a clinical screening tool to evaluate insomnia disorder. *BMJ Open, 45(1),* 29-49. https://doi.org/10.1136/bmjopen-2013

Fawcett, J. (1997). The detection and consequences of anxiety in clinical depression. *Journal of Clinical Psychiatry, 58(8),* 35–40. https://www.psychiatrist.com/read-pdf/4287/

Feng, Y., Bernier, J., McIntosh, C., & Orpana, H. (2009). Validation of disability categories derived from Health Utilities Index Mark 3 scores and Heather Orpana. *Health Reports, 20(2),* 43-50.

Fenn, M. K., & Byrne, D. M. (2013). The key principles of cognitive behavioural therapy. *InnovAiT: Education and inspiration for general practice, 6(9),* 579-585. Http://Dx.Doi.Org/10.1177/1755738012471029

Fitzgerald, S. (2021). Post-stroke emotionalism (PSE): Making sense of neurological, psychological and experiential factors. Doctoral thesis, University of East Anglia.

Fennell, M. J. V. (1997). Low self-esteem: A cognitive perspective. Behavioural and Cognitive Psychotherapy, 25(1), 1–25. https://doi.org/10.1017/S1352465800015368

Fleming, M. K., Smejka, T., Henderson Slater, D., Chiu, E. G., Demeyere, N., & Johansen-Berg, H. (2021). Self-Reported and Objective Sleep Measures in Stroke Survivors With Incomplete Motor Recovery at the Chronic Stage. *Neurorehabilitation and Neural Repair, 35(10),* 851-860. https://doi.org/10.1177/15459683211029889

Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the Beck anxiety inventory. *Journal of Anxiety Disorders, 6(1),* 55–61. https://doi.org/10.1016/0887-6185(92)90026-4

Gainotti, G., Azzoni, A., & Marra, C. (1999). Frequency, phenomenology and anatomical-clinical correlates of major post-stroke depression. *British Journal of Psychiatry, 175,* 163–167. https://doi.org/10.1192/bjp.175.2.163

Gainotti, G, Azzoni, A., Razzano, C., Lanzillotta, M., Marra, C., & Gasparini, F. (1997). The Post-Stroke Depression Rating Scale: a test specifically devised to investigate affective disorders of stroke patients. *Journal of Clinical and Experimental Neuropsychology, 19(3),* 340–356. https://doi.org/10.1080/01688639708403863

González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., & Cano-Vindel, A. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE, 13(2).* https://doi.org/10.1371/journal.pone.0193356

Grove, L. J., Loeb, S. J., & Penrod, J. (2009). Selective optimization with compensation: A model for elder health programming. *Clinical Nurse Specialist, 23(1),* 25–32. https://doi.org/10.1097/01.NUR.0000343080.57838.2F

Hackett, M. L., & Anderson, C. S. (2005). Predictors of Depression after Stroke. *Stroke, 36(10),* 2296–2301. https://doi.org/10.1161/01.STR.0000183622.75135.A4

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23(1),* 56–62. https://doi.org/10.1136/jnnp.23.1.56

House, A., Dennis, M., Mogridge, L., Warlow, C., Hawton, K., & Jones, L. (1991). Mood disorders in the year after first stroke. *The British Journal of Psychiatry, 158,* 83–92. https://doi.org/10.1192/bjp.158.1.83

Hu, X., Silva, T. M. De, Chen, J., & Faraci, F. M. (2017). Cerebral Vascular Disease and Neurovascular Injury in Ischemic Stroke. *Circulation Research, 120(3),* 449–471. https://doi.org/10.1161/CIRCRESAHA.116.308427

Hunt, G. E., O'Hara-Aarons, M., O'Connor, N., & Cleary, M. (2012). Why are some patients admitted to psychiatric hospital while others are not? A study assessing risk during the admission interview and relationship to outcome. International Journal of Mental Health Nursing, 21(2), 145–153. https://doi.org/10.1111/J.1447-0349.2011.00778.X

Jaywant, A., DelPonte, L., Kanellopoulos, D., O'Dell, M. W., & Gunning, F. M. (2022). The Structural and Functional Neuroanatomy of Post-Stroke Depression and Executive Dysfunction: A Review of Neuroimaging Findings and Implications for Treatment. *Journal of Geriatric Psychiatry and Neurology, 35(1),* 3–11. https://doi.org/10.1177/0891988720968270

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. *Frontiers in Psychology, 7,* 109. https://doi.org/10.3389/fpsyg.2016.00109

Jorge, R. E., Starkstein, S. E., & Robinson, R. G. (2010). Apathy following stroke. *The Canadian Journal of Psychiatry / La Revue Canadienne de Psychiatrie, 55(6),* 350–354. https://doi.org/10.1177/070674371005500603

Kim, E. S., & Yoon, M. (2011). *Testing Measurement Invariance: A Comparison of Multiple-Group Categorical CFA and IRT, 18(2),* 212–228. https://doi.org/10.1080/10705511.2011.557337

Kitson, A. L., Dow, C., Calabrese, J. D., Locock, L., & Athlin, Å. M. (2013). Stroke survivors' experiences of the fundamentals of care: A qualitative analysis. *International Journal of Nursing Studies, 50(3),* 392–403. https://doi.org/10.1016/J.IJNURSTU.2012.09.017

Kotila, M., Numminen, H., Waltimo, O., & Kaste, M. (1999). Post-stroke depression and functional recovery in a population-based stroke register. The Finnstroke study. *European Journal of Neurology, 6(3),* 309–312. https://doi.org/10.1046/j.1468-1331.1999.630309.x

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16(9),* 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Laidlaw, K., Thompson, L. W., Dick-Siskin, L., & Gallagher-Thompson, D. (2008). *Cognitive behaviour therapy with older people.* John Wiley & Sons. https://doi.org/10.1002/9780470713402

Lees, R., Stott, D. J., Quinn, T. J., & Broomfield, N. M. (2014). Feasibility and diagnostic accuracy of early mood screening to diagnose persisting clinical depression/anxiety disorder after stroke. *Cerebrovascular Diseases, 37(5),* 323–329. https://doi.org/10.1159/000360755

Lindén, T., Blomstrand, C., & Skoog, I. (2007). Depressive disorders after 20 months in elderly stroke patients: A case-control study. *Stroke, 38(6),* 1860–1863. https://doi.org/10.1161/STROKEAHA.106.471805

Lipsey, J. R., Spencer, W. C., Rabins, P. V., & Robinson, R. G. (1986). Phenomenological comparison of poststroke depression and functional depression. *American Journal of Psychiatry, 143(4),* 527–529. https://doi.org/10.1176/AJP.143.4.527

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33(3),* 335–343. https://doi.org/10.1016/0005-7967(94)00075-U

Ludman, E., Katon, W., Bush, T., Rutter, C., Lin, E., Simon, G., von Korff, M., & Walker, E. (2003). Behavioural factors associated with symptom outcomes in a primary care-based depression prevention intervention trial. *Psychological Medicine, 33(6),* 1061–1070. https://doi.org/10.1017/S003329170300816X

Ma, L. L., Wang, Y. Y., Yang, Z. H., Huang, D., Weng, H., & Zeng, X. T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: What are they and which is better? *Military Medical Research, 7(1),* 1–11. https://doi.org/10.1186/S40779-020-00238-8/TABLES/1

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology, 626(1),* 83–86. https://doi.org/10.1016/J.EJPHAR.2009.08.046

McAleese, N., Guzman, A., O'Rourke, S. J., & Gillespie, D. C. (2021). Post-stroke emotionalism: A qualitative investigation. *Disability and Rehabilitation: An International, Multidisciplinary Journal, 43(2),* 192–200. https://doi.org/10.1080/09638288.2019.1620876

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44(11 Suppl 3),* 69-77. https://doi.org/10.1097/01.MLR.0000245438.73837.89

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research, 14(1),* 1–10. https://doi.org/10.1186/S12913-014-0579-0/TABLES/7

Mitchell, A. J., Sheth, B., Gill, J., Yadegarfar, M., Stubbs, B., Yadegarfar, M., & Meader, N. (2017). Prevalence and predictors of post-stroke mood disorders: A meta-analysis and meta-regression of depression, anxiety and adjustment disorder. *General Hospital Psychiatry, 47,* 48–60. https://doi.org/10.1016/j.genhosppsych.2017.04.001

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry: The Journal of Mental Science, 134(4),* 382–389. https://doi.org/10.1192/BJP.134.4.382

Moorey, S. (2010). The Six Cycles Maintenance Model: Growing a "Vicious Flower" for Depression. *Behavioural and Cognitive Psychotherapy, 38(2),* 173–184. https://doi.org/10.1017/S1352465809990580

Morris, P. L. P., Robinson, R. G., & Raphael, B. (1993). Emotional Lability after Stroke. *Australian and New Zealand Journal of Psychiatry, 27(4),* 601–605. https://doi.org/10.3109/00048679309075822

National Heart Lung and Blood Institute (NHLBI). (2013). Study Quality Assessment Tools. https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Newberg, A. R., Davydow, D. S., & Lee, H. B. (2009). Cerebrovascular disease basis of depression: Post-stroke depression and vascular depression. *International Review of Psychiatry, 18(5),* 433–441. https://doi.org/10.1080/09540260600935447

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews, 5(1).* https://doi.org/10.1186/S13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ,* 372. https://doi.org/10.1136/BMJ.N71

Palagini, L., Ragno, G., Caccavale, L., Gronchi, A., Terzaghi, M., Mauri, M., Kyle, S., Espie, C. A., & Manni, R. (2015). Italian validation of the Sleep Condition Indicator: A clinical screening tool to evaluate Insomnia Disorder according to DSM-5 criteria. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology, 98(3 Pt 1),* 435–440. https://doi.org/10.1016/J.IJPSYCHO.2015.08.008

Pan, A., Keum, N., Okereke, O. I., Sun, Q., Kivimaki, M., Rubin, R. R., & Hu, F. B. (2012). Bidirectional association between depression and metabolic syndrome: A systematic review and meta-analysis of epidemiological studies. *Diabetes Care, 35(5),* 1171–1180. https://doi.org/10.2337/dc11-2055

Pickard, A. S., Dalal, M. R., & Bushnell, D. M. (2006). A Comparison of Depressive Symptoms in Stroke and Primary Care: Applying Rasch Models to Evaluate the Center for Epidemiologic Studies-Depression Scale. *Value in Health, 9(1),* 59–64. https://doi.org/10.1111/j.1524-4733.2006.00082.x

Pollock, A., & Berge, E. (2018). How to do a systematic review. *International Journal of Stroke, 13(2),* 138–156. https://doi.org/10.1177/1747493017743796

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews.* A

Product from the ESRC Methods Programme Peninsula Medical School, Universities of Exeter and Plymouth.

Protogerou, C., & Hagger, M. S. (2018). A Case For a Study Quality Appraisal in Survey Studies in Psychology. *Frontiers in Psychology, 9.* https://doi.org/10.3389/FPSYG.2018.02788

Radloff, L. S. (2016). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement, 1(3),* 385–401. https://doi.org/10.1177/014662167700100306

Reiswig, J. (2010). Mendeley. *Journal of the Medical Library Association: JMLA, 98(2),* 193. https://doi.org/10.3163/1536-5050.98.2.021

Robinson, R. G. (2006). *The clinical neuropsychiatry of stroke: Second edition.* Cambridge University Press. https://doi.org/10.1017/CBO9780511544231

Robinson, R. G., & Jorge, R. E. (2016). Post-stroke depression: A review. *The American Journal of Psychiatry, 173(3),* 221–231. https://doi.org/10.1176/appi.ajp.2015.15030363

Rosenberg, M. (1979). Conceiving the Self. New York: Basic Books.

Schöttke, H., & Giabbiconi, C. M. (2015). Post-stroke depression and post-stroke anxiety: Prevalence and predictors. *International Psychogeriatrics, 27(11),* 1805–1812. https://doi.org/10.1017/S1041610215000988

Schramke, C. J., Stowe, R. M., Ratcliff, G., Goldstein, G., & Condray, R. (1998). Poststroke depression and anxiety: Different assessment methods result in variations in incidence and severity estimates. *Journal of Clinical and Experimental Neuropsychology, 20(5),* 723–737. https://doi.org/10.1076/jcen.20.5.723.1117

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric depression scale (GDS) recent evidence and development of a shorter version. *Clinical Gerontologist, 5(1–2),* 165–173. https://doi.org/10.1300/J018v05n01_09

Shi, Y., Yang, D., Zeng, Y., & Wu, W. (2017). Risk factors for post-stroke depression: A meta-analysis. *Frontiers in Aging Neuroscience, 9,* 218. https://doi.org/10.3389/fnagi.2017.00218

Simon, G. E., Rutter, C. M., Peterson, D., Oliver, M., Whiteside, U., Operskalski, B., & Ludman, E. J. (2013). Does response on the PHQ-9 depression questionnaire predict subsequent suicide

attempt or suicide death? *Psychiatric Services, 64(12),* 1195–1202. https://doi.org/10.1176/APPI.PS.201200587/ASSET/IMAGES/LARGE/1195F2.JPEG

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research, 39(3),* 315–325. https://doi.org/10.1016/0022-3999(94)00125-O

Smith, A. B., Selby, P. J., Velikova, G., Stark, D., Wright, E. P., Gould, A., & Cull, A. (2002). Factor analysis of the Hospital Anxiety and Depression Scale from a large cancer population. *Psychology and Psychotherapy, 75(2),* 165–176. https://doi.org/10.1348/147608302169625

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression Inventory-II in Clinically Depressed Outpatients. *Journal Of Clinical Psychology, 55(1),* 117-128.

Stokes, E. K., O'Connell, C., & Murphy, B. (2011). *An investigation into fatigue post-stroke and its multidimensional nature. Advances in Physiotherapy, 13(1),* 2–10. https://doi.org/10.3109/14038196.2010.534175

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the beck depression inventory—second edition in a sample of college students. Depression and Anxiety, 19(3), 187–189. https://doi.org/10.1002/DA.20002

Tang, W.-K., Grace Lau, C., Mok, V., Ungvari, G. S., & Wong, K.-S. (2015). Insomnia and health-related quality of life in stroke. *Topics in Stroke Rehabilitation, 22(3),* 201–207. https://doi.org/10.1179/1074935714Z.0000000026

Taule, T., & Råheim, M. (2014). Life changed existentially: A qualitative study of experiences at 6-8 months after mild stroke. *Disability and Rehabilitation, 36(25),* 2107–2119. https://doi.org/10.3109/09638288.2014.904448

Taylor, G. H., Todman, J., & Broomfield, N. M. (2011). Post-stroke emotional adjustment: A modified social cognitive transition model. *Neuropsychological Rehabilitation, 21(6),* 808–824. https://doi.org/10.1080/09602011.2011.598403

Teasdale, T. W., & Engberg, A. W. (2010). Psychosocial consequences of stroke: A long-term population-based follow-up. *Brain Injury,* 19(12), 1049–1058. https://doi.org/10.1080/02699050500110421

Thomas, A. J., Kalaria, R. N., & O'Brien, J. T. (2004). Depression and vascular disease: What is the relationship? *Journal of Affective Disorders, 79(1–3),* 81–95. https://doi.org/10.1016/S0165-0327(02)00349-X

Thomson, H. J., & Thomas, S. (2013). The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods, 4(1),* 95–101. https://doi.org/10.1002/JRSM.1060

Tinakon, W., & Nahathai, W. (2012). A Comparison of Reliability and Construct Validity between the Original and Revised Versions of the Rosenberg Self-Esteem Scale. *Psychiatry Investigation, 9(1),* 54. https://doi.org/10.4306/PI.2012.9.1.54

Townend, E. (2005). Beliefs about "stroke" and 'its effects': a study of their association with emotional distress. [Doctoral dissertation, University of Edinburgh]. University of Edinburgh Research Repository. https://era.ed.ac.uk/handle/1842/30852

UK National Collaborating Centre for Mental Health. (2010). The classification of depression and depression rating scales/questionnaires. https://www-ncbi-nlm-nih-gov.uea.idm.oclc.org/books/NBK82926/

Vataja, R., Pohjasvaara, T., Mäntylä, R., Ylikoski, R., Leppävuori, A., Leskelä, M., Kalska, H., Hietanen, M., Aronen, H. J., Salonen, O., Kaste, M., & Erkinjuntti, T. (2003). MRI correlates of executive dysfunction in patients with ischaemic stroke. *European Journal of Neurology, 10(6),* 625–631. https://doi.org/10.1046/j.1468-1331.2003.00676.x

Vickery, C. D., Sepehri, A., & Evans, C. C. (2008). Self-esteem in an acute stroke rehabilitation sample: a control group comparison. *Clinical Rehabilitation, 22(2),* 179–187. https://doi.org/10.1177/0269215507080142

Wade, D. T., Wood, V. A., & Hewer, R. L. (1985). Recovery after stroke - The first 3 months. *Journal of Neurology Neurosurgery and Psychiatry, 48(1),* 7–13. https://doi.org/10.1136/jnnp.48.1.7

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67(6),* 361–370. https://doi.org/10.1111/J.1600-0447.1983.TB09716.X

## Chapter 3. Systematic Review Extended Methodology

### Inclusion and Exclusion Criteria

Studies exclusively investigating Transient Ischaemic Attacks (TIAs) or that mixed TIA and stroke samples were excluded because we were interested in capturing the complexities of the post-stroke experience in the context of permanent neurological injury.

The decision to include only studies with control groups, as opposed to studies investigating each population separately, was made as a means of controlling for demographic variables, such as nationality and the year of study. Studies featuring control groups with physical health difficulties were excluded with the justification of reducing heterogeneity.

Studies comparing symptom prevalence or severity were excluded if overall depression severity was not approximately controlled between groups. If depression was substantially different in severity between groups, it would have been difficult to determine if the difference in any individual symptom was caused by differences in depression phenomenology or more severe overall depression.

Comparisons of overall depression scores were excluded because we aimed to focus on symptomology; differences in depression prevalence and overall severity are well documented in the literature (e.g. Desmond et al., 2003; Lindén et al., 2007; Mitchell et al., 2017). In general, we limited the scope of our analysis to symptoms or domains that were contained within commonly used depression measures or diagnostic criteria (Bell, 1994), meaning other domains, such as quality of life or life satisfaction, were excluded; an exception was made for anxiety because of its theoretical links with the maintenance of depression and the high degree of comorbidity in stroke (Fennell, 1997; Schöttke and Giabbiconi, 2015). Specific measures were included in the search terms as a means of excluding articles without quantitative analysis of validated depression measures.

### Quality Appraisal and Risk of Bias

The NHLBI case-control checklist was also considered in the review, but this checklist lacked an item relating to outcome measure validity, which was deemed to be important because of the dependence of our findings on accurate assessment and the high variability of outcome measures used.

The item relating to sufficient timeframe between exposure (the stroke event) and outcome (the emergence of depression) in the cross-sectional NHLBI checklist was also deemed to be highly relevant, in light of the evidence for the poor predictive value of early stroke assessment for predicting the development of depression later in recovery (Lees et al., 2014) and processes of normal adjustment (Taylor et al., 2011). There needed to be some consideration of elapsed time since stroke so that the likelihood of capturing the presence of longer-term mood disorders was maximised. For these reasons, the more general cross-sectional checklist was selected, despite all included studies featuring control groups.

**Additional Scoring Information**

Because the risk of bias was the primary reason for quality assessment, ratings were applied to studies only for methods extracted, and not the quality or impact of the paper in its entirety. As such, quality ratings should not be interpreted as an indicator of the quality of the research, but as an indicator of the risk of bias for the specific analyses extracted.

For item 4 of the NHLBI checklist, relating to samples from similar populations, 'no' ratings were generally given to studies that reported uncontrolled demographic differences between groups. Gainotti et al. (1999) did not report demographic features of the control group, so a 'not reported' rating was given. Non-reporting contributed negatively to the overall appraisal.

For item 7, relating to sufficient timeframe between the exposure, the stroke event, and outcome (the emergence of mood disorders), studies that sampled patients under 2-4 weeks post-stroke were given 'no' ratings for the reasons outlined above. Studies that only provided vague indications of time since stroke were rated 'cannot determine' for this item.

For item 10, relating to the assessment of exposures more than once over time, this item was adapted to focus on the outcome, depression, being assessed on multiple occasions. This decision was made because stroke diagnosis does not generally require multiple assessments, unless degenerative processes are suspected, and repeated longitudinal measures of depression would reduce bias by supporting the separation of emotional adjustment processes and acute medical illness with long-term

depression. Only House et al. (1991) used a longitudinal design, and Gainotti et al. (1999) were the only authors to investigate multiple distinct groups at separate time points since the index stroke.

Consistency of rating for item 14, which related to the statistical adjustment for confounding variables, was challenging because several studies controlled for some demographic differences but not others (Lipsey et al., 1986). 'Yes' responses were generally given to studies that used some form of matching or adjustment for key demographic differences, and 'no' responses were given to studies that had significant population differences without any clear evidence of control for the specific analysis that was extracted. de Man-van Ginkel et al. (2015) controlled for demographic differences in some of their analyses, but not for the extracted profile comparison, resulting in a 'no' rating for item 14.

Two studies were given overall 'fair' ratings, despite performing well on the NHLBI checklist (Fleming et al., 2021; Stokes et al., 2011). Each of these papers used analyses that provided imprecise indications of the relationship between the symptom and depression; Fleming et al. reported only that group membership was not a significant contributor to the regression model, and Stokes et al. reported data relating to the point increase in fatigue scores caused by one-point changes in depression scores. These analyses offered less information and precision than correlations and did not support direct significance testing of correlation strengths. The imprecision of these analyses concerning the specific research question of the review was suspected to increase the risk of bias, despite the NHLBI not containing an item relating to the informativeness of the statistical approach adopted.

**Measures Included**

There was high heterogeneity of measures utilised by the included studies. Data from eight depression measures were synthesised in this review: the Centre for Epidemiologic Studies- Depression Scale (CES-D; Radloff, 2016), Geriatric Depression Scale (GDS; Sheikh and Yesavage, 1986), the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960), the Hospital Anxiety and Depression Scale (HADS; Zigmond and Snaith, 1983), the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001), the Montgomery-Asberg Depression Rating Scale (MADRS; Montgomery and Asberg, 1979), the Post-Stroke Depression Rating Scale (PDRS; Gainotti et al., 1997), and diagnostic interviews using the Present State Examination (PSE; Cooper, 1985). The abovementioned tools have

generally demonstrated adequate psychometric properties in stroke (Burton and Tyson, 2015), except for the PDRS, which has limited validity evidence.

The following additional measures were used in the analyses of comparative correlation strengths: the Visual Analogue Self-Esteem Scale (VASES; Brumfitt and Sheeran, 1999), the Rosenberg Self-Esteem Scale (RSES; Rosenberg, 1979), the Multidimensional Fatigue Inventory (MFI; Smets et al., 1995), the Beck Anxiety Inventory (BAI; Beck et al., 1988), and the Sleep Condition Indicator (SCI; Espie et al., 2014). The VASES and MFI have been validated in stroke populations (Bennett et al., 2006; Christensen et al., 2008), while only general population validity evidence exists for the BAI (Fydrich et al., 1992), SCI (Palagini et al., 2015), and RSES (Tinakon and Nahathai, 2012).

## Results

### *Combining Studies Across Different Methodologies*

**Grouping Variables.** For analyses of grouping variables, such as time since stroke, and their potential link to results, comparative correlation and profile comparison studies were combined. The results of these two methodologies were judged to broadly correspond because a trait with a higher degree of correlation with depression in one group might also be expected to have greater prevalence and severity in a depressed sample. Combining these studies at this stage was also necessary because neither methodological category alone could cover each of the levels of certain characteristics, such as time since stroke categories. The single DIF study (Pickard et al., 2006) was excluded from this analysis because differences in latent symptom severity were judged to be conceptually distinct from the association of symptoms with depression or their prevalence. For example, a symptom can be equally correlated with depression between groups, but it might be more sensitive to a higher band of depression severity in one group. It was, therefore, suspected that the inclusion of this study was suspected to add noise to the analysis because “more”, “less”, or “no difference” codes would have been potentially incompatible or conflicting between these design types, obscuring patterns of effect direction.

Furthermore, a finding of lower symptom severity of an item, such as the crying item, in the stroke group might be associated with greater prevalence of that symptom in the stroke population in

a profile comparison study, if its lower latent severity is an indicator of ease to endorse. This might have led to conflicts, whereby the DIF study would lead to a 'less' code and a profile comparison study leading to a 'more' code, despite an internally consistent underlying mechanism. DIF findings were therefore not combined with other methodologies for the analysis of grouping variables.

**Main Analyses of Symptom Differences.** Analyses of phenomenological differences were completed separately for each methodology because, despite broad epistemological similarities, methodological differences might have obscured more nuanced and detailed relationships. Comparisons of profiles are unable to indicate the relative degree of association of a symptom with depression, and therefore its specificity as an indicator of depressed mood, because greater prevalence and/or severity could, at least partially, be explained by causes other than depression; fatigue might be more prevalent after stroke, because of post-stroke fatigue (Ellis, 2014), but not necessarily due to the presence of depression or the correlation between the two.

## Chapter 4. Bridging Chapter

The systematic review identified similarities and differences in the phenomenology of depression between stroke and non-stroke controls. Differences were found in some domains, such as indicators of apathy, but, contrary to expectations, were not demonstrated in the somatic depression domain. It was hypothesised that extraneous bias caused observed differences in emotionalism and inhibition of return to work between groups and the influence of undetected bias for other symptom domains could not be ruled out.

A concern that arose from these findings was that, if measurement bias was indeed present, then the assumed general equivalence of latent (underlying) depression severity between groups in profile comparison studies might have been violated. Inflated work inhibition and emotionalism scores may, therefore, have led to an overestimation of general depression severity scores in the stroke group, making the groups appear equal in depression severity even if the non-stroke group had greater underlying severity. In cases where participants were specifically matched for depression severity by inspecting total scores or observing broadly overlapping profiles, there may have been significant differences in true depression if those overall scores are biased, as suspected.

Between-groups measurement invariance analysis is one method for compensating for this problem (Kim and Yoon, 2011). If item scores are significantly biased by extraneous sources, this can result in a) differences in measure dimensionality, b) the latent positioning of thresholds from which a person endorses a higher item category, c) the degree of depression predicted by the item, d) the baseline score of the item when controlling for latent depression severity, and e) the variance of the item (Yang et al., 2008). Confirmatory Factor Analysis (CFA) can model these above parameters and identify the presence of differences between groups using measurement invariance analysis (Wu and Estabrook, 2016). Furthermore, it has the capability of estimating differences in *latent depression*, which is the true underlying severity of depression, more free from bias from differences in the above parameters (Kim and Yoon, 2011). CFA can, therefore, directly address the above-mentioned concern by identifying whether these biases exist and if these biases influence depression total scores on measures. The empirical paper that follows, therefore, aimed to apply CFA and measurement

invariance analysis to a commonly used depression tool, the Patient Health Questionnaire-9 (PHQ-9), and identify whether stroke comorbidities lead to measurement bias.

## Chapter 5. The Factor Structure of the Patient Health Questionnaire-9 in Stroke: A Comparison with a Non-Stroke Population

Blake, J.[1]; Dr Munyomwe, T., PhD [2]; Dr Fischer, F., PhD [3]; Dr Levis, B., PhD [4]; Dr Benedetti, A., PhD [4]; Prof Thombs, B., PhD[4]; Dr Gracey, F., DClinPsy [1]; & Prof Broomfield, N., PhD, DClinPsy [1]

[1]Department of Clinical Psychology and Psychological Therapies, University of East Anglia, UK

[2]School of Medicine, University of Leeds, UK

[3]Charité Universitätsmedizin Berlin, Germany

[4]Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Canada

**Corresponding author**

Joshua Blake

Department of Clinical Psychology and Psychological Therapies

Norwich Medical School

University of East Anglia

Norwich Research Park

Norwich, NR4 7TJ

Email: joshua.blake@nhs.net

**Collaborating authors:** Dr de Man Van-Ginkel, J.; Dr Munholz, T. N.; Dr Tan, L.; Dr Koehler, S.; Dr Wang, J. L.; Dr Jeon, H. J.; Prof Whooley, M. A.; Dr Van Der Feltz-Cornelis, C.; Dr Hobfoll, S.; Dr Levin-Aspenson, H.; Dr Patten, S. B.; Dr Quinn, T.; & Dr Simning, A.

Abstract word count: 250

Manuscript word count (not including title page, abstract, tables, or references): 6044

This manuscript has been formatted for submission to the *Journal of Stroke & Cerebrovascular Diseases.*

**Keywords:** Depression, Stroke, Confirmatory Factor Analysis, Dimensionality Assessment, PHQ-9

## Abstract

**Objectives:** This paper aimed to investigate the factor structure of the Patient Health Questionnaire-9 (PHQ-9) depression measure in stroke, benchmarked against a non-stroke comparison sample, because of concerns that comorbid stroke sequelae may bias the measurement of somatic items.

**Materials and Methods:** Data were obtained from authors contributing to the DEPRESSD secondary data project. The sample constituted 787 stroke and 12,016 non-stroke participants. A subsample of 1,574 more demographically aligned non-stroke participants was selected via propensity matching. Dimensionality was assessed by comparing fit statistics of one-factor, two-factor, and bi-factor models. Measurement invariance analysis was performed to identify between-group differences in factor structure.

**Results:** A two-factor model, consisting of somatic and cognitive-affect latent variables, had a superior fit to a unidimensional model (CFI = .984 versus CFI =.974), but the high between-factor correlations indicated unidimensionality (r = .866). Configural invariance between stroke and non-stroke was supported (CFI = .983, RMSEA = .080), as were invariant item thresholds (p = .092) and loadings (p = .103). Strong invariance was violated (p < .000, ΔCFI = -.003), indicating non-invariant item intercepts. Follow-up analysis indicated between-group differences in tiredness and appetite intercepts, and latent depression was significantly overestimated in the stroke group using a summed score approach.

**Conclusions:** The PHQ-9 performed well in the stroke samples and was generally robust to interference from extraneous stroke comorbidities. However, the presence of non-invariant intercepts meant that stroke PHQ-9 total scores using a summed score approach were incomparable with non-stroke equivalents. Directions for future research are outlined.

## Introduction

Depression is a prevalent consequence of stroke, occurring in approximately 33% of survivors (Hackett and Pickles, 2014; Mitchell et al., 2017). Post-stroke depression is not only a concern because of its clear wellbeing implications, but also its interference with functional recovery (Pohjasvaara et al., 2001). It is, therefore, important that post-stroke depression is assessed accurately so that appropriate support can be provided.

For depression assessment post-stroke, screening tools are recommended because of their speed of administration and generally sufficient psychometric properties (Burton and Tyson, 2015). Accordingly, their usage is commonly recommended in national accreditation guidelines, such as Sentinel Stroke National Audit Programme (SSNAP; King's College London, 2020).

The Patient Health Questionnaire-9 (PHQ-9) is one of the most widely used depression screening tools in stroke (Burton and Tyson, 2015; Kroenke et al., 2001). Its popularity in stroke populations most likely rests on its ubiquity as a depression measure more broadly (Levis et al., 2019; Moriarty et al., 2015), the fact that it is free to use, its favourable reliability and validity in stroke (De Man-Van Ginkel et al., 2012b), and its generally acceptable diagnostic accuracy (Burton and Tyson, 2015; Levis et al., 2019).

Despite evidence for favourable diagnostic accuracy and convergent validity, concerns about the applicability of the PHQ-9 in stroke remain (Burton et al., 2013). A primary criticism suggests that several items contained within the PHQ-9, such as tiredness and difficulties with concentration, may capture experiences caused by the extraneous physical health complications that are common in stroke recovery, rather than by depression symptomology (Chilcot et al., 2013; Langhorne et al., 2000). For example, items relating to tiredness and sleep disruption may be heavily influenced by the presence of the clinically distinct phenomena of post-stroke fatigue (Acciarresi et al., 2014) and post-stroke insomnia (Baylan et al., 2020). Conversely, it is also possible that these items are a valid measure of depression, even if there is some degree of overlay attributable to such biasing factors (de Man-van Ginkel et al., 2015; Robinson, 2006).

Failure of PHQ-9 unidimensionality has been demonstrated in other health populations, such as spinal cord injury and palliative care (Chilcot et al., 2013; Krause et al., 2010, 2011). A two-factor solution, consisting of a somatic cluster and a cognitive-affective cluster, was favoured over a one-factor solution in both populations. Inspection of factor correlations indicated that a large proportion of variance could not be accounted for by a second-order depression factor in these groups. The findings of a two-factor solution in populations with health conditions contrasts with a primary care population, where unidimensionality has generally been indicated (Kim and Lee 2019; Boothroyd, Dagnan, and Muncer 2019; González-Blanch et al. 2018; Huang et al. 2006). These findings collectively indicate a possible vulnerability of the PHQ-9 to compromised construct validity in populations with health conditions, because of the interference of physical health symptoms.

Multidimensionality, if not accounted for, could undermine the psychometric effectiveness of the measure. For example, it may add noise to optimal cut-off estimations if there is a general pattern of score inflation because of background physical comorbidity. This, in turn, could result in inaccurate diagnosis and inappropriate treatment in subpopulations with significant physical health complaints (Kang, 2013). Similarly, multidimensionality may impact clinical trials that use the PHQ-9 as an outcome measure.

A recent publication found evidence that somatic symptoms may not substantially undermine the psychometric performance of the PHQ-9 in stroke (Katzan et al., 2021). The authors used Differential Item Functioning (DIF) to assess differences in item responses between three levels of disease severity in stroke, Parkinson's disease, migraine, and Amyotrophic Lateral Sclerosis (ALS) populations in a high-powered study. The authors found a negligible impact of stroke severity on item responses, indicating the possible robustness of PHQ-9 items to the sequelae of cerebrovascular accidents.

However, there are two limitations with these findings. First, the Stroke Impact Scale (Duncan et al., 1999) was used as an indicator of disease severity; while this is a valid scale of stroke severity, its scope is broad and it may not be a sensitive indicator of the particular hypothesised factors, such as fatigue, that could be problematic for the factor structure of the PHQ-9. Indeed, with

post-stroke fatigue so commonly reported in stroke (Acciarresi et al., 2014), fatigue-related interference could have been present in each of the three levels of stroke severity featured in the study. Second, the authors assessed DIF in only a one-factor model and did not perform a comparison with a two-factor model; therefore, the optimal factor structure of PHQ-9 in stroke remained unclear.

Another recently published paper investigated the dimensionality of the PHQ-8, with the suicide item removed, and found evidence for unidimensionality (Dong et al., 2022). The authors also assessed longitudinal measurement invariance concerning time since stroke, at 3-, 6-, and 12-months post-stroke and found evidence for non-invariance of loadings, attributable to the appetite item. Model fit of the single-factor model at 12 months, according to Root Mean Square Error of Approximation (RMSEA) values, approached insufficiency. The finding of poorer fit of the unidimensional model with increased time since stroke is surprising because, if somatic stroke symptoms indeed interfere with accurate depression measurement, the presence of functional recovery after a stroke (Kotila et al., 1999; Wade et al., 1985) might be expected to reduce the presence of somatic symptom complaints, and consequential interference of such symptoms with the unidimensional measurement of depression, rather than increase such interference in the direction of multidimensionality. The finding of borderline insufficient fit at 12 months post-stroke substantiates the need for further assessment of dimensionality, and there remains a clear rationale for assessing dimensionality of the full PHQ-9 measure with the suicide item included.

Even if unidimensionality was found to hold in stroke, it is possible that population factors, such as the elevated presence of fatigue in stroke (Acciarresi et al., 2014), interfere with measurement accuracy in more subtle ways, such as causing differences in item category thresholds, factor loadings, item intercepts and/or residual variances (Yang and Jones, 2008). Differences in these parameters between stroke and other populations could invalidate statistical comparisons between these groups (Meredith and Teresi, 2006). For example, unequal intercepts between populations can bias depression estimates using the traditional summed score approach, and potentially lead to a greater likelihood of type I errors (Kim and Yoon, 2011). Such differences for individual items would

also be of importance to clinicians that wish to interpret individual item scores. Therefore, any confound to measurement accuracy and comparability between groups must be understood.

In addition to concerns about the dimensionality of the PHQ-9, there has been considerable debate within the literature about the optimal factor positioning of certain items. For example, 7 and 8, pertaining to problems with concentration and slowed movement, respectively, have commonly been loaded onto the affective factor by some authors (Chilcot et al., 2013; González-Blanch et al., 2018) and onto the somatic factor by others (Villarreal-Zegarra et al., 2019). Direct comparisons of the two alternatives have rendered inconsistent results (Elhai et al., 2012; Patel et al., 2019). It remains unclear which factors account for these differences, given that one Peru-based study found evidence for measurement invariance for sex, age, education, socioeconomic status, marital status, and area of residence, although this study did not compare between-country differences (Villarreal-Zegarra et al., 2019).

In stroke, there are several additional factors that might be captured by items 7 and 8 and cause it to load onto the somatic factor; such factors include neurologically predicted executive dysfunction, attention deficits, slowed processing, dysphasia, dysarthria, post-stroke-fatigue, insomnia, and motor disruption (Baylan et al., 2020; Flowers et al., 2013; Loetscher and Lincoln, 2013; Low et al., 2017; Olney and Richards, 1996; Pohjasvaara et al., 2002; Rahamatali et al., 2021). Though there is likely to be an interdependence between these additional stroke sequelae and depression (Broomfield et al., 2011), undue influence from these other factors, which potentially load onto items 7 and 8 even in the absence of depression (de Man-van Ginkel et al., 2015), may result in these items covarying more strongly with somatic items. Understanding this optimal positioning in stroke could support the theoretical understanding of the measurement orientation of certain items by indicating whether problems with, for example, slowed movement or concentration difficulties covary more with cognitive-affective or somatic experiences of depression.

Based on the literature outlined above, it is therefore important that the dimensionality of the PHQ-9 in stroke is further appraised. Furthermore, it is important that the factor structure of the PHQ-

9 in stroke is compared with other populations to understand population-level effects. We, therefore, address the following aims in the present study:

1) To provide an assessment of the dimensionality of the PHQ-9 in a stroke population
2) To identify differences in factor structure that might be attributable to stroke via measurement invariance comparison with non-stroke samples
3) To identify the optimal positioning of items 7 and 8 in a two-factor model

## Methods

### Design

We conducted a cross-sectional study, using secondary data from multiple investigators contributing to the DEPRESSD individual participant meta-analysis (*The DEPRESSD Project*, 2021). All studies contained item-level PHQ-9 data, pseudonymised demographic characteristics, and binary stroke presence status. Studies spanned multiple nations and languages (Table 3).

### Participants

The stroke group was composed of data from five studies (De Man-Van Ginkel et al., 2012b; Prisnie et al., 2016; Simning et al., 2018; Thombs et al., 2008; Quinn, 2022). These data produced a combined sample of 796 stroke participants. The non-stroke group was comprised of data from eight studies (Hobfoll et al., 2011; Janssen et al., 2016; Kim et al., 2017; Levin-Aspenson and Watson, 2018; Liu and Wang, 2015; Santos et al., 2013; Simning et al., 2012; Volker et al., 2016), with a total of 12,016 participants. The non-stroke comparison group were recruited from a variety of settings, generally free of significant health morbidities. Most used random sampling of the general population ( Kim et al., 2017; Levin-Aspenson and Watson, 2018; Liu and Wang, 2015; Santos et al., 2013), but three of the smaller samples were recruited from specific groups or contexts (Hobfoll et al., 2011; Simning et al., 2012; Volker et al., 2016). These participants are, therefore, referred to as the non-stroke or comparison group.

Demographic data for each cluster, the stroke group, and the non-stroke comparison group, are summarised in Table 3. Substantial differences were found between stroke and non-stroke

comparison samples in sex, $X^2$ *(1, N=12,798) = 56.9, p < .001, w = .067*; nationality, $X^2$ *(6, N=12,798) = 2669.6, p < .001, w = .46;* language, $X^2$*(5, N=12,798) = 347.8, p < .001, w = .17*; age, *t(12,796) = -39.3, p< .001, d= 1.46;* and PHQ-9 total score, *t(12,796) = -14.9, p< .001, d= .62*.

Data relating to stroke characteristics, such as type, location, and elapsed time since index stroke were not available at the individual level. However, where available, the broad range of elapsed time since the index stroke event for each cluster is provided in Table 3. Participants from Prisnie et al. (2016) and de Man-van Ginkel et al. (2015) were sampled within one-year post-stroke, and participants from Quinn et al. (2022) were sampled within the first two weeks of stroke. The data were insufficient for between-groups analyses of different post-stroke timepoints.

**Measures**

The PHQ-9 consists of nine items, which correspond to the nine criteria for depression in the DSM-IV (Bell, 1994). Participants respond based on their experiences of mood problems over the previous two weeks on a four-point ordinal scale, with higher ordinal categories denoting the increased frequency of the respective symptom. The selection of a higher-level ordinal category, therefore, does not indicate increased symptom intensity directly, but increased depression severity is commonly assumed from higher overall scores. Item scores were added to calculate a measure total, with a maximum overall score of 27. Suggested cut-off scores for depression categorisation in stroke vary in the literature, but many studies agree on an optimal cut-off of ≥10 (Burton and Tyson, 2015; De Man-Van Ginkel et al., 2012b; Levis et al., 2019; Williams et al., 2005). The PHQ-9 has demonstrated high classification accuracy in stroke and primary care (Burton and Tyson, 2015; Levis et al., 2019).

**Ethics**

The Research Ethics Committee of the Jewish General Hospital recommended that the DEPRESSD project did not require research ethics approval because all data are from secondary sources. In keeping with local university ethical policy, an application for ethical approval was made to the University of East Anglia Faculty of Medicine and Health Research Ethics Committee (UEA FMH REC) on 10th November 2020 and approval was confirmed on 11th August 2021. Data transfer procedures required that the original authors send their respective DEPRESSD formatted de-identified

data to the authors directly. Documentation of ethical approval and an explanation of the data transfer process is provided in Appendix F.

**Table 3** Demographic data, stratified by cluster

| Study | n | % female | Age (SD) | Country | Language | Population description | Mean PHQ-9 (SD) | Approximate time since stroke (months) |
|---|---|---|---|---|---|---|---|---|
| **Stroke** | | | | | | | | |
| De Man-Van Ginkel (2012) | 382 | 45.8 | 69.2 (14.5) | Netherlands | Dutch | Community stroke patients | 6.6 (5.6) | 0.3 – 2.6 (M: 1.58) |
| Prisnie (2016) | 114 | 56.1 | 59.6 (15.5) | Canada | English | Outpatient community stroke | 4.8 (5.2) | 2.2 – 10.6 (M: 3.6) |
| Quinn (2022) | 135 | 47.7 | 68.4 (12.7) | UK | English | Inpatient acute stroke | 6.7 (6.3 | 0 – 0.45 |
| Simning (2018) | 21 | 47.6 | 70.0 (6.5) | US | English | Older adults in public housing | 4.7 (4.4) | Unavailable |
| Thombs (2008) | 144 | 16.7 | 69.7 (10.1) | US | English | People with Coronary Artery Disease in the community | 5.8 (5.6) | Unavailable |
| **Total** | 796 | 42.3 | 67.8 (13.9) | | | | 6.1 (5.6) | |
| **Non-stroke Comparison Samples** | | | | | | | | |
| Hobfoll (2011) | 144 | 57.7 | 41.6 (15.2) | Israel | Hebrew and Arabic | Jewish and Palestinian residents of Jerusalem exposed to war | 5.9 (5.9) | |
| Kim (2017) | 3071 | 56.6 | 38.8 (12.2) | South Korea | Korean | Randomly selected adults, via S. Korean census | 2.2 (2.2) | |
| Liu (2015) | 4182 | 55.0 | 44.7 (10.0) | Canada | English | Working population | 3.2 (4.0) | |
| Janssen (2016) | 3502 | 55.3 | 59.5 (8.6) | Netherlands | Dutch | Population-based cohort study, recruited from multiple sources | 2.7 (3.2) | |
| Levin-Aspenson (2018) | 408 | 68.6 | 45.0 (13.4) | US | English | General population community-based adults | 6.5 (6.6) | |
| Santos (2013) | 447 | NA | 43.8 (15.1) | Brazil | Portuguese | General population via random household sampling | 5.0 (5.1) | |
| Simning (2012) | 169 | 59.2 | 67.3 (6.6) | US | English | Older adults in public housing | 5.1 (4.3) | |
| Volker (2016) | 93 | 50.5 | 46.4 (10.9) | Netherlands | Dutch | Employees on sickness leave in an occupational health setting | 8.5 (7.6) | |
| **Total** | 12016 | 56.0 | 47.8 (13.5) | | | | 3.1 (4.0) | |

## Analysis

### *Initial Data Processing and Demographic Matching*

Stroke clusters were combined into one set because each cluster had an insufficient sample size to be modelled separately. Statistically, accounting for clustering was impractical because of the number of clusters, the small size of each cluster, and the limitations of current software capabilities (Yang, 2019).

Because of the large demographic differences between stroke and non-stroke comparison clusters, propensity score matching was used to select a sample from the non-stroke dataset that was more demographically aligned to the stroke sample (Caliendo and Kopeinig, 2008). Propensity score matching is commonly used in observational designs to compare intervention and control groups, where non-random sampling exists (Muruet et al., 2018). The procedure involves the logistic regression of specified covariates, such as gender and age, onto group membership. Propensity scores, the percentage likelihood of a person to be categorised into the target group based on their covariate profile, are calculated for the target and control groups. Individuals are subsequently matched between the groups, based on their propensity scores.

Propensity score matching was conducted using the MatchIt package in R (Ho et al., 2011). Matching was completed with differing ratios of control to target participants, and the most closely matched sample was selected. Matching was completed for age, sex, country, and PHQ-9 total score.

The DEPRESSD datasets contained only complete PHQ-9 records, but demographic data were missing in a small percent of cases. Propensity matching required complete data for all demographic variables selected for matching. Participant data that contained missing values for these variables were, therefore, excluded listwise from any analysis after the matching process.

### *Assessment of Dimensionality*

Confirmatory Factor Analysis (CFA) was modelled in R using Lavaan (Rosseel, 2012) and SEMTools (semTools Contributors, 2016) packages. A single-factor CFA model, which represented one latent factor of generalised depression, was evaluated alongside two variations of a two-factor model and a

bi-factor model, consistent with previous dimensionality research (Chilcot et al., 2013; Krause et al., 2010). The two-factor models consisted of a somatic factor and cognitive-affective factor (Chilcot et al., 2013; Krause et al., 2010). Because of inconsistencies in the literature about the optimal specification of the two-factor model (Chilcot et al., 2013; Elhai et al., 2012; Krause et al., 2010; Patel et al., 2019), we fitted two alternative two-factor models with items 7 and 8 loaded onto each factor.

The bi-factor model consisted of a general factor and two specific factors, cognitive-affective and somatic (Fischer et al., 2021). The factor location of items seven and eight in the bifactor model was determined by examining which of the two-factor models had a superior fit in each group (Elhai et al., 2012). Bifactor models specify uncorrelated factors; this means that the global factor should measure latent depression, and each specific factor should measure covariance between the items contained within each specific factor that is ***not*** explained by latent depression (Fischer et al., 2021). Bifactor models can be used to test for dimensionality by comparing latent depression scores between the bifactor and one-factor models. In the case of unidimensionality, a high degree of correlation would be expected between latent depression scores from the one-factor model and latent depression scores (the global factor) from the bi-factor model, because each respective factor would act as a specific measure of depression (Dunn & McCray, 2020). If multidimensionality were present, the global depression factor in the bifactor model would lose significant variance to the specific, uncorrelated, factors because a large proportion of somatic item variance would not be due to depression (Reise et al., 2007). In this case, the bifactor global depression factor would vary less than the latent factor in the one-factor model, and a poorer correlation would be observed (Dunn & McCray, 2020).

Each CFA model was fitted using Diagonally Weighted Least Squares (DWLS) estimation, which is suitable for ordinal-level data and cases where multivariate normality is violated (Forero et al., 2009). Robust standard errors, and mean and variance adjusted test statistics, were computed by Lavaan using the full weight matrix (Rosseel, 2012). Robust equivalents of Root Mean Square Error of Approximation (RMSEA; values of <.08 interpreted as acceptable fit) and Comparative Fit Index (CFI; >.96 interpreted as acceptable fit) were used to evaluate model fit (Li, 2016). Model fit of the

single, two-factor and bifactor models was compared to identify whether one or two latent factors best described the covariances in item responses in the stroke sample (Chilcot et al., 2013; Fischer et al., 2021; Krause et al., 2010).

***Measurement Invariance***

The stroke and comparison groups were assessed for measurement invariance, using the procedures proposed by Wu and Estabrook (2016) for ordinal measures, to evaluate possible differences in factor structure. This involved equality testing of four parameters between groups: item thresholds, loadings, intercepts, and residual variances. Item thresholds can be understood as the point on the item's latent continuum in which people, on average, start to endorse each higher ordinal category, loadings as the degree of association of an item with its respective factor, item intercepts as the expected value of the item if the latent factor (depression) is set to 0, and residual variances as variance that is not accounted for by the latent factors (Finch and French, 2015). If the groups do not significantly differ in any of these parameters, the groups are considered to be invariant (Schmitt and Kuljanin, 2008). Invariance of thresholds is henceforth referred to as 'threshold invariance', invariance of loadings as 'weak invariance' or 'metric invariance', invariance of item intercepts as 'strong invariance' or 'scalar invariance', and invariance of item residuals as 'full invariance' (Putnick and Bornstein, 2016).

The procedure for measurement invariance testing, using the Wu and Estabrook (2016) methodology, is as follows: 1) a baseline multi-group CFA model, referred to as a configural model, is fit to the data. The configural model constrains factor means to 0 and item intercepts to 0, and factor and residual variances to 1. Providing the configural model has sufficient fit, referred to as configural invariance, progressive restraints are applied and tested against the previous model. Thresholds are constrained to be equal across the groups first, followed by loadings, intercepts and residual variances. Unnecessary constraints are progressively dropped as new constraints are added. This methodology is described in more detail, in a similar clinical context, by Fischer et al. (2018).

Each model was compared to the previous, to identify whether the imposition of the new constraint was responsible for a significant reduction in model fit. In cases of non-invariance, a change in the direction of poorer fit would be expected because the model specifies equivalences that

are not reflected in the data. Thus, changes to fit after each stage were examined, via two methods: a one-way ANOVA significance test in chi-square fit statistics, using the Satorra (2000) method, and by inspecting changes to the CFI and RMSEA. Increases in $X^2$ and RMSEA values and decreases in CFI values are indicative of reduced model fit. Changes in CFI >.001 are commonly interpreted as an indicator of measurement non-invariance (Khademi et al., 2021).

Chi-square difference tests are sample-size dependent and likely to find significant differences among small non-meaningful effects in large samples (Davidov et al., 2014). However, caution is also advised with interpreting changes to CFI and RMSEA when using ordered data and DWLS estimation (Sass et al., 2014). A pragmatic approach was, therefore, adopted; non-significant changes to chi-square values were assumed to be robust indicators of invariance, given the large samples. In cases where a significant p-value of chi-square difference was observed, detailed exploration of invariance violation was explored to identify the meaningfulness of the observed differences.

### *Sample Size*

There is divided opinion about the minimum sample requirements for accurate CFA modelling in the literature, and the accuracy of parameter estimations depend on the number of estimated parameters, degrees of freedom, number of items, the scale of measurement, and other factors (Kyriazos, 2018). MacCallum et al. (1999) suggest sample sizes of 300-500 are robust to low communalities and loadings, with a minimum of 200. Based on the data available in the present study, our sample size in each group was therefore sufficient for accurate and robust parameter estimation.

## Results

See Appendix G for R code used in the below analyses.

### Data Processing and Matching

Eight cases of missing data were identified in the stroke sample and removed before propensity matching, resulting in a final stroke sample of 787. Equally, five missing cases from the non-stroke comparison sample were identified and removed prior to matching. Missing cases constituted just .11% of the original dataset and were, therefore, not anticipated to bias findings.

Propensity score matching was modelled with a 1:1 and 2:1 ratio of non-stroke to stroke participants, respectively. Inspection of demographic differences between each non-stroke sample and the stroke sample indicated that the 2:1 ratio sample had similar differences in age, gender, and country, but smaller differences in PHQ-9 total score. Accordingly, the 2:1 ratio sample was selected for analysis.

Demographic details of the stroke group and propensity-matched comparison group, including significance tests of demographic differences, are summarised in Table 4. Compared with the pre-matched sample, the demographic differences in the matched sample were substantially reduced for each variable, with non-significant gender and PHQ-9 total score differences. Significant differences remained for nationality, language, and age, with medium to large effect sizes. Despite findings of significant differences in nationality, most participants in both groups were from western developed nations (94.9% in non-stroke and 100% in stroke) and may, therefore, have represented broadly similar cultural backgrounds.

Post-hoc significance testing, using adjusted residual statistics (Sharpe, 2015), indicated significant proportional differences in nationality between stroke and comparison groups for the UK, Canada, Brazil, Israel, and South Korea. Significant proportional language differences were only observed for Portuguese and Korean.

**Table 4** Demographic overview of the matched sample

| | | Comparison Group (n = 1574) | Stroke (n = 787) | Diff | Test statistic | p | Effect statistic | Effect size descriptor |
|---|---|---|---|---|---|---|---|---|
| **% Female** | | 41.4% | 42.3% | -0.9% | .17 ($X^2$) | .679 | .01 ($\phi$) | Non-significant |
| **Age** | | 61.8 (SD 11.3) | 67.8 (SD 13.9) | 6.04 | -10.56 (t) | < .000 | .48 (*d*) | Medium |
| **Country** | UK | 0.0% | 16.0% | -16.0% | 349.37 ($X^2$) | < .000 | .39 (*V*) | Large |
| | US | 17.1% | 21.0% | -3.9% | | | | |
| | Canada | 30.3% | 14.5% | 15.8% | | | | |
| | Netherlands | 47.5% | 48.5% | -1.0% | | | | |
| | Brazil | 2.8% | 0.0% | 2.8% | | | | |
| | Israel | 1.2% | 0.0% | 1.2% | | | | |
| | South Korea | 1.1% | 0.0% | 1.1% | | | | |
| | **Total** | 100% | 100% | | | | | |
| **Language** | English | 47.4% | 51.5% | -4.1% | 42.41 ($X^2$) | < .000 | .13 (*V*) | Medium |
| | Dutch | 47.5% | 48.5% | -1.0% | | | | |
| | Portuguese | 2.8% | 0.0% | 2.8% | | | | |
| | Hebrew | 0.7% | 0.0% | 0.7% | | | | |
| | Arabic | 0.5% | 0.0% | 0.5% | | | | |
| | Korean | 1.1% | 0.0% | 1.1% | | | | |
| | **Total** | 100% | 100% | | | | | |
| **PHQ-9 total** | | 5.9 (6.0) | 6.1 (5.6) | .24 | -.95 (t) | .343 | .04 (*d*) | Non-significant |

**The Dimensionality of the PHQ-9 in Stroke**

Fit statistics of the one-factor model, each two-factor model, and the bifactor model are summarised in Table 5 for stroke and non-stroke groups. All models had sufficient fit to the data, based on CFI values >.96 in both groups. All stroke group models had sufficient RMSEA fit of <.08, but the non-stroke one-factor model did not.

Of the two alternate forms of the two-factor model, the model that specified problems with concentration and moving slowly in the somatic factor, two-factor A, had a superior fit in both groups, which suggests that items 7 and 8 covary more strongly with the somatic items 3, 4, and 5. A high

degree of correlation was observed between factors in each two-factor model for each group, which is indicative of unidimensionality.

The bifactor model was the best-fitting model in both groups. Superior fit was expected because the specific factors (cognitive/affective and somatic) in bifactor models are designed to capture additional variance not explained by the global depression factor. To further assess dimensionality in the stroke group, global depression factor scores were calculated from the bifactor model and plotted against latent depression scores from the one-factor model (see Appendix H). The correlation between latent depression scores was .99 in the stroke group, which indicated substantial shared variance in factor-derived scores. This provided strong support for the unidimensionality of the PHQ-9 in stroke.

**Table 5** Specification and fit statistics of stroke CFA models

| | | | | Stroke group | | | | Comparison group | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Model** | **Factors** | **Model specification** | **Model parameters** | **X2 (df)** | **Robust CFI** | **Robust RMSEA** | **Factor correlation** | **X2 (df)** | **Robust CFI** | **Robust RMSEA** | **Factor correlation** |
| One-factor | 1 | All items onto a single factor | 36 | 129.04 (27) | .974 | .069 | | 355.9 (27) | .982 | .088 | |
| Two-factor A | 2 | Cognitive/affective: 1,2,6 9<br>Somatic: 3, 4, 5, 7, 8 | 37 | 88.61 (26) | .984 | .055 | .866 | 212.1 (26) | .990 | .067 | .910 |
| Two-factor B | 2 | Cognitive/affective: 1,2,6, 7, 8, 9<br>Somatic: 3, 4, 5 | 37 | 108.27 (26) | .979 | .063 | .865 | 252.1 (26) | .988 | .074 | .908 |
| Bifactor | 3 | Global depression: all items<br>Cognitive/affective: 1,2,6 9<br>Somatic: 3, 4, 5, 7, 8<br>Factor variances are freely estimated. Correlations between factors set to 0 | 45 | 39.36 (18) | .995 | .039 | Set to 0 | 121.8 (18) | .994 | .061 | Set to 0 |

*$X^2$= Chi-squared, df = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation*

**Measurement Invariance Testing**

With unidimensionality established in the stroke group, measurement invariance testing between stroke and comparison groups was performed to identify the existence of subtler differences in factor structure. Findings are summarised in Table 6. A unidimensional multi-group configural model was first specified, which was found to have sufficient model fit (CFI > .96, RMSEA = .080). This indicated a similar overall factor structure between the groups.

**Table 6** Unstandardised $X^2$ fit statistics of multi-group CFA at each level of constraint

| **Model** | ***X*² (df)** | **CFI scaled** | **RMSEA scaled** | ***ΔX*² (*Δ df*)** | **p** | ***ΔCFI*** | ***ΔRMSEA*** |
|---|---|---|---|---|---|---|---|
| Configural model | 268.6 (54) | .983 | .080 | | | | |
| Constrained thresholds | 272.3 (63) | .981 | .077 | 15.0 (9) | .092 | -.001 | -.003 |
| Constrained loadings | 284.2 (71) | .982 | .072 | 13.2 (8) | .103 | .000 | -.005 |
| Constrained intercepts | 403.8 (79) | .979 | .074 | 57.8 (8) | <.000* | -.003 | .001 |

*$X^2$= Chi-squared, df = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, Δ= change in the associated fit statistic,* the p-value corresponds to the significance of change of chi-square of model fit.

The constraining of item thresholds and loadings to be equal between groups led to non-significant increases in unstandardised chi-square statistics and only marginal changes to CFI and RMSEA values, which indicated that such impositions did not substantially reduce model fit. Metric invariance was, therefore, found, which implied equal thresholds and that each item contributed to the latent construct to a similar degree across groups.

A statistically significant reduction of model fit was observed after the constraint of item intercepts, as indicated by the p-value for the chi-square difference. This finding indicated the presence of significant between-group differences in intercepts and, therefore, scalar non-invariance.

To identify the intercepts responsible for the violation of scalar invariance, nine CFA models were specified with item intercept constraints released one by one. See Appendix I for a summary of intercept differences. Inspection of intercepts indicated substantial between-group differences for item 4, relating to tiredness, and 5, relating to appetite. The tiredness intercept was substantially greater in the stroke group, with a relative difference of .446, and the appetite intercept was significantly greater

in the matched comparison group, with a relative difference of .346. By contrast, the average absolute magnitude of intercept differences of the remaining items was .07. These findings indicated that tiredness scores were higher in the stroke group and appetite disruption scores were higher in the comparison group when controlling for latent depression severity.

A partially invariant model was specified to confirm that between-groups differences in intercepts for items 4 and 5 were responsible for failed scalar invariance. The partially invariant model estimated the intercepts of items 4 and 5 freely and maintained constraints for the remaining intercepts. The fit of the partially constrained model ($X^2$=296.2, CFI= .985, RMSEA = .064) was statistically compared to the constrained loadings model ($X^2$ = 284.2, CFI = .982, RMSEA = .072), with no significant reduction in model fit observed, $X^2$ diff = 7.69, df diff = 6, p = .262. This finding confirmed the responsibility of items 4 and 5 for failed scalar invariance.

To identify the meaningfulness of the scalar invariance violation, the effect sizes of between groups comparisons using a traditional summed score approach and using model-derived scores were compared (see Table 7). The stroke and non-stroke groups did not significantly differ when using the traditional summed score approach, because of group matching. However, model-implied depression scores indicated significantly greater latent depression in the non-stroke group. The observed effect size was .117 larger for the partially invariant model than for the fully invariant model, which suggested that falsely assumed scalar invariance would have resulted in moderate underestimation of between-groups differences.

**Table 7** Between-groups differences in depression scores, derived from a sum score approach and model-derived latent factor estimations for fully and partially invariant models

| Scoring approach | Stroke M | Non-stroke M | t/z statistic | p | Effect size (Cohen's d) | Descriptor |
|---|---|---|---|---|---|---|
| Sum score | 6.141 | 5.9 | -0.94 (t) | .255 | -.045 | Non-significant |
| Fully invariant model | 0 | .272 | 4.80 (z) | < .000 | .272 | Small |
| Partially invariant model | 0 | .389 | 5.91 (z) | <.000 | .389 | Small |

*Negative signs indicate higher estimated depression in the stroke group, and positive signs indicate higher depression in the non-stroke group.*

The effect size of between-groups comparisons of depression scores derived from the partially invariant model was .434 larger than that observed when using the summed score method, which is a difference that equates to a small-to-medium effect. The effect of unequal intercepts was, therefore, substantial, which indicated that depression scores, using a summed score approach, were incomparable between groups.

## Discussion

This study aimed to assess the dimensionality of the PHQ-9 in stroke, to identify possible differences in factor structure to those in the wider population and identify whether items 7 and 8 fit better onto the somatic or cognitive-affective latent factors. Despite two-factor models demonstrating better fit than the one-factor model, we found there to be a strong indication of unidimensionality in stroke. A high correlation between latent factors was observed, with a strong association of bifactor model-derived global depression scores with one-factor derived scores. This finding contributes further evidence for PHQ-9 in favour of good psychometric performance in stroke (Burton and Tyson, 2015) and general robustness to the sequelae of stroke (Katzan et al., 2021).

Items 7 and 8, relating to difficulties with concentration and moving slowly/feeling restless, respectively, were found to load better onto the somatic factor. This finding was also observed in the comparison group; for this reason, we did not interpret the superior fit of two-factor model A to be caused by stroke-specific factors, such as the interference of physical health consequences. A superior model fit of two-factor A is consistent with findings from some studies (Elhai et al., 2012; Krause et al., 2011; Villarreal-Zegarra et al., 2019) but inconsistent with the optimal factor structure reported by other authors (Chilcot et al., 2013; Kocalevent et al., 2013; Patel et al., 2019). This finding indicated that in both our stroke and non-stroke samples, the items relating to feeling slowed down and having difficulties concentrating covaried more strongly with somatic depression symptoms.

From the comparisons of factor structure with the non-stroke sample, the two groups were found to possess invariant thresholds and factor loadings. This implies that the factor correlations of items are broadly equivalent between populations and that the thresholds in which patients move to

endorse a higher response rating occur at approximately equal points on the items' latent continuous severity scales.

Differences in intercepts were, however, observed; specifically, a large positive intercept of tiredness was observed in the stroke group and a large positive intercept of appetite disruption in the comparison group. This finding implies that stroke patients are more likely to positively endorse the tiredness item, and non-stroke participants the appetite item, even when latent depression is 0.

The violation of scalar invariance was substantial, with measurable implications. A small-to-medium between-groups difference in latent depression was obscured by the inequalities in intercepts, which suggests that the two intercept differences do not cancel out and that the tiredness intercept significantly inflates the stroke PHQ-9 total score. This, in turn, suggests that depression scores between stroke and non-stroke populations cannot be reliably compared using a summed score approach. Even though group-matching for PHQ-9 total scores made depression severity appear similar between groups, the non-stroke sample was, on average, more depressed.

There are several possible explanations for the greater tiredness intercept in the stroke group. First, it is possible that, as hypothesised, post-stroke fatigue and other relevant stroke sequelae lead to greater loading onto the tiredness item, even when latent depression is 0 (Acciarresi et al., 2014; Lai et al., 2002). It is also possible that the difference can be explained by imperfect matching, such as age or nationality-related effects. If the tiredness item is, indeed, biased by the presence of post-stroke fatigue, this might add noise to optimal cut-off points for depression in stroke samples, and potentially lead to an overdiagnosis of depression for those with post-stroke fatigue and an underdiagnosis for those without.

The possible explanations for the greater appetite intercept in the non-stroke group are unclear. One possibility is the observed age differences between groups, as a negative association between age and appetite disruption on the PHQ-9 in primary care has been previously reported by the primary authors of one of the included data clusters (de Man-van Ginkel et al., 2015). The intercept for appetite disruption may, therefore, be age-dependent, although this was not identified by

previous measurement invariance studies (Villarreal-Zegarra et al., 2019). It is also possible that differences in nationality and language are relevant, but sample sizes were too small within each country and each stroke status group to assess measurement invariance of these factors. Identifying possible causes for this finding should, therefore, be a focus of future research.

**Strengths and Limitations**

A major strength of this research is that it was one of the first studies to compare the factor structure of the PHQ-9 between stroke and non-stroke comparison groups. We have provided additional confirmation of the general robustness of the measure's overall dimensionality to physical health problems and other stroke sequelae (Dong et al., 2022; Katzan et al., 2021). The finding of noninvariant intercepts is also of importance because it suggests that comparisons of PHQ-9 scores between stroke and other groups might be invalid, even if valid for comparisons within these groups.

Despite these strengths, there are several limitations inherent in the current study. First, it was difficult to statistically account for nested data because this would have required multiple measurement invariance calculations between groups and because Lavaan has limited functionality for completing measurement invariance on nested data. Unaccounted clustering can lead to biased parameter estimations and standard errors because of the presence of item covariations within clusters that are not accounted for by the latent variable (Dyer et al., 2005). Second, the large demographic variation between clusters and groups meant that demographic matching was imperfect, which placed limitations on the interpretation of measurement invariance testing. Finally, data on the amount of time elapsed since the index stroke event were not available. As such, the stroke participants were likely to have a substantial variance in their position in stroke recovery. Time since the stroke event may also be important because a recent publication has demonstrated metric non-invariance of the appetite item loading as a factor of time since stroke (Dong et al., 2022) and because significant improvement in physical functioning can be observed after six months and sometimes up to one year (Kwakkel et al., 2004; Studenski et al., 2001) in addition to the emergence of cognitive decline and greater impairment in some (Mijajlović et al., 2017). The variability in time since stroke may, therefore, have introduced noise to the analysis. These abovementioned limitations mean that

confidence in the conclusions is compromised by the quality of available evidence and highlight avenues for future research.

**Clinical Implications**

Based on our findings of unidimensionality, clinicians are encouraged to continue using the PHQ-9 in stroke practice. However, clinicians should be mindful that stroke patients may report higher baseline tiredness on item 4 of the PHQ-9, and that those with significant post-stroke fatigue may have inflated total scores compared to those who do not. This may lead to reduced accuracy of the cut-off point for depression if optimal cut-offs are indeed based on samples that include a mixture of both sub-groups. Clinicians should also be mindful of recent findings of measurement non-invariance associated with time since stroke, which implies that the measure may function less as a unified measure of global depression for those later in the recovery trajectory. Caution in the interpretation of patient scores near the boundary of cut-off points is, therefore, advised until a sensitive analysis of sub-groups and interaction with stroke recovery, is completed.

**Future Research**

Several avenues for future research emerge from the current study. The causes of the differences in intercepts must be investigated in more detail; new hypotheses have now been raised for why the appetite intercept may have been greater in the non-stroke group. It is also important that measurement invariance is assessed between people with post-stroke fatigue and those without and if non-invariance is observed, the impact on diagnostic accuracy should be explored. Further investigation into the effects of age on factor structure is also needed; while Villarreal-Zegarra et al. (2019) found measurement invariance of age, there may be an interaction between age and nationality, so it is important that this finding is replicated in other countries.

The recent findings of only partial invariance in factor structure with increased time since stroke, with notable loading non-invariance for the appetite item, and the possible failure of unidimensionality above 12 months post-stroke (Dong et al., 2022) warrant further investigation into the effects of stroke recovery. It is important that a mixed two-factor approach to measurement

invariance is adopted, whereby group-level and longitudinal-level measurement invariance are simultaneously assessed so that the reasons for changes to factor structure can be better understood.

Finally, a future analysis of the item structure of the PHQ-9 using Item Response Theory (IRT) techniques is recommended because it would provide greater insights into the ability of items to discriminate between mild and severe depression, as well as provide a clearer picture of the latent severity of items.

**Conclusion**

The results of the current study provide strong support for unidimensionality in a stroke population. Differences in intercepts between stroke and non-stroke groups were observed, with a higher tiredness intercept in the stroke group and a higher appetite disruption intercept in the comparison group. This violation of measurement invariance is clinically significant. Researchers should interpret mean differences in PHQ-9 total scores between stroke and other groups with extreme caution and consider using CFA model-derived depression scores as an alternative. Further research is needed to identify whether there are differences in factor structure throughout stroke recovery.

## References

Acciarresi, M., Bogousslavsky, J., & Paciaroni, M. (2014). Post-Stroke Fatigue: Epidemiology, Clinical Characteristics and Treatment. *European Neurology, 72(5–6),* 255–261. https://doi.org/10.1159/000363763

Baylan, S., Griffiths, S., Grant, N., Broomfield, N. M., Evans, J. J., & Gardani, M. (2020). Incidence and prevalence of post-stroke insomnia: A systematic review and meta-analysis. *Sleep Medicine Reviews, 49,* 101222. https://doi.org/10.1016/j.smrv.2019.101222

Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research, 271,* 532–534. https://doi.org/10.1016/j.psychres.2018.12.048

Burke, D. L., Ensor, J., & Riley, R. D. (2017). Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Statistics in Medicine, 36(5),* 855. https://doi.org/10.1002/SIM.7141

Burton, L-J. & Tyson, S. (2015). Screening for mood disorders after stroke: A systematic review of psychometric properties and clinical utility. *Psychological Medicine, 45(1),* 29–49. https://doi.org/10.1017/S0033291714000336

Burton, L-J., Tyson, S., & McGovern, A. (2013). Staff perceptions of using outcome measures in stroke rehabilitation. *Disability and Rehabilitation, 35(10),* 828–834. https://doi.org/10.3109/09638288.2012.709305

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys, 22(1),* 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Chilcot, J., Rayner, L., Lee, W., Price, A., Goodwin, L., Monroe, B., Sykes, N., Hansford, P., & Hotopf, M. (2013). The factor structure of the PHQ-9 in palliative care. *Journal of Psychosomatic Research, 75(1),* 60–64. https://doi.org/10.1016/j.jpsychores.2012.12.012

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement Equivalence in Cross-National Research. *Annual Review of Sociology, 40,* 55–75. https://doi.org/10.1146/ANNUREV-SOC-071913-043137

de Man-van Ginkel, J. M., Hafsteinsdóttir, T. B., Lindeman, E., Geerlings, M. I., Grobbee, D. E., & Schuurmans, M. J. (2015). Clinical manifestation of depression after stroke: Is it different from depression in other patient populations? *PLoS ONE, 10(12).* https://doi.org/10.1371/journal.pone.0144450

de Man-van Ginkel, J. M., Hafsteinsdóttir, T., Lindeman, E., Burger, H., Grobbee, D., & Schuurmans, M. (2012b). An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke*, *43(3)*, 854–856. https://doi.org/10.1161/STROKEAHA.111.640276

Dong, L., Williams, L. S., Briceno, E., Morgenstern, L. B., & Lisabeth, L. D. (2022). Longitudinal assessment of depression during the first year after stroke: Dimensionality and measurement invariance. *Journal of Psychosomatic Research, 153*, 110689. https://doi.org/10.1016/J.JPSYCHORES.2021.110689

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale Version 20: Evaluation of reliability, validity, and sensitivity to change. *Stroke*, *30(10)*, 2131–2140. https://doi.org/10.1161/01.STR.30.10.2131

Dunn, K. J., & McCray, G. (2020). The Place of the Bifactor Model in Confirmatory Factor Analysis Investigations Into Construct Dimensionality in Language Testing. *Frontiers in Psychology*, *11*, 1357. https://doi.org/10.3389/fpsyg.2020.01357

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16(1)*, 149–167. https://doi.org/10.1016/J.LEAQUA.2004.09.009

Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Prescott, M. R., Shirley, E., Chan, P. K., Slembarski, R., Liberzon, I., Galea, S., & Calabrese, J. R. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research, 199(3)*, 169–173. https://doi.org/10.1016/j.psychres.2012.05.018

Finch, W. H., & French, B. F. (2015). Latent variable modeling with R. https://doi.org/10.4324/9781315869797/LATENT-VARIABLE-MODELING-HOLMES-FINCH-BRIAN-FRENCH

Fischer, F., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Leplège, A. (2018). Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Quality of Life Research, 27(4)*, 999–1014. https://doi.org/10.1007/s11136-018-1785-8

Fischer, F., Levis, B., Falk, C., Sun, Y., Ioannidis, J. P. A., Cuijpers, P., Shrier, I., Benedetti, A., & Thombs, B. D. (2021). Comparison of different scoring methods based on latent variable models of the PHQ-9: an individual participant data meta-analysis. *Psychological Medicine*, 1–12. https://doi.org/10.1017/s0033291721000131

Flowers, H. L., Silver, F. L., Fang, J., Rochon, E., & Martino, R. (2013). The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic stroke. *Journal of Communication Disorders, 46(3)*, 238–248. https://doi.org/10.1016/j.jcomdis.2013.04.001

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 16(4)*, 625–641. https://doi.org/10.1080/10705510903203573

González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., & Cano-Vindel, A. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE, 13(2)*. https://doi.org/10.1371/journal.pone.0193356

Hackett, M. L., & Pickles, K. (2014). Part I: Frequency of depression after stroke: An updated systematic review and meta-analysis of observational studies. *International Journal of Stroke, 9(8)*, 1017–1025. https://doi.org/10.1111/ijs.12357

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. Journal of Statistical Software, 42(8), 1–28. https://doi.org/10.18637/JSS.V042.I08

Hobfoll, S. E., Canetti, D., Hall, B. J., Brom, D., Palmieri, P. A., Johnson, R. J., Pat-Horenczyk, R., & Galea, S. (2011). Are Community Studies of Psychological Trauma's Impact Accurate? A Study Among Jews and Palestinians. *Psychological Assessment, 23(3),* 599–605. https://doi.org/10.1037/A0022817

Huang, F. Y., Chung, H., Kroenke, K., Delucchi, K. L., & Spitzer, R. L. (2006). Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal of General Internal Medicine, 21(6),* 547–552. https://doi.org/10.1111/j.1525-1497.2006.00409.x

Janssen, E. P. C. J., Köhler, S., Stehouwer, C. D. A., Schaper, N. C., Dagnelie, P. C., Sep, S. J. S., Henry, R. M. A., van der Kallen, C. J. H., Verhey, F. R., & Schram, M. T. (2016). The Patient Health Questionnaire-9 as a Screening Tool for Depression in Individuals with Type 2 Diabetes Mellitus: The Maastricht Study. *Journal of the American Geriatrics Society, 64(11),* e201–e206. https://doi.org/10.1111/JGS.14388

Kang, H. (2013). A guide on the use of factor analysis in the assessment of construct validity. *Journal of Korean Academy of Nursing, 43(5),* 587–594. https://doi.org/10.4040/jkan.2013.43.5.587

Katzan, I. L., Lapin, B., Griffith, S., Jehi, L., Fernandez, H., Pioro, E., Tepper, S., & Crane, P. K. (2021). Somatic symptoms have negligible impact on Patient Health Questionnaire-9 depression scale scores in neurological patients. *European Journal of Neurology, 28(6),* 1812–1819. http://10.0.4.87/ene.14822

Khademi, A., Wells, C. S., & Oliveri, M. E. (2021). Examining Appropriacy of CFI/TLI Cutoff Value in Multiple-group CFA Test of Measurement Invariance to Enhance Accuracy of Test Score Interpretation. https://doi.org/10.31234/OSF.IO/Y36UV

Kim, D. J., Kim, K., Lee, H. W., Hong, J. P., Cho, M. J., Fava, M., Mischoulon, D., Heo, J. Y., & Jeon, H. J. (2017). Internet Game Addiction, Depression, and Escape from Negative

Emotions in Adulthood: A Nationwide Community Sample of Korea. *Journal of Nervous and Mental Disease, 205(7),* 568–573. https://doi.org/10.1097/NMD.0000000000000698

Kim, E. S., & Yoon, M. (2011). Testing Measurement Invariance: A Comparison of Multiple-Group Categorical CFA and IRT. *Structural Equation Modelling: A Multidisciplinary Journal, 18(2),* 212-228. https://doi.org/10.1080/10705511.2011.557337

Kim, Y. E., & Lee, B. (2019). The psychometric properties of the patient health questionnaire-9 in a sample of Korean university students. *Psychiatry Investigation, 16(12),* 904–910. https://doi.org/10.30773/pi.2019.0226

King's College London. (2020). SSNAP. Retrieved from: https://www.strokeaudit.org/About-SSNAP.aspx

Kocalevent, R. D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry, 35(5),* 551–555. https://doi.org/10.1016/J.GENHOSPPSYCH.2013.04.006

Kotila, M., Numminen, H., Waltimo, O., & Kaste, M. (1999). Post-stroke depression and functional recovery in a population-based stroke register. The Finnstroke study. *European Journal of Neurology, 6(3),* 309–312. https://doi.org/10.1046/j.1468-1331.1999.630309.x

Krause, J. S., Reed, K. S., & McArdle, J. J. (2010). Factor structure and predictive validity of somatic and nonsomatic symptoms from the patient health questionnaire-9: A longitudinal study after spinal cord injury. *Archives of Physical Medicine and Rehabilitation, 91(8),* 1218–1224. https://doi.org/10.1016/j.apmr.2010.04.015

Krause, J. S., Saunders, L. L., Bombardier, C., & Kalpakjian, C. (2011). Confirmatory Factor Analysis of the Patient Health Questionnaire-9: A Study of the Participants From the Spinal Cord Injury Model Systems. *PM and R, 3(6),* 533–540. https://doi.org/10.1016/j.pmrj.2011.03.003

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16(9),* 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kwakkel, G., Kollen, B., & Lindeman, E. (2004). Understanding the pattern of functional recovery

after stroke: Facts and theories. *Restorative Neurology and Neuroscience, 22(3-5),* 281-299.

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology, 9(8),* 2207–2230. http://www.scirp.org/journal/doi.aspx?DOI=10.4236/psych.2018.98126

Lai, S. M., Studenski, S., Duncan, P. W., & Perera, S. (2002). Persisting consequences of stroke measured by the stroke impact scale. *Stroke, 33(7),* 1840–1844. https://doi.org/10.1161/01.STR.0000019289.15440.F2

Langhorne, P., Stott, D. J., Robertson, L., MacDonald, J., Jones, L., McAlpine, C., Dick, F., Taylor, G. S., & Murray, G. (2000). Medical Complications After Stroke. *Stroke, 31(6),* 1223–1229. https://doi.org/10.1161/01.STR.31.6.1223

Levin-Aspenson, H. F., & Watson, D. (2018). Mode of administration effects in psychopathology assessment: Analyses of gender, age, and education differences in self-rated versus interview-based depression. *Psychological Assessment, 30(3),* 287–295. https://doi.org/10.1037/PAS0000474

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: Individual participant data meta-analysis. *The BMJ,* 365. https://doi.org/10.1136/bmj.l1476

Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. Psychological Methods, 21(3), 369–387. https://doi.org/10.1037/MET0000093

Liu, Y., & Wang, J. L. (2015). Validity of the Patient Health Questionnaire-9 for DSM-IV major depressive disorder in a sample of Canadian working population. *Journal of Affective Disorders, 187,* 122–126. https://doi.org/10.1016/J.JAD.2015.07.044

Loetscher, T., & Lincoln, N. B. (2013). Cognitive rehabilitation for attention deficits following stroke. *Cochrane Database of Systematic Reviews, 2013(5).* https://doi.org/10.1002/14651858.CD002842.pub2

Low, E., Crewther, S. G., Ong, B., Perre, D., & Wijeratne, T. (2017). Compromised motor dexterity confounds processing speed task outcomes in stroke patients. *Frontiers in Neurology*, *8*, 21. https://doi.org/10.3389/fneur.2017.00484

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4(1),* 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44(11 Suppl 3).* https://doi.org/10.1097/01.MLR.0000245438.73837.89

Mijajlović, M. D., Pavlović, A., Brainin, M., Heiss, W.-D., Quinn, T. J., Ihle-Hansen, H. B., Hermann, D. M., Assayag, E. Ben, Richard, E., Thiel, A., Kliper, E., Shin, Y.-I., Kim, Y.-H., Choi, S., Jung, S., Lee, Y.-B., Sinanović, O., Levine, D. A., Schlesinger, I., … Bornstein, N. M. (2017). Post-stroke dementia - a comprehensive review. *BMC Medicine, 15(1),* 11. https://doi.org/10.1186/s12916-017-0779-7

Mitchell, A. J., Sheth, B., Gill, J., Yadegarfar, M., Stubbs, B., Yadegarfar, M., & Meader, N. (2017). Prevalence and predictors of post-stroke mood disorders: A meta-analysis and meta-regression of depression, anxiety and adjustment disorder. *General Hospital Psychiatry*, *47*, 48–60. https://doi.org/10.1016/j.genhosppsych.2017.04.001

Moriarty, A. S., Gilbody, S., McMillan, D., & Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. *General Hospital Psychiatry, 37(6),* 567–576. https://doi.org/10.1016/j.genhosppsych.2015.06.012

Murad, H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence Based Medicine, 21(4),* 125-127. https://doi.org/10.1136/ebmed

Muruet, W., Rudd, A., Wolfe, C. D. A., & Douiri, A. (2018). Long-term survival after intravenous thrombolysis for ischemic stroke: A propensity score-matched cohort with up to 10-year follow-up. *Stroke, 49(3),* 607–613. https://doi.org/10.1161/STROKEAHA.117.019889

Olney, S. J., & Richards, C. (1996). Hemiparetic gait following stroke. Part I: Characteristics. *Gait and Posture, 4(2),* 136–148. Elsevier B.V. https://doi.org/10.1016/0966-6362(96)01063-6

Patel, J. S., Oh, Y., Rand, K. L., Wu, W., Cyders, M. A., Kroenke, K., & Stewart, J. C. (2019). Measurement Invariance of the Patient Health Questionnaire-9 (PHQ-9) Depression Screener in U.S. Adults Across Sex, Race/Ethnicity, and Education Level: NHANES 2005-2016. *Depression and Anxiety*, *36(9)*, 813. https://doi.org/10.1002/DA.22940

Pohjasvaara, T., Leskela, M., Vataja, R., Kalska, H., Ylikoski, R., Hietanen, M., Leppavuori, A., Kaste, M., & Erkinjuntti, T. (2002). Post-stroke depression, executive dysfunction and functional outcome. *European Journal of Neurology*, *9(3)*, 269–275. https://doi.org/10.1046/j.1468-1331.2002.00396.x

Pohjasvaara, T., Vataja, R., Leppävuori, A., Kaste, M., & Erkinjuntti, T. (2001). Depression is an independent predictor of poor long-term functional outcome post-stroke. *European Journal of Neurology*, *8(4)*, 315–319. https://doi.org/10.1046/j.1468-1331.2001.00182.x

Prisnie, J. C., Fiest, K. M., Coutts, S. B., Patten, S. B., Atta, C. A. M., Blaikie, L., Bulloch, A. G. M., Demchuk, A., Hill, M. D., Smith, E. E., & Jetté, N. (2016). Validating screening tools for depression in stroke and transient ischemic attack patients. *International Journal of Psychiatry in Medicine*, *51(3)*, 262–277. https://doi.org/10.1177/0091217416652616

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. Developmental Review: *DR*, *41*, 71. https://doi.org/10.1016/J.DR.2016.06.004

Rahamatali, M., De Bont, N., Valet, M., Halkin, V., Hanson, P., Deltombe, T., Lejeune, T., & Selves, C. (2021). Post-stroke fatigue: how it relates to motor fatigability and other modifiable factors in people with chronic stroke. *Acta Neurologica Belgica*, *121(1)*, 181–189. https://doi.org/10.1007/s13760-020-01453-9

Quinn, T. J., Taylor-Rowan, M., Elliott, E., Drozdowska, B., McMahon, D., Broomfield, N. M., Barber, M., MacLeod, M. J., Cvoro, V., Byrne, A., Ross, S., Crow, J., Slade, P., Dawson, J., & Langhorne, P. (2022). Research protocol – Assessing Post-Stroke Psychology Longitudinal Evaluation (APPLE) study: A prospective cohort study in stroke. *Cerebral Circulation - Cognition and Behavior*, *3*, 100042. https://doi.org/10.1016/J.CCCB.2022.100042

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16(SUPPL. 1),* 19–31. https://doi.org/10.1007/s11136-007-9183-7

Robinson, R. G. (2006). *The clinical neuropsychiatry of stroke: Second edition.* Cambridge University Press. https://doi.org/10.1017/CBO9780511544231

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48(2),* 1 - 46. https://doi.org/10.18637/JSS.V048.I02

Santos, I. S., Tavares, B. F., Munhoz, T. N., de Almeida, L. S. P., da Silva, N. T. B., Tams, B. D., Patella, A. M., & Matijasevich, A. (2013). Sensibilidade e especificidade do Patient Health Questionnaire-9 (PHQ-9) entre adultos da população geral. *Cadernos de Saude Publica, 29(8),* 1533–1543. https://doi.org/10.1590/0102-311X00144612

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Structural Equation Modeling, 21(2),* 167–180. https://doi.org/10.1080/10705511.2014.882658/SUPPL_FILE/HSEM_A_882658_SM7145.DOCX

Satorra, A. (2000). *Scaled and Adjusted Restricted Tests in Multi-Sample Analysis of Moment Structures.* In: Heijmans, R.D.H., Pollock, D.S.G., Satorra, A. (eds) Innovations in Multivariate Statistical Analysis. Advanced Studies in Theoretical and Applied Econometrics, vol 36. Springer, Boston, MA. https://doi.org/10.1007/978-1-4615-4603-0_17

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18(4),* 210–222. https://doi.org/10.1016/J.HRMR.2008.03.003

semTools Contributors. (2016). semTools: Useful tools for structural equation modeling. R package version 0.4-14. Retrieved from: https://cran.r-project.org/package=semTools

Sharpe, D. (2015). Chi-Square Test is Statistically Significant: Now What? Practical Assessment, *Research, and Evaluation Practical Assessment,* 20. https://doi.org/10.7275/tbfa-x148

Simning, A., Seplaki, C. L., & Conwell, Y. (2018). The association of a heart attack or stroke with depressive symptoms stratified by the presence of a close social contact: findings from the National Health and Aging Trends Study Cohort. *International Journal of Geriatric Psychiatry, 33(1),* 96–103. https://doi.org/10.1002/gps.4684

Simning, A., Van Wijngaarden, E., Fisher, S. G., Richardson, T. M., & Conwell, Y. (2012). Mental healthcare need and service utilization in older adults living in public housing. *The American Journal of Geriatric Psychiatry: Official Journal of the American Association for Geriatric Psychiatry, 20(5),* 441–451. https://doi.org/10.1097/JGP.0B013E31822003A7

Studenski, S. A., Wallace, D., Duncan, P. W., Rymer, M., & Lai, S. M. (2001). Predicting Stroke Recovery: Three and Six-Month Rates of Patient-Centered Functional Outcomes Based on the Orpington Prognostic Scale. *Journal of the American Geriatrics Society, 49(3),* 308–312. https://doi.org/10.1046/j.1532-5415.2001.4930308.x

The DEPRESSD Project. (2021). https://www.depressd.ca/

Thombs, B. D., Ziegelstein, R. C., & Whooley, M. A. (2008). Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *Journal of General Internal Medicine, 23(12),* 2014–2017. https://doi.org/10.1007/S11606-008-0802-Y

Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabé-Ortiz, A., Melendez-Torres, G. J., & Bazo-Alvarez, J. C. (2019). Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics. *PLoS ONE, 14(9).* https://doi.org/10.1371/JOURNAL.PONE.0221717

Volker, D., Zijlstra-Vlasveld, M. C., Brouwers, E. P. M., Homans, W. A., Emons, W. H. M., & van der Feltz-Cornelis, C. M. (2016). Validation of the Patient Health Questionnaire-9 for Major Depressive Disorder in the Occupational Health Setting. *Journal of Occupational Rehabilitation, 26(2),* 237–244. https://doi.org/10.1007/S10926-015-9607-0

Wade, D. T., Wood, V. A., & Hewer, R. L. (1985). Recovery after stroke - The first 3 months. *Journal of Neurology Neurosurgery and Psychiatry, 48(1),* 7–13. https://doi.org/10.1136/jnnp.48.1.7

Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika, 81(4),* 1014. https://doi.org/10.1007/S11336-016-9506-0

Yang, F. M., & Jones, R. N. (2008). Measurement differences in depression: chronic health-related and sociodemographic effects in older Americans. *Psychosomatic Medicine, 70(9),* 993–1004. https://doi.org/10.1097/PSY.0b013e31818ce4fa

Yang, L. (2019). Testing Measurement Invariance in Multilevel Data with Unequal Cross-Level Factor Structures. [Doctoral Thesis from University of Arkansas]. Graduate Theses and Dissertations Retrieved from https://scholarworks.uark.edu/etd/3294

# Chapter 6. Discussion and Critical Analysis

This thesis portfolio aimed to critically examine the concept of post-stroke depression and explore the evidence for (Aizenstein et al., 2014, 2016; Crowe et al., 2016; Gainotti et al., 1999; Pan et al., 2012), and against (Cumming et al., 2010; de Man-van Ginkel et al., 2015) its distinct phenomenology. Two approaches were adopted to address this aim: (1) a systematic review and synthesis of studies that compared the symptomatology of post-stroke depression with depression in the general population, and (2) an exploration of the factor structure of one of the most used depression screening tools, the Patient Health Questionnaire-9 (Kroenke et al., 2001).

In this chapter, the findings of each paper are briefly summarised, followed by a broader discussion of the significance of the portfolio in its entirety. A critical analysis follows, with considerations of alternative ontological and epistemological perspectives. Finally, areas for future research are outlined.

## Research Summary

### *Systematic Review*

Results of profile comparison studies have been commonly cited in papers that reviewed the construct of post-stroke depression (Llorca et al., 2015; Robinson, 2006; Robinson and Jorge, 2016). However, no systematic approach to synthesising these results, alongside evidence available from alternative methodologies, had yet been conducted. We identified two additional methodologies: comparisons of correlation strengths between a symptom and depression, and Differential Item Functioning (DIF) of latent symptom severity, but both were underutilised. Longitudinal explorations of changes in symptomatology over the recovery span, which offer powerful insights into the moderating influences of adjustment and physical recovery (House et al., 1991), were also underutilised.

High heterogeneity of included studies obscured any clear patterns, and analysis of between-groups differences in the time since stroke, study quality, choice of methodology and comparison group setting were all found to influence findings substantially. Despite this limitation, we found evidence of broad similarities in symptom profiles, with comparable symptom prevalence and severity

found for negative affect, somatic features of depression, negative cognitions, cognitive difficulty, anxiety, and suicidal ideation between the stroke and non-stroke general populations.

Anhedonia/apathy appeared to be more prevalent in the general population, and correlation studies provided tentative indication for a greater relatedness of anxiety and fatigue in the depressive experience of people in the general population. Positive responses to appetite disruption and crying were indicative of more severe depression in primary care than in stroke (Pickard et al., 2006). Higher latent symptom severity of crying in primary care might be indicative of the interference of post-stroke emotionalism, which causes uncontrollable crying in stroke patients without depression (Broomfield et al., 2021; Calvert et al., 1998; Fitzgerald et al., 2021). Reasons for differences in appetite remain unclear.

Greater work inhibition and emotional variation were observed in the stroke samples. No other consistent findings emerged in support of greater prevalence, severity, or construct "closeness" of a symptom in the stroke group. The biasing of emotional variation and work inhibition symptoms by post-stroke emotionalism and physical and cognitive barriers to working was hypothesised (Calvert et al., 1998; Wolfenden and Grace, 2009).

The possible presence of extraneous bias made it difficult to compare 'like for like' depression between groups. In other words, the appearance of similar profiles when visually plotted, or with matched total scores, does not mean that latent depression is indeed the same if biasing factors exist. This problem established a clear rationale for using a methodology, such as factor analytic techniques, that would be more robust to this effect.

***Empirical Paper***

The aims of the empirical paper were three-fold: (1) to test the hypothesis that post-stroke sequelae sufficiently bias the PHQ-9 so that it measures two latent factors, (2) to identify the optimal positioning of items 7 and 8, relating to problems with concentration and slowed movement, and (3) to identify differences in factor structure with a general population group. As above, factor analysis

has the capability of capturing incidences where items are biased by other sources, via item intercepts, and measuring a latent depression while accounting for this problem.

Though a two-factor model had a superior fit to the one-factor model, the correlation between factors was sufficiently high that a clear second order of depression was indicated. This conclusion was buttressed by the near-perfect correlation of global depression on a bifactor model with depression scores from the one-factor model. This means that the PHQ-9 does not measure depression and physical disability as two uncorrelated clusters in stroke, it measures depression.

However, the presence of unidimensionality does not mean that the measure is free from the influence of the effects of stroke. While factor loadings and thresholds were broadly similar, the intercept for tiredness was significantly greater in the stroke group than the general population group, indicating higher baseline tiredness/energy complaints, independent of depression. The effects of post-stroke fatigue (Acciarresi et al., 2014) and insomnia (Baylan et al., 2020; Nguyen et al., 2019) were hypothesised explanatory factors for the intercept differences in tiredness. The intercept of the tiredness item was large enough that it biased PHQ-9 total scores to the extent that a difference of small-to-medium effect size, in the direction of primary care being more depressed, was masked. Therefore, the PHQ-9 does, indeed, appear to be biased by stroke but only for the tiredness item but not to the extent that it forces multidimensionality. A surprising finding was that the intercept for appetite was higher in the general population sample. Reasons for differences in the appetite intercept remain unclear, but age and ethnicity could be factors (de Man-van Ginkel et al., 2015).

We found evidence that concentration difficulties and slowed movement covary more with somatic symptoms of depression than cognitive-affective. This was observed in both groups, suggesting that this observation is not specific to stroke.

**Synthesis of Findings**

Combined, the findings of the two papers presented in this portfolio suggest that there are broad similarities in the phenomenology of depression after stroke compared to non-stroke general population. Profile comparison studies indicated that most symptoms, except for anhedonia and

apathy, have comparable prevalence and severity, and the factor structure of the PHQ-9 in stroke is comparable to the general population, with only the intercepts of two items differing between groups.

Partial agreement between the two papers was found for comparative correlation strengths of depressive symptoms with depression. Factor loadings, which share theoretical similarities to correlations between a symptom and a depression measure, were found to be invariant between groups. This, in turn, indicates similarities between groups in the "closeness" of each respective symptom to the depression construct. Loading non-invariance was consistent with factor correlation studies investigating insomnia, but not fatigue. The PHQ-9 does not directly measure self-esteem or anxiety, meaning these comparative correlation findings from the systematic review cannot be compared with the factor loading findings of the empirical paper. It must be considered that the non-invariance of factor loadings is only an indication of overall difference and does not provide a detailed and sensitive evaluation of differences at the symptom level.

***Extraneous Bias***

A core focus of this thesis portfolio has been to consider the relationship between post-stroke depression and numerous other post-stroke sequelae, such as physical and cognitive disability, functional impairment, personality changes, neurovascular alterations, fatigue, insomnia, and post-stroke emotionalism (Duncan, 1994; Hu et al., 2017; Teasdale and Engberg, 2010). In each of the two original papers presented in this portfolio, evidence in support of the loading of these factors onto depression items has been found. Concluding that between-group differences are the result of extraneous bias is, however, challenging because such difficulties are also likely to be relevant to the psychological maintenance of depression and potentially covary with depression severity (Broomfield et al., 2011).

Examples of suspected bias were potential interference attributable to post-stroke emotionalism and psychological adjustment in emotional variation measures (Calvert et al., 1998; Townend et al., 2010) and the interference of physical and cognitive barriers to returning to work, unaccounted age differences, or decisions to refocus priorities and take early retirement as factors

hypothesised to interfere with work inhibition items (Saeki et al., 1995; Wang et al., 2014; Wolfenden and Grace, 2009).

Somatic symptoms of depression were of particular interest and scrutiny in each respective paper. The finding of a high correlation between somatic and affective factors in the empirical study, indicating the absence of significant bias of somatic symptoms post-stroke, is consistent with broadly equal prevalence and severity of the somatic domain between groups reported in the systematic review. This indicates that somatic items are less influenced by extraneous sources than hypothesised. Exceptions were noted for fatigue/tiredness and appetite.

In the systematic review, the severity and prevalence of tiredness appeared broadly equal between groups, but fatigue was found to correlate more weakly with depression. In the empirical paper, we found evidence for a biasing of the tiredness item in the stroke group, as indicated by intercept differences, which would suggest higher scores on the tiredness item when controlling for latent depression.

A possible explanation for these findings might be that there is, indeed, a weaker association between fatigue and depression in stroke, but higher baseline fatigue leads to a cancelling out effect, leading to observations of no overall difference in prevalence or severity of fatigue between groups. An overlay of depression-related fatigue onto an elevated baseline of fatigue in stroke populations has, indeed, been observed (de Man-van Ginkel et al., 2015). This hypothesis would require an insensitivity of the measurement invariance testing to specific differences in tiredness item loadings.

Another possibility is that stroke patients exhibit significantly more variance/noise because post-stroke fatigue does not affect all stroke survivors and has varying severity in those who do experience it. The presence of elevated variance in fatigue in stroke might lead to weaker observed correlations with depression (Field, 2014), and heteroskedasticity in between-groups analyses. The relationship between fatigue and depression, and the interaction of the experiences of post-stroke fatigue in some, but not all, stroke survivors is complex, and further study is warranted to fully understand it.

Differences in the intercepts for appetite disruption were observed in the empirical paper, indicating a higher baseline prevalence of appetite disruption in the general population group, but no clear differences in severity or prevalence were observed for this symptom in the profile comparison studies. Interestingly, the single IRT DIF study included in the review suggested that appetite disruption had significantly higher latent symptom severity in the general population, meaning that positive endorsement of this item was indicative of more severe depression (Pickard et al., 2006). Synthesising these findings is challenging because a higher intercept might lead to the expectation of lower relative symptom severity, given that higher intercepts suggest higher endorsement of appetite disruption when depression is set to 0. It could be that stroke events may protect against appetite disruption in some way, perhaps through the careful regulation of dietary intake during hospital-based rehabilitation or the comorbidity of diabetes (Chen et al., 2016; Sami et al., 2017), but this remains a hypothesis. It must be noted that demographic differences existed in both the IRT DIF study and the empirical paper, so explanations extraneous to stroke status cannot be ruled out.

The presence of differences in intercepts and their consequences for the accurate estimation of overall depression severity was suspected from the results of the systematic review and confirmed by the findings of the empirical paper. It is, therefore, important that any analyses with an assumed equivalence of depression severity between stroke and the general population are interpreted with caution.

**Summary and Theoretical Relevance**

In summary, evidence for both similarities and differences in the phenomenology of depression between stroke and the general population was found. Though the groups were invariant for item thresholds and loadings, and non-significant differences in most depression symptom domains were found in the review, subtler differences in phenomenology, such as the prevalence and severity of anhedonia, have been highlighted and greater exploration of this is warranted. The biasing effects of other sequelae do appear to be relevant, but perhaps less dominating than hypothesised.

Given the adoption of a critical realist position and acknowledgement that differences in the expression of depression are inevitable between contexts, establishing a cut-off point for concluding

overall similarity or difference in phenomenology between groups would have been logically incoherent. Under the adopted perspective, differences between groups are inevitable consequences of changes in context, not as bias to be removed. Instead, the value of big-data approaches to phenomenological comparison, such as those employed in the two papers presented here, is in commenting on the nature and magnitude of any observed contextual differences and how this may support understanding of human experiences in different contexts.

When integrating current findings with findings from previous studies that compare stroke depressed to stroke non-depressed patients (de Man-van Ginkel et al., 2015; Robinson, 2006), there is evidence that, even in cases where stroke comorbidities load onto depression items, substantial item variance is captured by depression and there is a significant overlay of mood on even somatic items. It is also possible that some additional variance is explained by an interaction between depression and these comorbid factors; for example, post-stroke fatigue induced 'boom and bust' cycles leading to reduced activity, stamina, and mood (Goudsmit et al., 2012; King et al., 2020).

A finding of critical significance to researchers is that there appears to be a violation of strong invariance between the general population and our stroke sample. This indicates that total PHQ-9 scores are not comparable between groups when using the traditional summed score methods. Researchers using between-populations testing of PHQ-9 total scores are, therefore, recommended to interpret any with extreme caution.

It is hoped that this portfolio will highlight a broader range of methodological tools available to researchers for investigating this fascinating research area. It has been shown that factor analysis and IRT DIF have utilities far broader than psychometric validation, and can provide distinct phenomenological insights, such as the hierarchical positioning of latent symptom severity, the presence of intercepts, factor correlations, and the correlations/discrimination of symptoms as indicators of depression. The symptomatology of any psychological construct is far more complex than simply the relative severity and prevalence of its component symptoms, and researchers must grapple with concepts such as the theoretical "closeness" of a symptom to depressive experience to engage with the complexity. The complexity of this topic, of course, goes far deeper still and

engagement with issues of ontology and epistemology must be demonstrated. Such issues will be discussed in the following section.

## Implications for Clinical Practice

### *Assessment and Formulation*

Based on the findings of unidimensionality in the empirical paper, clinicians can be reassured that the PHQ-9 measures depression and that each item sufficiently captures depression variance and continued use of the measure is recommended. Despite recommended use, clinicians are advised to interpret cut-off points with caution, because item 4 scores, relating to tiredness could be inflated by post-stroke fatigue. This does not mean that the item or the measure is invalid as a measurement of depressed mood or to monitor intervention outcomes, only that cut-off points may be less reliable. Though an optimal cut-off of 10 has been established in stroke (de Man-van Ginkel et al., 2012a; De Man-Van Ginkel et al., 2012b), the existence of noise relating to post-stroke fatigue and other sources may result in overdiagnosis in patients with severe post-stroke fatigue and patients and underdiagnosis in those with mild or no post-stroke fatigue, using the PHQ-9. Cautious interpretation is, therefore, advised in cases where scores are close to this cut-off, particularly if this cut-off is the basis of any healthcare decision, such as the decision to prescribe or not prescribe antidepressants (Gillham et al., 2011). Of course, if accuracy of cut-off points is a priority, then another measure should be considered until further research evaluates this issue.

Clinicians are recommended to be careful not to overly attribute presentations of significant fatigue or other depression symptoms to neurological origins alone, as it is likely that high scores on, for example, the tiredness items are indicative of mood disturbance, at least to some extent, particularly if other item scores are also high. Similarly, the role of depression should not be overlooked if there is positive endorsement of the concentration difficulty item from a patient presenting with mild-to-moderate cognitive difficulty. An overlay (Bhome et al., 2019; Kay et al., 1992) or biopsychosocial (FitzGerald et al., 2012; Ormstad and Eilertsen, 2015) model of understanding and formulating a person's presentation is, therefore, recommended, because these

approaches are consistent with the evidence presented in this thesis. Further exploration of measurement invariance concerning varying levels of cognitive disability is, however, still warranted.

### *Intervention*

The interpretation that the many comorbid effects of stroke are relevant and will bring individual significance to a person, but with broadly similar symptomatology to other populations, corroborates the rationale for an augmented form of Cognitive Behavioural Therapy (CBT) after stroke (Broomfield et al., 2011). Given the potential loading of psychological adjustment processes onto items relating to emotion dysregulation (Gainotti et al., 1999; Gainotti et al., 1997; Pickard et al., 2006), our findings also buttress the importance of facilitating psychological adjustment in stroke populations (Gracey et al., 2009; Taylor et al., 2011), which has previously been recommended as part of an augmented approach (Broomfield et al., 2011).

Furthermore, techniques such as values-based Behavioural Activation (Hooker et al., 2020; Okifuji et al., 2015) and Selective Optimisation with Compensation (Baltes, 1997; Broomfield et al., 2011; Grove et al., 2009) may be appropriate interventions for patients presenting with low mood in the context of occupation or role loss (Laidlaw et al., 2008). Third-wave CBT approaches that consider personal values may also be relevant to people undergoing a period of significant personal transformation (Harris, 2009; Majumdar and Morris, 2019).

## Critical Evaluation

In this section, specific strengths and weaknesses of this research will be briefly considered, followed by an exploration of the wider issues faced by phenomenology research.

### *Strengths*

A key strength of the systematic review was its open search criteria regarding study methodology, which permitted the identification of a greater diversity of sources of evidence in examining the research questions than previous non-systematic reviews (e.g. Llorca et al., 2015). Another strength of this work was the engagement with the complexity associated with identifying evidence for and against the extraneous influence of stroke sequelae and the consideration of depression alongside

other psychological phenomena, such as emotional adjustment. The empirical paper held strengths in its novelty, its sufficient sample size and the importance of its findings for understanding phenomenology, the influence of other stroke sequelae, and the incomparability of PHQ-9 total scores between stroke and general population groups.

A core strength of this portfolio more broadly is its methodological harmony. The empirical study was able to examine and test a methodological weakness of the review; that is, concerns about whether latent depression was truly equal between groups in the review, despite similar profiles, could be addressed by the ability of factor analysis to estimate latent depression. The empirical paper indeed found that summed depression scores are incomparable between groups. The methodologies also complement each other by assessing phenomenology in two distinct ways, which supported a rich synthesis of the findings.

***Weaknesses***

The systematic review had several limitations. To obtain a sufficient focus for the scope of the project, comparison groups with other health populations and studies that compared people with post-stroke depression to people without depression after stroke were excluded. Such studies would have supported a larger sample and provided richer information about the unique variance contributed by depression. Furthermore, the high heterogeneity forced several analytical decisions, among multiple valid alternatives. These decisions are likely to have influenced findings and the number of influencing factors was likely to be disproportionately high for the available data.

The empirical paper also had several limitations. Despite a large pool of general population data, the matching process was imperfect and medium effect sizes were observed for age, nationality, and language. The presence of these differences limited the ability to attribute findings to stroke status. Furthermore, accounting for the clustered nature of the data was impractical, meaning model accuracy may have been compromised.

Both papers featured samples primarily from Western developed nations and was unable to investigate hypotheses relating to ethnicity or culture. Accordingly, the findings of this thesis may not be generalisable to populations outside of the demographics and ethnicities.

**Wider Issues in Phenomenology Research**

As outlined in the introduction, a *Critical Realist* philosophical orientation was adopted, which means that a fundamental reality of post-stroke depression was assumed, but that the social contextual nature of this reality means that observing it directly and achieving a fundamental context-independent understanding is not possible. In the two studies presented in this portfolio, the assumption that depression can be measured equally between contexts, and that depression itself is independent of context, was rejected. The attempts to investigate between-group measurement differences and differences in symptomatology exemplify this stance.

By demonstrating the contextual dependence of symptomatologic experiences, such as through the observation that indicators of anhedonia severity/prevalence differed between groups, the existence of depression as a fundamental construct, independent of human context, has therefore been rejected. The validity of the assumption that depression can be measured in the same way for all people in all contexts has, therefore, also been rejected. This rejection was made despite the common assumption that, for example, diagnostic interviews are considered the "gold standard" of depression diagnosis in stroke and other health populations (de Man-van Ginkel et al., 2015; Gainotti et al., 1999; Gainotti et al., 1997; House et al., 1991) and that questionnaires can be validated simply by testing their agreement with said standard (Burton and Tyson, 2015). Indeed, because the findings of the systematic review were dependent on papers that made this assumption, the review also possessed this limitation. Factor analytic and IRT tools are important to phenomenology research for those adopting a contextualist approach because they support the testing of assumptions of equivalence of a broader construct by *deconstructing* it into its component symptoms and exploring differences between human contexts. The implications of this thesis were, therefore, directed at those who adopt a general acceptance of the depression construct, but acknowledge that there is a need to refine how it is understood between contexts. For example, this perspective has been reflected in cognitive-

behavioural models of depression in older people, which acknowledge contextual differences inherent in this population (Laidlaw et al., 2008). Attempts in this thesis, therefore, to compare groups whose reports indicate similar overall distress experiences, so that context-related differences in the experiences within this broader construct could be identified, are viewed as consistent with this perspective.

Despite the critical stance regarding depression as a context-independent construct adopted here, an acceptance of the existence of some form of construct of depressive experience has still been assumed. The methodologies utilised in the research presented assumed that there exist covarying clusters of depression symptoms and that there is, therefore, likely a biopsychosocial reality to the phenomenon of depression (Alexopoulos et al., 1997; Castanheira et al., 2019; Zeng et al., 2021).

Though the *Critical Realist* stance adopted in this portfolio embraced variation in the presentation of depression between contexts, this raises the issue of the arbitrary point in which the variations in expression become so large that the term "depression" becomes broad to the point of meaninglessness. Indeed, many psychotherapeutic models reject the existence of depression as a diagnostic entity and support the notion that depressive symptoms are idiosyncratic, person-specific, and transcend the bounds of any depression diagnostic category (Johnstone, 2018; Johnstone and Dallos, 2006). For example, an influential model of low self-esteem offers a transdiagnostic mechanism by which depressive and anxious symptoms interact (Fennell, 1997). In such a model, the concept of depression is de-emphasised in favour of a formulation-driven model for the maintenance of distress, with no clear boundary between when depression ends and other difficulties start. Within this perspective, attempts to control for overall depression severity between contexts, as was present in the systematic review and empirical paper, are undermined in validity because the set of symptoms used to identify group-level differences is arbitrary.

Similarly, a stance that there are clear and well-defined bounds between when stroke sequelae might be relevant and intertwined with depressive experience and when their loading onto depression items is purely extraneous has been assumed. It is probable that, in reality, there is a fuzzy boundary between these two positions and that there are complex and obfuscated interactions between these

numerous effects of stroke, the psychological internalisation of them, and the interface between them and the depression measure (Broomfield et al., 2011). Even in cases where there was no observation of bias or item loading from stroke sequelae, it is possible that these effects existed but remained unobserved, because of epistemological limitations. For example, if post-stroke fatigue interacts with physical disability (Cuesta, 2013), negative cognitions (Telfer, 2014), interpersonal attachment style (Bifulco et al., 2002; Li et al., 2008), personality traits (Klein et al., 2011), coping mechanisms (Visser et al., 2015), caregiver support and coping (Boerboom et al., 2014; Jaracz et al., 2012), cultural attitudes to disability and mental health (Sheikh and Furnham, 2000), cohort beliefs (Laidlaw et al., 2008), rapport with the assessor (Thompson and McCabe, 2012), response biases (Wetzel et al., 2016), and the context of hospital rehabilitation (Walker et al., 2013), it would not be unexpected for this complex system of exchanges to remain undetected by a simple between-groups symptom prevalence comparison, factor loading, or item intercept. Thus, it is acknowledged that the attempt to make clear distinctions between genuine population-level differences in phenomenology versus differences that result purely from extraneous bias should be considered a limitation.

This perspective, of course, has implications for the meaningfulness of the methodology in the systematic review and empirical paper; if the separation of somatic distress, such as problems with fatigue, from psychological distress, such as guilt, is meaningless, then the concepts of depression being *overestimated* by extraneous factors in stroke were also meaningless. Instead, higher average fatigue scores could be viewed as a distressing experience that itself can be formulated with the individual and, regardless of whether this experience mechanistically interacts with other depressive symptoms, may benefit from focus of psychological support.

Finally, if this research were to be critically appraised through a relativistic or social constructionist lens, as is more commonly adopted in qualitative research (Welch and Patton, 1992), the comparison between groups at even the ***symptom level***, a step closer to relativism than investigations of the disorder level, could be considered a limitation. Qualitative research affords rich narratives into the phenomenology of person-centred experiences, which is lost by quantitative attempts at measurement (Braun and Clarke, 2013). Those adopting social constructionist perspectives

might argue that there are limitations to comparisons of broadly defined symptoms between groups if the fundamental way in which these symptoms are understood and constructed is specific to an individual's context, and dependent on individual, group, cultural, and historical factors. For example, self-blame or guilt in the context of requiring physical health care from family members, which can be mediated by cultural concepts of familism, collectivism, and traditions of care responsibility (Crowe et al., 2016; Knight et al., 2002; Losada et al., 2007), may be seen by some as fundamentally incomparable to guilt experienced in the context of adult mental health and complex attachment histories (Unthank, 2019). Attempts to measure guilt in such different contexts are, therefore, arguably invalid if these experiences, and the ways in which individuals between contexts interact with any form of measurement, are different. This, of course, would mean that 'controlling' for overall severity in order to measure differences in specific symptoms, as performed in this thesis, is also invalid.

Furthermore, the low ethnic and socioeconomic diversity of the samples included in this portfolio and absence of specific analysis exploring the role of ethnicity may have resulted in findings that are potentially incompatible with, or even damaging to, individuals from cultures that construct depression or problems with low mood in different ways and should therefore be applied to different cultures with caution. While the methodologies in this portfolio have been adopted to try and bridge this gap between broad and arbitrary diagnostic categories and the richness of individual phenomenological narratives by understanding symptom-level relationships, this gap remains large and epistemologically limited.

**Future Research**

Many questions remain unanswered in post-stroke depression phenomenology research. Regarding profile comparison studies, we have highlighted a greater need for the use of more longitudinal studies in symptomatology, so that effects of stroke recovery and adjustment can be better tested (House et al., 1991). Preferably, this study would use a measure with a large diversity of items, such as the Beck Depression Inventory-II (BDI-II; Beck and Steer, 1984), and compare profiles of

depressed and non-depressed stroke and general population participants, similar to de Man-van Ginkel et al. (2015).

It is also important that further research aims to compare associations of symptoms with the latent construct of depression and investigates differences in the latent severity of those symptoms as indicators. Item Response Theory (IRT) modelling can accomplish both aims, via the assessment of item discrimination and the analysis of DIF (Pickard et al., 2006). So far, this aspect of phenomenology research in stroke is greatly underutilised.

The factor analysis methodology used in the empirical paper has been available for less than a decade (Wu and Estabrook, 2016), and development in the area of measurement invariance testing is ongoing. As advancements are made, a replication of this empirical paper using multilevel modelling, and with a greater statistical understanding of the accuracy of fit statistics in the context of categorical variables, will be required (Sass et al., 2014). Furthermore, there remains a need to combine a between-groups measurement variance approach of stroke and non-stroke, as has been completed here, with a longitudinal within-subjects approach of assessing measurement invariance with increased time since the index stroke (Dong et al., 2022). This combined approach would offer greater insights into reasons behind findings of measurement non-invariance and reduced unidimensional model fit with increased time since stroke.

Regarding particular symptoms, unexpected findings were reported for tiredness and appetite disruption. Hypotheses about the differences in tiredness and appetite intercepts, in the context of an absence of differences in the somatic domain reported in the systematic review, warrant careful investigation so that a mechanism can be identified. For example, an exploration of whether neurologically implied post-stroke fatigue accounts for differences in intercepts, and whether there is measurement non-invariance between post-stroke fatigue groups, is recommended.

Group differences in indicators of emotional expression and tearfulness (Gainotti et al., 1999; Gainotti et al., 1997; Pickard et al., 2006) should also be explored to identify the differential contribution of emotionalism (Calvert et al., 1998), depression (Vingerhoets et al., 2007), and

psychological adjustment to loss (Townend, 2005). Recent developments in the assessment of post-stroke emotionalism (Broomfield et al., 2021), and the use of longitudinal designs (House et al., 1991), will support such differentiation.

**Conclusion**

This thesis portfolio aimed to consolidate existing research into the comparative phenomenology of post-stroke depression and contribute new understanding by applying methodologies that aim to address previous research limitations. This portfolio has found evidence in support of a broad similarity of the symptomatology of post-stroke depression with depression in the general population. There is some evidence that post-stroke depression is characterized by comparatively less severe and prevalent experiences of anhedonia. Contrary to expectation, assessment of depression in stroke, using a commonly utilised screening tool, was robust to measurement interference attributable to the numerous stroke sequelae and demonstrated unidimensionality. However, there does appear to be moderate bias in the assessment of tiredness/fatigue, emotional variability, and work inhibition attributable to post-stroke fatigue, post-stroke emotionalism, psychological adjustment, and physical and cognitive impairment. Avenues for future research are outlined.

## References

Acciarresi, M., Bogousslavsky, J., & Paciaroni, M. (2014). Post-Stroke Fatigue: Epidemiology, Clinical Characteristics and Treatment. *European Neurology*, *72*(5–6), 255–261. https://doi.org/10.1159/000363763

Adams, K. B., Matto, H. C., & Sanders, S. (2004). Confirmatory factor analysis of the geriatric depression scale. *Gerontologist*, *44*(6), 818–826. https://doi.org/10.1093/geront/44.6.818

Aizenstein, H. J., Baskys, A., Boldrini, M., Butters, M. A., Diniz, B. S., Jaiswal, M. K., Jellinger, K. A., Kruglov, L. S., Meshandin, I. A., Mijajlovic, M. D., Niklewski, G., Pospos, S., Raju, K., Richter, K., Steffens, D. C., Taylor, W. D., & Tene, O. (2016). Vascular depression consensus report – a critical update. *BMC Medicine 2016 14:1*, *14*(1), 1–16. https://doi.org/10.1186/S12916-016-0720-5

Aizenstein, H. J., Khalaf, A., Walker, S. E., & Andreescu, C. (2014). Magnetic resonance imaging predictors of treatment response in late-life depression. *Journal of Geriatric Psychiatry and Neurology*, *27*(1), 24–32. https://doi.org/10.1177/0891988713516541

Alexopoulos, G. S., Meyers, B. S., Young, R. C., Campbell, S., Silbersweig, D., & Charlson, M. (1997). "Vascular depression" hypothesis. In *Archives of General Psychiatry* (Vol. 54, Issue 10, pp. 915–922). Arch Gen Psychiatry. https://doi.org/10.1001/archpsyc.1997.01830220033006

Andersen, G., Vestergaard, K., & Ingeman-Nielsen, M. (1995). Post-stroke pathological crying: frequency and correlation to depression. *European Journal of Neurology*, *2*(1), 45–50. https://doi.org/10.1111/J.1468-1331.1995.TB00092.X

Ang, Y. S., Lockwood, P., Apps, M. A. J., Muhammed, K., & Husain, M. (2017). Distinct Subtypes of Apathy Revealed by the Apathy Motivation Index. *PLOS ONE*, *12*(1), e0169938. https://doi.org/10.1371/JOURNAL.PONE.0169938

Annis, D. B. (1978). A Contextualist Theory of Epistemic Justification. *American Philosophical Quarterly*, *15*(3). https://philpapers.org/rec/ANNACT

Baltes, P. B. (1997). On the Incomplete Architecture of Human Ontogeny: Selection, Optimization, and Compensation as Foundation of Developmental Theory. *American Psychologist*, *52*(4), 366–380. https://doi.org/10.1037/0003-066X.52.4.366

Baylan, S., Griffiths, S., Grant, N., Broomfield, N. M., Evans, J. J., & Gardani, M. (2020). Incidence and prevalence of post-stroke insomnia: A systematic review and meta-analysis. In *Sleep Medicine Reviews* (Vol. 49, p. 101222). W.B. Saunders Ltd. https://doi.org/10.1016/j.smrv.2019.101222

Beck, A. T. (1979). *Cognitive Therapy of Depression*. Guildford Press, New York. https://books.google.com.sg/books?id=L09cRS0xWj0C&printsec=frontcover&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false

Beck, Aaron T., Epstein, N., Brown, G., & Steer, R. A. (1988). An Inventory for Measuring Clinical Anxiety: Psychometric Properties. *Journal of Consulting and Clinical Psychology*, *56*(6), 893–897. https://doi.org/10.1037/0022-006X.56.6.893

Beck, Aaron T., & Steer, R. A. (1984). Internal consistencies of the original and revised beck depression inventory. *Journal of Clinical Psychology*, *40*(6), 1365–1367. https://doi.org/10.1002/1097-4679(198411)40:6<1365::AID-JCLP2270400615>3.0.CO;2-D

Bell, C. C. (1994). DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. *JAMA: The Journal of the American Medical Association*, *272*(10), 828. https://doi.org/10.1001/jama.1994.03520100096046

Bennett, H. E., Thomas, S. A., Austen, R., Morris, A. M. S., & Lincoln, N. B. (2006). Validation of screening measures for assessing mood in stroke patients. *British Journal of Clinical Psychology*, *45*(3), 367–376. https://doi.org/10.1348/014466505X58277

Bernier, J. (2009). *Article Validation of disability categories derived from Health Utilities Index Mark 3 scores and Heather Orpana*. *July*.

Bhome, R., Huntley, J. D., Price, G., & Howard, R. J. (2019). Clinical presentation and neuropsychological profiles of Functional Cognitive Disorder patients with and without co-morbid depression. *Cognitive Neuropsychiatry*, *24*(2), 152–164. https://doi.org/10.1080/13546805.2019.1590190/SUPPL_FILE/PCNP_A_1590190_SM4055.DOCX

Bifulco, A., Moran, P. M., Ball, C., & Bernazzani, O. (2002). Adult attachment style. I: Its relationship to clinical depression. *Social Psychiatry and Psychiatric Epidemiology 2002 37:2*,

*37*(2), 50–59. https://doi.org/10.1007/S127-002-8215-0

Boerboom, W., Jacobs, E. A. C., Khajeh, L., van Kooten, F., Ribbers, G. M., & Heijenbrok-Kal, M. H. (2014). The relationship of coping style with depression, burden, and life dissatisfaction in caregivers of patients with subarachnoid haemorrhage. *Journal of Rehabilitation Medicine*, *46*(4), 321–326. https://doi.org/10.2340/16501977-1273

Boothroyd, L., Dagnan, D., & Muncer, S. (2019). PHQ-9: One factor or two? *Psychiatry Research*, *271*, 532–534. https://doi.org/10.1016/j.psychres.2018.12.048

Braun, V., & Clarke, V. (2013). Some very important starting information. In *Successful Qualitative Research a practical guide for beginners* (p. 400). SAGE Publications. http://www.uk.sagepub.com/books/Book233059

Broomfield, N. M., Laidlaw, K., Hickabottom, E., Murray, M. F., Pendrey, R., Whittick, J. E., & Gillespie, D. C. (2011a). Post-stroke depression: the case for augmented, individually tailored cognitive behavioural therapy. *Clinical Psychology & Psychotherapy*, *18*(3), 202–217. https://doi.org/10.1002/cpp.711

Broomfield, N. M., Laidlaw, K., Hickabottom, E., Murray, M. F., Pendrey, R., Whittick, J. E., & Gillespie, D. C. (2011b). Post-stroke depression: The case for augmented, individually tailored cognitive behavioural therapy. *Clinical Psychology and Psychotherapy*, *18*(3), 202–217. https://doi.org/10.1002/cpp.711

Broomfield, N. M., West, R., House, A., Munyombwe, T., Barber, M., Gracey, F., Gillespie, D. C., & Walters, M. (2021). Psychometric evaluation of a newly developed measure of emotionalism after stroke (TEARS-Q). *Clinical Rehabilitation*, *35*(6), 894–903. https://doi.org/10.1177/0269215520981727

Brumfitt, S. M., & Sheeran, P. (1999). The development and validation of the visual analogue self-esteem scale (VASES). *British Journal of Clinical Psychology*, *38*(4), 387–400. https://doi.org/10.1348/014466599162980

Burton, L.-J. J., & Tyson, S. (2015). Screening for mood disorders after stroke: A systematic review of psychometric properties and clinical utility. *Psychological Medicine*, *45*(1), 29–49. https://doi.org/10.1017/S0033291714000336

Burton, L. J., Tyson, S., & McGovern, A. (2013). Staff perceptions of using outcome measures in stroke rehabilitation. *Disability and Rehabilitation*, *35*(10), 828–834. https://doi.org/10.3109/09638288.2012.709305

Burvill, P., Johnson, G., Jamrozik, K., Anderson, C., & Stewart-Wynne, E. (1997). Risk factors for post-stroke depression. *International Journal of Geriatric Psychiatry*, *12*(2), 219–226. https://doi.org/10.1002/(sici)1099-1166(199702)12:2<219::aid-gps581>3.0.co;2-e

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Calvert, T., Knapp, P., & House, A. (1998). Psychological associations with emotionalism after stroke. *Journal of Neurology Neurosurgery and Psychiatry*, *65*(6), 928–929. https://doi.org/10.1136/jnnp.65.6.928

Castanheira, L., Silva, C., Cheniaux, E., & Telles-Correia, D. (2019). Neuroimaging Correlates of Depression—Implications to Clinical Practice. *Frontiers in Psychiatry*, *10*, 703. https://doi.org/10.3389/FPSYT.2019.00703/BIBTEX

Chen, R., Ovbiagele, B., & Feng, W. (2016). Diabetes and Stroke: Epidemiology, Pathophysiology, Pharmaceuticals and Outcomes. *The American Journal of the Medical Sciences*, *351*(4), 380. https://doi.org/10.1016/J.AMJMS.2016.01.011

Chilcot, J., Rayner, L., Lee, W., Price, A., Goodwin, L., Monroe, B., Sykes, N., Hansford, P., & Hotopf, M. (2013). The factor structure of the PHQ-9 in palliative care. *Journal of Psychosomatic Research*, *75*(1), 60–64. https://doi.org/10.1016/j.jpsychores.2012.12.012

Christensen, D., Johnsen, S. P., Watt, T., Harder, I., Kirkevold, M., & Andersen, G. (2008). Dimensions of post-stroke fatigue: a two-year follow-up study. *Cerebrovascular Diseases (Basel, Switzerland)*, *26*(2), 134–141. https://doi.org/10.1159/000139660

Clara, I. P., Cox, B. J., & Enns, M. W. (2001). Confirmatory Factor Analysis of the Depression-Anxiety-Stress Scales in Depressed and Anxious Patients. *Journal of Psychopathology and Behavioral Assessment*, *23*(1), 61–67. https://doi.org/10.1023/A:1011095624717

Cooper, J. E. (1985). The Development of the Present State Examination (P.S.E.). *Clinical*

*Psychopathology Nomenclature and Classification*, 133–139. https://doi.org/10.1007/978-1-4899-5049-9_20

Crowe, C., Coen, R. F., Kidd, N., Hevey, D., Cooney, J., & Harbison, J. (2016). A qualitative study of the experience of psychological distress post-stroke. *Journal of Health Psychology*, *21*(11), 2572–2579. https://doi.org/10.1177/1359105315581067

Cuesta, G. M. (2013). Assessment in acute stroke rehabilitation. In L. D. Ravdin & H. L. Katzen (Eds.), *Handbook on the neuropsychology of aging and dementia.* (pp. 295–312). Springer Science + Business Media. https://doi.org/10.1007/978-1-4614-3106-0_20

Cumming, T. B., Churilov, L., Skoog, I., Blomstrand, C., & Linden, T. (2010). Little evidence for different phenomenology in poststroke depression. *Acta Psychiatrica Scandinavica*, *121*(6), 424–430. https://doi.org/10.1111/j.1600-0447.2010.01558.x

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. In *Annual Review of Sociology* (Vol. 40, pp. 55–75). Annual Reviews. https://doi.org/10.1146/annurev-soc-071913-043137

De Coster, L., Leentjens, A. F. G., Lodder, J., & Verhey, F. R. J. (2005). The sensitivity of somatic symptoms in post-stroke depression: a discriminant analytic approach. *International Journal of Geriatric Psychiatry*, *20*(4), 358–362. http://10.0.3.234/gps.1290

De Man-Van Ginkel, J. M., Gooskens, F., Schepers, V. P. M., Schuurmans, M. J., Lindeman, E., & Hafsteinsdóttir, T. B. (2012). Screening for poststroke depression using the patient health questionnaire. *Nursing Research*, *61*(5), 333–341. https://doi.org/10.1097/NNR.0B013E31825D9E9E

De Man-Van Ginkel, J. M., Gooskens, F., Schepers, V. P. M., Schuurmans, M. J., Lindeman, E., Hafsteinsdóttir, T. B., JM, de M. G., Gooskens, F., VP, S., MJ, S., Lindeman, E., TB, H., De Man-Van Ginkel, J. M., Gooskens, F., Schepers, V. P. M., Schuurmans, M. J., Lindeman, E., & Hafsteinsdóttir, T. B. (2012). Screening for poststroke depression using the Patient Health Questionnaire. *Nursing Research*, *61*(5), 333–341. https://doi.org/10.1097/NNR.0b013e31825d9e9e

de Man-van Ginkel, J. M., Hafsteinsdóttir, T. B., Lindeman, E., Geerlings, M. I., Grobbee, D. E., &

Schuurmans, M. J. (2015). Clinical manifestation of depression after stroke: Is it different from depression in other patient populations? *PLoS ONE*, *10*(12). https://doi.org/10.1371/journal.pone.0144450

de Man-van Ginkel, J. M., Hafsteinsdóttir, T., Lindeman, E., Burger, H., Grobbee, D., Schuurmans, M., JM, de M. G., Hafsteinsdóttir, T., Lindeman, E., Burger, H., Grobbee, D., Schuurmans, M., de Man-van Ginkel, J. M., Hafsteinsdóttir, T., Lindeman, E., Burger, H., Grobbee, D., & Schuurmans, M. (2012). An efficient way to detect poststroke depression by subsequent administration of a 9-item and a 2-item Patient Health Questionnaire. *Stroke (00392499)*, *43*(3), 854–856. https://doi.org/10.1161/STROKEAHA.111.640276

Deng, L., Sun, X., Qiu, S., Xiong, Y., Li, Y., Wang, L., Wei, Q., Wang, D., & Liu, M. (2017). Interventions for management of post-stroke depression: A Bayesian network meta-analysis of 23 randomized controlled trials. *Scientific Reports 2017 7:1*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-16663-0

Desmond, D. W., Remien, R. H., Moroney, J. T., Stern, Y., Sano, M., & Williams, J. B. W. (2003). Ischemic stroke and depression. *Journal of the International Neuropsychological Society : JINS*, *9*(3), 429–439. https://doi.org/10.1017/S1355617703930086

Dong, L., Williams, L. S., Briceno, E., Morgenstern, L. B., & Lisabeth, L. D. (2022). Longitudinal assessment of depression during the first year after stroke: Dimensionality and measurement invariance. *Journal of Psychosomatic Research*, *153*, 110689. https://doi.org/10.1016/J.JPSYCHORES.2021.110689

Douiri, A., Rudd, A. G., & Wolfe, C. D. A. (2013). Prevalence of poststroke cognitive impairment: South London stroke register 1995-2010. *Stroke*, *44*(1), 138–145. https://doi.org/10.1161/STROKEAHA.112.670844

Dowswell, G., Lawler, J., Dowswell, T., Young, J., Forster, A., & Hearn, J. (2000). Investigating recovery from stroke: A qualitative study. In *Journal of Clinical Nursing* (Vol. 9, Issue 4). https://doi.org/10.1046/j.1365-2702.2000.00411.x

Duncan, P. W. (1994). Stroke Disability. *Physical Therapy*, *74*(5), 399–407. https://doi.org/10.1093/PTJ/74.5.399

Duncan, P. W., Wallace, D., Lai, S. M., Johnson, D., Embretson, S., & Laster, L. J. (1999). The Stroke Impact Scale Version 20: Evaluation of reliability, validity, and sensitivity to change. *Stroke*, *30*(10), 2131–2140. https://doi.org/10.1161/01.STR.30.10.2131

Dunn, K. J., & McCray, G. (2020). The Place of the Bifactor Model in Confirmatory Factor Analysis Investigations Into Construct Dimensionality in Language Testing. *Frontiers in Psychology*, *11*, 1357. https://doi.org/10.3389/fpsyg.2020.01357

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*(1), 149–167. https://doi.org/10.1016/J.LEAQUA.2004.09.009

Elhai, J. D., Contractor, A. A., Tamburrino, M., Fine, T. H., Prescott, M. R., Shirley, E., Chan, P. K., Slembarski, R., Liberzon, I., Galea, S., & Calabrese, J. R. (2012). The factor structure of major depression symptoms: A test of four competing models using the Patient Health Questionnaire-9. *Psychiatry Research*, *199*(3), 169–173. https://doi.org/10.1016/j.psychres.2012.05.018

Ellis, C. J. (2014). Measurement issues in poststroke depression. *Journal of Neuroscience Nursing*, *46*(4), 198. https://doi.org/10.1097/JNN.0000000000000070

Espárrago Llorca, G., Castilla-Guerra, L., Fernández Moreno, M. C., Ruiz Doblado, S., & Jiménez Hernández, M. D. (2015). Post-stroke depression: an update. *Neurologia (Barcelona, Spain)*, *30*(1), 23–31. https://doi.org/10.1016/j.nrl.2012.06.008

Espie, C. A., Kyle, S. D., Hames, P., Gardani, M., Fleming, L., & Cape, J. (2014). The Sleep Condition Indicator: A clinical screening tool to evaluate insomnia disorder. *BMJ Open*, *4*(3). https://doi.org/10.1136/bmjopen-2013-004183

Fenn, K., & Byrne, M. (2013). The key principles of cognitive behavioural therapy. *InnovAiT: Education and Inspiration for General Practice*, *6*(9), 579–585. https://doi.org/10.1177/1755738012471029

Fennell, M. J. V. (2005). *Low Self-Esteem BT - Encyclopedia of Cognitive Behavior Therapy* (A. Freeman, S. H. Felgoise, C. M. Nezu, A. M. Nezu, & M. A. Reinecke (Eds.); pp. 236–240). Springer US. https://doi.org/10.1007/0-306-48581-8_66

Field, A. (2014). *Discovering Statistics Using IBM SPSS Statistics*. SAGE Publications.

http://www.lavoisier.fr/notice/frMJO3AKKA6RIXLO.html%5Cnpapers2://publication/uuid/01872072-935D-4D89-AB3D-D7F41393EC7D

Finch, W. H., & French, B. F. (2015). Latent variable modeling with R. *Latent Variable Modeling with R*, 1–321. https://doi.org/10.4324/9781315869797/LATENT-VARIABLE-MODELING-HOLMES-FINCH-BRIAN-FRENCH

Fischer, F., Gibbons, C., Coste, J., Valderas, J. M., Rose, M., & Leplège, A. (2018). Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. *Quality of Life Research*, *27*(4), 999–1014. https://doi.org/10.1007/s11136-018-1785-8

Fischer, F., Levis, B., Falk, C., Sun, Y., Ioannidis, J. P. A., Cuijpers, P., Shrier, I., Benedetti, A., & Thombs, B. D. (2021). Comparison of different scoring methods based on latent variable models of the PHQ-9: an individual participant data meta-analysis. *Psychological Medicine*, 1–12. https://doi.org/10.1017/s0033291721000131

FitzGerald, M. C. C., Carton, S., O'Keeffe, F., Coen, R. F., & Dockree, P. M. (2012). Impaired self-awareness following acquired brain injury: Current theory, models and anatomical understanding. *Irish Journal of Psychology*, *33*(2–3), 78–85. https://doi.org/10.1080/03033910.2012.708898

Fitzgerald, S. (2021). *Post-stroke emotionalism (PSE): Making sense of neurological, psychological and experiential factors*.

Fitzgerald, S., Gracey, F., & Broomfield, N. (2021). Post-stroke emotionalism (PSE): a qualitative longitudinal study exploring individuals' experience with PSE. *Disability and Rehabilitation*. https://doi.org/10.1080/09638288.2021.2002439

Fleming, M. K., Smejka, T., Henderson Slater, D., Chiu, E. G., Demeyere, N., & Johansen-Berg, H. (2021). Self-Reported and Objective Sleep Measures in Stroke Survivors With Incomplete Motor Recovery at the Chronic Stage. *Neurorehabilitation and Neural Repair*, *35*(10), 851–860. https://doi.org/10.1177/15459683211029889

Flowers, H. L., Silver, F. L., Fang, J., Rochon, E., & Martino, R. (2013). The incidence, co-occurrence, and predictors of dysphagia, dysarthria, and aphasia after first-ever acute ischemic

stroke. *Journal of Communication Disorders*, *46*(3), 238–248. https://doi.org/10.1016/j.jcomdis.2013.04.001

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor Analysis with Ordinal Indicators: A Monte Carlo Study Comparing DWLS and ULS Estimation. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 625–641. https://doi.org/10.1080/10705510903203573

Fydrich, T., Dowdall, D., & Chambless, D. L. (1992). Reliability and validity of the beck anxiety inventory. *Journal of Anxiety Disorders*, *6*(1), 55–61. https://doi.org/10.1016/0887-6185(92)90026-4

Gainotti, G., Azzoni, A., & Marra, C. (1999). Frequency, phenomenology and anatomical-clinical correlates of major post-stroke depression. *British Journal of Psychiatry*, *175*(AUG.), 163–167. https://doi.org/10.1192/bjp.175.2.163

Gainotti, G, Azzoni, A., Razzano, C., Lanzillotta, M., Marra, C., & Gasparini, F. (1997). The Post-Stroke Depression Rating Scale: a test specifically devised to investigate affective disorders of stroke patients. *Journal of Clinical and Experimental Neuropsychology*, *19*(3), 340–356. https://doi.org/10.1080/01688639708403863

Gainotti, Guido, Azzoni, A., Razzano, C., Lanzillotta, M., Marra, C., & Gasparini, F. (1997). The Post-Stroke Depression Rating Scale: a test specifically devised to investigate affective disorders of stroke patients. *Journal of Clinical and Experimental Neuropsychology*, *19*(3), 340–356. https://doi.org/10.1080/01688639708403863

Gillham, S., Clark, L., Blake, N., Carey, T., Hewitt, A., Holden, K., House, A., Knapp, P., Kneebone, I., Knights, P., Lincoln, N., Macniven, J., Reich, S., Swinburn, K., & Watkins, C. (2011). *Psychological care after stroke: Improving stroke services for people with cognitive and mood disorders*. www.improvement.nhs.uk/stroke

González-Blanch, C., Medrano, L. A., Muñoz-Navarro, R., Ruíz-Rodríguez, P., Moriana, J. A., Limonero, J. T., Schmitz, F., & Cano-Vindel, A. (2018). Factor structure and measurement invariance across various demographic groups and over time for the PHQ-9 in primary care patients in Spain. *PLoS ONE*, *13*(2). https://doi.org/10.1371/journal.pone.0193356

Goudsmit, E. M., Nijs, J., Jason, L. A., & Wallman, K. E. (2012). Disability and Rehabilitation Pacing as a strategy to improve energy management in myalgic encephalomyelitis/ chronic fatigue syndrome: a consensus document. *Disability & Rehabilitation*, *34*(13), 1140–1147. https://doi.org/10.3109/09638288.2011.635746

Gracey, F., Evans, J. J., & Malley, D. (2009). Capturing process and outcome in complex rehabilitation interventions: A "Y-shaped" model. *Neuropsychological Rehabilitation*, *19*(6), 867–890. https://doi.org/10.1080/09602010903027763

Grasso, M. G., Pantano, P., & Intiso, D. (1994). *Mesial temporal cortex hypoperfusion is associated with depression in subcortical stroke*. https://doi.org/10.1161/01.STR.25.5.980

Grove, L. J., Loeb, S. J., & Penrod, J. (2009). Selective optimization with compensation: A model for elder health programming. *Clinical Nurse Specialist*, *23*(1), 25–32. https://doi.org/10.1097/01.NUR.0000343080.57838.2F

Hackett, M. L., & Anderson, C. S. (2005). Predictors of Depression after Stroke. *Stroke*, *36*(10), 2296–2301. https://doi.org/10.1161/01.STR.0000183622.75135.A4

Hackett, M. L., & Pickles, K. (2014). Part I: Frequency of depression after stroke: An updated systematic review and meta-analysis of observational studies. *International Journal of Stroke*, *9*(8), 1017–1025. https://doi.org/10.1111/ijs.12357

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, *23*(1), 56–62. https://doi.org/10.1136/jnnp.23.1.56

Harris, R. (2009). *ACT made simple an easy-to-read primer on acceptance and commitment therapy*. New Harbinger Publications, Inc. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=421259

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, *42*(8), 1–28. https://doi.org/10.18637/JSS.V042.I08

Hobfoll, S. E., Canetti, D., Hall, B. J., Brom, D., Palmieri, P. A., Johnson, R. J., Pat-Horenczyk, R., & Galea, S. (2011). Are Community Studies of Psychological Trauma's Impact Accurate? A Study

Among Jews and Palestinians. *Psychological Assessment*, *23*(3), 599–605. https://doi.org/10.1037/A0022817

Hooker, S. A., Slattengren, A. H., Boyle, L., & Sherman, M. D. (2020). Values-Based Behavioral Activation for Chronic Pain in Primary Care: A Pilot Study. *Journal of Clinical Psychology in Medical Settings*, *27*(4), 633–642. https://doi.org/10.1007/S10880-019-09655-X/TABLES/2

House, A., Dennis, M., Mogridge, L., Warlow, C., Hawton, K., & Jones, L. (1991). Mood disorders in the year after first stroke. *British Journal of Psychiatry*, *158*(JAN.), 83–92. https://doi.org/10.1192/bjp.158.1.83

House, A., Dennis, M., Molyneux, A., Warlow, C., & Hawton, K. (1989). Emotionalism after stroke. *BMJ (Clinical Research Ed.)*, *298*(6679), 991–994. https://doi.org/10.1136/bmj.298.6679.991

Hu, X., Silva, T. M. De, Chen, J., & Faraci, F. M. (2017). Cerebral Vascular Disease and Neurovascular Injury in Ischemic Stroke. *Circulation Research*, *120*(3), 449–471. https://doi.org/10.1161/CIRCRESAHA.116.308427

Huang, F. Y., Chung, H., Kroenke, K., Delucchi, K. L., & Spitzer, R. L. (2006). Using the patient health questionnaire-9 to measure depression among racially and ethnically diverse primary care patients. *Journal of General Internal Medicine*, *21*(6), 547–552. https://doi.org/10.1111/j.1525-1497.2006.00409.x

Hunt, G. E., O'Hara-Aarons, M., O'Connor, N., & Cleary, M. (2012). Why are some patients admitted to psychiatric hospital while others are not? A study assessing risk during the admission interview and relationship to outcome. *International Journal of Mental Health Nursing*, *21*(2), 145–153. https://doi.org/10.1111/J.1447-0349.2011.00778.X

Janssen, E. P. C. J., Köhler, S., Stehouwer, C. D. A., Schaper, N. C., Dagnelie, P. C., Sep, S. J. S., Henry, R. M. A., van der Kallen, C. J. H., Verhey, F. R., & Schram, M. T. (2016). The Patient Health Questionnaire-9 as a Screening Tool for Depression in Individuals with Type 2 Diabetes Mellitus: The Maastricht Study. *Journal of the American Geriatrics Society*, *64*(11), e201–e206. https://doi.org/10.1111/JGS.14388

Jaracz, K., Grabowska-Fudala, B., & Kozubski, W. (2012). Caregiver burden after stroke: towards a structural model. *Neurologia i Neurochirurgia Polska*, *46*(3), 224–232.

https://doi.org/10.5114/ninp.2012.29130

Jaywant, A., DelPonte, L., Kanellopoulos, D., O'Dell, M. W., & Gunning, F. M. (2022). The Structural and Functional Neuroanatomy of Post-Stroke Depression and Executive Dysfunction: A Review of Neuroimaging Findings and Implications for Treatment. In *Journal of Geriatric Psychiatry and Neurology* (Vol. 35, Issue 1, pp. 3–11). https://doi.org/10.1177/0891988720968270

Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample Size Requirements for Estimation of Item Parameters in the Multidimensional Graded Response Model. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.00109

Johnstone, L. (2018). Psychological Formulation as an Alternative to Psychiatric Diagnosis. *Journal of Humanistic Psychology*, *58*(1), 30–46. https://doi.org/10.1177/0022167817722230

Johnstone, L., & Dallos, R. (2006). *Formulation in psychology and psychotherapy : making sense of people's problems*. 241.

Jorge, R. E., Starkstein, S. E., & Robinson, R. G. (2010). Apathy following stroke. *The Canadian Journal of Psychiatry / La Revue Canadienne de Psychiatrie*, *55*(6), 350–354. https://doi.org/10.1177/070674371005500603

Juhasz, G., Eszlari, N., Pap, D., & Gonda, X. (2012). Cultural differences in the development and characteristics of depression. *Neuropsychopharmacol Hung*, *14*(4), 259–265. https://doi.org/10.5706/nph201212007

Kang, H. (2013). A guide on the use of factor analysis in the assessment of construct validity. *Journal of Korean Academy of Nursing*, *43*(5), 587–594. https://doi.org/10.4040/jkan.2013.43.5.587

Katzan, I. L., Lapin, B., Griffith, S., Jehi, L., Fernandez, H., Pioro, E., Tepper, S., & Crane, P. K. (2021). Somatic symptoms have negligible impact on Patient Health Questionnaire-9 depression scale scores in neurological patients. *European Journal of Neurology*, *28*(6), 1812–1819. http://10.0.4.87/ene.14822

Kauhanen, M. L., Korpelainen, J. T., Hiltunen, P., Brusin, E., Mononen, H., Määttä, R., Nieminen, P., Sotaniemi, K. A., & Myllylä, V. V. (1999). Poststroke depression correlates with cognitive impairment and neurological deficits. *Stroke*, *30*(9), 1875–1880.

https://doi.org/10.1161/01.STR.30.9.1875

Kauhanen, Marja Liisa, Korpelainen, J. T., Hiltunen, P., Nieminen, P., Sotaniemi, K. A., & Myllylä, V. V. (2000). Domains and determinants of quality of life after stroke caused by brain infarction. *Archives of Physical Medicine and Rehabilitation*, *81*(12), 1541–1546. https://doi.org/10.1053/apmr.2000.9391

Kay, T., Newman, B., Cavallo, M., Ezrachi, O., & Resnick, M. (1992). Toward a Neuropsychological Model of Functional Disability After Mild Traumatic Brain Injury. *Neuropsychology*, *6*(4), 371–384. https://doi.org/10.1037/0894-4105.6.4.371

Khademi, A., Wells, C. S., & Oliveri, M. E. (2021). *Examining Appropriacy of CFI/TLI Cutoff Value in Multiple-group CFA Test of Measurement Invariance to Enhance Accuracy of Test Score Interpretation*. https://doi.org/10.31234/OSF.IO/Y36UV

Kim, D. J., Kim, K., Lee, H. W., Hong, J. P., Cho, M. J., Fava, M., Mischoulon, D., Heo, J. Y., & Jeon, H. J. (2017). Internet Game Addiction, Depression, and Escape from Negative Emotions in Adulthood: A Nationwide Community Sample of Korea. *Journal of Nervous and Mental Disease*, *205*(7), 568–573. https://doi.org/10.1097/NMD.0000000000000698

Kim, E. S., & Yoon, M. (2011). Testing Measurement Invariance: A Comparison of Multiple-Group Categorical CFA and IRT. *Http://Dx.Doi.Org.Uea.Idm.Oclc.Org/10.1080/10705511.2011.557337*, *18*(2), 212–228. https://doi.org/10.1080/10705511.2011.557337

Kim, Y. E., & Lee, B. (2019). The psychometric properties of the patient health questionnaire-9 in a sample of Korean university students. *Psychiatry Investigation*, *16*(12), 904–910. https://doi.org/10.30773/pi.2019.0226

King's College London. (2020). *SSNAP - About SSNAP*. https://www.strokeaudit.org/About-SSNAP.aspx

King, E., Beynon, M., Chalder, T., Sharpe, M., & White, P. D. (2020). Patterns of daytime physical activity in patients with chronic fatigue syndrome. *Journal of Psychosomatic Research*, *135*, 110154. https://doi.org/10.1016/J.JPSYCHORES.2020.110154

Kitson, A. L., Dow, C., Calabrese, J. D., Locock, L., & Athlin, Å. M. (2013). Stroke survivors'

experiences of the fundamentals of care: A qualitative analysis. *International Journal of Nursing Studies*, *50*(3), 392–403. https://doi.org/10.1016/J.IJNURSTU.2012.09.017

Klein, D. N., Kotov, R., & Bufferd, S. J. (2011). Personality and Depression: Explanatory Models and Review of the Evidence. *Annual Review of Clinical Psychology*, *7*, 269. https://doi.org/10.1146/ANNUREV-CLINPSY-032210-104540

Knight, B. G., Robinson, G. S., Flynn Longmire, C. V., Chun, M., Nakao, K., & Kim, J. H. (2002). Cross cultural issues in caregiving for persons with dementia: Do familism values reduce burden and distress? *Ageing International 2002 27:3*, *27*(3), 70–94. https://doi.org/10.1007/S12126-003-1003-Y

Kocalevent, R. D., Hinz, A., & Brähler, E. (2013). Standardization of the depression screener patient health questionnaire (PHQ-9) in the general population. *General Hospital Psychiatry*, *35*(5), 551–555. https://doi.org/10.1016/J.GENHOSPPSYCH.2013.04.006

Kotila, M., Numminen, H., Waltimo, O., & Kaste, M. (1999). Post-stroke depression and functional recovery in a population-based stroke register. The Finnstroke study. *European Journal of Neurology*, *6*(3), 309–312. https://doi.org/10.1046/j.1468-1331.1999.630309.x

Krause, J. S., Reed, K. S., & McArdle, J. J. (2010). Factor structure and predictive validity of somatic and nonsomatic symptoms from the patient health questionnaire-9: A longitudinal study after spinal cord injury. *Archives of Physical Medicine and Rehabilitation*, *91*(8), 1218–1224. https://doi.org/10.1016/j.apmr.2010.04.015

Krause, J. S., Saunders, L. L., Bombardier, C., & Kalpakjian, C. (2011). Confirmatory Factor Analysis of the Patient Health Questionnaire-9: A Study of the Participants From the Spinal Cord Injury Model Systems. *PM and R*, *3*(6), 533–540. https://doi.org/10.1016/j.pmrj.2011.03.003

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Kwakkel, G., Kollen, B., & Lindeman, E. (2004). *Understanding the pattern of functional recovery after stroke: Facts and theories*. Restorative Neurology and Neuroscience.

https://eds.b.ebscohost.com/eds/pdfviewer/pdfviewer?vid=1&sid=7a6fe5f8-96a2-4b7c-9145-e671f68fe6c4%40pdc-v-sessmgr06

Kyriazos, T. A. (2018). Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. *Psychology*, *09*(08), 2207–2230. http://www.scirp.org/journal/doi.aspx?DOI=10.4236/psych.2018.98126

Lai, S. M., Studenski, S., Duncan, P. W., & Perera, S. (2002). Persisting consequences of stroke measured by the stroke impact scale. *Stroke*, *33*(7), 1840–1844. https://doi.org/10.1161/01.STR.0000019289.15440.F2

Laidlaw, K., Thompson, L. W., Dick-Siskin, L., & Gallagher-Thompson, D. (2008). Cognitive behaviour therapy with older people. *Cognitive Behaviour Therapy with Older People*, 1–215. https://doi.org/10.1002/9780470713402

Langhorne, P., Stott, D. J., Robertson, L., MacDonald, J., Jones, L., McAlpine, C., Dick, F., Taylor, G. S., & Murray, G. (2000). Medical Complications After Stroke. *Stroke*, *31*(6), 1223–1229. https://doi.org/10.1161/01.STR.31.6.1223

Lees, R., Stott, D. J., Quinn, T. J., & Broomfield, N. M. (2014). Feasibility and diagnostic accuracy of early mood screening to diagnose persisting clinical depression/anxiety disorder after stroke. *Cerebrovascular Diseases*, *37*(5), 323–329. https://doi.org/10.1159/000360755

Levin-Aspenson, H. F., & Watson, D. (2018). Mode of administration effects in psychopathology assessment: Analyses of gender, age, and education differences in self-rated versus interview-based depression. *Psychological Assessment*, *30*(3), 287–295. https://doi.org/10.1037/PAS0000474

Levis, B., Benedetti, A., & Thombs, B. D. (2019). Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: Individual participant data meta-analysis. *The BMJ*, *365*. https://doi.org/10.1136/bmj.l1476

Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods*, *21*(3), 369–387. https://doi.org/10.1037/MET0000093

Li, T., Li, J., & Dai, Q. (2008). Adult attachment, social support, and depression level of poststroke

patients. *Social Behavior and Personality*, *36*(10), 1341–1352. https://doi.org/10.2224/sbp.2008.36.10.1341

Lindén, T., Blomstrand, C., & Skoog, I. (2007). Depressive disorders after 20 months in elderly stroke patients: A case-control study. *Stroke*, *38*(6), 1860–1863. https://doi.org/10.1161/STROKEAHA.106.471805

Lipsey, J. R., Spencer, W. C., Rabins, P. V., & Robinson, R. G. (1986). Phenomenological comparison of poststroke depression and functional depression. *American Journal of Psychiatry*, *143*(4), 527–529. https://doi.org/10.1176/AJP.143.4.527

Lipsey, John R, Robinson, R. G., Pearlson, G. D., Rao, K., & Price, T. R. (1983). Mood change following bilateral hemisphere brain injury. *The British Journal of Psychiatry*, *143*, 266–273. https://doi.org/10.1192/bjp.143.3.266

Lipsey, John R, Spencer, W. C., Rabins, P. V, & Robinson, R. G. (1986). Phenomenological comparison of poststroke depression and functional depression. *The American Journal of Psychiatry*, *143*(4), 527–529. https://doi.org/10.1176/ajp.143.4.527

Liu, Y., & Wang, J. L. (2015). Validity of the Patient Health Questionnaire-9 for DSM-IV major depressive disorder in a sample of Canadian working population. *Journal of Affective Disorders*, *187*, 122–126. https://doi.org/10.1016/J.JAD.2015.07.044

Loetscher, T., & Lincoln, N. B. (2013). Cognitive rehabilitation for attention deficits following stroke. In *Cochrane Database of Systematic Reviews* (Vol. 2013, Issue 5). John Wiley and Sons Ltd. https://doi.org/10.1002/14651858.CD002842.pub2

Losada, A., Robinson Shurgot, G., Knight, B. G., Márquez, M., Montorio, I., Izal, M., & Ruiz, M. A. (2007). Cross-cultural study comparing the association of familism with burden and depressive symptoms in two samples of Hispanic dementia caregivers. *Https://Doi.Org/10.1080/13607860500307647*, *10*(1), 69–76. https://doi.org/10.1080/13607860500307647

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343. https://doi.org/10.1016/0005-

7967(94)00075-U

Low, E., Crewther, S. G., Ong, B., Perre, D., & Wijeratne, T. (2017). Compromised motor dexterity confounds processing speed task outcomes in stroke patients. *Frontiers in Neurology*, *8*(SEP), 21. https://doi.org/10.3389/fneur.2017.00484

Ludman, E., Katon, W., Bush, T., Rutter, C., Lin, E., Simon, G., von Korff, M., & Walker, E. (2003). Behavioural factors associated with symptom outcomes in a primary care-based depression prevention intervention trial. *Psychological Medicine*, *33*(6), 1061–1070. https://doi.org/10.1017/S003329170300816X

Ma, L. L., Wang, Y. Y., Yang, Z. H., Huang, D., Weng, H., & Zeng, X. T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: What are they and which is better? *Military Medical Research*, *7*(1), 1–11. https://doi.org/10.1186/S40779-020-00238-8/TABLES/1

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Majumdar, S., & Morris, R. (2019). Brief group-based acceptance and commitment therapy for stroke survivors. *British Journal of Clinical Psychology*, *58*(1), 70–90. https://doi.org/10.1111/bjc.12198

Marazziti, D., Consoli, G., Picchetti, M., Carlini, M., & Faravelli, L. (2010). Cognitive impairment in major depression. *European Journal of Pharmacology*, *626*(1), 83–86. https://doi.org/10.1016/J.EJPHAR.2009.08.046

Maul, A., Torres Irribarra, D., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, *79*, 311–320. https://doi.org/10.1016/J.MEASUREMENT.2015.11.001

McAleese, N., Guzman, A., O'Rourke, S. J., & Gillespie, D. C. (2021). Post-stroke emotionalism: A qualitative investigation. *Disability and Rehabilitation: An International, Multidisciplinary Journal*, *43*(2), 192–200. https://doi.org/10.1080/09638288.2019.1620876

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care*, *44*(11 Suppl 3). https://doi.org/10.1097/01.MLR.0000245438.73837.89

Methley, A. M., Campbell, S., Chew-Graham, C., McNally, R., & Cheraghi-Sohi, S. (2014). PICO, PICOS and SPIDER: A comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. *BMC Health Services Research*, *14*(1), 1–10. https://doi.org/10.1186/S12913-014-0579-0/TABLES/7

Mijajlović, M. D., Pavlović, A., Brainin, M., Heiss, W.-D., Quinn, T. J., Ihle-Hansen, H. B., Hermann, D. M., Assayag, E. Ben, Richard, E., Thiel, A., Kliper, E., Shin, Y.-I., Kim, Y.-H., Choi, S., Jung, S., Lee, Y.-B., Sinanović, O., Levine, D. A., Schlesinger, I., … Bornstein, N. M. (2017). Post-stroke dementia - a comprehensive review. *BMC Medicine*, *15*(1), 11. https://doi.org/10.1186/s12916-017-0779-7

Mitchell, A. J., Sheth, B., Gill, J., Yadegarfar, M., Stubbs, B., Yadegarfar, M., & Meader, N. (2017). Prevalence and predictors of post-stroke mood disorders: A meta-analysis and meta-regression of depression, anxiety and adjustment disorder. *General Hospital Psychiatry*, *47*(October 2015), 48–60. https://doi.org/10.1016/j.genhosppsych.2017.04.001

Montgomery, S. A., & Asberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry : The Journal of Mental Science*, *134*(4), 382–389. https://doi.org/10.1192/BJP.134.4.382

Moorey, S. (2010). The Six Cycles Maintenance Model: Growing a "Vicious Flower" for Depression. *Behavioural and Cognitive Psychotherapy*, *38*(2), 173–184. https://doi.org/10.1017/S1352465809990580

Moriarty, A. S., Gilbody, S., McMillan, D., & Manea, L. (2015). Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): A meta-analysis. *General Hospital Psychiatry*, *37*(6), 567–576. https://doi.org/10.1016/j.genhosppsych.2015.06.012

Morris, P. L. P., Robinson, R. G., & Raphael, B. (1993). Emotional Lability after Stroke. *Australian and New Zealand Journal of Psychiatry*, *27*(4), 601–605. https://doi.org/10.3109/00048679309075822

Muruet, W., Rudd, A., Wolfe, C. D. A., & Douiri, A. (2018). Long-term survival after intravenous thrombolysis for ischemic stroke: A propensity score-matched cohort with up to 10-year follow-up. *Stroke*, *49*(3), 607–613. https://doi.org/10.1161/STROKEAHA.117.019889

National Heart Lung and Blood Institute (NHLBI). (2013). *Study Quality Assessment Tools* . https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools

Newberg, A. R., Davydow, D. S., & Lee, H. B. (2009). Cerebrovascular disease basis of depression: Post-stroke depression and vascular depression. *Http://Dx.Doi.Org/10.1080/09540260600935447*, *18*(5), 433–441. https://doi.org/10.1080/09540260600935447

Nguyen, S., Wong, D., McKay, A., Rajaratnam, S. M. W., Spitz, G., Williams, G., Mansfield, D., & Ponsford, J. L. (2019). Cognitive behavioural therapy for post-stroke fatigue and sleep disturbance: A pilot randomised controlled trial with blind assessment. *Neuropsychological Rehabilitation*, *29*(5), 723–738. https://doi.org/10.1080/09602011.2017.1326945

Okifuji, A., Dennis, •, Turk, C., & Turk, D. C. (2015). Behavioral and Cognitive-Behavioral Approaches to Treating Patients with Chronic Pain: Thinking Outside the Pill Box. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *33*, 218–238. https://doi.org/10.1007/s10942-015-0215-x

Olney, S. J., & Richards, C. (1996). Hemiparetic gait following stroke. Part I: Characteristics. In *Gait and Posture* (Vol. 4, Issue 2, pp. 136–148). Elsevier B.V. https://doi.org/10.1016/0966-6362(96)01063-6

Ormstad, H., & Eilertsen, G. (2015). A biopsychosocial model of fatigue and depression following stroke. *Medical Hypotheses*, *85*(6), 835–841. https://doi.org/10.1016/j.mehy.2015.10.001

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan-a web and mobile app for systematic reviews. *Systematic Reviews*, *5*(1). https://doi.org/10.1186/S13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/BMJ.N71

Palagini, L., Ragno, G., Caccavale, L., Gronchi, A., Terzaghi, M., Mauri, M., Kyle, S., Espie, C. A.,

& Manni, R. (2015). Italian validation of the Sleep Condition Indicator: A clinical screening tool to evaluate Insomnia Disorder according to DSM-5 criteria. *International Journal of Psychophysiology : Official Journal of the International Organization of Psychophysiology*, *98*(3 Pt 1), 435–440. https://doi.org/10.1016/J.IJPSYCHO.2015.08.008

Pan, A., Keum, N., Okereke, O. I., Sun, Q., Kivimaki, M., Rubin, R. R., & Hu, F. B. (2012). Bidirectional association between depression and metabolic syndrome: A systematic review and meta-analysis of epidemiological studies. *Diabetes Care*, *35*(5), 1171–1180. https://doi.org/10.2337/dc11-2055

Patel, J. S., Oh, Y., Rand, K. L., Wu, W., Cyders, M. A., Kroenke, K., & Stewart, J. C. (2019). Measurement Invariance of the Patient Health Questionnaire-9 (PHQ-9) Depression Screener in U.S. Adults Across Sex, Race/Ethnicity, and Education Level: NHANES 2005-2016. *Depression and Anxiety*, *36*(9), 813. https://doi.org/10.1002/DA.22940

Picardi, A., Battisti, F., de Girolamo, G., Morosini, P., Norcio, B., Bracco, R., & Biondi, M. (2008). Symptom structure of acute mania: A factor study of the 24-item Brief Psychiatric Rating Scale in a national sample of patients hospitalized for a manic episode. *Journal of Affective Disorders*, *108*(1–2), 183–189. https://doi.org/10.1016/j.jad.2007.09.010

Pickard, A. S., Dalal, M. R., Bushnell, D. M., AS, P., MR, D., DM, B., Pickard, A. S., Dalal, M. R., & Bushnell, D. M. (2006). A Comparison of Depressive Symptoms in Stroke and Primary Care: Applying Rasch Models to Evaluate the Center for Epidemiologic Studies-Depression Scale. *Value in Health*, *9*(1), 59–64. https://doi.org/10.1111/j.1524-4733.2006.00082.x

Pilgrim, D. (2013). Some implications of critical realism for mental health research. *Social Theory & Health 2013 12:1*, *12*(1), 1–21. https://doi.org/10.1057/STH.2013.17

Pohjasvaara, T., Leskela, M., Vataja, R., Kalska, H., Ylikoski, R., Hietanen, M., Leppavuori, A., Kaste, M., & Erkinjuntti, T. (2002). Post-stroke depression, executive dysfunction and functional outcome. *European Journal of Neurology*, *9*(3), 269–275. https://doi.org/10.1046/j.1468-1331.2002.00396.x

Pohjasvaara, T., Vataja, R., Leppävuori, A., Kaste, M., & Erkinjuntti, T. (2001). Depression is an independent predictor of poor long-term functional outcome post-stroke. *European Journal of*

*Neurology*, *8*(4), 315–319. https://doi.org/10.1046/j.1468-1331.2001.00182.x

Pohjasvaara, Tarja, Leppävuori, A., Siira, I., Vataja, R., Kaste, M., & Erkinjuntti, T. (1998). Frequency and clinical determinants of poststroke depression. *Stroke*, *29*(11), 2311–2317. https://doi.org/10.1161/01.STR.29.11.2311

Pollock, A., & Berge, E. (2018). How to do a systematic review. In *International Journal of Stroke* (Vol. 13, Issue 2, pp. 138–156). https://doi.org/10.1177/1747493017743796

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K., & Duffy, S. (2006). *Guidance on the Conduct of Narrative Synthesis in Systematic Reviews A Product from the ESRC Methods Programme Peninsula Medical School, Universities of Exeter and Plymouth*.

Prisnie, J. C., Fiest, K. M., Coutts, S. B., Patten, S. B., Atta, C. A. M., Blaikie, L., Bulloch, A. G. M., Demchuk, A., Hill, M. D., Smith, E. E., & Jetté, N. (2016). Validating screening tools for depression in stroke and transient ischemic attack patients. *International Journal of Psychiatry in Medicine*, *51*(3), 262–277. https://doi.org/10.1177/0091217416652616

Protogerou, C., & Hagger, M. S. (2018). A Case For a Study Quality Appraisal in Survey Studies in Psychology. *Frontiers in Psychology*, *9*(JAN). https://doi.org/10.3389/FPSYG.2018.02788

Putnick, D. L., & Bornstein, M. H. (2016). Measurement Invariance Conventions and Reporting: The State of the Art and Future Directions for Psychological Research. *Developmental Review : DR*, *41*, 71. https://doi.org/10.1016/J.DR.2016.06.004

Quinn, T. J., Taylor-Rowan, M., Elliott, E., Drozdowska, B., McMahon, D., Broomfield, N. M., Barber, M., MacLeod, M. J., Cvoro, V., Byrne, A., Ross, S., Crow, J., Slade, P., Dawson, J., & Langhorne, P. (2022). Research protocol – Assessing Post-Stroke Psychology Longitudinal Evaluation (APPLE) study: A prospective cohort study in stroke. *Cerebral Circulation - Cognition and Behavior*, *3*, 100042. https://doi.org/10.1016/J.CCCB.2022.100042

Radloff, L. S. (2016). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Http://Dx.Doi.Org.Uea.Idm.Oclc.Org/10.1177/014662167700100306*, *1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Rahamatali, M., De Bont, N., Valet, M., Halkin, V., Hanson, P., Deltombe, T., Lejeune, T., & Selves,

C. (2021). Post-stroke fatigue: how it relates to motor fatigability and other modifiable factors in people with chronic stroke. *Acta Neurologica Belgica*, *121*(1), 181–189. https://doi.org/10.1007/s13760-020-01453-9

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(SUPPL. 1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Reiswig, J. (2010). Mendeley. *Journal of the Medical Library Association : JMLA*, *98*(2), 193. https://doi.org/10.3163/1536-5050.98.2.021

Robinson, R. G. (2006). The clinical neuropsychiatry of stroke: Second edition. In *The Clinical Neuropsychiatry of Stroke*. https://doi.org/10.1017/CBO9780511544231

Robinson, R. G., & Jorge, R. E. (2016). Post-stroke depression: A review. *The American Journal of Psychiatry*, *173*(3), 221–231. https://doi.org/10.1176/appi.ajp.2015.15030363

Robinson, R. G., Kubos, K. L., Starr, L. B., Rao, K., & Price, T. R. (1983). Mood changes in stroke patients: Relationship to lesion location. *Comprehensive Psychiatry*, *24*(6), 555–566. https://doi.org/10.1016/0010-440X(83)90024-X

Robinson, R. G., & Spalletta, G. (2010). Poststroke depression: A review. *The Canadian Journal of Psychiatry / La Revue Canadienne de Psychiatrie*, *55*(6), 341–349. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2011-13263-002&site=ehost-live

Rosenberg, M. (1965). Rosenberg self-esteem scale. *Journal of Religion and Health*. https://psycnet.apa.org/doiLanding?doi=10.1037/t01038-000

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*. https://doi.org/10.18637/JSS.V048.I02

Saeki, S., Ogata, H., Okubo, T., Takahashi, K., & Hoshuyama, T. (1995). Return to Work After Stroke. *Stroke*, *26*(3), 399–401. https://doi.org/10.1161/01.STR.26.3.399

Sami, W., Ansari, T., Butt, N. S., & Hamid, M. R. A. (2017). Effect of diet on type 2 diabetes mellitus: A review. *International Journal of Health Sciences*, *11*(2), 65. /pmc/articles/PMC5426415/

Santos, I. S., Tavares, B. F., Munhoz, T. N., de Almeida, L. S. P., da Silva, N. T. B., Tams, B. D., Patella, A. M., & Matijasevich, A. (2013). Sensibilidade e especificidade do Patient Health Questionnaire-9 (PHQ-9) entre adultos da população geral. *Cadernos de Saude Publica*, *29*(8), 1533–1543. https://doi.org/10.1590/0102-311X00144612

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating Model Fit With Ordered Categorical Data Within a Measurement Invariance Framework: A Comparison of Estimators. *Structural Equation Modeling*, *21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658/SUPPL_FILE/HSEM_A_882658_SM7145.DOCX

Satorra, A. (2000). *Scaled and Adjusted Restricted Tests in Multi-Sample Analysis of Moment Structures*. 233–247. https://doi.org/10.1007/978-1-4615-4603-0_17

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, *18*(4), 210–222. https://doi.org/10.1016/J.HRMR.2008.03.003

Schöttke, H., & Giabbiconi, C. M. (2015). Post-stroke depression and post-stroke anxiety: Prevalence and predictors. *International Psychogeriatrics*, *27*(11), 1805–1812. https://doi.org/10.1017/S1041610215000988

Schramke, C. J., Stowe, R. M., Ratcliff, G., Goldstein, G., & Condray, R. (1998). Poststroke depression and anxiety: Different assessment methods result in variations in incidence and severity estimates. *Journal of Clinical and Experimental Neuropsychology*, *20*(5), 723–737. https://doi.org/10.1076/jcen.20.5.723.1117

semTools Contributors. (2016). *semTools: Useful tools for structural equation modeling. R package version 0.4-14*. https://cran.r-project.org/package=semTools

Sharpe, D. (2015). Chi-Square Test is Statistically Significant: Now What? *Practical Assessment, Research, and Evaluation Practical Assessment*, *20*. https://doi.org/10.7275/tbfa-x148

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric depression scale (GDS) recent evidence and development of a shorter version. *Clinical Gerontologist*, *5*(1–2), 165–173. https://doi.org/10.1300/J018v05n01_09

Sheikh, S., & Furnham, A. (2000). A cross-cultural study of mental health beliefs and attitudes towards seeking professional help. *Social Psychiatry and Psychiatric Epidemiology*, *35*(7), 326–334. https://doi.org/10.1007/S001270050246

Shi, Y., Yang, D., Zeng, Y., & Wu, W. (2017). Risk factors for post-stroke depression: A meta-analysis. In *Frontiers in Aging Neuroscience* (Vol. 9, Issue JUL, p. 218). Frontiers Media S.A. https://doi.org/10.3389/fnagi.2017.00218

Simning, A., Seplaki, C. L., & Conwell, Y. (2018). The association of a heart attack or stroke with depressive symptoms stratified by the presence of a close social contact: findings from the National Health and Aging Trends Study Cohort. *International Journal of Geriatric Psychiatry*, *33*(1), 96–103. https://doi.org/10.1002/gps.4684

Simning, A., Van Wijngaarden, E., Fisher, S. G., Richardson, T. M., & Conwell, Y. (2012). Mental healthcare need and service utilization in older adults living in public housing. *The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry*, *20*(5), 441–451. https://doi.org/10.1097/JGP.0B013E31822003A7

Simon, G. E., Rutter, C. M., Peterson, D., Oliver, M., Whiteside, U., Operskalski, B., & Ludman, E. J. (2013). Does response on the PHQ-9 depression questionnaire predict subsequent suicide attempt or suicide death? *Psychiatric Services*, *64*(12), 1195–1202. https://doi.org/10.1176/APPI.PS.201200587/ASSET/IMAGES/LARGE/1195F2.JPEG

Smets, E. M. A., Garssen, B., Bonke, B., & De Haes, J. C. J. M. (1995). The Multidimensional Fatigue Inventory (MFI) psychometric qualities of an instrument to assess fatigue. *Journal of Psychosomatic Research*, *39*(3), 315–325. https://doi.org/10.1016/0022-3999(94)00125-O

Smith, A. B., Selby, P. J., Velikova, G., Stark, D., Wright, E. P., Gould, A., & Cull, A. (2002). Factor analysis of the Hospital Anxiety and Depression Scale from a large cancer population. *Psychology and Psychotherapy*, *75*(Pt 2), 165–176. https://doi.org/10.1348/147608302169625

Starkstein, S. E., & Robinson, R. G. (1989). Affective disorders and cerebral vascular disease. *The British Journal of Psychiatry : The Journal of Mental Science*, *154*(FEB.), 170–182. https://doi.org/10.1192/BJP.154.2.170

Steer, R. A., Ball, R., Ranieri, W. F., & Beck, A. T. (1999). Dimensions of the Beck Depression

Inventory-II in Clinically Depressed Outpatients. In *JOURNAL OF CLINICAL PSYCHOLOGY* (Vol. 55, Issue 1).

Stokes, E. K., O'Connell, C., & Murphy, B. (2011). An investigation into fatigue post-stroke and its multidimensional nature. *Advances in Physiotherapy*, *13*(1), 2–10. https://doi.org/10.3109/14038196.2010.534175

Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the beck depression inventory—second edition in a sample of college students. *Depression and Anxiety*, *19*(3), 187–189. https://doi.org/10.1002/DA.20002

Studenski, S. A., Wallace, D., Duncan, P. W., Rymer, M., & Lai, S. M. (2001). Predicting Stroke Recovery: Three and Six-Month Rates of Patient-Centered Functional Outcomes Based on the Orpington Prognostic Scale. *Journal of the American Geriatrics Society*, *49*(3), 308–312. https://doi.org/10.1046/j.1532-5415.2001.4930308.x

Tang, W.-K., Grace Lau, C., Mok, V., Ungvari, G. S., & Wong, K.-S. (2015). Insomnia and health-related quality of life in stroke. *Topics in Stroke Rehabilitation*, *22*(3), 201–207. https://doi.org/10.1179/1074935714Z.0000000026

Taule, T., & Råheim, M. (2014). Life changed existentially: A qualitative study of experiences at 6-8 months after mild stroke. *Disability and Rehabilitation*, *36*(25), 2107–2119. https://doi.org/10.3109/09638288.2014.904448

Taylor, G. H., Todman, J., & Broomfield, N. M. (2011). Post-stroke emotional adjustment: A modified social cognitive transition model. *Neuropsychological Rehabilitation*, *21*(6), 808–824. https://doi.org/10.1080/09602011.2011.598403

Teasdale, T. W., & Engberg, A. W. (2010). Psychosocial consequences of stroke: A long-term population-based follow-up. *Https://Doi.Org/10.1080/02699050500110421*, *19*(12), 1049–1058. https://doi.org/10.1080/02699050500110421

Telfer, R. E. (2014). *The experience and expression of social cognition changes for stroke survivors*. https://search.ebscohost.com/login.aspx?direct=true&db=ddu&AN=EB2ADE1E50B7039F&site=ehost-live

*The DEPRESSD Project*. (2021). https://www.depressd.ca/

Thomas, A. J., Kalaria, R. N., & O'Brien, J. T. (2004). Depression and vascular disease: What is the relationship? *Journal of Affective Disorders*, *79*(1–3), 81–95. https://doi.org/10.1016/S0165-0327(02)00349-X

Thombs, B. D., Ziegelstein, R. C., & Whooley, M. A. (2008). Optimizing detection of major depression among patients with coronary artery disease using the patient health questionnaire: data from the heart and soul study. *Journal of General Internal Medicine*, *23*(12), 2014–2017. https://doi.org/10.1007/S11606-008-0802-Y

Thompson, L., & McCabe, R. (2012). The effect of clinician-patient alliance and communication on treatment adherence in mental health care: A systematic review. *BMC Psychiatry*, *12*(1), 1–12. https://doi.org/10.1186/1471-244X-12-87/TABLES/3

Thomson, H. J., & Thomas, S. (2013). The effect direction plot: visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods*, *4*(1), 95–101. https://doi.org/10.1002/JRSM.1060

Tinakon, W., & Nahathai, W. (2012). A Comparison of Reliability and Construct Validity between the Original and Revised Versions of the Rosenberg Self-Esteem Scale. *Psychiatry Investigation*, *9*(1), 54. https://doi.org/10.4306/PI.2012.9.1.54

Townend, E. (2005). *Beliefs about "stroke" and 'its effects' : a study of their association with emotional distress*. https://search.ebscohost.com/login.aspx?direct=true&db=ddu&AN=822260FDE532F712&site=ehost-live

Townend, E., Tinson, D., Kwan, J., & Sharpe, M. (2010). 'Feeling sad and useless': An investigation into personal acceptance of disability and its association with depression following stroke. *Clinical Rehabilitation*, *24*(6), 555–564. https://doi.org/10.1177/0269215509358934

UK National Collaborating Centre for Mental Health. (2010). *The classification of depression and depression rating scales/questionnaires*. https://www-ncbi-nlm-nih-gov.uea.idm.oclc.org/books/NBK82926/

Ulbricht, C. M., Chrysanthopoulou, S. A., Levin, L., & Lapane, K. L. (2018). The use of latent class analysis for identifying subtypes of depression: A systematic review. *Psychiatry Research*, *266*,

228–246. https://doi.org/10.1016/J.PSYCHRES.2018.03.003

Unthank, K. W. (2019). How self-blame empowers and disempowers survivors of interpersonal trauma: An intuitive inquiry. *Qualitative Psychology*, *6*(3), 359–378. https://doi.org/10.1037/QUP0000136

Vataja, R., Pohjasvaara, T., Mäntylä, R., Ylikoski, R., Leppävuori, A., Leskelä, M., Kalska, H., Hietanen, M., Aronen, H. J., Salonen, O., Kaste, M., & Erkinjuntti, T. (2003). MRI correlates of executive dysfunction in patients with ischaemic stroke. *European Journal of Neurology*, *10*(6), 625–631. https://doi.org/10.1046/j.1468-1331.2003.00676.x

Vickery, C. D., Sepehri, A., Evans, C. C., CD, V., Sepehri, A., CC, E., Vickery, C. D., Sepehri, A., Evans, C. C., CD, V., Sepehri, A., & CC, E. (2008). Self-esteem in an acute stroke rehabilitation sample: a control group comparison. *Clinical Rehabilitation*, *22*(2), 179–187. https://doi.org/10.1177/0269215507080142

Vickery, C. D., Sherer, M., Evans, C. C., Gontkovsky, S. T., Lee, J. E., CD, V., Sherer, M., CC, E., ST, G., JE, L., Vickery, C. D., Sherer, M., Evans, C. C., Gontkovsky, S. T., & Lee, J. E. (2008). The relationship between self-esteem and functional outcome in the acute stroke-rehabilitation setting. *Rehabilitation Psychology*, *53*(1), 101–109. https://doi.org/10.1037/0090-5550.53.1.101

Villarreal-Zegarra, D., Copez-Lonzoy, A., Bernabé-Ortiz, A., Melendez-Torres, G. J., & Bazo-Alvarez, J. C. (2019). Valid group comparisons can be made with the Patient Health Questionnaire (PHQ-9): A measurement invariance study across groups by demographic characteristics. *PLoS ONE*, *14*(9). https://doi.org/10.1371/JOURNAL.PONE.0221717

Vingerhoets, A. J. J. M., Rottenberg, J., Cevaal, A., & Nelson, J. K. (2007). Is there a relationship between depression and crying? A review. In *Acta Psychiatrica Scandinavica* (Vol. 115, Issue 5, pp. 340–351). https://doi.org/10.1111/j.1600-0447.2006.00948.x

Visser, M. M., Heijenbrok-Kal, M. H., Spijker, A. van't, Oostra, K. M., Busschbach, J. J., & Ribbers, G. M. (2015). Coping, Problem Solving, Depression, and Health-Related Quality of Life in Patients Receiving Outpatient Stroke Rehabilitation. *Archives of Physical Medicine & Rehabilitation*, *96*(8), 1492–1498. https://doi.org/10.1016/j.apmr.2015.04.007

Volker, D., Zijlstra-Vlasveld, M. C., Brouwers, E. P. M., Homans, W. A., Emons, W. H. M., & van

der Feltz-Cornelis, C. M. (2016). Validation of the Patient Health Questionnaire-9 for Major Depressive Disorder in the Occupational Health Setting. *Journal of Occupational Rehabilitation*, *26*(2), 237–244. https://doi.org/10.1007/S10926-015-9607-0

Wade, D. T., Wood, V. A., & Hewer, R. L. (1985). Recovery after stroke - The first 3 months. *Journal of Neurology Neurosurgery and Psychiatry*, *48*(1), 7–13. https://doi.org/10.1136/jnnp.48.1.7

Walker, M. F., Sunnerhagen, K. S., & Fisher, R. J. (2013). Evidence-based community stroke rehabilitation. *Stroke*, *44*(1), 293–297. https://doi.org/10.1161/STROKEAHA.111.639914

Wang, Y. C., Kapellusch, J., & Garg, A. (2014). Important factors influencing the return to work after stroke. *Work*, *47*(4), 553–559. https://doi.org/10.3233/WOR-131627

Welch, J. K., & Patton, M. Q. (1992). Qualitative Evaluation and Research Methods. *The Modern Language Journal*, *76*(4), 543. https://doi.org/10.2307/330063

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). *Response biases*. University of Kent.

Wolfenden, B., & Grace, M. (2009). Returning to work after stroke: A review. In *International Journal of Rehabilitation Research* (Vol. 32, Issue 2, pp. 93–97). https://doi.org/10.1097/MRR.0b013e328325a358

Wu, H., & Estabrook, R. (2016). Identification of Confirmatory Factor Analysis Models of Different Levels of Invariance for Ordered Categorical Outcomes. *Psychometrika*, *81*(4), 1014. https://doi.org/10.1007/S11336-016-9506-0

Yang, F. M., & Jones, R. N. (2008). Measurement differences in depression: chronic health-related and sociodemographic effects in older Americans. *Psychosomatic Medicine*, *70*(9), 993–1004. https://doi.org/10.1097/PSY.0b013e31818ce4fa

Yang, F. M., Jones, R. N., & Jones Scd, R. N. (2008). Measurement differences in depression: chronic health-related and sociodemographic effects in older Americans. *Psychosomatic Medicine*, *70*(9), 993–1004. https://doi.org/10.1097/PSY.0b013e31818ce4fa

Yang, L. (2019). *Testing Measurement Invariance in Multilevel Data with Unequal Cross-Level Factor Structures*. https://scholarworks.uark.edu/etd/3294%0AThis

Zeng, Y.-Y., Cheng, H.-R., Cheng, L., Huang, G., Chen, Y.-B., Tang, W.-J., & He, J.-C. (2021).

Comparison of poststroke depression between acute ischemic and hemorrhagic stroke patients. *International Journal of Geriatric Psychiatry*, *36*(4), 493–499. https://doi.org/10.1002/gps.5444

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. https://doi.org/10.1111/J.1600-0447.1983.TB09716.X

## Appendices

## Appendix A: Submission Guidelines for *Neuropsychological Rehabilitation*

**COVID-19 impact on peer review**
As a result of the significant disruption that is being caused by the COVID-19 pandemic we understand that many authors and peer reviewers will be making adjustments to their professional and personal lives. As a result they may have difficulty in meeting the timelines associated with our peer review process. Please let the journal editorial office know if you need additional time. Our systems will continue to remind you of the original timelines but we intend to be flexible.

Thank you for choosing to submit your paper to us. These instructions will ensure we have everything required so your paper can move through peer review, production and publication smoothly. Please take the time to read and follow them as closely as possible, as doing so will ensure your paper matches the journal's requirements.

For general guidance on every stage of the publication process, please visit our Author Services website.

For editing support, including translation and language polishing, explore our Editing Services website

This journal uses ScholarOne Manuscripts (previously Manuscript Central) to peer review manuscript submissions. Please read the guide for ScholarOne authors before making a submission. Complete guidelines for preparing and submitting your manuscript to this journal are provided below.

This title utilises format-free submission. Authors may submit their paper in any scholarly format or layout. References can be in any style or format, so long as a consistent scholarly citation format is applied. For more detail see the format-free submission section below.

About the Journal

*Neuropsychological Rehabilitation* is an international, peer-reviewed journal publishing high-quality, original research. Please see the journal's Aims & Scope for information about its focus and peer-review policy.

Please note that this journal only publishes manuscripts in English.

*Neuropsychological Rehabilitation* accepts the following types of article: original articles, scholarly reviews, book reviews.

Open Access

You have the option to publish open access in this journal via our Open Select publishing program. Publishing open access means that your article will be free to access online immediately on publication, increasing the visibility, readership and impact of your research. Articles published Open Select with Taylor & Francis typically receive 32% more citations* and over 6 times as many downloads** compared to those that are not published Open Select.

Your research funder or your institution may require you to publish your article open access. Visit our Author Services website to find out more about open access policies and how you can comply with these.

You will be asked to pay an article publishing charge (APC) to make your article open access and this cost can often be covered by your institution or funder. Use our APC finder to view the APC for this journal.

Please visit our Author Services website or contact openaccess@tandf.co.uk if you would like more information about our Open Select Program.

*Citations received up to Jan 31st 2020 for articles published in 2015-2019 in journals listed in Web of Science®.
**Usage in 2017-2019 for articles published in 2015-2019.

## Peer Review and Ethics

Taylor & Francis is committed to peer-review integrity and upholding the highest standards of review. Once your paper has been assessed for suitability by the editor, it will then be single blind peer reviewed by independent, anonymous expert referees. Find out more about what to expect during peer review and read our guidance on publishing ethics.

## Preparing Your Paper

All authors submitting to medicine, biomedicine, health sciences, allied and public health journals should conform to the Uniform Requirements for Manuscripts Submitted to Biomedical Journals, prepared by the International Committee of Medical Journal Editors (ICMJE).

**Clinical trials:** must conform to the Consort guidelines http://www.consort-statement.org. Submitted papers should include a checklist confirming that all of the Consort requirements have been met, together with the corresponding page number of the manuscript where the information is located. In addition, trials must be pre-registered on a site such as clinicaltrials.gov or equivalent, and the manuscript should include the reference number to the relevant pre-registration.

**Systematic reviews:** submitted papers should follow PRISMA http://www.prisma-statement.org/ guidelines and submission should also be accompanied by a completed PRISMA checklist, together with the corresponding page number of the manuscript where the information is located.

**Single-case studies:** submitted papers should follow SCRIBE guidelines ( http://psycnet.apa.org/fulltext/2016-17384-001.html ) and include a completed SCRIBE checklist together with the corresponding page number of the manuscript where the information is located.

**Observational studies:** submitted papers should follow the STROBE guidelines ( https://www.strobe-statement.org/index.php?id=strobe-home) and also include a completed checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

**Qualitative studies:** should follow the COREQ guidelines ( http://www.equator-network.org/reporting-guidelines/coreq/) and be accompanied by a completed COREQ checklist of compliance, together with the corresponding page number of the manuscript where the information is located.

The EQUATOR Network (Enhancing the Quality and Transparency of Health Research) website provides further information on available guidelines.

## Structure

Your paper should be compiled in the following order: title page; abstract; keywords; main text introduction, materials and methods, results, discussion; acknowledgments; declaration of interest statement; references; appendices (as appropriate); table(s) with caption(s) (on individual pages); figures; figure captions (as a list).

## Word Limits

Please include a word count for your paper. There are no word limits for papers in this journal.

## Style Guidelines

Please refer to these quick style guidelines when preparing your paper, rather than any published articles or a sample copy.

Please use American spelling style consistently throughout your manuscript.

Please use single quotation marks, except where ‘a quotation is “within” a quotation’.

Please note that long quotations should be indented without quotation marks.

## Alt Text

This journal is now including Alt Text (alternative text), a short piece of text that can be attached to your figure to convey to readers the nature or contents of the image. It is typically used by systems such as pronouncing screen readers to make the object accessible to people that cannot read or see the object, due to a visual impairment or print disability.

Alt text will also be displayed in place of an image, if said image file cannot be loaded. Alt Text can also provide better image context/descriptions to search engine crawlers, helping them to index an image properly.

To include Alt Text in your article, please follow our Guidelines

## Format-Free Submission

Authors may submit their paper in any scholarly format or layout. Manuscripts may be supplied as single or multiple files. These can be Word, rich text format (rtf), open document format (odt), or PDF files. Figures and tables can be placed within the text or submitted as separate documents. Figures should be of sufficient resolution to enable refereeing.

There are no strict formatting requirements, but all manuscripts must contain the essential elements needed to evaluate a manuscript: abstract, author affiliation, figures, tables, funder information, and references. Further details may be requested upon acceptance.

References can be in any style or format, so long as a consistent scholarly citation format is applied. Author name(s), journal or book title, article or chapter title, year of publication, volume and issue (where appropriate) and page numbers are essential. All bibliographic entries must contain a corresponding in-text citation. The addition of DOI (Digital Object Identifier) numbers is recommended but not essential.

The journal reference style will be applied to the paper post-acceptance by Taylor & Francis.

Spelling can be US or UK English so long as usage is consistent.

Note that, regardless of the file format of the original submission, an editable version of the article must be supplied at the revision stage.

## Taylor & Francis Editing Services

To help you improve your manuscript and prepare it for submission, Taylor & Francis provides a range of editing services. Choose from options such as English Language Editing, which will ensure that your article is free of spelling and grammar errors, Translation, and Artwork Preparation. For more information, including pricing, visit this website.

## Checklist: What to Include

**Author details.** Please ensure everyone meeting the International Committee of Medical Journal Editors (ICMJE) requirements for authorship is included as an author of your paper. All authors of a manuscript should include their full name and affiliation on the cover page of the manuscript. Where available, please also include ORCiDs and social media handles (Facebook, Twitter or LinkedIn). One author will need to be identified as the corresponding author, with their email address normally displayed in the article PDF (depending on the journal) and the online article. Authors' affiliations are the affiliations where the research was conducted. If any of the named co-authors moves affiliation during the peer-review process, the new affiliation can be given as a footnote. Please note that no changes to affiliation can be made after your paper is accepted. Read more on authorship.

Should contain an unstructured abstract of 200 words.

You can opt to include a **video abstract** with your article. Find out how these can help your work reach a wider audience, and what to think about when filming.

Between 5 and 5 **keywords**. Read making your article more discoverable, including information on choosing a title and search engine optimization.

**Funding details.** Please supply all details required by your funding and grant-awarding bodies as follows:
*For single agency grants*
This work was supported by the [Funding Agency] under Grant [number xxxx].
*For multiple agency grants*
This work was supported by the [Funding Agency #1] under Grant [number xxxx]; [Funding Agency #2] under Grant [number xxxx]; and [Funding Agency #3] under Grant [number xxxx].

**Disclosure statement.** This is to acknowledge any financial interest or benefit that has arisen from the direct applications of your research. Further guidance on what is a conflict of interest and how to disclose it.

**Data availability statement.** If there is a data set associated with the paper, please provide information about where the data supporting the results or analyses presented in the paper can be found. Where applicable, this should include the hyperlink, DOI or other persistent identifier associated with the data set(s). Templates are also available to support authors.

**Data deposition.** If you choose to share or make the data underlying the study open, please deposit your data in a recognized data repository prior to or at the time of submission. You will be asked to provide the DOI, pre-reserved DOI, or other persistent identifier for the data set.

**Geolocation information.** Submitting a geolocation information section, as a separate paragraph before your acknowledgements, means we can index your paper's study area accurately in JournalMap's geographic literature database and make your article more discoverable to others. More information.

**Supplemental online material.** Supplemental material can be a video, dataset, fileset, sound file or anything which supports (and is pertinent to) your paper. We publish supplemental material online via Figshare. Find out more about supplemental material and how to submit it with your article.

**Figures.** Figures should be high quality (1200 dpi for line art, 600 dpi for grayscale and 300 dpi for colour, at the correct size). Figures should be supplied in one of our preferred file formats: EPS, PS, JPEG, TIFF, or Microsoft Word (DOC or DOCX) files are acceptable for figures that have been drawn in Word. For information relating to other file types, please consult our Submission of electronic artwork document.

**Tables.** Tables should present new information rather than duplicating what is in the text. Readers should be able to interpret the table without reference to the text. Please supply editable files.

**Equations.** If you are submitting your manuscript as a Word document, please ensure that equations are editable. More information about mathematical symbols and equations.

**Units.** Please use SI units (non-italicized).

Using Third-Party Material in your Paper

You must obtain the necessary permission to reuse third-party material in your article. The use of short extracts of text and some other types of material is usually permitted, on a limited basis, for the purposes of criticism and review without securing formal permission. If you wish to include any material in your paper for which you do not hold copyright, and which is not covered by this informal agreement, you will need to obtain written permission from the copyright owner prior to submission. More information on requesting permission to reproduce work(s) under copyright.

Disclosure Statement

Please include a disclosure statement, using the subheading "Disclosure of interest." If you have no interests to declare, please state this (suggested wording: *The authors report no conflict of interest*). For all NIH/Wellcome-funded papers, the grant number(s) must be included in the declaration of interest statement. Read more on declaring conflicts of interest.

Clinical Trials Registry

In order to be published in a Taylor & Francis journal, all clinical trials must have been registered in a public repository at the beginning of the research process (prior to patient enrolment). Trial registration numbers should be included in the abstract, with full details in the methods section. The registry should be publicly accessible (at no charge), open to all prospective registrants, and managed by a not-for-profit organization. For a list of registries that meet these requirements, please visit the WHO International Clinical Trials Registry Platform (ICTRP). The registration of all clinical trials facilitates the sharing of information among clinicians, researchers, and patients, enhances public confidence in research, and is in accordance with the ICMJE guidelines.

Complying With Ethics of Experimentation

Please ensure that all research reported in submitted papers has been conducted in an ethical and responsible manner, and is in full compliance with all relevant codes of experimentation and legislation. All papers which report in vivo experiments or clinical trials on humans or animals must include a written statement in the Methods section. This should explain that all work was conducted with the formal approval of the local human subject or animal care committees (institutional and national), and that clinical trials have been registered as legislation requires. Authors who do not have formal ethics review committees should include a statement that their study follows the principles of the Declaration of Helsinki.

Consent

All authors are required to follow the ICMJE requirements on privacy and informed consent from patients and study participants. Please confirm that any patient, service user, or participant (or that person's parent or legal guardian) in any research, experiment, or clinical trial described in your paper has given written consent to the inclusion of material pertaining to themselves, that they acknowledge that they cannot be identified via the paper; and that you have fully anonymized them. Where someone is deceased, please ensure you have written consent from the family or estate. Authors may use this Patient Consent Form, which should be completed, saved, and sent to the journal if requested.

## Health and Safety

Please confirm that all mandatory laboratory health and safety procedures have been complied with in the course of conducting any experimental work reported in your paper. Please ensure your paper contains all appropriate warnings on any hazards that may be involved in carrying out the experiments or procedures you have described, or that may be involved in instructions, materials, or formulae.

Please include all relevant safety precautions; and cite any accepted standard or code of practice. Authors working in animal science may find it useful to consult the International Association of Veterinary Editors' Consensus Author Guidelines on Animal Ethics and Welfare and Guidelines for the Treatment of Animals in Behavioural Research and Teaching. When a product has not yet been approved by an appropriate regulatory body for the use described in your paper, please specify this, or that the product is still investigational.

## Submitting Your Paper

This journal uses ScholarOne Manuscripts to manage the peer-review process. If you haven't submitted a paper to this journal before, you will need to create an account in ScholarOne. Please read the guidelines above and then submit your paper in the relevant Author Centre, where you will find user guides and a helpdesk.

Please note that *Neuropsychological Rehabilitation* uses Crossref™ to screen papers for unoriginal material. By submitting your paper to *Neuropsychological Rehabilitation* you are agreeing to originality checks during the peer-review and production processes.

On acceptance, we recommend that you keep a copy of your Accepted Manuscript. Find out more about sharing your work.

## Data Sharing Policy

This journal applies the Taylor & Francis Basic Data Sharing Policy. Authors are encouraged to share or make open the data supporting the results or analyses presented in their paper where this does not violate the protection of human subjects or other valid privacy or security concerns.

Authors are encouraged to deposit the dataset(s) in a recognized data repository that can mint a persistent digital identifier, preferably a digital object identifier (DOI) and recognizes a long-term preservation plan. If you are uncertain about where to deposit your data, please see this information regarding repositories.

Authors are further encouraged to cite any data sets referenced in the article and provide a Data Availability Statement.

At the point of submission, you will be asked if there is a data set associated with the paper. If you reply yes, you will be asked to provide the DOI, pre-registered DOI, hyperlink, or other persistent identifier associated with the data set(s). If you have selected to provide a pre-registered DOI, please be prepared to share the reviewer URL associated with your data deposit, upon request by reviewers.

Where one or multiple data sets are associated with a manuscript, these are not formally peer reviewed as a part of the journal submission process. It is the author's responsibility to ensure the soundness of data. Any errors in the data rest solely with the producers of the data set(s).

## Publication Charges

There are no submission fees, publication fees or page charges for this journal.

Colour figures will be reproduced in colour in your online article free of charge. If it is necessary for the figures to be reproduced in colour in the print version, a charge will apply.

Charges for colour figures in print are £300 per figure ($400 US Dollars; $500 Australian Dollars; €350). For more than 4 colour figures, figures 5 and above will be charged at £50 per figure ($75 US Dollars; $100 Australian Dollars; €65). Depending on your location, these charges may be subject to local taxes.

### Copyright Options

Copyright allows you to protect your original material, and stop others from using your work without your permission. Taylor & Francis offers a number of different license and reuse options, including Creative Commons licenses when publishing open access. Read more on publishing agreements.

### Complying with Funding Agencies

We will deposit all National Institutes of Health or Wellcome Trust-funded papers into PubMedCentral on behalf of authors, meeting the requirements of their respective open access policies. If this applies to you, please tell our production team when you receive your article proofs, so we can do this for you. Check funders' open access policy mandates here. Find out more about sharing your work.

### My Authored Works

On publication, you will be able to view, download and check your article's metrics (downloads, citations and Altmetric data) via My Authored Works on Taylor & Francis Online. This is where you can access every article you have published with us, as well as your free eprints link, so you can quickly and easily share your work with friends and colleagues.

We are committed to promoting and increasing the visibility of your article. Here are some tips and ideas on how you can work with us to promote your research.

### Article Reprints

You will be sent a link to order article reprints via your account in our production system. For enquiries about reprints, please contact the Taylor & Francis Author Services team at reprints@tandf.co.uk. You can also order print copies of the journal issue in which your article appears.

### Queries

Should you have any queries, please visit our Author Services website or contact us here.

## Appendix B: Full Search Strategy for the Systematic Review

Search strategy

Academic Search Complete, AMED (The Allied and Complementary Medicine Database), APA PsycArticles, APA PsycInfo, CINAHL Complete (Cumulative Index of Nursing and Allied Health Literature), MEDLINE Complete, OpenDissertations.

S1. “stroke” OR “cerebrovascular accident*”OR “post-stroke” OR “subarachnoid hemorrhage” OR “cerebral infarct*” OR “lacunar infarct*” OR “lacunar stroke” OR “cerebral hemorrhage” OR “Hypoxia-ischemia, Brain” OR “brain infarction”

S2. “low mood” OR “depress*” OR “mood” OR “wellbeing” OR “distress*” OR “affect” OR “psychological distress” OR “Stress, psychological” OR “psychological distress” OR “mental depression”

S3 “indicators” OR “predictors” OR “correlate*” OR “factors” or “control-group” OR "control group" OR “non-stroke” OR "healthy control" OR "neurotypical"

S4. S1 OR S2

This search produced high sensitivity and low precision. Too many articles were identified to reasonably consider

S5. S1 AND S2

This search had improved precision but still produced too many results, most of which appeared irrelevant

S6. S1 AND S2 AND S3

The addition of the design constraint narrowed results considerably, with concerns of poor sensitivity

S7. “phq-9” OR “phq-2” OR “phq9” OR “patient health questionnaire-9” OR “patient health questionnaire” OR “patient health questionnaire-2” OR “Geriatric Depression Scale” OR “GDS” OR “GDS-15” OR “hospital anxiety and depression scale” OR “HADS” OR “Center for Epidemiologic Studies Depression Scale” OR “CES-D” OR “Beck Depression Inventory” OR “Beck Depression Inventory-II” OR “BDI-II” OR “BDI” OR “Structured Clinical Interview for DSM-IV” OR “SCID” OR “SCID-II” OR “The Structured Clinical Interview for DSM-5” OR “Composite International Diagnostic Interview” OR “CIDI” OR “Diagnostic Interview Schedule” OR “Mini-International Neuropsychiatric Interview” OR “MINI” OR “M.I.N.I” OR “Aphasia Depression Rating Scale” OR “ADRS” OR “Brief Assessment Schedule Depression Cards” OR “BASDEC” OR “Montgomery–Asberg Depression Rating Scale” OR “MADRS” OR “Psychiatric Assessment System” OR “Schedule for Affective Disorders and Schizophrenia” OR “SADS” OR “Schedules for Clinical Assessment in Neuropsychiatry” OR “Signs of Depression Scale” OR “SODS” OR “Visual Analogue Mood Scale” OR “VAMS” OR “Hamilton Depression Rating Scale” OR “HAM-D”

A scope of results from S4 and S5 indicated that many irrelevant articles did not use quantitative assessment of depression measures. As this was a requirement for the review, constraining articles to include mention of psychometric depression measures was judged to be a valid way of improving precision with low cost to sensitivity. Depression measures were selected from the extensive review of depression measures in stroke (L.-J. J. Burton & Tyson, 2015).

S7. S1 AND S2 AND S7

## Appendix C: References of Articles that were Examined in Full-text Screening and Excluded (N=46) for the Systematic Review

The below articles are structured by exclusion reason. Extra detail is provided in italics in some cases.

### No Interpretable Comparisons Between Groups

Abdulla, F. A., Al-Khamis, F. A., Alsulaiman, A. A., & Alshami, A. M. (2019). Psychometric properties of an Arabic version of the fatigue severity scale in patients with stroke. Topics in Stroke Rehabilitation, 26(6), 448–455. https://doi.org/10.1080/10749357.2019.1628465

Brodaty, H., Liu, Z., Withall, A., & Sachdev, P. S. (2013). The longitudinal course of post-strike apathy over five years. The Journal of Neuropsychiatry and Clinical Neurosciences, 25(4), 283–291. https://doi.org/10.1176/appi.neuropsych.12040080

Brodaty, H., Withall, A., Altendorf, A., & Sachdev, P. S. (2007). Rates of depression at 3 and 15 months poststroke and their relationship with cognitive decline: The Sydney Stroke Study. The American Journal of Geriatric Psychiatry, 15(6), 477–486. https://doi.org/10.1097/JGP.0b013e3180590bca

Cooper, C. L., Phillips, L. H., Johnston, M., Whyte, M., & MacLeod, M. J. (2015). The role of emotion regulation on social participation following stroke. British Journal of Clinical Psychology, 54(2), 181–199. https://doi.org/10.1111/bjc.12068

de Bruijn, M. A. A. M., Synhaeve, N. E., van Rijsbergen, M. W. A., de Leeuw, F.-E., Mark, R. E., Jansen, B. P. W., & de Kort, P. L. M. (2015). Quality of Life after Young Ischemic Stroke of Mild Severity Is Mainly Influenced by Psychological Factors. Journal of Stroke & Cerebrovascular Diseases, 24(10), 2183–2188. https://doi.org/10.1016/j.jstrokecerebrovasdis.2015.04.040

Egerton, T., Hokstad, A., Askim, T., Bernhardt, J., & Indredavik, B. (2015). Prevalence of fatigue in patients 3 months after stroke and association with early motor activity: A prospective study comparing stroke patients with a matched general population cohort. BMC Neurology, 15. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-46562-001&site=ehost-live

Ellis-Hill, C. S., Horn, S., Ellis-Hill, C. S., & Horn, S. (2000). Change in identity and self-concept: a new theoretical approach to recovery following a stroke. Clinical Rehabilitation, 14(3), 279–287. https://doi.org/10.1191/026921500671231410

Em, S., Bozkurt, M., Karakoç, M., Çağlayan, M., Akdeniz, D., Oktayoğlu, P., Varol, S., & Nas, K. (2015). Determining quality of life and associated factors in patients with stroke. Turkiye Fiziksel Tip ve Rehabilitasyon Dergisi, 61(2), 148–154. https://doi.org/10.5152/tftrd.2015.80090

Goh, H.-T., Tan, M.-P., Mazlan, M., Abdul-Latif, L., & Subramaniam, P. (2019). Social Participation Determines Quality of Life Among Urban-Dwelling Older Adults With Stroke in a Developing Country. Journal of Geriatric Physical Therapy, 42(4), E77–E84. https://doi.org/10.1519/JPT.0000000000000196

Ingles, J. L., Eskes, G. A., & Phillips, S. J. (1999). Fatigue after stroke. Archives of Physical Medicine and Rehabilitation, 80(2), 173–178. https://doi.org/10.1016/S0003-9993(99)90116-8

Khedr, E. M., Abdelrahman, A. A., Desoky, T., Zaki, A. F., & Gamea, A. (2020). Post-stroke depression: frequency, risk factors, and impact on quality of life among 103 stroke patients—hospital-based study. Egyptian Journal of Neurology, Psychiatry & Neurosurgery, 56(1), 1–8. http://10.0.4.162/s41983-020-00199-8

Kong, K. H., & Yang, S. Y. (2006). Health-related quality of life among chronic stroke survivors attending a rehabilitation clinic. Singapore Medical Journal, 47(3), 213–218. https://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=16518556&site=ehost-live

Maaijwee, N. A. M. M., Arntz, R. M., Rutten-Jacobs, L. C. A., Schaapsmeerders, P., Schoonderwaldt, H. C., van Dijk, E. J., & de Leeuw, F.-E. (2015). Post-stroke fatigue and its association with

poor functional outcome after stroke in young adults. Journal of Neurology, Neurosurgery & Psychiatry, 86(10), 1120–1126. https://doi.org/10.1136/jnnp-2014-308784

Myhre, P., Radakovic, R., & Ford, C. (2022). Validation of the self-rated Dimensional Apathy Scale in community stroke survivors. Journal of the Neurological Sciences, 434, 120103. https://doi.org/10.1016/J.JNS.2021.120103

Naess, H., Waje-Andreassen, U., Thomassen, L., Nyland, H., Myhr, K.-M., Naess, H., Waje-Andreassen, U., Thomassen, L., Nyland, H., & Myhr, K.-M. (2006). Health-related quality of life among young adults with ischemic stroke on long-term follow-up. Stroke (00392499), 37(5), 1232–1236. https://doi.org/10.1161/01.str.0000217652.42273.02

Ojagbemi, A., Bello, T., & Elugbadebo, F. (2019). Suicidal thoughts and contexts in Black African stroke survivors. Journal of Geriatric Psychiatry and Neurology, 32(2), 74–80. https://doi.org/10.1177/0891988718824035

Oni, O. D., Aina, O. F., Ojini, F. I., & Olisah, V. O. (2016). Quality of life and associated factors among poststroke clinic attendees at a University Teaching Hospital in Nigeria. Nigerian Medical Journal, 57(5), 290–298. https://doi.org/10.4103/0300-1652.190602

Pearce, S. C., Stolwyk, R. J., New, P. W., & Anderson, C. (2016). Sleep disturbance and deficits of sustained attention following stroke. Journal of Clinical and Experimental Neuropsychology, 38(1), 1–11. https://doi.org/10.1080/13803395.2015.1078295

Powell, J., Kitchen, N., Heslin, J., & Greenwood, R. (2004). Psychosocial outcomes at 18 months after good neurological recovery from aneurysmal subarachnoid haemorrhage. Journal of Neurology, Neurosurgery & Psychiatry, 75(8), 1119–1124. http://10.0.4.112/jnnp.2002.000414

Sterr, A., Herron, K., Dijk, D.-J., & Ellis, J. (2008). Time to wake-up: Sleep problems and daytime sleepiness in long-term stroke survivors. Brain Injury, 22(7/8), 575–579. http://10.0.4.56/02699050802189727

Todorov, V., Dimitrova, M., Todorova, V., & Mihaylova, E. (2020). Assessment of Anxiety and Depressive Symptoms in the Early Post-stroke Period. Folia Medica, 62(4), 695–702. https://doi.org/10.3897/folmed.62.e49453

Vuletić, V., Lezaić, Z., & Morović, S. (2011). Post-stroke fatigue. Acta Clinica Croatica, 50(3), 341–344. https://search.ebscohost.com/login.aspx?direct=true&db=mdc&AN=22384767&site=ehost-live

Yilmaz, H., Gumus, H., Yilmaz, S. D., Akkurt, H. E., & Odabas, F. O. (2017). The evaluation of sexual function in women with stroke. Neurology India, 65(2), 271–276. https://doi.org/10.4103/neuroindia.NI_1102_15

Yule, M. (2013). Emotion regulation, executive functioning and quality of life following stroke : a research portfolio.

**Comparison Group Had Physical Or Neurological Health Problems**

Batur, E. B., Ozyemisçi-Taskiran, O., Yuksel, S., Cengiz, M., & Karatas, G. K. (2021). Validity and reliability of the fatigue impact scale in stroke. Topics in Stroke Rehabilitation. https://doi.org/10.1080/10749357.2021.1978629

Cully, J. A., Gfeller, J. D., Heise, R. A., Ross, M. J., Teal, C. R., & Kunik, M. E. (2005). Geriatric depression, medical diagnosis, and functional recovery during acute rehabilitation. Archives of Physical Medicine and Rehabilitation, 86(12), 2256–2260. https://doi.org/10.1016/j.apmr.2005.07.292

Dam, H. (2001). Depression in stroke patients 7 years following stroke. Acta Psychiatrica Scandinavica, 103(4), 287–293. http://10.0.4.10/j.1600-0447.2001.103004287.x

Francoeur, R. B. (2016). Symptom profiles of subsyndromal depression in disease clusters of diabetes, excess weight, and progressive cerebrovascular conditions: a promising new type of finding from a reliable innovation to estimate exhaustively specified multiple indicators-multi. Diabetes, Metabolic Syndrome and Obesity : Targets and Therapy, 9, 391–416. https://doi.org/10.2147/DMSO.S118432

Mast, B. T. (2004). Cerebrovascular disease and late-life depression: a latent-variable analysis of depressive symptoms after stroke. The American Journal of Geriatric Psychiatry : Official Journal of the American Association for Geriatric Psychiatry, 12(3), 315–322. https://doi.org/10.1176/appi.ajgp.12.3.315

Quattropani, M. C., Geraci, A., Lenzo, V., Chiaie, R. D., & Filastro, A. (2018). Post-stroke anxiety and depression: Relationships to cognitive rehabilitation outcome. Clinical Neuropsychiatry: Journal of Treatment Evaluation, 15(1), 12–18. https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2018-11717-002&site=ehost-live

Yang, F. M., & Jones, R. N. (2008). Measurement differences in depression: chronic health-related and sociodemographic effects in older Americans. Psychosomatic Medicine, 70(9), 993–1004. https://doi.org/10.1097/PSY.0b013e31818ce4fa

**Pooled Stroke and Non-Stroke Data**

Brodaty, H., Sachdev, P. S., Withall, A., Altendorf, A., VALENZUELA, M. J., Lorentz, L., PS, S., Withall, A., Altendorf, A., MJ, V., Lorentz, L., Sachdev, P. S., Withall, A., Altendorf, A., VALENZUELA, M. J., Lorentz, L., PS, S., Withall, A., Altendorf, A., … Lorentz, L. (2005). Frequency and clinical, neuropsychological and neuroimaging correlates of apathy following stroke--the Sydney Stroke Study. *Psychological Medicine*, *35*(12), 1707–1716. https://doi.org/10.1017/S0033291705006173

Kellett, N., Drummond, A., Palmer, T., Munshi, S., & Lincoln, N. (2014). Impact of transient ischaemic attack and minor stroke on daily life. *International Journal of Therapy & Rehabilitation*, *21*(7), 318–323. http://10.0.50.168/ijtr.2014.21.7.318

Milak, M. S., Aniskin, D. B., Eisenberg, D. P., Prikhojan, A., Cohen, L. J., Yard, S. S., & Galynker, I. I. (2007). The negative syndrome as a dimension: Factor analyses of PANSS in major depressive disorder and organic brain disease compared with negative syndrome structures found in the schizophrenia literature. *Cognitive and Behavioral Neurology*, *20*(2), 113–120. https://doi.org/10.1097/WNN.0b013e3180653c35

Mimentza, N., González-Fraile, E., Arango-Lasprilla, J. C., Ortiz-Marqués, N., Berrios, G. E., González-Pinto, A., & Quemada, J. I. (2020). Assessing irritability in patients with stroke: Psychometric properties of the Irritability Questionnaire. *Brain Injury*, *34*(1), 115–121. https://doi.org/10.1080/02699052.2019.1681513

Smith, O. R. F., van den Broek, K. C., Renkens, M., & Denollet, J. (2008). Comparison of Fatigue Levels in Patients with Stroke and Patients with End-Stage Heart Failure: Application of the Fatigue Assessment Scale. *Journal of the American Geriatrics Society*, *56*(10), 1915–1919. https://doi.org/10.1111/j.1532-5415.2008.01925.x

**Data Impossible to Extract**

Andersen, G., Vestergaard, K., Riis, J., & Lauritzen, L. (1994). Incidence of post-stroke depression during the first year in a large unselected stroke population determined using a valid standardized rating scale. *Acta Psychiatrica Scandinavica*, *90*(3), 190–195. https://doi.org/10.1111/j.1600-0447.1994.tb01576.x

da Rocha e Silva, C. E., alves Brasil, M. a., Matos do Nascimento, E., de Bragança Pereira, B., & andré, C. (2013). Is Poststroke Depression a Major Depression? *Cerebrovascular Diseases*, *35*(4), 385–391. http://10.0.4.135/000348852

Ho, L. Y. W., Lai, C. K. Y., & Ng, S. S. M. (2021). Testing the psychometric properties of the Chinese version of the Neurological Fatigue Index-Stroke. *Clinical Rehabilitation*, *35*(9), 1329–1340. https://doi.org/10.1177/02692155211001684

*The Ho et al. study reported median fatigue scores for depressed (GDS≥8) and non-depressed (GDS<8) in each group. This offers similar information to a correlation, but was excluded because of concerns that this is not a robust enough indicator of comparative correlation strength. This is because interquartile ranges were different between groups, so the size of fatigue difference between depressed and non-depressed could not be interpreted as an indicator of the strength of correlation.*

Kontou, E., Thomas, S. A., & Lincoln, N. B. (2012). Psychometric properties of a revised version of the

Li, X., & Wang, X. (2020). Relationships between stroke, depression, generalized anxiety disorder and physical disability: Some evidence from the Canadian Community Health Survey-Mental Health. *Psychiatry Research*, *290*. https://doi.org/10.1016/j.psychres.2020.113074

**No Control Group**

Coelho Gonçalves, P. R., Frias, A. C., & Dias da Fonseca, A. (2014). Emotional adjustment of patients surgically treated for aneurysmal Subarachnoid Haemorrhage (SAH). *Revista de Enfermagem Referência*, *4*(1), 15–23.

Jackson, J. (2008). *Tackling measurement issues in health predictors and outcomes using item response theory*.

**No Control for Depression Severity**

Desmond, D. W., Remien, R. H., Moroney, J. T., Stern, Y., Sano, M., & Williams, J. B. W. (2003). Ischemic stroke and depression. *Journal of the International Neuropsychological Society : JINS*, *9*(3), 429–439. https://doi.org/10.1017/S1355617703930086

**No Stroke Participants**

Jones, A., Smakowski, A., O'Connell, N., Chalder, T., & David, A. S. (2020). Functional stroke symptoms: A prospective observational case series. *Journal of Psychosomatic Research*, *132*, N.PAG-N.PAG. https://doi.org/10.1016/j.jpsychores.2020.109972

**Unavailable Full Text**

Lower, N. K. (1995). Psychological adjustment with respect to cognitive and physical loss [ProQuest Information & Learning]. In *Dissertation Abstracts International: Section B: The Sciences and Engineering* (Vol. 55, Issues 8-B).

## Appendix D: Quality rating by study for the Systematic Review

| | Gainotti et al. (1999) | Gainotti et al. (1997) | House et al. (1991) | Lipsey et al. (1986) | Cumming et al. (2010) | de Man-van Ginkel et al. (2015) | Bennett et al. (2006) | Fleming et al. (2021) | Schramke et al. (1998) | Stokes et al. (2011) | Vickery et al. (2008) | Pickard et al. (2006) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Was the study question or objective clearly stated? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 2. Was the study population clearly specified and defined? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 3. Was the participation rate of eligible persons at least 50%? | Yes | NR | Yes | NR | Yes | NR | NR | Yes | NR | NR | NR | NR |
| 4. Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion crtieria for being in the study prespecified and applied uniformly to all participants? | NR | Yes | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | No |
| 5. Was the sample size justification, power description, or variance and effect estimates provided? | No | No | No | No | No | No | No | No | No | No | No | Yes |
| 6. For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured? [Exposure = Stroke] | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 7. Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed? | Yes | Yes | Yes | Yes | Yes | Yes | No | Yes | CD | Yes | No | Yes |
| 8. For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (e.g., categories of exposure, or exposure measured as continuous variable)? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 9. Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 10. Was the exposure(s) assessed more than once over time? [replaced with outcome assessed over time] | No | Yes | Yes | No | No | No | No | No | No | No | No | No |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11. Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? | No | No | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| 12. Were the outcome assessors blinded to the exposure status of participants? | Yes | NR | NR | NR | NR | NR | NR | NR | NR | NR | NR | Yes |
| 13. Was loss to follow-up after baseline 20% or less? | N/a | N/a | Yes | N/a | N/a | N/a | N/a | N/a | N/a | N/a | N/a | N/a |
| 14. Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)? | No | No | Yes | Yes | Yes | No | No | Yes | Yes | Yes | Yes | No |
| **Quality rating (Good, Fair, Poor)** | Fair to Poor | Fair to Poor | Good | Fair | Good | Fair | Fair | Fair | Fair | Fair | Fair | Fair |

| | |
|---|---|
| | Yes |
| | No |
| | CD (cannot determine) |
| | NA (not applicable) |
| | NR (not reported) |

**Guidance for Assessing the Quality of Observational Cohort and Cross-Sectional Studies**
(National Heart Lung and Blood Institute (NHLBI; 2013)

The guidance document below is organized by question number from the tool for quality assessment of observational cohort and cross-sectional studies.

Question 1. Research question

Did the authors describe their goal in conducting this research? Is it easy to understand what they were looking to find? This issue is important for any scientific paper of any type. Higher quality scientific research explicitly defines a research question.

Questions 2 and 3. Study population

Did the authors describe the group of people from which the study participants were selected or recruited, using demographics, location, and time period? If you were to conduct this study again, would you know who to recruit, from where, and from what time period? Is the cohort population free of the outcomes of interest at the time they were recruited?

An example would be men over 40 years old with type 2 diabetes who began seeking medical care at Phoenix Good Samaritan Hospital between January 1, 1990 and December 31, 1994. In this example, the population is clearly described as: (1) who (men over 40 years old with type 2 diabetes); (2) where (Phoenix Good Samaritan Hospital); and (3) when (between January 1, 1990 and December 31, 1994). Another example is women ages 34 to 59 years of age in 1980 who were in the nursing profession and had no known coronary disease, stroke, cancer, hypercholesterolemia, or diabetes, and were recruited from the 11 most populous States, with contact information obtained from State nursing boards.

In cohort studies, it is crucial that the population at baseline is free of the outcome of interest. For example, the nurses' population above would be an appropriate group in which to study incident coronary disease. This information is usually found either in descriptions of population recruitment, definitions of variables, or inclusion/exclusion criteria.

You may need to look at prior papers on methods in order to make the assessment for this question. Those papers are usually in the reference list.

If fewer than 50% of eligible persons participated in the study, then there is concern that the study population does not adequately represent the target population. This increases the risk of bias.

Question 4. Groups recruited from the same population and uniform eligibility criteria

Were the inclusion and exclusion criteria developed prior to recruitment or selection of the study population? Were the same underlying criteria used for all of the subjects involved? This issue is related to the description of the study population, above, and you may find the information for both of these questions in the same section of the paper.

Most cohort studies begin with the selection of the cohort; participants in this cohort are then measured or evaluated to determine their exposure status. However, some cohort studies may recruit or select exposed participants in a different time or place than unexposed participants, especially retrospective cohort studies–which is when data are obtained from the past (retrospectively), but the analysis examines exposures prior to outcomes. For example, one research question could be whether diabetic men with clinical depression are at higher risk for cardiovascular disease than those without clinical depression. So, diabetic men with depression might be selected from a mental health clinic, while diabetic men without depression might be selected from an internal medicine or endocrinology clinic. This study recruits groups from different clinic populations, so this example would get a "no."

However, the women nurses described in the question above were selected based on the same inclusion/exclusion criteria, so that example would get a "yes."

Question 5. Sample size justification

Did the authors present their reasons for selecting or recruiting the number of people included or analyzed? Do they note or discuss the statistical power of the study? This question is about whether or not the study had enough participants to detect an association if one truly existed.

A paragraph in the methods section of the article may explain the sample size needed to detect a hypothesized difference in outcomes. You may also find a discussion of power in the discussion section (such as the study had 85 percent power to detect a 20 percent increase in the rate of an outcome of interest, with a 2-sided alpha of 0.05). Sometimes estimates of variance and/or estimates of effect size are given, instead of sample size calculations. In any of these cases, the answer would be "yes."

However, observational cohort studies often do not report anything about power or sample sizes because the analyses are exploratory in nature. In this case, the answer would be "no." This is not a "fatal flaw." It just may indicate that attention was not paid to whether the study was sufficiently sized to answer a prespecified question–i.e., it may have been an exploratory, hypothesis-generating study.

Question 6. Exposure assessed prior to outcome measurement

This question is important because, in order to determine whether an exposure causes an outcome, the exposure must come before the outcome.

For some prospective cohort studies, the investigator enrolls the cohort and then determines the exposure status of various members of the cohort (large epidemiological studies like Framingham used this approach). However, for other cohort studies, the cohort is selected based on its exposure status, as in the example above of depressed diabetic men (the exposure being depression). Other examples include a cohort identified by its exposure to fluoridated drinking water and then compared to a cohort living in an area without fluoridated water, or a cohort of military personnel exposed to combat in the Gulf War compared to a cohort of military personnel not deployed in a combat zone.

With either of these types of cohort studies, the cohort is followed forward in time (i.e., prospectively) to assess the outcomes that occurred in the exposed members compared to nonexposed members of the cohort. Therefore, you begin the study in the present by looking at groups that were exposed (or not) to some biological or behavioral factor, intervention, etc., and then you follow them forward in time to examine outcomes. If a cohort study is conducted properly, the answer to this question should be "yes," since the exposure status of members of the cohort was determined at the beginning of the study before the outcomes occurred.

For retrospective cohort studies, the same principal applies. The difference is that, rather than identifying a cohort in the present and following them forward in time, the investigators go back in time (i.e., retrospectively) and select a cohort based on their exposure status in the past and then follow them forward to assess the outcomes that occurred in the exposed and nonexposed cohort members. Because in retrospective cohort studies the exposure and outcomes may have already occurred (it depends on how long they follow the cohort), it is important to make sure that the exposure preceded the outcome.

Sometimes cross-sectional studies are conducted (or cross-sectional analyses of cohort-study data), where the exposures and outcomes are measured during the same timeframe. As a result, cross-sectional analyses provide weaker evidence than regular cohort studies regarding a potential causal relationship between exposures and outcomes. For cross-sectional analyses, the answer to Question 6 should be "no."

Question 7. Sufficient timeframe to see an effect

Did the study allow enough time for a sufficient number of outcomes to occur or be observed, or enough time for an exposure to have a biological effect on an outcome? In the examples given above, if clinical depression has a biological effect on increasing risk for CVD, such an effect may take years. In the other example, if higher dietary sodium increases BP, a short timeframe may be sufficient to assess its association with BP, but a longer timeframe would be needed to examine its association with heart attacks.

The issue of timeframe is important to enable meaningful analysis of the relationships between exposures and outcomes to be conducted. This often requires at least several years, especially when looking at health outcomes, but it depends on the research question and outcomes being examined.

Cross-sectional analyses allow no time to see an effect, since the exposures and outcomes are assessed at the same time, so those would get a "no" response.

Question 8. Different levels of the exposure of interest

If the exposure can be defined as a range (examples: drug dosage, amount of physical activity, amount of sodium consumed), were multiple categories of that exposure assessed? (for example, for drugs: not on the medication, on a low dose, medium dose, high dose; for dietary sodium, higher than average U.S. consumption, lower than recommended consumption, between the two). Sometimes discrete categories of exposure are not used, but instead exposures are measured as continuous variables (for example, mg/day of dietary sodium or BP values).

In any case, studying different levels of exposure (where possible) enables investigators to assess trends or dose-response relationships between exposures and outcomes–e.g., the higher the exposure, the greater the rate of the health outcome. The presence of trends or dose-response relationships lends credibility to the hypothesis of causality between exposure and outcome.

For some exposures, however, this question may not be applicable (e.g., the exposure may be a dichotomous variable like living in a rural setting versus an urban setting, or vaccinated/not vaccinated with a one-time vaccine). If there are only two possible exposures (yes/no), then this question should be given an "NA," and it should not count negatively towards the quality rating.

Question 9. Exposure measures and assessment

Were the exposure measures defined in detail? Were the tools or methods used to measure exposure accurate and reliable–for example, have they been validated or are they objective? This issue is important as it influences confidence in the reported exposures. When exposures are measured with less accuracy or validity, it is harder to see an association between exposure and outcome even if one exists. Also as important is whether the exposures were assessed in the same manner within groups and between groups; if not, bias may result.

For example, retrospective self-report of dietary salt intake is not as valid and reliable as prospectively using a standardized dietary log plus testing participants' urine for sodium content. Another example is measurement of BP, where there may be quite a difference between usual care, where clinicians measure BP however it is done in their practice setting (which can vary considerably), and use of trained BP assessors using standardized equipment (e.g., the same BP device which has been tested and calibrated) and a standardized protocol (e.g., patient is seated for 5 minutes with feet flat on the floor, BP is taken twice in each arm, and all four measurements are averaged). In each of these cases, the former would get a "no" and the latter a "yes."

Here is a final example that illustrates the point about why it is important to assess exposures consistently across all groups: If people with higher BP (exposed cohort) are seen by their providers

more frequently than those without elevated BP (nonexposed group), it also increases the chances of detecting and documenting changes in health outcomes, including CVD-related events. Therefore, it may lead to the conclusion that higher BP leads to more CVD events. This may be true, but it could also be due to the fact that the subjects with higher BP were seen more often; thus, more CVD-related events were detected and documented simply because they had more encounters with the health care system. Thus, it could bias the results and lead to an erroneous conclusion.

Question 10. Repeated exposure assessment

Was the exposure for each person measured more than once during the course of the study period? Multiple measurements with the same result increase our confidence that the exposure status was correctly classified. Also, multiple measurements enable investigators to look at changes in exposure over time, for example, people who ate high dietary sodium throughout the followup period, compared to those who started out high then reduced their intake, compared to those who ate low sodium throughout. Once again, this may not be applicable in all cases. In many older studies, exposure was measured only at baseline. However, multiple exposure measurements do result in a stronger study design.

Question 11. Outcome measures

Were the outcomes defined in detail? Were the tools or methods for measuring outcomes accurate and reliable–for example, have they been validated or are they objective? This issue is important because it influences confidence in the validity of study results. Also important is whether the outcomes were assessed in the same manner within groups and between groups.

An example of an outcome measure that is objective, accurate, and reliable is death–the outcome measured with more accuracy than any other. But even with a measure as objective as death, there can be differences in the accuracy and reliability of how death was assessed by the investigators. Did they base it on an autopsy report, death certificate, death registry, or report from a family member? Another example is a study of whether dietary fat intake is related to blood cholesterol level (cholesterol level being the outcome), and the cholesterol level is measured from fasting blood samples that are all sent to the same laboratory. These examples would get a "yes." An example of a "no" would be self-report by subjects that they had a heart attack, or self-report of how much they weigh (if body weight is the outcome of interest).

Similar to the example in Question 9, results may be biased if one group (e.g., people with high BP) is seen more frequently than another group (people with normal BP) because more frequent encounters with the health care system increases the chances of outcomes being detected and documented.

Question 12. Blinding of outcome assessors

Blinding means that outcome assessors did not know whether the participant was exposed or unexposed. It is also sometimes called "masking." The objective is to look for evidence in the article that the person(s) assessing the outcome(s) for the study (for example, examining medical records to determine the outcomes that occurred in the exposed and comparison groups) is masked to the exposure status of the participant. Sometimes the person measuring the exposure is the same person conducting the outcome assessment. In this case, the outcome assessor would most likely not be blinded to exposure status because they also took measurements of exposures. If so, make a note of that in the comments section.

As you assess this criterion, think about whether it is likely that the person(s) doing the outcome assessment would know (or be able to figure out) the exposure status of the study participants. If the answer is no, then blinding is adequate. An example of adequate blinding of the outcome assessors is

to create a separate committee, whose members were not involved in the care of the patient and had no information about the study participants' exposure status. The committee would then be provided with copies of participants' medical records, which had been stripped of any potential exposure information or personally identifiable information. The committee would then review the records for prespecified outcomes according to the study protocol. If blinding was not possible, which is sometimes the case, mark "NA" and explain the potential for bias.

Question 13. Followup rate

Higher overall followup rates are always better than lower followup rates, even though higher rates are expected in shorter studies, whereas lower overall followup rates are often seen in studies of longer duration. Usually, an acceptable overall followup rate is considered 80 percent or more of participants whose exposures were measured at baseline. However, this is just a general guideline. For example, a 6-month cohort study examining the relationship between dietary sodium intake and BP level may have over 90 percent followup, but a 20-year cohort study examining effects of sodium intake on stroke may have only a 65 percent followup rate.

Question 14. Statistical analyses

Were key potential confounding variables measured and adjusted for, such as by statistical adjustment for baseline differences? Logistic regression or other regression methods are often used to account for the influence of variables not of interest.

This is a key issue in cohort studies, because statistical analyses need to control for potential confounders, in contrast to an RCT, where the randomization process controls for potential confounders. All key factors that may be associated both with the exposure of interest and the outcome–that are not of interest to the research question–should be controlled for in the analyses.

For example, in a study of the relationship between cardiorespiratory fitness and CVD events (heart attacks and strokes), the study should control for age, BP, blood cholesterol, and body weight, because all of these factors are associated both with low fitness and with CVD events. Well-done cohort studies control for multiple potential confounders.

**Some general guidance *for determining the overall quality rating of observational cohort and cross-sectional studies***

The questions on the form are designed to help you focus on the key concepts for evaluating the internal validity of a study. They are not intended to create a list that you simply tally up to arrive at a summary judgment of quality.

Internal validity for cohort studies is the extent to which the results reported in the study can truly be attributed to the exposure being evaluated and not to flaws in the design or conduct of the study–in other words, the ability of the study to draw associative conclusions about the effects of the exposures being studied on outcomes. Any such flaws can increase the risk of bias.

Critical appraisal involves considering the risk of potential for selection bias, information bias, measurement bias, or confounding (the mixture of exposures that one cannot tease out from each other). Examples of confounding include co-interventions, differences at baseline in patient characteristics, and other issues throughout the questions above. High risk of bias translates to a rating of poor quality. Low risk of bias translates to a rating of good quality. (Thus, the greater the risk of bias, the lower the quality rating of the study.)

In addition, the more attention in the study design to issues that can help determine whether there is a causal relationship between the exposure and outcome, the higher quality the study. These include exposures occurring prior to outcomes, evaluation of a dose-response gradient, accuracy of

measurement of both exposure and outcome, sufficient timeframe to see an effect, and appropriate control for confounding–all concepts reflected in the tool.

Generally, when you evaluate a study, you will not see a "fatal flaw," but you will find some risk of bias. By focusing on the concepts underlying the questions in the quality assessment tool, you should ask yourself about the potential for bias in the study you are critically appraising. For any box where you check "no" you should ask, "What is the potential risk of bias resulting from this flaw in study design or execution?" That is, does this factor cause you to doubt the results that are reported in the study or doubt the ability of the study to accurately assess an association between exposure and outcome?

The best approach is to think about the questions in the tool and how each one tells you something about the potential for bias in a study. The more you familiarize yourself with the key concepts, the more comfortable you will be with critical appraisal. Examples of studies rated good, fair, and poor are useful, but each study must be assessed on its own based on the details that are reported and consideration of the concepts for minimizing bias

## Appendix E: Submission Guidelines for the Journal of Stroke and Cardiovascular Diseases

**Introduction**

**Editors:**
José Biller, MD, FACP, FAAN, FAHA
*Journal of Stroke and Cerebrovascular Diseases*
Loyola University Chicago
Stritch School of Medicine
Department of Neurology, Bldg. 105
2160 S. 1st Avenue
Maywood, IL 60153

The goal of the *Journal of Stroke and Cerebrovascular Diseases* is to provide its readership with the highest quality material possible through a process of careful peer review and editorial comment. The Journal seeks to publish original papers on basic and clinical science related to the fields of stroke and cerebrovascular disease including review articles, controversies, methods and technical notes, selected case reports, and other original articles of a special nature. Our editorial mission is to focus on prevention and management of cerebrovascular disease. Thus, the scientific disciplines welcomed for publication will span from epidemiology to rehabilitation medicine. Another mission is to publish experimental studies from the test tube to the in vivo model whenever these approaches are applied to an understanding of the mechanisms of injury or repair of the brain and its circulation. The Journal will emphasize the physiopathology and molecular mechanisms of ischemia and hemorrhagic cell damage. Clinical papers will emphasize medical, surgical, and endovascular aspects of stroke, clinical trials and design, epidemiology, stroke care delivery systems and outcomes, imaging sciences, and rehabilitation of stroke.

**Submissions**

Authors should adhere to the following instructions for submission of manuscripts to the *Journal of Stroke and Cerebrovascular Diseases*. All manuscripts should be submitted electronically, uploading documents to the submission website (https://www.editorialmanager.com/jscvd). The system will convert documents to PDF files. Authors are encouraged to submit manuscripts in Microsoft Word. Any manuscript determined to be improperly prepared or edited can be returned to the authors without review.

All correspondence, including the Editor's decision and request for revisions, will be by e-mail. Authors may send queries concerning the submission process, manuscript status, or journal procedures to the Editorial Office at jscvd.rbiller@gmail.com. Authors unable to submit an electronic version should contact the Editorial Office to discuss alternatives.

**Ethics in publishing**

Please see our information on Ethics in publishing.

**Studies in humans and animals**

If the work involves the use of human subjects, the author should ensure that the work described has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans. The manuscript should be in line with the Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals and aim for the inclusion of representative human populations (sex, age and ethnicity) as per those recommendations. The terms sex and gender should be used correctly.

Authors should include a statement in the manuscript that informed consent was obtained for experimentation with human subjects. The privacy rights of human subjects must always be observed. All animal experiments should comply with the ARRIVE guidelines and should be carried out in accordance with the U.K. Animals (Scientific Procedures) Act, 1986 and associated guidelines, EU Directive 2010/63/EU for animal experiments, or the National Research Council's Guide for the Care and Use of Laboratory Animals and the authors should clearly indicate in the manuscript that such guidelines have been followed. The sex of animals must be indicated, and where appropriate, the influence (or association) of sex on the results of the study.

**Informed consent and patient details**

Studies on patients or volunteers require ethics committee approval and informed consent, which should be documented in the paper. Appropriate consents, permissions and releases must be obtained where an author wishes to include case details or other personal information or images of patients and any other individuals in an Elsevier publication. Written consents must be retained by the author but copies should not be provided to the journal. Only if specifically requested by the journal in exceptional circumstances (for example if a legal issue arises) the author must provide copies of the consents or evidence that such consents have been obtained. For more information, please review the Elsevier Policy on the Use of Images or Personal Information of Patients or other Individuals. Unless you have written permission from the patient (or, where applicable, the next of kin), the personal details of any patient included in any part of the article and in any supplementary materials (including all illustrations and videos) must be removed before submission.

**Declaration of interest**

All authors must disclose any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work. Examples of potential competing interests include employment, consultancies, stock ownership, honoraria, paid expert testimony, patent applications/registrations, and grants or other funding. Authors must disclose any interests in two places: 1. A summary declaration of interest statement in the title page file (if double anonymized) or the manuscript file (if single anonymized). If there are no interests to declare then please state this: 'Declarations of interest: none'. 2. Detailed disclosures as part of a separate Declaration of Interest form, which forms part of the journal's official records. It is important for potential interests to be declared in both places and that the information matches. More information.

**Use of inclusive language**

Inclusive language acknowledges diversity, conveys respect to all people, is sensitive to differences, and promotes equal opportunities. Content should make no assumptions about the beliefs or commitments of any reader; contain nothing which might imply that one individual is superior to another on the grounds of age, gender, race, ethnicity, culture, sexual orientation, disability or health condition; and use inclusive language throughout. Authors should ensure that writing is free from bias, stereotypes, slang, reference to dominant culture and/or cultural assumptions. We advise to seek gender neutrality by using plural nouns ("clinicians, patients/clients") as default/wherever possible to avoid using "he, she," or "he/she." We recommend avoiding the use of descriptors that refer to personal attributes such as age, gender, race, ethnicity, culture, sexual orientation, disability or health condition unless they are relevant and valid. When coding terminology is used, we recommend to avoid offensive or exclusionary terms such as "master", "slave", "blacklist" and "whitelist". We suggest using alternatives that are more appropriate and (self-) explanatory such as "primary", "secondary", "blocklist" and "allowlist". These guidelines are meant as a point of reference to help identify appropriate language but are by no means exhaustive or definitive.

**Changes to authorship**
Authors are expected to consider carefully the list and order of authors **before** submitting their manuscript and provide the definitive list of authors at the time of the original submission. Any addition, deletion or rearrangement of author names in the authorship list should be made only **before** the manuscript has been accepted and only if approved by the journal Editor. To request such a change, the Editor must receive the following from the **corresponding author**: (a) the reason for the change in author list and (b) written confirmation (e-mail, letter) from all authors that they agree with the addition, removal or rearrangement. In the case of addition or removal of authors, this includes confirmation from the author being added or removed.
Only in exceptional circumstances will the Editor consider the addition, deletion or rearrangement of authors **after** the manuscript has been accepted. While the Editor considers the request, publication of the manuscript will be suspended. If the manuscript has already been published in an online issue, any requests approved by the Editor will result in a corrigendum.

*Registration of clinical trials*
Registration in a public trials registry is a condition for publication of clinical trials in this journal in accordance with International Committee of Medical Journal Editors recommendations. Trials must register at or before the onset of patient enrolment. The clinical trial registration number should be included at the end of the abstract of the article. A clinical trial is defined as any research study that prospectively assigns human participants or groups of humans to one or more health-related interventions to evaluate the effects of health outcomes. Health-related interventions include any intervention used to modify a biomedical or health-related outcome (for example drugs, surgical procedures, devices, behavioural treatments, dietary interventions, and process-of-care changes). Health outcomes include any biomedical or health-related measures obtained in patients or participants, including pharmacokinetic measures and adverse events. Purely observational studies (those in which the assignment of the medical intervention is not at the discretion of the investigator) will not require registration.

**Copyright**
Upon acceptance of an article, authors will be asked to complete a 'Journal Publishing Agreement' (see more information on this). An e-mail will be sent to the corresponding author confirming receipt of the manuscript together with a 'Journal Publishing Agreement' form or a link to the online version of this agreement.
Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission of the Publisher is required for resale or distribution outside the institution and for all other derivative works, including compilations and translations. If excerpts from other copyrighted works are included, the author(s) must obtain written permission from the copyright owners and credit the source(s) in the article. Elsevier has preprinted forms for use by authors in these cases.
For gold open access articles: Upon acceptance of an article, authors will be asked to complete a 'License Agreement' (more information). Permitted third party reuse of gold open access articles is determined by the author's choice of user license.

***Author rights***
As an author you (or your employer or institution) have certain rights to reuse your work. More information.
*Elsevier supports responsible sharing*
Find out how you can share your research published in Elsevier journals.

**Role of the funding source**
You are requested to identify who provided financial support for the conduct of the research and/or preparation of the article and to briefly describe the role of the sponsor(s), if any, in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication. If the funding source(s) had no such involvement then this should be stated.

**Open access**
Please visit our Open Access page for more information.
*Elsevier Researcher Academy*
Researcher Academy is a free e-learning platform designed to support early and mid-career researchers throughout their research journey. The "Learn" environment at Researcher Academy offers several interactive modules, webinars, downloadable guides and resources to guide you through the process of writing for research and going through peer review. Feel free to use these free resources to improve your submission and navigate the publication process with ease.
*Language (usage and editing services)*
Please write your text in good English (American or British usage is accepted, but not a mixture of these). Authors who feel their English language manuscript may require editing to eliminate possible grammatical or spelling errors and to conform to correct scientific English may wish to use the English Language Editing service available from Elsevier's Author Services.

**Peer review**
This journal operates a single anonymized review process. All contributions will be initially assessed by the editor for suitability for the journal. Papers deemed suitable are then typically sent to a minimum of two independent expert reviewers to assess the scientific quality of the paper. The Editor is responsible for the final decision regarding acceptance or rejection of articles. The Editor's decision is final. Editors are not involved in decisions about papers which they have written themselves or have been written by family members or colleagues or which relate to products or services in which the editor has an interest. Any such submission is subject to all of the journal's usual procedures, with peer review handled independently of the relevant editor and their research groups. More information on types of peer review.

**Submission Requirements**
**Cover letter**
The author(s) should provide the aforementioned written assurances. In addition, author(s) are welcome to suggest three to five referees for their paper whom have expertise on the topic. Full names and professional/academic email addresses are required. Suggested reviewers may not have any conflict of interests or be affiliated with any of the authors' institutions. Editors cannot guarantee assignment of a particular reviewer to a paper. Any unusual circumstances surrounding the research or explanations for deviations from standard procedures or format should be explained in the cover letter.

**Abbreviations**
When using abbreviations, write the full name of the abbreviated item followed by the abbreviation in parentheses at the point of first mention within the body of the manuscript. Do not use abbreviations in the abstract or title of the manuscript.

**Title page**

The title page should contain the following information: (1) Full title of the manuscript, not to exceed 120 characters. (2) Full name, degrees, professional email, academic affiliation, and phone number for each author. Authors should be limited to those individuals who contributed in an important manner to the study design, data collection and analysis, or writing of the paper. (3) Department and institution where work was performed. (4) Grant support. (5) In addition, one author should be designated as a Corresponding Author to whom all communications regarding the manuscript should be directed. A full postal address should also be provided for the Corresponding Author. (6) A shortened version of the title that is 45 characters or less to be used as a running title. (7) Keywords (between 4 and 8).

**Abstract**

Original Articles, Reviews, Short Communications, and Case Studies should include a structured abstract with the following headings: Objectives, Materials and Methods, Results, Conclusions (max 250 words).

**Tables and Figures**

Figures, Graphics, and Photos should not be embedded in the manuscript text file. Please upload separate figure files. Preferred formats are EPS, TIFF, and JPEG. Please exclude it from the manuscript text file or it will appear twice in the PDF. For each Figure and Table, please include the appropriate Figure/Table number in the description field. This will allow the Editor to identify the Figure/Table in the PDF.

**Tables**

Tables must be cited in text and numbered according to order of appearance. Explanatory matter and source notations for borrowed or adapted tables should be placed in a table footnote, not in the title or table body.

**Figure legends**

Figure legends should allow interpretation of the figures without reference to the text.

**Figures**

Figures—All figures must be cited in the text and numbered in order of appearance. Computer-generated figures should use solid fills or cross-hatching, not tonal shading. Figure legends should be presented separately and placed in the manuscript after the list of references. Figure legends should be brief and not repetitive of description in the text.

**Highlights**

Highlights are optional yet highly encouraged for this journal, as they increase the discoverability of your article via search engines. They consist of a short collection of bullet points that capture the novel results of your research as well as new methods that were used during the study (if any). Please have a look at the examples here: example Highlights.

Highlights should be submitted in a separate editable file in the online submission system. Please use 'Highlights' in the file name and include 3 to 5 bullet points (maximum 85 characters, including spaces, per bullet point).

**Artwork**

*Electronic artwork*

*General points*

• Make sure you use uniform lettering and sizing of your original artwork.

• Embed the used fonts if the application provides that option.
• Aim to use the following fonts in your illustrations: Arial, Courier, Times New Roman, Symbol, or use fonts that look similar.
• Number the illustrations according to their sequence in the text.
• Use a logical naming convention for your artwork files.
• Provide captions to illustrations separately.
• Size the illustrations close to the desired dimensions of the published version.
• Submit each illustration as a separate file.
• Ensure that color images are accessible to all, including those with impaired color vision.
A detailed guide on electronic artwork is available.

**You are urged to visit this site; some excerpts from the detailed information are given here.**
*Formats*
If your electronic artwork is created in a Microsoft Office application (Word, PowerPoint, Excel) then please supply 'as is' in the native document format.
Regardless of the application used other than Microsoft Office, when your electronic artwork is finalized, please 'Save as' or convert the images to one of the following formats (note the resolution requirements for line drawings, halftones, and line/halftone combinations given below):
EPS (or PDF): Vector drawings, embed all used fonts.
TIFF (or JPEG): Color or grayscale photographs (halftones), keep to a minimum of 300 dpi.
TIFF (or JPEG): Bitmapped (pure black & white pixels) line drawings, keep to a minimum of 1000 dpi.
TIFF (or JPEG): Combinations bitmapped line/half-tone (color or grayscale), keep to a minimum of 500 dpi.

**Please do not:**
• Supply files that are optimized for screen use (e.g., GIF, BMP, PICT, WPG); these typically have a low number of pixels and limited set of colors;
• Supply files that are too low in resolution;
• Submit graphics that are disproportionately large for the content.
*Color artwork*

Please make sure that artwork files are in an acceptable format (TIFF (or JPEG), EPS (or PDF), or MS Office files) and with the correct resolution. If, together with your accepted article, you submit usable color figures then Elsevier will ensure, at no additional charge, that these figures will appear in color online (e.g., ScienceDirect and other sites) regardless of whether or not these illustrations are reproduced in color in the printed version. **For color reproduction in print, you will receive information regarding the costs from Elsevier after receipt of your accepted article**. Please indicate your preference for color: in print or online only. Further information on the preparation of electronic artwork.

*Illustration services*
Elsevier's Author Services offers Illustration Services to authors preparing to submit a manuscript but concerned about the quality of the images accompanying their article. Elsevier's expert illustrators can produce scientific, technical and medical-style images, as well as a full range of charts, tables and graphs. Image 'polishing' is also available, where our illustrators take your image(s) and improve them to a professional standard. Please visit the website to find out more.

**References**

Cite references in text in order of appearance using Arabic numerals in parentheses for citations. Place the reference list at the end of the final text page. References should be listed in text citation number order and must be double spaced. Include the names of all authors up to a total of three before resorting to the use of "et al." All published material, including brief communications and Letters to the Editor, must be cited in the References section. References to unpublished material, such as personal communications and unpublished data, must be placed within the text and not cited in the References section. Personal communications and unpublished data must include the individual's name, location, and month and year of communication as appropriate. In the reference list, use only abbreviations approved for use in the latest edition of Index Medicus and conform style and punctuation to the requirements listed below:

*Journal article:*
Bontia R, Ford MA, Stewart AW. Predicting survival after stroke: A three-year follow-up. Stroke 1988;19:669-673.
*Book chapter:*
Whyte J, Robinson KM. Pharmacologic management. In: Glenn MB, Whyte J, eds. The practical management of spasticity in children and adults. Philadelphia: Lea & Febiger, 1990:201-226.
*Complete book:*
Brooks VB. The neural basis of motor control. New York: Oxford University Press, 1986.
*Special type of article:*
Schmidt R, Fazekas F, Horner S, et al. Lipoprotein (a) serum levels of normals are not associated with carotid atherosclerosis and microangiopathy-related cerebral damage. J Stroke Cerebrovasc Dis 1995;5:116 (abstr).

**Supplementary material**

Supplementary material such as applications, images and sound clips, can be published with your article to enhance it. Submitted supplementary items are published exactly as they are received (Excel or PowerPoint files will appear as such online). Please submit your material together with the article and supply a concise, descriptive caption for each supplementary file. If you wish to make changes to supplementary material during any stage of the process, please make sure to provide an updated file. Do not annotate any corrections on a previous version. Please switch off the 'Track Changes' option in Microsoft Office files as these will appear in the published version.

**Special Sections**

**Special sections in the Journal of Stroke and Cerebrovascular Diseases**

***Short Communications:*** A Short Communication is any short, concise report which contributes significantly to the existing literature. Please note that it differs from a Rapid Communication in that it is not expedited for peer review or publication. The length is generally 2,500 words, and should be no more than 3,500 words. There should be only a few figures or tables, and references should average around 10. There is no structured abstract required, and the Methods, Results and Discussion may be combined in a single section to help stay within the word limitation. This differs from an Original Article which requires a structured abstract, a specific format with sections, and has no specific word, reference, or figure/table limit.

***Controversies:*** When appropriate, a Controversies Section will appear several times yearly and serve as a medium for communicating controversy in the field of stroke and cerebrovascular disease.

Authors will be selected for their expertise or outspoken positions or for their objectivity and analysis of a chosen subject. The Editor of the Controversy Section will oversee the preparation of the submitted manuscripts. They or an invited Guest Editor will add their commentary.

***Rapid Communications:*** The Editor will provide the most rapid turnaround time possible in the review process. Authors may request rapid review of material found to be of critical importance to the field. The authors should accompany their manuscript with a letter stating clearly why they consider their work is appropriate for that request. Once accepted, these manuscripts will be published in the next subsequent issue of the Journal at the Editor's discretion.

***Case Reports:*** Reports of clinical interest should contain no more than 400 words of text and one Table or Figure. Only unique Case Reports which contribute to existing literature will be considered for review and possible publication. The reference list should be limited to 10. The authors should also limit their case reports to new materials, rare clinical conditions, or cases that provide insight into diagnostic or therapeutic alternatives of a disease process. They should provide their reason for the case report in a separate cover letter to the author. Most Case Reports will be e-published only.

***Letters to the Editor:*** Letters to the Editor will be considered only when it relates to original articles recently published in JSCVD. A Letter to the Editor should provide new insight, suggest corrections, offer alternative theories, or request clarification about content printed in JSCVD. Once received by JSCVD and at the direction of the Editor, selected Letters will be sent to the author (of the article being addressed) for a response. Letters and responses will be published together.

***Methodologic and Technical Reports:*** The Journal will accept the full description of selective randomized clinical trial protocols prior to publication of results. The manuscript must include comprehensive details of methodologic design. Publications are expected to enhance the science of clinical trials design and to provide extensive detail of these methods. It is hoped that authors will follow up with publication of some, if not, all aspects of the results of the clinical trial.
The Technical Note section will permit publication of innovative techniques in neurology, neurosurgery, and interventional radiology as regards the stroke and cerebrovascular disease sciences. The techniques should be novel and extensively described.

**Editorial commentary**
Every issue will be introduced by a section in which Guest Editors will comment on papers in the current Journal. These articles will be summarized, highlighting items of special interest. The commentary will discuss the importance of the paper and its relevance to the field. Controversial aspects of the selected articles will be emphasized.

**Online proof correction**
To ensure a fast publication process of the article, we kindly ask authors to provide us with their proof corrections within two days. Corresponding authors will receive an e-mail with a link to our online proofing system, allowing annotation and correction of proofs online. The environment is similar to MS Word: in addition to editing text, you can also comment on figures/tables and answer questions from the Copy Editor. Web-based proofing provides a faster and less error-prone process by allowing you to directly type your corrections, eliminating the potential introduction of errors.
If preferred, you can still choose to annotate and upload your edits on the PDF version. All

instructions for proofing will be given in the e-mail we send to authors, including alternative methods to the online version and PDF.

We will do everything possible to get your article published quickly and accurately. Please use this proof only for checking the typesetting, editing, completeness and correctness of the text, tables and figures. Significant changes to the article as accepted for publication will only be considered at this stage with permission from the Editor. It is important to ensure that all corrections are sent back to us in one communication. Please check carefully before replying, as inclusion of any subsequent corrections cannot be guaranteed. Proofreading is solely your responsibility.
Visit the Elsevier Support Center to find the answers you need. Here you will find everything from Frequently Asked Questions to ways to get in touch.
You can also check the status of your submitted article or find out when your accepted article will be published

## Appendix F: Details of Ethical Approval for the Empirical Paper

Original approval was given to access data from DEPRESSD and Greater Glasgow and Clyde (GGC) NHS Foundation Trust. Data were not collected from GGC for the purpose of this thesis, because of COVID disruption and time limitations, but this will be considered in future. Nonetheless, evidence of this approval is included below because UEA Faculty of Medicine and Health Research Ethics Committee (FMH REC) make mention of this in the amendment and approval letters.

### Greater Glasgow and Clyde (GGC) Clinical Audit Approval

Provided as PDF. Available on request.

Dr Joanne Robertson Data Protection Officer
Joanne.Robertson@ggc.scot.nhs.uk Information Governance Department
NHS Greater Glasgow & Clyde
2nd Floor, 1 Smithhills Street
Paisley PA1 1EB
Date: 28/08/2020
Enquiries to: Isobel Brown
Tel: 0141 355 2020

Email: Isobel.Brown@ggc.scot.nhs.uk

Dear Dr Robertson

**Re: Evaluation of the psychometric properties of the Patient Health Questionnaire-9 in a stroke population, using Item Response Theory (IRT) and Factor Analytic techniques**

Thank you for your Caldicott application received on 27/08/2020 regarding your proposed Service Improvement.

I have reviewed this application and can confirm that I am happy to approve this application on behalf of the Caldicott Guardian.

Please note that this approval only covers access to NHSGGC patients.

Please find attached a signed copy of your application for your records.

Yours sincerely

Isobel Brown

Data Protection Officer
Information Governance

**FMH Preliminary Approval Subject to Amendments**

Faculty of Medicine and Health Sciences Research Ethics Committee

Joshua Blake

Norwich Medical School

University of East Anglia
Norwich Research Park
Norwich
NR4 7TJ

NORWICH MEDICAL SCHOOL

Bob Champion Research & Educational Building

Rosalind Franklin Road

University of East Anglia

Norwich Research Park

4th December 2020

Dear Joshua

**Title: Evaluation of the psychometric properties of the Patient Health Questionnaire-9 in a stroke population, using Item Response Theory (IRT) and Factor Analytic techniques**

**Reference: 2020/21-046**

The submission of your research proposal was considered by the Faculty Research Ethics Committee at its meeting on 26th November 2020.

The Committee is happy to approve your application in principle but has the following concerns which it would like you to address please:

1. The Committee needs to see the ethics approvals for both studies – Greater Glasgow and DEPRESSED.
2. The statement from the DEPRESSED data controllers as to the status of the data is insufficient, please provide further detail.

Please write to me once you have addressed the above issues. The Committee has requested that you detail the changes below the relevant point on the text in this letter and also include your amendments as tracked changes within your application. The revisions can be considered by Chair's action so please email them to fmh.ethics@uea.ac.uk at any time.

As your project does not have ethics approval until the above issues have been resolved, I want to remind you that you should not be undertaking your research project until you have ethical approval from the Faculty Research Ethics Committee. Planning on the project or literature-based elements can still take place but not the research involving these ethical issues. This is to ensure that you and your research are insured by the University and that your research is undertaken within the University's 'Guidelines on Good Practice in Research' approved by Senate in July 2015.

Yours sincerely

**Dr Jackie Buck**
**Chair, FMH Research Ethics Committee**

## Confirmation from Dr Levis of the DEPRESSD Team that Ethical Approval was not Required



### Re: PHQ-9 validity in stroke query

Brooke Levis <brooke.levis@mail.mcgill.ca>

Mon 01/03/2021 22:33

To:
Josh Blake (MED - Postgraduate Researcher) <Joshua.Blake@uea.ac.uk>

Cc:
Brett Thombs, Dr. <brett.thombs@mcgill.ca>;
Sheryl Sun <ying.sun2@mail.mcgill.ca>;
Andrea Benedetti, Dr. <andrea.benedetti@mcgill.ca>;
Fergus Gracey (MED - Staff) <F.Gracey@uea.ac.uk>;
Niall Broomfield (MED - Staff) <N.Broomfield@uea.ac.uk>;
Theresa Munyombwe <T.Munyombwe@leeds.ac.uk>

1 attachments (132 KB)
McGill PHQ-9 Research Protocol final_BL.docx;

Hi Josh,I started to make some comments in the protocol (attached), but before we proceed with more detailed feedback, we would like to clarify a few points. Firstly, regarding **ethics**: On our end, we have ethics for large pooled analyses (i.e. IPDMAs). See this statement from a recent manuscript using DEPRESSD data:

***As this study involved secondary analysis of anonymized previously collected data, the Research Ethics Committee of the Jewish General Hospital declared that this project did not require research ethics approval. However, for each included dataset, we confirmed that the original study received ethics approval and that all patients provided informed consent.***

Based on our understanding, you are seeking data from 15 PHQ studies with stroke participants, 14 PHQ studies with participants recruited from non-medical settings, and 11 HADS studies with stroke patients. If you are indeed intending to pool the data in these 3 sets of studies, then we can treat it like an IPDMA. If your institution requires additional ethical approval for each study beyond our general waiver, however, then you will have to contact the primary study authors directly to request this.

On a related note, it is important that the data be analyzed as an IPDMA, accounting for study clustering in the models, which I did not see in the protocol. This needs to be addressed.Second, regarding the **specific data** requested, we are still unclear as to which specific variables you need (e.g., PHQ-9 items, diagnostic classification, other variables (e.g., country)), and which variables are to be included in which analytic models. For instance, you mentioned in the protocol that our databases have major depression diagnoses, but it is unclear whether and how they are intended to be included in any analyses.

Thank you for clarifying these points.

Best wishes,
-Brooke
--Brooke Levis, PhD
Postdoctoral Research Fellow
Centre for Prognosis Research
School of MedicineRoom 1.103,
David Weatherall Building
Keele University
Staffordshire,
UK ST5 5BG

**Confirmation of Full Ethical Approval from FMH REC**

Faculty of Medicine and Health Sciences Research Ethics Committee

Joshua Blake
Trainee Clinical Psychologist
Department of Clinical Psychology and Psychological Therapies (CPPT)
Norwich Medical School
University of East Anglia, Norwich Research Park
Norwich, United Kingdom
NR4 7TJ

NORWICH MEDICAL SCHOOL

Bob Champion Research & Educational Building

Rosalind Franklin Road

University of East Anglia

Norwich Research Park

11 Aug 2021

Dear Josh

**Project Title: Evaluation of the psychometric properties of the Patient Health Questionnaire-9 in a stroke population, using Item Response Theory (IRT) and Factor Analytic techniques**

**Reference: 2020/21-046**

Thank you for your email of 04 Aug 2021 notifying us of the amendments you would like to make to your above proposal. These have been considered and I can confirm that your amendments have been approved.

Please can you ensure that any further amendments to either the protocol or documents submitted are notified to us in advance, and that any adverse events which occur during your project are reported to the Committee.

Approval by the FMH Research Ethics Committee should not be taken as evidence that your study is compliant with GDPR and the Data Protection Act 2018. If you need guidance on how to make your study GDPR compliant, please contact your institution's Data Protection Officer.

Please can you arrange to send us a report once your project is completed.

Yours sincerely

Dr Jackie Buck
Chair
FMH Research Ethics Committee

**Explanation of Data Transfer Process**

The DEPRESSD IPDMA exists as a 'harmonised' dataset, whereby data contributed by multiple primary authors has been processed and formatted onto a single database. This harmonised dataset is owned by the DEPRESSD team, based at McGill University.

The DEPRESSD team only had permission from the primary authors in their data sharing agreements to send their harmonised dataset to those employed by the team and those who contributed data. As such, data transfer required the primary authors to send their respective DEPRESSD-harmonised samples directly to our research team. Some authors required the signing of data transfer agreements, which we obliged with and signed. Others freely shared their data without this requirement. All data were sent transferred via email. The data do not contain directly identifiable information, such as names or dates of birth, but they do contain age, a participant ID number, sex, and PHQ-9, country of origin, and other personal information. It is highly unlikely that any intercepting party would be able to identify the individuals involved in the study. The transfer of these data was deemed to be of low ethical risk because of assurances from the DEPRESSD team that all primary data had third-party sharing permission. UEA FMH ethics were aware of this agreement and approved data transfer.

No incidences of data breach have occurred since transfer, and there has been no concern of any personal data breach of the primary author. The data have only been sent to Dr Theresa Munyombwe, one of the research team, via email, and have otherwise not been sent elsewhere. The data will be permanently deleted upon the project's completion and will only be used for the purposes set out in our ethical application and for this project. It is possible that further analysis using IRT modelling will be completed after this portfolio is submitted, but this has already been approved by UEA FMH REC.

## Appendix G: R Code for Analyses in the Empirical Paper

Some inspection of data was performed in SPSS and Excel. Below are all analyses performed in R.

**Propensity Score Matching**

```
library(lavaan)
library(semPlot)
library(tidyverse)
library(haven)
library(MatchIt)
library(dplyr)
library(ggplot2)
devtools::install_github("simsem/semTools/semTools")
library(semTools)
library(writexl)
library(WriteXLS)

PHQ_Stroke_Patten <- read_sav("PHQ_Stroke_Patten.sav")
PHQ_Stroke_Quinn <- read_sav("PHQ_Stroke_Quinn.sav")
PHQ_Stroke_Simning <- read_sav("PHQ_Stroke_Simning.sav")
PHQ_Stroke_Whooley <- read_sav("PHQ_Stroke_Whooley.sav")
PHQ_Stroke_Janneke<- read_sav("PHQ_Stroke_Janneke.sav")
PHQ_Stroke_Yamada<-read_sav("PHQ_Stroke_Yamada.sav")

PHQ_GenPop_Jeon <- read_sav("PHQ_GenPop_Jeon.sav")
PHQ_GenPop_JianLiWang <- read_sav("PHQ_GenPop_JianLiWang.sav")
PHQ_GenPop_LevinAspenson <- read_sav("PHQ_GenPop_LevinAspenson.sav")
PHQ_GenPop_Santos <- read_excel("PHQ_GenPop_Santos.xlsx")
Simning_Stroke_and_non_med_combined <- read_sav("Simning Stroke and non med combined.sav")
PHQ_GenPop_VanDerWeltz <- read_sav("PHQ_GenPop_VanDerWeltz.sav")
PHQ_GenPop_Koehler <- read_sav("PHQ_GenPop_Koehler.sav")
PHQ_GenPop_Hobfoll<-read_sav("PHQ_GenPop_Hobfoll.sav")

PHQ_GenPop_Simning<-
Simning_Stroke_and_non_med_combined[!(Simning_Stroke_and_non_med_combined$STROKE==
1),]
PHQ_GenPop_Koehler<-PHQ_GenPop_Koehler[!(PHQ_GenPop_Koehler$DIABETES==1),]
remove(Simning_Stroke_and_non_med_combined)

########## Reducing datasets into lists for a tidy environment

### Stroke whole dataset
PHQ_Stroke_Patten <- PHQ_Stroke_Patten %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Patten")
PHQ_Stroke_Quinn<- PHQ_Stroke_Quinn %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Quinn")
PHQ_Stroke_Whooley<- PHQ_Stroke_Whooley %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Whooley")
PHQ_Stroke_Simning<- PHQ_Stroke_Simning%>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Simning Stroke")
PHQ_Stroke_Janneke<- PHQ_Stroke_Janneke %>% add_column(Group="Janneke")

PHQ_Stroke_Combined<-rbind(PHQ_Stroke_Janneke,PHQ_Stroke_Patten,PHQ_Stroke_Quinn,
PHQ_Stroke_Whooley, PHQ_Stroke_Simning)
```

```
### GenPop Whole dataset
PHQ_GenPop_Jeon<-PHQ_GenPop_Jeon %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Jeon")
PHQ_GenPop_JianLiWang<-PHQ_GenPop_JianLiWang %>% add_column(TimeStr_Diag=NA)
%>% add_column(Group="JianLiWang")
PHQ_GenPop_LevinAspenson<-PHQ_GenPop_LevinAspenson %>%
add_column(TimeStr_Diag=NA) %>% add_column(Group="LevinAspenson")
PHQ_GenPop_Santos<-PHQ_GenPop_Santos %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Santos")
PHQ_GenPop_Simning<-PHQ_GenPop_Simning %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Simning")
PHQ_GenPop_VanDerWeltz<-PHQ_GenPop_VanDerWeltz %>% add_column(TimeStr_Diag=NA)
%>% add_column(Group="VanDerWeltz")
PHQ_GenPop_Koehler<-PHQ_GenPop_Koehler %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Koehler")
PHQ_GenPop_Hobfoll<- PHQ_GenPop_Hobfoll %>% add_column(TimeStr_Diag=NA) %>%
add_column(Group="Hobfoll")

PHQ_GenPop_Combined<-rbind(PHQ_GenPop_Jeon, PHQ_GenPop_JianLiWang,
PHQ_GenPop_LevinAspenson, PHQ_GenPop_Santos,
PHQ_GenPop_Simning,PHQ_GenPop_VanDerWeltz, PHQ_GenPop_Koehler,
PHQ_GenPop_Hobfoll)

### Whole dataset for PHQ
PHQ_All_Combined<-rbind(PHQ_Stroke_Janneke,PHQ_Stroke_Patten,PHQ_Stroke_Quinn,
PHQ_Stroke_Whooley, PHQ_Stroke_Simning, PHQ_GenPop_Jeon, PHQ_GenPop_JianLiWang,
PHQ_GenPop_LevinAspenson, PHQ_GenPop_Santos,
PHQ_GenPop_Simning,PHQ_GenPop_VanDerWeltz, PHQ_GenPop_Koehler,
PHQ_GenPop_Hobfoll)

## Replace stroke NAs with 0 - performed in excel for manual inspection
writexl::write_xlsx(PHQ_All_Combined, path="C:/Users/Josh/Desktop/DClinPsy/Thesis/Empirical
paper/Data/Primary R directory/Primary R Directory/PHQ_All_Combined.xlsx")
PHQ_All_Combined<-read_xlsx(path="C:/Users/Josh/Desktop/DClinPsy/Thesis/Empirical
paper/Data/Primary R directory/Primary R Directory/PHQ_All_Combined.xlsx")

## Demographic Inspection
library(ggplot2)
Plot <- ggplot(PHQ_All_Combined, aes(fill = factor(STROKE))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete("STROKE")
Plot + aes(x = EDUCATION_LEVEL)
Plot + aes(x = SEX)
Plot + aes(x = NUMBER_MEDICAL_DX)
Plot + aes(x = MARITAL_STATUS)
Plot + aes(x = COUNTRY)


#PHQ_CombinedData$STROKE<-dummy_cols(PHQ_CombinedData$STROKE)

ggplot(PHQ_All_Combined, aes(x = AGE_CONT, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")
```

```
ggplot(PHQ_All_Combined, aes(x = TOTAL, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")

## Removing cases with NAs
PHQ_All_Combined_Removed_NAs<-PHQ_All_Combined %>% drop_na(AGE_CONT,SEX)
colSums(is.na(PHQ_All_Combined_Removed_NAs ))

### Matching ratio 1
MatchedWholeData <-matchit(STROKE ~ AGE_CONT + SEX + TOTAL + COUNTRY, data =
PHQ_All_Combined_Removed_NAs)

## Ratio 2 matching
MatchedWholeData_Ratio2 <-matchit(STROKE ~ AGE_CONT + SEX + TOTAL + COUNTRY,
data = PHQ_All_Combined_Removed_NAs, ratio=2)

## Match Ratio 1 summary statistics
MatchedWholeDataSum<-summary(MatchedWholeData,standardize = TRUE)
plot(MatchedWholeData, type = "jitter", interactive = FALSE)
plot(MatchedWholeData, type = "hist")

plot(MatchedWholeDataSum)
plot(MatchedWholeData, which.xs = c("AGE_CONT","TOTAL"))

## Match Ratio 2 summary statistics
MatchedWholeDataSum_Ratio2<-summary(MatchedWholeData_Ratio2,standardize = TRUE)
plot(MatchedWholeData_Ratio2, type = "jitter", interactive = FALSE)
plot(MatchedWholeData_Ratio2, type = "hist")

plot(MatchedWholeDataSum_Ratio2)
plot(MatchedWholeData_Ratio2, which.xs = c("AGE_CONT","TOTAL"))

## Converting to matched dataframe
Matched_Data_Final<-match.data(MatchedWholeData)
Matched_Data_Final_Ratio2<-match.data(MatchedWholeData_Ratio2)

## Examining matched dataset (Ratio 1)
library(ggplot2)
Plot <- ggplot(Matched_Data_Final, aes(fill = factor(STROKE))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete("STROKE")
Plot + aes(x = EDUCATION_LEVEL)
Plot + aes(x = SEX)
Plot + aes(x = NUMBER_MEDICAL_DX)
Plot + aes(x = MARITAL_STATUS)
Plot + aes(x = COUNTRY)

ggplot(Matched_Data_Final, aes(x = AGE_CONT, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")

ggplot(Matched_Data_Final, aes(x = TOTAL, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")
```

```
writexl::write_xlsx(Matched_Data_Final, path="C:/Users/Josh/Desktop/DClinPsy/Thesis/Empirical paper/Data/Primary R directory/Primary R Directory/Matched_Data_Final.xlsx")

## Examining matched dataset (Ratio 2)
library(ggplot2)
Plot <- ggplot(Matched_Data_Final_Ratio2, aes(fill = factor(STROKE))) +
  geom_bar(position = "dodge") +
  scale_fill_discrete("STROKE")
Plot + aes(x = EDUCATION_LEVEL)
Plot + aes(x = SEX)
Plot + aes(x = NUMBER_MEDICAL_DX)
Plot + aes(x = MARITAL_STATUS)
Plot + aes(x = COUNTRY)

ggplot(Matched_Data_Final_Ratio2, aes(x = AGE_CONT, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")

ggplot(Matched_Data_Final_Ratio2, aes(x = TOTAL, fill = factor(STROKE))) +
  geom_histogram(position = "identity") +
  scale_fill_discrete("STROKE")

writexl::write_xlsx(Matched_Data_Final_Ratio2,
path="C:/Users/Josh/Desktop/DClinPsy/Thesis/Empirical paper/Data/Primary R directory/Primary R Directory/Matched_Data_Final_GenPop_Ratio2.xlsx")

###### Separating datasets for individual group analysis
## GenPop group only
Matched_Data_Final_GenPop<-Matched_Data_Final[Matched_Data_Final$STROKE==0,]
Matched_Data_Final_Stroke<-Matched_Data_Final[Matched_Data_Final$STROKE==1,]

Matched_Data_Final_GenPop_Ratio2<-
Matched_Data_Final_Ratio2[Matched_Data_Final_Ratio2$STROKE==0,]
Matched_Data_Final_Stroke_Ratio2<-
Matched_Data_Final_Ratio2[Matched_Data_Final_Ratio2$STROKE==1,]
```

**Dimensionality of Each Group Separately**

```
library(lavaan)
library(semPlot)
library(tidyverse)
library(haven)
library(MatchIt)
library(dplyr)
library(ggplot2)
devtools::install_github("simsem/semTools/semTools")
library(semTools)
library(writexl)
library(WriteXLS)


#######creating the models
## One factor
CFA_Models$GlobalDepression<-'
```

```
Depression=~PHQ9_Q1+PHQ9_Q2+PHQ9_Q3+PHQ9_Q4+PHQ9_Q5+PHQ9_Q6+PHQ9_Q7+PH
Q9_Q8+PHQ9_Q9
'
## 2 FACTOR MODEL
CFA_Models$TwoFactor<-'
    somatic=~PHQ9_Q3+PHQ9_Q4+PHQ9_Q5+PHQ9_Q7+PHQ9_Q8
    cognitive.affective=~PHQ9_Q1+PHQ9_Q2+PHQ9_Q6+PHQ9_Q9
'
## 2 alternative factor
CFA_Models$AlternateTwoFactor<-'
    somatic=~PHQ9_Q3+PHQ9_Q4+PHQ9_Q5
    cognitive.affective=~PHQ9_Q1+PHQ9_Q2+PHQ9_Q6+PHQ9_Q7+PHQ9_Q8+PHQ9_Q9
'
## Bifactor model
CFA_Models$BiFactor<-'
global=~PHQ9_Q1+PHQ9_Q2+PHQ9_Q3+PHQ9_Q4+PHQ9_Q5+PHQ9_Q6+PHQ9_Q7+PHQ9_Q
8+PHQ9_Q9
    somatic=~PHQ9_Q3+PHQ9_Q4+PHQ9_Q5+PHQ9_Q7+PHQ9_Q8
    cognitive.affective=~PHQ9_Q1+PHQ9_Q2+PHQ9_Q6+PHQ9_Q9
    global ~~ 1*global
    cognitive.affective ~~ 1*cognitive.affective
    somatic ~~ 1*somatic
    global ~~ 0*somatic
    global ~~ 0*cognitive.affective
    somatic ~~ 0*cognitive.affective
'


###### Running each model in stroke group to identify best fit
PHQ_Matched_R2_Stroke_OneFactor<-cfa(CFA_Models$GlobalDepression ,data =
Matched_Data_Final_Stroke_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_R2_Stroke_OneFactor,fit.measures=TRUE, standardized=T)

PHQ_Matched_R2_Stroke_TwoFactor<-cfa(CFA_Models$TwoFactor ,data =
Matched_Data_Final_Stroke_Ratio2, ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_R2_Stroke_TwoFactor,fit.measures=TRUE, standardized=T)

PHQ_Matched_R2_Stroke_AlternateTwoFactor<-cfa(CFA_Models$AlternateTwoFactor ,data =
Matched_Data_Final_Stroke_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_R2_Stroke_AlternateTwoFactor,fit.measures=TRUE, standardized=T)

PHQ_Matched_R2_Stroke_BiFactor<-cfa(CFA_Models$BiFactor ,data =
Matched_Data_Final_Stroke_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_R2_Stroke_BiFactor,fit.measures=TRUE, standardized=T)

############  Stroke predicted scores
Unidimensional_stroke_scores<-predict(PHQ_Matched_R2_Stroke_OneFactor)# Produces scores for
uni model
Two_factor_stroke_scores<-predict(PHQ_Matched_R2_Stroke_TwoFactor)# Produces scores for
two-factor model
```

```
Bifactor_stroke_scores<-predict(PHQ_Matched_R2_Stroke_BiFactor)# Produces scores for two-factor model

Model_predicted_stroke_scores<-
cbind(Unidimensional_stroke_scores,Two_factor_stroke_scores,Bifactor_stroke_scores)
column_names<-c("One_factor", "Two_Factor_Somatic", "Two_Factor_Affective",
"BiFactor_Global","BiFactor_Somatic","BiFactor_Affect")

colnames(Model_predicted_stroke_scores)<-column_names

plot(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,4])# Scatter plot association uni vs bifactor_global

cor.test(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,4]) # r association

plot(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,5])# Scatter plot association uni vs bifactor somatic

cor.test(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,5]) # r association

plot(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,6])# Scatter plot association uni vs bifactor affect

cor.test(Model_predicted_stroke_scores[,1],Model_predicted_stroke_scores[,6]) # r association

########## Testing matched GenPop
PHQ_Matched_Data_GenPop_OneFactor_R2<-cfa(CFA_Models$GlobalDepression ,data =
Matched_Data_Final_GenPop_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_Data_GenPop_OneFactor_R2,fit.measures=TRUE, standardized=T)

PHQ_Matched_Data_GenPop_TwoFactor_R2<-cfa(CFA_Models$TwoFactor ,data =
Matched_Data_Final_GenPop_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_Data_GenPop_TwoFactor_R2,fit.measures=TRUE, standardized=T)

PHQ_Matched_Data_GenPop_AlternateTwoFactor_R2<-cfa(CFA_Models$AlternateTwoFactor
,data = Matched_Data_Final_GenPop_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE,
std.lv = TRUE)
summary(PHQ_Matched_Data_GenPop_AlternateTwoFactor_R2,fit.measures=TRUE,
standardized=T)

PHQ_Matched_Data_GenPop_BiFactor_R2<-cfa(CFA_Models$BiFactor ,data =
Matched_Data_Final_GenPop_Ratio2,  ordered = paste0("PHQ9_Q", 1:9), std.ov = TRUE, std.lv =
TRUE)
summary(PHQ_Matched_Data_GenPop_BiFactor_R2,fit.measures=TRUE, standardized=T)
```

**Measurement Invariance Testing**

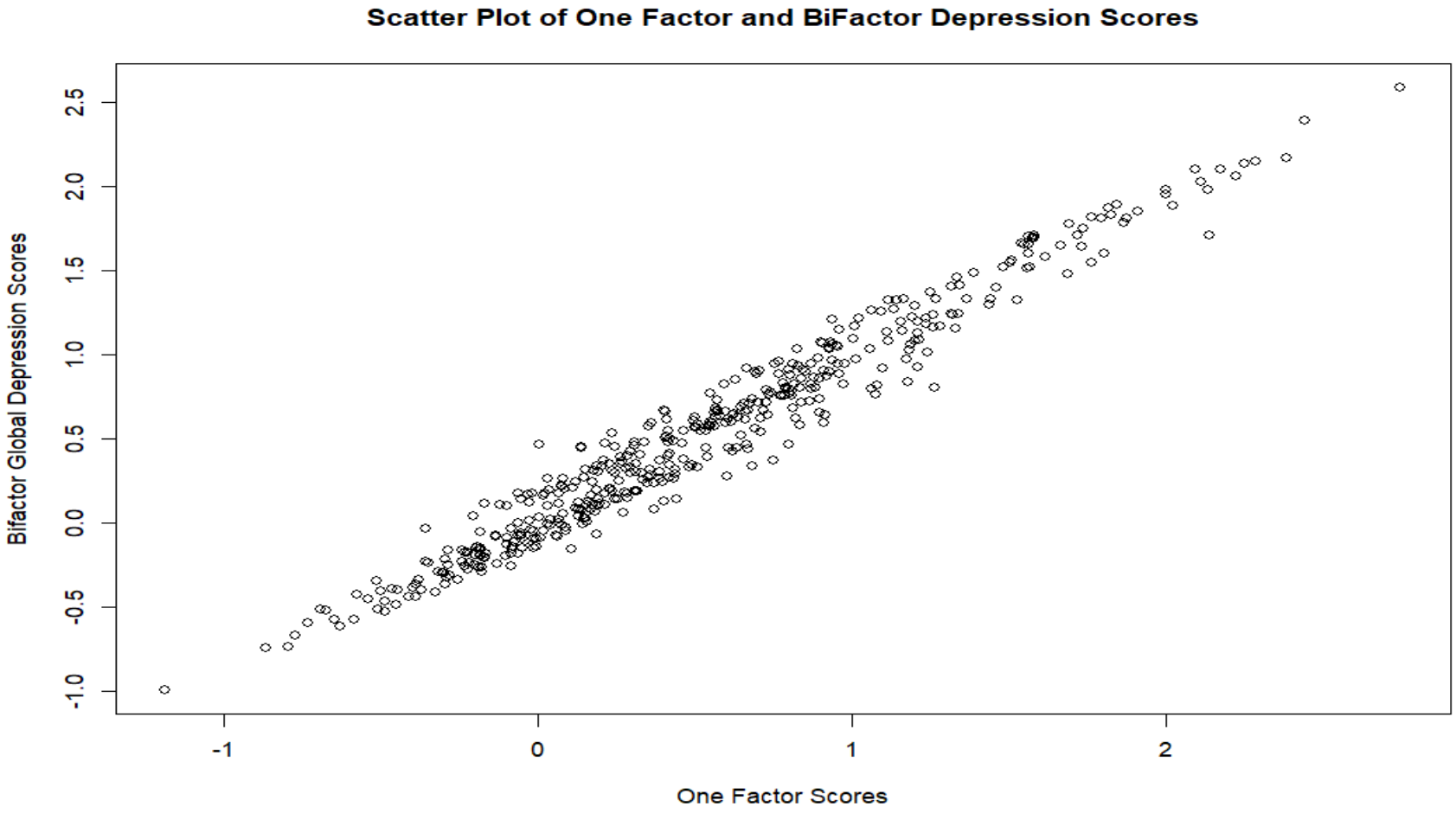
```
##### Run propensity matching script to obtain "Matched_Data_Final_Ratio 2"

### configural model
fit.config.uni_r2<- measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", return.fit=TRUE)

sum.fit.config.2<-summary(fit.config.uni_r2, fit.measures=TRUE, standardized=T)

### threshold invariance

fit.thresh.uni_r2<- measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = "thresholds",
return.fit=TRUE)

summary(fit.thresh.uni, fit.measures=TRUE, standardized=T)
anova(fit.config.uni_r2,fit.thresh.uni_r2) # MI test  ## ANOVA test of configural vs threshold

## metric invariance (weak invariance)

fit.metric.uni_r2<- measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings"),
return.fit=TRUE)

summary(fit.metric.uni_r2, fit.measures=TRUE, standardized=T)

anova(fit.thresh.uni_r2,fit.metric.uni_r2) # Measurement Invariance  (MI)ANOVA test
anova(fit.config.uni_r2,fit.metric.uni_r2)

## scalar invariance (equal intercepts)

fit.scalar.uni_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), return.fit=TRUE)

summary(fit.scalar.uni_r2, fit.measures=TRUE, standardized=T)

anova(fit.metric.uni_r2,fit.scalar.uni_r2) # MI test


## Measurement invariance failed at scalar level - unequal intercepts
Measurement_Invariance_Output.uni_r2<-
compareFit(fit.config.uni_r2,fit.thresh.uni_r2,fit.metric.uni_r2, fit.scalar.uni_r2)
Measurement_Invariance_Output_Summary.uni_r2<-
summary(Measurement_Invariance_Output.uni_r2)

MI_fit_table<-as.tibble(Measurement_Invariance_Output_Summary.uni_r2[[4]])
write_xlsx(Measurement_Invariance_Output_Summary.uni_r2[[5]],
path="C:/Users/Josh/Desktop/DClinPsy/Thesis/Empirical paper/Data/Primary R directory/Primary R
Directory/Measurement_invariance_ouput_unidimensional.xlsx")
```

```
################  Invariant intercepts analysis

Intercepts_1_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q1 ~ 1", return.fit = T)
summary(Intercepts_1_r2, fit.measures=TRUE, standardized=T)

Intercepts_2_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q2 ~ 1", return.fit = T)
summary(Intercepts_2_r2, fit.measures=TRUE, standardized=T)

Intercepts_3_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q3 ~ 1", return.fit = T)
summary(Intercepts_3_r2, fit.measures=TRUE, standardized=T)

Intercepts_4_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q4 ~ 1", return.fit = T)
summary(Intercepts_4_r2, fit.measures=TRUE, standardized=T)  ### might be suspect

Intercepts_5_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q5 ~ 1", return.fit = T)
summary(Intercepts_5_r2, fit.measures=TRUE, standardized=T) ## might be suspect

Intercepts_6_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q6 ~ 1", return.fit = T)
summary(Intercepts_6_r2, fit.measures=TRUE, standardized=T)

Intercepts_7_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q7 ~ 1", return.fit = T)
summary(Intercepts_7_r2, fit.measures=TRUE, standardized=T)

Intercepts_8_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q8 ~ 1", return.fit = T)
summary(Intercepts_8_r2, fit.measures=TRUE, standardized=T)

Intercepts_9_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression, data =
Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9), ID.fac =
"std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds", "loadings",
"intercepts"), group.partial = "PHQ9_Q9 ~ 1", return.fit = T)
```

```
summary(Intercepts_9_r2, fit.measures=TRUE, standardized=T)

summary(Interceptscompare)

Suspect_Intercepts_Free_r2<-measEq.syntax(configural.model = CFA_Models$GlobalDepression,
data = Matched_Data_Final_Ratio2, parameterization = "theta", ordered = paste0("PHQ9_Q", 1:9),
ID.fac = "std.lv", ID.cat="Wu.Estabrook.2016", group="STROKE", group.equal = c("thresholds",
"loadings", "intercepts"), group.partial = c("PHQ9_Q4 ~ 1","PHQ9_Q5 ~ 1"), return.fit = T)
summary(Suspect_Intercepts_Free_r2, fit.measures=TRUE, standardized=T)

anova(fit.metric.uni_r2,Suspect_Intercepts_Free_r2)  ## MI test to test the partial model against the
constrained loadings model

Measurement_Invariance_Output.uni_r2<-
compareFit(fit.config.uni_r2,fit.thresh.uni_r2,fit.metric.uni_r2, fit.scalar.uni_r2,
Suspect_Intercepts_Free_r2)
summary(Measurement_Invariance_Output.uni_r2)


#### Model implied scores

group_names<-c("Stroke","GenPop")

Partial_Invariant_Scores<-predict(Suspect_Intercepts_Free_r2)# Produces scores for partial model
names(Partial_Invariant_Scores)<-group_names

Partial_Invariant_Scores_Df_Stroke<-Partial_Invariant_Scores[[1]] %>%
as_tibble(Partial_Invariant_Scores_Df_Stroke)# Partial scores Stroke only
colnames(Partial_Invariant_Scores_Df_Stroke)<-"Partial_Scores"

Partial_Invariant_Scores_Df_GenPop<-Partial_Invariant_Scores[[2]] %>%
as_tibble(Partial_Invariant_Scores_Df_GenPop) # Partial scores GenPop only
colnames(Partial_Invariant_Scores_Df_GenPop)<-"Partial_Scores"

Fully_Invariant_Scores<-predict(fit.scalar.uni_r2) # Produces scores from fully invariant model
names(Fully_Invariant_Scores)<-group_names

Fully_Invariant_Scores_Df_Stroke<-Fully_Invariant_Scores[[1]] %>%
as_tibble(Fully_Invariant_Scores_Df_Stroke)# Fully inv scores Stroke only
colnames(Fully_Invariant_Scores_Df_Stroke)<-"Fully_Scores"
Fully_Invariant_Scores_Df_GenPop<-Fully_Invariant_Scores[[2]] %>%
as_tibble(Fully_Invariant_Scores_Df_GenPop)# Fully inv scores GenPop only
colnames(Fully_Invariant_Scores_Df_GenPop)<-"Fully_Scores"

Partial_and_Fully_Stroke<-
cbind(Partial_Invariant_Scores_Df_Stroke,Fully_Invariant_Scores_Df_Stroke) # Stroke partial and
fully
plot(Partial_and_Fully_Stroke) # Scatter plot association

Partial_and_Fully_GenPop<-
cbind(Partial_Invariant_Scores_Df_GenPop,Fully_Invariant_Scores_Df_GenPop) # GenPop partial
and fully
plot(Partial_and_Fully_GenPop)# Scatter plot association
cor(Partial_and_Fully_GenPop) # r association
```

```
## adding data to the main dataset
Tempgenpop<-Matched_Data_Final_Ratio2[Matched_Data_Final_Ratio2$STROKE==0,]
Tempstroke<-Matched_Data_Final_Ratio2[Matched_Data_Final_Ratio2$STROKE==1,]

Tempgenpop<-Tempgenpop %>% add_column(Partial_Invariant_Scores_Df_GenPop) %>%
add_column(Fully_Invariant_Scores_Df_GenPop)
Tempstroke<-Tempstroke %>% add_column(Partial_Invariant_Scores_Df_Stroke) %>%
add_column(Fully_Invariant_Scores_Df_Stroke)

Matched_R2_New<-bind_rows(Tempstroke,Tempgenpop)

######## Testing differences in model-implied depression scores between groups for each model - t
test
library(effsize)
?effsize

## Partial
t.test(Matched_R2_New$Partial_Scores ~ Matched_R2_New$STROKE, var.equal=F)
cohen.d(Matched_R2_New$Partial_Scores ~ Matched_R2_New$STROKE)
## Fully
t.test(Matched_R2_New$Fully_Scores ~ Matched_R2_New$STROKE, var.equal=F)
cohen.d(Matched_R2_New$Fully_Scores ~ Matched_R2_New$STROKE)
## Sum score
t.test(Matched_R2_New$TOTAL ~ Matched_R2_New$STROKE, var.equal=F)
cohen.d(Matched_R2_New$TOTAL ~ Matched_R2_New$STROKE)


with(Matched_Data_Final_Ratio2, sd(TOTAL[STROKE==0]))
with(Matched_Data_Final_Ratio2, sd(TOTAL[STROKE==1]))
cohen.d(Matched_Data_Final_Ratio2$TOTAL ~ Matched_Data_Final_Ratio2$STROKE) # d= -
.04054242

## Partial model testing and effect size
t.test(Partial_Invariant_Scores[[1]],Partial_Invariant_Scores[[2]], var.equal = F)
cohen.d(d=Partial_Invariant_Scores[[1]],f=Partial_Invariant_Scores[[2]]) # d= -.1935826

## Fully inv model testing and effect size
t.test(Fully_Invariant_Scores[[1]],Fully_Invariant_Scores[[2]], var.equal = F)
cohen.d(d=Fully_Invariant_Scores[[1]],f=Fully_Invariant_Scores[[2]]) # d= -.1125794

## Sum score difference
t.test(Matched_Data_Final_Ratio2$TOTAL ~ Matched_Data_Final_Ratio2$STROKE, var.equal=F)
with(Matched_Data_Final_Ratio2, sd(TOTAL[STROKE==0]))
with(Matched_Data_Final_Ratio2, sd(TOTAL[STROKE==1]))
cohen.d(Matched_Data_Final_Ratio2$TOTAL ~ Matched_Data_Final_Ratio2$STROKE) # d= -
.04054242
```

**Appendix H: Scatter Plot of One-factor and Bi-factor Depression scores in the Empirical Paper**




r = .9898 (CI: .9882 - .9911), p < .001,

## Appendix I: Printouts for Free-loaded Intercept CFA Models in the Empirical Paper

Only Intercept of group 2 included due to length of printouts

### PHQ_9 Q1 Intercept Freed

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (n.1.) | 0.073 | 0.075 | 0.968 | 0.333 | 0.073 | 0.044 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.053 | 0.055 | 0.959 | 0.338 | 0.052 | 0.052 |

### PHQ_9 Q2 Intercept Freed

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (n.3.) | 0.072 | 0.051 | 1.419 | 0.156 | 0.072 | 0.081 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.040 | 0.055 | 0.723 | 0.469 | 0.039 | 0.039 |

## PHQ_9 Q3 Intercept Freed

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (n.3.) | 0.072 | 0.051 | 1.419 | 0.156 | 0.072 | 0.081 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |

Factor of the PHQ-9 in Stroke

| | | | | | | |
|---|---|---|---|---|---|---|
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.040 | 0.055 | 0.723 | 0.469 | 0.039 | 0.039 |

### PHQ_9 Q4 Intercept Freed
> summary(Intercepts_4_r2, fit.measures=TRUE, standardized=T)  ### might be suspect

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (n.4.) | -0.426 | 0.061 | -7.031 | 0.000 | -0.426 | -0.446 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.205 | 0.054 | 3.799 | 0.000 | 0.221 | 0.221 |

PHQ-9 Q5 Intercept Freed

> summary(Intercepts_5_r2, fit.measures=TRUE, standardized=T) ## might be suspect

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (n.5.) | 0.358 | 0.068 | 5.278 | 0.000 | 0.358 | 0.346 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.034 | 0.054 | 0.630 | 0.529 | 0.034 | 0.034 |

PHQ-9 Q6 Intercept Freed

> summary(Intercepts_6_r2, fit.measures=TRUE, standardized=T)

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (n.6.) | 0.203 | 0.082 | 2.477 | 0.013 | 0.203 | 0.134 |

Factor of the PHQ-9 in Stroke

| | | | | | | |
|---|---|---|---|---|---|---|
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.049 | 0.054 | 0.899 | 0.368 | 0.049 | 0.049 |

## PHQ-9 Q7 Intercept Freed

> summary(Intercepts_7_r2, fit.measures=TRUE, standardized=T)

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (n.7.) | -0.143 | 0.072 | -1.989 | 0.047 | -0.143 | -0.140 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.074 | 0.054 | 1.365 | 0.172 | 0.073 | 0.073 |

## PHQ-9 Q8 Intercept Freed
> summary(Intercepts_8_r2, fit.measures=TRUE, standardized=T)
0.693 1.000

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (n.8.) | -0.017 | 0.099 | -0.177 | 0.859 | -0.017 | -0.017 |
| .PHQ9_Q9 (nu.9) | 0.000 | | | | 0.000 | 0.000 |
| Deprssn (a.1.) | 0.062 | 0.054 | 1.139 | 0.255 | 0.061 | 0.061 |

PHQ-9 Q9 Intercept Freed

> summary(Intercepts_9_r2, fit.measures=TRUE, standardized=T)

Group 2 [0]:

Intercepts:

| | Estimate | Std.Err | z-value | P(>\|z\|) | Std.lv | Std.all |
|---|---|---|---|---|---|---|
| .PHQ9_Q1 (nu.1) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q2 (nu.2) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q3 (nu.3) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q4 (nu.4) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q5 (nu.5) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q6 (nu.6) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q7 (nu.7) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q8 (nu.8) | 0.000 | | | | 0.000 | 0.000 |
| .PHQ9_Q9 (n.9.) | -0.037 | 0.166 | -0.224 | 0.823 | -0.037 | -0.027 |
| Deprssn (a.1.) | 0.061 | 0.054 | 1.128 | 0.259 | 0.060 | 0.060 |