Developing Ultra-Short Screening Tools for Depression in Stroke Survivors

Alyssa Jennifer Marion Herzig

A Thesis

In the Department

of

Psychology

Presented in Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy (Psychology) at

Concordia University

Montréal, Québec, Canada

October 2020

© Alyssa Jennifer Marian Herzig, 2020

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Alyssa Herzig

Entitled: Developing Ultra-Short Screening Tools for Depression in Stroke

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Psychology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Michael Sacher

_____ External Examiner
Dr. Lisa Koski

_____ External to Program
Dr. Robert Kilgour

_____ Examiner
Dr. Andrew Ryder

_____ Examiner
Dr. Mark Ellenbogen

_____ Thesis Co-Supervisor
Dr. Syd Miller

Approved by _____
Dr. Andrew Chapman, Graduate Program Director

November 26, 2020

_____
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Developing Ultra-Short Screening Tools for Depression in Stroke Survivors

Alyssa J. M. Herzig, Ph.D.

Concordia University, 2020

Background: National stroke guidelines recommend that all stroke survivors be screened for depression multiple times following stroke. However, compliance with screening guidelines is poor and persons with post stroke depression (PSD) remain under diagnosed and undertreated. Ultra-short PSD screening tools are recommended to increase feasibility of routine screening. Existing ultra-short depression screening tools were developed for primary care and have not shown sufficient sensitivity for screening PSD relative to gold-standard reference tests. **Objectives**: The principal objective was to determine what brief combinations of self-report mood questions can be used in ultra-short screening measures to accurately and feasibly screen for depression in stroke survivors, relative to a gold standard reference test for depression at different times post stroke. **Method:** This prospective study included a sample of 121 adults recruited from consecutive hospital admissions. At five time points within the first year following a stroke, participants were administered the current Major Depressive Episode diagnostic interview of the SCID-I/NP and completed questionnaires containing items assessing mood state, including The Stroke Specific Geriatric Depression Scale, The Mental Health Index, The Stroke Impact Scale version 3.0, and The EuroQOL five dimensions questionnaire. This author developed and followed a novel protocol for systematically comparing single items and two-item combinations for their abilities to accurately detect diagnoses of major, minor, and any depression according to the SCID-I/NP gold-standard measure. **Results and conclusions**: Results indicate that two-item screening tools containing items assessing mood state can accurately screen for depressive disorders and that the accuracy of ultra-short index tests for PSD are time-dependent. Results are consistent with previous studies suggesting that different screening questions are required at different time points following stroke. This study concludes that continued development of screening tools can contribute to minimizing the substantial negative impacts and costs of PSD through closing the gap between actual practice and best practice in PSD screening.

Acknowledgements

I would like to begin by expressing deep gratitude to my supervisor, Sydney Miller. I am thankful for your continual faith in me, your support, encouragement, and guidance, and for granting me the opportunity to pursue my passion and life's work. I view you as an ally and I am deeply grateful for your respect of my rhythms and direction throughout this process. I am deeply appreciative to Nancy Mayo at McGill University who welcomed me into her epidemiology laboratory, offered me insight into multi-disciplinary perspectives, and met me with guidance, energy, and enthusiasm throughout my Ph.D.

I would like to further express gratitude to my many clinical supervisors and professors for their faith in me and for their inspiration, encouragement, and guidance throughout my educational journey. I am also thankful to my colleagues and co-students for our shared support and encouragement as we pursued our paths together.

To my parents, Anne-Carol Herzig and Michael Herzig, and to my brother Elliot, I am deeply grateful for your endless love, encouragement, and support. To Tammer El-Sheikh and Jonathan Braunstein, thank you for your insight and for believing in me so deeply: you have been a source of strength in the tough moments of my doctoral journey.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

Acronym	Meaning
APA	American Psychiatric Association
BDI	Beck Depression Inventory
BI	Barthel Index
Brief-MMSE	Brief Version Mini-Mental State Examination
CES-D	Center for Epidemiologic Studies Depression Scale
CI	Confidence Interval
CNS	Canadian Neurological Scale
CSBPR	Canadian Stroke Best Practice Recommendations
DD-NOS	Depressive Disorder Not Otherwise Specified
DSM-IV-TR	Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, 5th ed.
EQ-5D	The EuroQOL five dimensions questionnaire
FN	False Negative
FP	False Positive
GDS	Geriatric Depression Scale
GDS-15	15-item Geriatric Depression Scale
ICD-10	International Classification of Diseases-10
K-W test	Kruskal-Wallis test
LOS	length of hospital stay
MUHC	McGill University Health Center
MDD	Major depressive disorder
MDE	Major depressive episode
MHI-5	The Mental Health Index
MIND	Minor depressive episode
MOS	Medical Outcomes Trust Study
NPV	Negative Predictive Value
OARS	Older Americans Resources and Services
OR	Odds ratio
PHQ-2	2 item Patient Health Questionnaire
PHQ-9	9 item Patient Health Questionnaire

PPV	Positive Predictive Value
PSD	Post stroke depression
QUAID	Question-understanding aid, software tool. Website: http://quaid.cohmetrix.com/
RR	Relative Risk
SAS	Statistical Analysis Software
SCID-I/NP	Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Non-Patient Edition
SD	Standard deviation
Se	Sensitivity
SF-36	Medical Outcomes Trust Study (MOS) 36-item Short Form Questionnaire
SIS	The Stroke Impact Scale
SS-GDS	The Stroke Specific Geriatric Depression Scale
STARD	Standards for Reporting Diagnostic accuracy studies
Sp	Specificity
T1…T5	Time 1 … Time 5
TN / TNR	True Negative / True Negative Rate
TP / TPR	True Positive / True Positive Rate
WHO	World Health Organization
WMW	Wilcoxon Mann-Whiney U test
Y-BOCS	Yale-Brown obsessive-compulsive scale
χ^2	Chi square

Chapter 1: Introduction

Post Stroke Depression: Scientific and Clinical Background

Post stroke depression (PSD) is a term used to describe depression occurring after a stroke event, and is a common and serious condition. The World Health Organization (2002) states that 15 million people suffer stroke worldwide each year. In 2009, the Public Health Agency of Canada reported that 50,000 Canadians suffer stroke annually, and recent research estimates that 405,000 individuals in Canada are living with the effects of stroke (Krueger et al., 2015). At any time up to five years after a stroke event, a pooled frequency of 31% of stroke survivors (95% confidence interval 28% to 35%) have been diagnosed with PSD, as determined in an updated systematic review and meta-analysis (Hackett and Pickles 2014) of 61 prospective, consecutive, observational studies of 25,488 adult stroke survivors.[1]

Impacts of Post Stroke Depression. Depression after stroke results in numerous negative outcomes for stroke survivors and their families. PSD is considered the strongest negative predictor of Quality of Life (QoL) in stroke survivors (Kim, Warren, Madill, & Hadley, 1999; King, 1996). Not only does a stroke increase the risk for depression, but depression in turn increases the disabilities associated with stroke and is considered the most critical obstacle to stroke rehabilitation (e.g., Adams & Hurwitz, 1963; Kotila, Waltimo, Niemi, Laaksonen, & Lempinen, 1984; Parikh et al., 1990). Numerous negative outcomes have been associated with PSD, including increased disability (Herrmann, Black, Lawrence, Szekely, & Szalai, 1998; Kotila, Numminen, Waltimo, & Kaste, 1999; Pohjasvaara, Vataja, Leppävuori, Kaste, & Erkinjuntti, 2001; Ramasubbu, Robinson, Flint, Kosier, & Price, 1998; Schwartz et al., 1993), increased cognitive impairment (Kauhanen et al., 1999), greater short and long term mortality (House, Knapp, Bamford, & Vail, 2001; Morris, Robinson, Andrzejewski, Samuels, & Price, 1993; Schulz et al., 2000; Williams, Ghose, & Swindle, 2004), longer hospital stays and poorer functional recovery from stroke (Clark & Smith, 1998), heightened risk of falling (Jørgensen, Engstad, & Jacobsen, 2002), less motivation to undergo stroke rehabilitation (Reynolds, 1992), poorer stroke rehabilitation outcome (Gillen, Tennen, McKee, Gernert-Dott, & Affleck, 2001; Paolucci et al., 1999, 2001; Sinyor et al., 1986; Van de Weg, Kuik, & Lankhorst, 1999), and failure to engage in recreational and social pursuits (Diller & Bishop, 1995).

Stroke is a primary source of adult disability in Canada (Krueger et al., 2015) and research in China has shown that stroke reduces social benefits and the labour force making it a

serious economic burden to patients, their families, and the government (Huang, 2010). Each year in Canada, accumulated days spent by stroke patients in acute care in Canadian hospitals is reported to be over 639,000 days, with 4.5 million days in residential care facilities (Canadian Stroke Network, 2011). The annual economic cost of stroke in Canada is approximately $3.6 billion, when both healthcare costs and loss of economic output are considered (Krueger et al., 2012). The increased disability and poorer stroke rehabilitation caused by PSD contribute to long term economic and social burden in Canada and internationally. Considering the large proportion of the aging population at risk for PSD, and the negative impact of PSD on patients and society, it is imperative that measures be taken to minimize the impact of PSD.

Risk factors associated with PSD. In general, all stroke patients are considered at high risk of developing PSD. Canadian national guidelines for stroke screening therefore require that all patients be screened, rather than only reserving screening for some stroke survivors identified as at greater risk (Lanctôt et al., 2019; Eskes et al., 2015). However, there are risk factors that are more consistently associated with PSD. Kutlubaev and Hackett (2014) published an updated systematic review and meta-analysis of 23 studies (n=18,374) with prospective consecutive recruitment of stroke survivors. The authors assessed variables associated with, or predictive of, the development of depression, and found that PSD was associated with pre-stroke depression, more severe neurological deficit, as well as physical disability in both the acute and later phases after stroke. Their research did not identify demographic parameters or stroke features (e.g., lesion location) to be consistently predictive of PSD. Ayerbe, Ayis, Wolfe and Rudd (2013) completed a systematic review and meta-analysis of 50 studies published between 1983 and 2011, in which 10 of the studies (n=16,045) reported predictors of depression after stroke. The systemic review found the major predictors of post stroke depression to include disability, pre-stroke depression, cognitive impairment, stroke severity, and anxiety. The time of the assessment ranged from the acute phase to five years post stroke.

Minimizing the impact of PSD through identification and treatment. Identification of PSD can allow stroke survivors to receive effective treatments. To date, there are several promising treatments for depression in stroke survivors. For example, evidence suggests that talk-based therapy in combinations with pharmacotherapy can promote the reduction of depressive symptoms (Alexopoulos et al., 2012; Lincoln & Flanaghan, 2003; Mitchell et al., 2009). Pharmacotherapy treatments for PSD have also been associated with reduction of

depressive symptoms (Xu et al., 2016; Chen, Guo, Zhan, & Patel, 2006; Hackett, Anderson, House, & Xia, 2008). In addition, although not yet widely used, acupuncture has been shown to effectively treat PSD. In their meta-analysis including results of 15 randomized controlled trials, Zhang, Chen, Yip, Ng, and Wong (2010) found that acupuncture was more effective than pharmacotherapy in leading to remission or recovery from PSD (OR¼1.48, 95% CI 1.10–1.97). Meta-analyses on repetitive transcranial magnetic stimulation (Shen et al., 2017) and physical activity (Eng & Reime, 2014) have shown that these approaches improve post stroke depression symptoms. Research on the effectiveness of light therapy (Søndergaard, Jarden, Martiny, Andersen, & Bech, 2006) and music therapy (Jun, Roh, & Kim, 2013; Särkämö, Tervaniemi, Laitinen, Forsblom, Soinila et al., 2008) for use in PSD are in the early stages. Untreated depression in the acute stage post stroke is associated with depression at 12 and 18 months after a stroke (Berg, Palomäki, Lehtihalmes, Lönnqvist, & Kaste, 2003).

Some researchers have proposed treating all stroke survivors for depression as a preventative measure (e.g., Jorge, Robinson, Arndt, & Starkstein, 2003). Indeed, evidence suggests that prophylactic antidepressant medication can prevent PSD symptoms in some stroke patients (Salter, Foley, Zhu, Jutai, & Teasall, 2013; Yi, Liu & Zhai, 2010). However, the Canadian Stroke Best Practice Recommendations (CSBPR) states that routine use of antidepressants to prevent post-stroke depression for all people who have experienced a stroke is not recommended because the benefits may not outweigh the risks[2], and because timing and duration of treatment has not been clarified (Eskes et al., 2015; Lanctôt et al., 2019). Similarly, the Scottish Intercollegiate Guidelines Network (SIGN, 2010) does not recommend preventative use of antidepressants for PSD. The American Veterans Health Administration and the Department of Defense state that there is conflicting evidence regarding the use of routine pharmacotherapy to prevent PSD (Management of Stroke Rehabilitation Working Group, 2010). Other current national stroke guidelines simply do not include the preventative option in their recommendations. Instead, CSBPR does recommend treatment for stroke patients diagnosed with PSD (Lanctôt et al., 2019).

Treatment of post stroke depression also plays an important role in reducing the burden on stroke survivors, their families, and society. For example, treatment and subsequent remission of PSD has been associated with improved cognitive recovery (Narushima, Chan, Kosier, & Robinson, 2003), physical recovery (Chemerinski, Robinson, Arndt, & Kosier, 2001;

Gainotti, Antonucci, Marra, & Paolucci, 2001; Robinson et al., 2000; Van de Weg et al., 1999) and functional recovery from stroke (Chemerinski et al., 2001), as well as decreased mortality (Jorge et al., 2003).

Despite the enormous prevalence of PSD internationally and the positive impact of identifying and treating the condition, PSD remains under-diagnosed and under-treated (Salter, McClure, Mahon, Foley, and Teasell, 2012; Hackett & Anderson, 2005). A key reason for the under-diagnosis of PSD is insufficient screening of PSD following stroke. Research describing insufficient PSD screening is discussed below in the section entitled 'actual and PSD screening relation to guidelines'. Therefore, to reduce negative impact of PSD through treatment, research efforts can be aimed at improving screening for PSD.

Defining PSD

In this study, PSD is defined as clinically significant depression that is identified at any time following a stroke event[3], including during the acute phase, sub-acute phase, or during the long-term follow-up. The research supporting the definition employed in this study is discussed in the section below entitled 'heterogeneous nature of PSD'. The common features of clinically significant depression, or depressive disorders in the DSM-5, is the presence of sad or irritable mood or emptiness, accompanied by somatic and cognitive changes that significantly affect the individual's capacity to function (American Psychiatric Association, 2013).

Variability in operational definition of PSD. There is substantial variability in how PSD is measured and operationally defined across studies. Different research groups employ diverse tools with varying cut-off scores to identify cases of PSD, and therefore they use different criteria to distinguish PSD from non-PSD. At least 20 distinct tools are used to assess PSD and some of the most frequently used tools have strikingly low correlations in stroke survivors (Roger & Johnson-Greene, 2009). PSD and non-PSD cases defined differently across studies contribute to variability in study results. Several sources claim that the high variability in the reported prevalence of PSD (e.g., 25% to 79% in Whyte & Mulsant, 2002; 5% to 63% in Johnson, Minarik, Nyström, Bautista, & Gorman, 2006) is partly due to different assessment tools (Andersen, Vestergaard, & Lauritzen, 1994; Collin, Tinson, & Lincoln, 1987; Stein, Sliwinski, Gordon, & Hibbard, 1996). Heterogeneity in the operational definition of PSD leads to difficulty interpreting findings across multiple studies, confused communication among practitioners, and reduced clarity in identifying which symptoms require treatment. Using

standardized diagnostic criteria for PSD may, as it does for all disorders, increase diagnostic precision, facilitate communication among health care team members, aid in the assessment of disorder severity and measurement of response to treatment, and reduce confounds in research. It is therefore a useful goal to move towards the standardization of diagnostic criteria for PSD.

Standardizing the operational definition of PSD. Employment of shared criteria and gold-standard assessment tools improves the standardization of the operational definition of PSD. There are currently two widely established systems that classify and operationally define criteria for mental health disorders. These are the Diagnostic and Statistical Manual of Mental Disorders (DSM) produced by the American Psychiatric Association (APA), and chapter 5 of the International Classification of Diseases (ICD-10) produced by World Health Organization (WHO).[4] The DSM serves as the principal authority for psychiatric diagnoses in the United States and Canada and has become a standard reference for clinical practice in the mental health field. The criteria provided by the DSM-5 are considered the gold-standard for the assessment of PSD (Hackett & Anderson, 2005; Salter, Bhogal, Foley, Jutai, & Teasell, 2007; Spalletta & Robinson 2010). There are several specific depressive disorders, with defined criteria, listed in the fifth and most recent edition of the DSM, the DSM-5 (APA, 2013). PSD can therefore be defined as meeting diagnostic criteria for a depressive disorder listed in the DSM (or ICD) in the context of stroke.

Debated criteria for PSD. The question of which criteria should be used to operationally define and diagnose PSD has been disputed for many years in PSD literature. Authors have debated whether traditional diagnostic criteria (e.g., DSM criteria) for depressive disorders should be used to operationally define and diagnose PSD. In their editorial '*How should depression be diagnosed in patients with stroke?*', Spalletta and Robinson (2010) review theory and evidence and conclude that standardized DSM-5 criteria for depressive disorders should be employed for identifying PSD with stroke patients. The authors note that diagnostic classifications are assessed with studies testing the reliability and validity of operationally defined criteria (symptoms, duration, and impairment). They review studies assessing DSM-5 diagnostic criteria for stroke patients and conclude that the phenomenology of depressive disorders appears similar regardless of etiology. Further they highlight the utility of employing gold standard criteria for depression (DSM-5) and emphasize that any alterations to standard criteria for defining depression after stroke would require substantial validation to be accepted.

Does stroke need to directly cause PSD? Robinson and Jorge (2015) propose that PSD should be defined as depression that is a direct pathophysiological consequence of a stroke. A problem with this requirement is that research has not established a distinct group of persons who become depressed post stroke with clearly biological etiological pathways. In fact, research suggests a multifactorial origin of PSD. This research is discussed below in the section on the 'multifactorial etiology of PSD'. As Spalletta and Robinson (2010) state in their editorial, the diagnostic nosology for mental disorders is not based on etiology because the etiology of psychiatric disorders is highly complex and still largely unknown. The authors specifically point out that this is the case for depressive disorders with and without stroke. In this dissertation, a defining characteristic of PSD is that it may be, but is not required to be, a direct pathophysiological consequence of stroke.

Heterogeneous Nature of PSD

The following sections provide a background on the heterogeneity of PSD in terms of time course, presentation, and etiology. This background informs our understanding of factors contributing to the effective screening of PSD.

Heterogeneity across studies. Notably, while research has been aimed at identifying meaningful and reliable statistics and patterns of association in the heterogeneous features of PSD, design variability across studies contributes to conflicting results. For example, in their updated systematic review and meta-analysis of 61 studies of PSD in adults, Hackett and Pickles (2014) found the frequency of depression varied across studies from 5% at two to five days after stroke, to 84% at three months after stroke. Substantial variability in identified rates of PSD across studies is partially attributable to design and methodological differences across studies, including different diagnostic criteria for depression (diverse definitions of depression as measured by separate tests with dissimilar cut off scores), varied assessment times (the length of time between stroke and assessment), disparate samples (e.g., different sampled populations, patient eligibility criteria, and sample size), and diverse analytic procedures. Variability in PSD rates across studies is also partially attributable to spectrum effect, meaning that because each study has a unique mix of patients, the performance (or accuracy) of a diagnostic test would also be expected to vary across studies. Irrespective of methodological differences, the expression and trajectory of PSD varies importantly across stroke survivors, as do the numerous factors that contribute to its development and maintenance.

Prevalence of PSD. Results from Hackett's and Pickles' (2014) updated systematic review indicate that prevalence rates of PSD peaked within the first-year post stroke, and although prevalence declined significantly thereafter, it remained a consistent concern within the first five years following stroke. Specifically, depression was present in 33% of participants at one-year after stroke (95% CI, 26%–39%), 25% of participants from one to under five years (95% CI, 16%–33%), and 23% of participants at five years after the stroke event (95% CI, 14%–31%), where depression was variously defined and diagnosed with a range of assessment tools. Robinson (2003) reported the pooled prevalence of PSD was 19.3% for major depression and 18.5% for minor depression among stroke patients in acute and rehabilitation hospitals based on studies published worldwide. In another meta-analysis of 43 studies, including 20,293 patients, Ayerbe, Ayis, Crichton, Wolfe, and Rudd (2013) found that the cumulative percent of patients who developed a depressive disorder within the first five years following stroke ranged from 39% to 52%.

Onset time and trajectory of PSD. Onset time of PSD is heterogeneous across participants; however, some patterns have emerged. In their systematic review of 14 cross-sectional prevalence studies, Whyte and Mulsant (2002) reported that most cases of PSD (variously defined) develop between one to six months post stroke, while fewer new cases develop after 12 months post stroke. In contrast, some studies reported finding the highest rates of depression in the first month following stroke (Aben et al., 2003, 2006; Andersen, Vestergaard, Ingemann-Nielsen, & Lauritzen, 1995; Bhogal, Teasell, Foley, & Speechley, 2004; Morrison, Pollard, Johnston, & MacWalter, 2005).

Studies with longitudinal designs provide important insights into the onset and trajectory of PSD. Paolucci et al. (2005) followed 1,064 consecutive stroke patients registered in the longitudinal DESTRO study (the Italian multicenter observational study depression in stroke). The authors reported that 36% of stroke survivors were diagnosed with depression during the nine months following stroke onset, where depression was defined as dysthymic (81%), major depression (3%), or an adaptation disorder (8%). Of the patients diagnosed with PSD, 61% first met the criteria within two to six weeks of stroke, 19% within two to three months of stroke, 13% within four to six months of stroke, and 6% within seven to nine months following stroke. Ayerbe et al. (2013) followed 4,022 patients registered in the South London Stroke Register between 1995 and 2009, assessing for depression at three months, one year, and then annually

for up to 15 years following the stroke event. The authors reported that the prevalence of depression ranged from 29% to 39% at each assessment point, whereas the cumulative incidence was 55% over the 15 years. Furthermore, the authors found that of all patients with PSD, depression was first detected at three months in 33% of participants, and most PSD cases had onset within the first year after the stroke. In addition, for all patients identified with PSD at three months, 50% had remitted for the first time at year one, 14% at year two, 8% at year three, and the remainder slowly remitted for the first time between four to nine years following a stroke event.

In their previous publication of 3,689 patients registered in the South London Stroke Register between 1995 to 2006, Ayerbe, Ayis, Rudd, Heuschmann, and Wolfe (2011) reported that 15-20% of individuals identified as depressed at each follow-up (up to five years post stroke) were new cases. Ostir, Berges, Ottenbacher, and Ottenbacher (2011) examined patterns of change of depressive symptomology in 544 survivors of a first stroke, age 65 and older. Rates of depression over the three measured time points were 28% at discharge, 22% at three months, and 18% at 12 months post stroke. The authors identified various trajectory patterns for PSD. Of the 150 patients depressed at discharge, 22% remained depressed at 3 and 12 months, 46% did not meet criteria at either follow up point, and the remainder moved in and out of depression. Of the 394 patients classified as not depressed at discharge, 4% met criteria at 3 and 12 months, 12% met criteria at 3 months only, and 7% met criteria at 12 months only.

In general, research suggests that time of PSD onset as well as the length of its course is heterogeneous across stroke survivors, with onset of first-time PSD cases highest within the first year of stroke. However, first PSD onset can occur after the first year in a smaller proportion of cases and remains a concern several years following stroke. In summary, research to date supports screening for depression at multiple time points post stroke, particularly throughout the first year of the stroke event.

Multifactorial etiology of PSD. A substantial body of research into numerous suspected etiological mechanisms of PSD has yielded conflicting findings. Currently, it is thought that numerous biological, psychological, and social/environmental factors interact to contribute to the development, expression, and maintenance of the disorder.

It has long been assumed that post stroke depression in some patients is a psychological response to disability and loss (reviewed by Spalletta et al., 2006). For example, while severity

of functional impairment is a consistent risk factor for PSD (Kutlubaev & Hackett, 2014), patients' perceptions of experienced changes (as opposed to objectively measured changes) has been described as triggering emotional and social difficulties, including loss of initiative, irritability (Hochstenbach, Prigatano, & Mulder, 2005), isolation, depression, anxiety, and frustration (Fonda, Wallace, & Herzog, 2001; Haun, Rittman, & Sberna, 2008; Turnbull, Evans, & Owen, 2005; Zawadzka & Domańska, 2014). In their systematic review of psychological etiologies of PSD, Ouimet, Primeau, and Cole (2001) reported that PSD is predicted by degree of disability, non-fluent aphasia, cognitive impairment, functional impairment, post stroke social isolation, living alone, dysphasia, personality development, neuroticism, lack of social support, disruption of social roles, dependence on others for activities of daily living, negative life events, and personal and familial history of affective and anxiety disorders. A large body of research has also forwarded at least four biological etiologies for PSD including lesion location, neurotransmitters, inflammatory cytokines, and gene polymorphism (reviewed by Fang & Cheng, 2009). In their review of research investigating biological and psychosocial mechanisms for PSD, Fang and Cheng (2009) concluded that research to date neither supports nor refutes a strictly biological model nor a strictly psychosocial model for PSD. Instead, both biological and psychosocial mechanisms appear to play a role in a multifactorial etiology of PSD. Their conclusion supports the use of a biopsychosocial model of mental illness.

Heterogeneous etiology of PSD. Several authors have interpreted that the etiological pathogenesis of PSD differs across patients (Bhogal et al., 2004; Fang & Cheng, 2009; Spalletta et al., 2006; Whyte & Mulsant, 2002). For example, in their 2006 review of research investigating the inflammatory cytokines mechanism for PSD, Spalletta et al. (2006) interpret conflicting results as support for their hypothesis that inflammatory cytokines are etiologically responsible for some, but not all, cases of PSD. In their review, Fang and Cheng (2009) further suggest that biological mechanisms may play a greater role in the development of PSD in the early stage of stroke, while psychological factors may become increasingly important in later onset PSD. Consistently, other researchers have hypothesized that biological mechanisms may be more implicated in the etiology of depression in the first months post stroke, such as damaged cerebral regions or pathways involved in the regulation of emotion. Psychological factors, on the other hand, such as difficulty adjusting to the consequences of stroke, may play a greater role in

later onset PSD (Iacoboni, Padovani, Di Piero, & Lenzi, 1995; Robinson-Smith, Johnston, & Allen, 2000).

In their systematic review, Bhogal et al. (2004) found that associations across studies between lesion location and PSD varied with patient samples, where PSD was associated with left hemisphere stroke in hospital samples, but right hemisphere stroke in community samples. The authors also reported that time of assessment after stroke moderated the relationship between lesion location and PSD. Specifically, depression always occurred in patients with left hemisphere stroke when assessed 28 days post stroke, whereas PSD assessed between one to four months post stroke occurred more frequently with right hemisphere lesions. Furthermore, PSD at six months post stroke was associated with right hemisphere lesions. In a more recent systematic review and meta-analysis on PSD and brain-lesions, Wei et al. (2015) also reported associations between a right hemisphere stroke and risk of depression in studies conducted one to six months post stroke, but not in studies conducted at other time points. In contradiction to the hypothesis that psychosocial factors play a greater role in development of depression later after stroke, Tsouna-Hadjis, Vemmos, Zakapoulos, and Stamatelopoulos (2000) found that with Greek stroke survivors, high perceived social support was significantly associated with lower depression scores at one and three months post stroke, and was not associated at six months. Overall, research suggests a multifactorial origin of PSD, with the relative contribution of factors differing between people. Further research is required to clarify the roles of moderating factors, such as assessment time and patient population subgroups, in the relationships among etiological factors. Research is also required to further examine relationships between PSD onset time and different etiological pathways.

In summary, PSD is shown to be dynamic in nature and heterogeneous across individuals. A substantial body of research indicates that multiple features of PSD are heterogeneous across individuals. These include the time point post stroke when PSD first emerges, the trajectory or time course of PSD, as well as the etiology and severity of the condition. Effective PSD screening protocols take into consideration the heterogeneous nature of PSD. Additionally, as discussed in the next section on assessment time, ongoing research suggests that the presence of PSD at different times post stroke may be associated with diverse etiological factors as well as reported symptoms, and that screening tests for PSD may perform dissimilarly based on assessment time post stroke.

Assessment Time

Assessment time and depressive symptoms. Researchers have long suggested that assessment time post stroke may be associated with different symptom profiles for PSD. Theoretically, if different onset times for PSD were associated with different etiological pathways for the condition, then onset time may also be associated with different experienced and reported symptom profiles. Paradiso, Ohkubo, and Robinson (1997) reported that major depression was characterized more by vegetative and autonomic symptoms of depression in the acute phase post stroke, yet was characterized by vegetative and psychological symptoms two years post stroke. Tateno, Kimura, and Robinson (2002) concluded that early-onset (three to six months post stroke) major depression was associated with more vegetative symptoms and larger lesion volume compared to late-onset (12 to 24 months post stroke) major depression. The authors concluded that early onset minor depression was associated with poorer social functioning, higher frequency of melancholic, vegetative, and psychological symptoms compared to late-onset minor depression. Tateno and colleagues interpreted their findings to mean that phenomenological characteristics of PSD change with time after the stroke event. Applying confirmatory factor analysis and principal components analyses with longitudinal and cross-sectional data, Bush (1998) found The Center for Epidemiological Studies-Depression Scale factor structure to change throughout the first year post stroke, and to have different factors at three months, six months, and one year post stroke. Although research to date remains speculative, and clear patterns have not yet been demonstrated, there is both evidence and theory suggesting that significantly different PSD symptoms may be associated with different assessment times after stroke. It follows then that tools used to screen for PSD may function differently at different times post stroke.

Time of assessment and PSD test accuracy. Several studies provide support for the concern that PSD screening tests and their items may perform differently at diverse times post stroke onset. Berg, Lönnqvist, Palomäki, and Kaste (2009) evaluated the accuracy of the Beck Depression Inventory (BDI) for identifying major depressive disorder (MDD) according to the DSM-III-R (American Psychiatric Association, 1987), with various cut-off scores and items in a longitudinal follow up study in 100 consecutive admissions with first stroke, at five time points post stroke onset. The authors found that sensitivity of the BDI using different cut-off scores varied with time in the period following stroke, and concluded that reliability and sensitivity of

the BDI was higher in late follow up than in the acute phase, and that lower cut-off points would be required at six months post stroke. Berg and colleagues also examined whether items from the BDI differed in their abilities to discriminate cases of MDD in the acute phase post stroke as compared to at 18 months post stroke. The authors found that in acute phase post stroke, the items "discouraged about the future," "feeling like a failure," "feeling guilty," and "looking unattractive" were important discriminators for MDD. In comparison, at 18 months post stroke, the items "discouraged about the future," "sadness," "dissatisfaction," "feeling disappointed," "loss of interest in people," and "difficulty with decisions" discriminated MDD. In an earlier study, Paradiso et al. (1997) reported that anxious foreboding, loss of libido, irritability, self-deprecation, suicidal ideation, and feelings of guilt discriminated the depressed (MDD; DSM-IV) from non-depressed patients early after stroke. Conversely, the authors reported that early morning awakening and social withdrawal better characterized depression later after the stroke event. Jackson and Alshekhlee (2015) assessed the two-item Patient Health Questionnaire (PHQ-2) and the nine-item Patient Health Questionnaire (PHQ-9) depression screening tools and reported a significant positive correlation between PHQ-2 results and the post stroke day when the tool was used. Research to date supports the argument that accurate PSD screening tools validated at one time post stroke onset should not be automatically assumed to generalize to other times post stroke.

In consideration of the aforementioned research, it follows that PSD screening should occur at multiple times, particularly throughout the first year of stroke, using screening tools validated for the time or stage after stroke when they are used. It should not be expected that a PSD screening test will perform consistently at different times after stroke. Unfortunately, Burton's and Tyson's (2015) review of diagnostic accuracy studies for PSD assessment tools revealed that a large majority of studies only examined tool accuracy at one time point, most often in the acute phase post stroke. In fact, only two of the 27 identified studies tested screening tools at more than one time point. Several of the studies identified by Burton and Tyson assessed screening tool accuracy in samples that ranged widely in post stroke assessment time, from under one month to upwards of over two years post stroke (e.g., Healey, Kneebone, Carroll, & Anderson, 2008; Parikh, Eden, Price, & Robinson, 1988; Sivrioglu et al., 2009; Turner et al., 2012). Several other studies did not report the amount of time after stroke when tools were

evaluated (Lightbody, Auton, & Baldwin, 2007; Lincoln, Nicholl, Flannaghan, Leonard, & Van der Gucht, 2003; Tham, Kwan, & Ang, 2012).

Considering that PSD assessed at different times post stroke may be associated with differing etiological mechanisms, symptom profiles, and accuracy of assessment tools, it can be argued that different questions may be required at specific time points post stroke to accurately screen for depression.

Towards Better Screening of PSD

Purpose of PSD screening in a two-step assessment process. In a two-step assessment procedure, in accordance with the CSBPR recommendations (Lanctôt et al., 2019; Eskes et al., 2015), the first step involves screening for PSD and the second step involves an assessment by a trained professional. Screening and diagnostic tests have different functions in the two-step diagnostic process. The function of screening tests is to identify individuals more likely to have a certain condition. In step one, all stroke patients are screened for possible depression at multiple times or transition points after stroke using screening tools validated with PSD samples. The tests are designed for routine mass use and should therefore be highly resource efficient. In step two, persons identified with possible depression by screening tools are thoroughly evaluated and diagnosed by a qualified health-care professional. The function of a diagnostic test is to provide a definitive diagnosis through a more extensive and more reliable evaluation. The procedures involved in the differential diagnosis, assessment of severity, and questions aimed at informing treatment plans are the business of a more complex and resource demanding assessment procedure.

Because generally only stroke patients obtaining positive results on screening tests are referred for further diagnostic evaluation and treatment, it is valuable for a PSD screening test to yield positive results for as many cases of actual PSD as possible. It is also important that a screening test yields negative results for as many non-depressed patients as possible because non-depressed patients incorrectly picked up by a screening test are also referred for further diagnostic evaluation, which further burdens the health-care system. Sensitivity (Se) and specificity (Sp) are frequently used test statistics for describing the accuracy of a screening test. The Se of a screening test is the probability that the screening test result will be positive when PSD is present, while Sp of the test is the probability that the test result will be negative when PSD is absent. For PSD screening tests, higher levels of Se are often prioritized over higher

levels of Sp, because the costs of incorrectly missing actual PSD cases (i.e., false negatives) are recognized as more detrimental than incorrectly picking up cases without PSD (i.e., false positives).

Guidelines for effective screening of PSD. PSD is underdiagnosed and undertreated, which contributes to its burden on patients, their families, and health care systems in Canada and internationally. To successfully address the stroke care issue in Canada, the Canadian Stroke Network,[5] in partnership with the Heart & Stroke charity, collaborate in producing and publishing The Canadian Stroke Best Practice Recommendations (CSBPR), which presents patient-care guidelines for implementation by health-care providers.[6] The CSBPR reviews and evaluates evidence and presents a comprehensive set of clinical practice guidelines pertaining to the screening, assessment and management of PSD (Lanctôt et al., 2019; Eskes et al., 2015). Recommended guidelines are the result of a rigorous framework outlined by Eskes and colleagues, and are each assigned a level of evidence. The CSBPR's guidelines state that diagnosis of PSD should occur in a two-step approach, first involving screening and then assessment. The CSBPR specifies that PSD screening with validated tools should occur for all stroke patients and repeated screening may be required. Screening should therefore be completed at multiple points or stages post stroke (Lanctôt et al.; Eskes et al.). Please refer to Appendix A for the list of 2015 and 2019 post stroke depression best practice recommendations. [7]

In their systematic review for the CSBPR, Lanctôt and Swartz (2019) identified and summarized six additional best practice guidelines published since 2010 which address the screening and assessment of post stroke depression. These include two guidelines published by the American Heart Association and American Stroke Association (Powers, Rabinstein, Ackerson, Adeoye, Bambakidis et al., 2018; Winstein, Stein, Arena, Bates, Cherney et al., 2016), one guideline published by the American Veterans Health Administration and the Department of Defense (Management of Stroke Rehabilitation Working Group, 2010), two guidelines published by the National Stroke Foundation in Australia (National Stroke Foundation 2017; 2010), and one published by the Intercollegiate Stroke Working Party in London (2016). Four of these six guidelines recommend stroke survivors be routinely screened for depression (Power et al.; Winstein et al.; National Stroke Foundation, 2010; Management of Stroke Rehabilitation Working Group), two of the six guidelines recommend periodic reassessment of depression

symptoms in stroke patients (Winstein et al.; Management of Stroke Rehabilitation Working Group, 2010), while one guideline states that the optimal time for PSD screening is uncertain (Powers et al.).

Actual PSD screening in relation to guidelines. As discussed here, available research indicates that actual PSD screening practices in Canada do not meet the CSBPR standards, suggesting that many stroke survivors are not being adequately screened. The CSBPR's summary of recent reports on the quality of stroke services across Canada revealed inconsistent PSD screening and monitoring in both urban and rural settings, resulting in problematic delays in its comprehensive assessment and management (reviewed by Eskes & Salter, 2013). Salter and colleagues (2012) questioned clinical staff and reviewed hospital data to identify gaps between recommendations (by the CSBPR) and actual screening, assessment, and treatment practices for PSD in 123 patients discharged from a specialized inpatient rehabilitation program in Ontario over 6 months. Salter and colleagues found poor coherence between actual practices and clinical guidelines and that the quantity of actual screening practices was very low. While 32% of patients were prescribed antidepressants at discharge, the authors found evidence of PSD screening for only 4.9% of patients, while another 9.8% of patients were referred for psychological consultation.

Notably, studies conducted in other countries also show similar gaps between national guidelines for PSD screening and actual practice. For example, the UK National Clinical Guidelines for Stroke (Intercollegiate Stroke Working Party, 2008) require that stroke patients be screened for PSD within the first month following a stroke, and that their mood be reviewed thereafter. The UK's National Sentinel Audit for the years 2001 – 2002 found that 88% of 145 hospitals with stroke services had a protocol for psychological assessment following stroke (Bowen, Knapp, Hoffman, & Lowe, 2005). However, compliance with guidelines was low as screening for PSD was shown to only be carried out with approximately 50% of stroke patients (Bowen et al., 2005; Intercollegiate Stroke Working Party, 2002, 2004). Although UK screening rates for PSD have improved between 2001 and 2010 (Bowen et al., 2005; National Audit Office, 2010), an estimated 20% of appropriate stroke patients in the UK were reportedly not screened for PSD (NSSA, 2011). Hart and Morris (2008) examined professional compliance with the UK's National Clinical Guidelines for Stroke, based on 75 hospital staff members in 16 stroke units across the UK (20% acute care, 63% rehab unit, and 17% mixed). Of the 75 participants,

only 35% were very likely to screen for PSD, 32% were very unlikely to screen, and 33% were intermediate.

In Australia, White, Towers, Turner, and Hambridge (2013) conducted semi-structured interviews on seven stroke clinicians (five neurologists and two rehabilitation physicians) and found that the clinicians rarely discussed or screened for PSD.

Barriers to PSD screening. The CSBPR guidelines require that all stroke survivors be screened, and they recommend this to occur at multiple points post stroke. Further, the CSBPR recommends that PSD screening be considered during transition points in care. Taken together, this means that screening will be required for patients in a range of settings including in acute inpatient care (stroke units and other units), in inpatient rehabilitation settings, and after discharge in other community settings. Most patients (58%) are discharged home, 19% are transferred to an inpatient rehabilitation facility, and 10% are transferred to long-term care (Canadian Stroke Network, 2011). PSD screening is completed by a variety of health care providers, including but not limited to doctors, nurses, occupational therapists, and other rehabilitation clinicians, who are stationed in various settings, and who vary in their education and training on stroke care, best practice guidelines, and PSD screening.

Research into the barriers to routine PSD screening helps to clarify the reasons for the gaps between standards and practice for PSD screening. Research conducted with health care staff in Canada and internationally has been aimed at identifying obstacles to adequately screening PSD in acute and rehabilitation settings. Some identified impediments are as follows:

1) **Time consuming screening procedures**. Hart and Morris (2008) found that time constraint was the primary negative predictor of actual PSD screening behaviour among 75 stroke clinicians in 16 stroke units across the UK. Reported time constraints was shown to reduce actual screening behaviour in staff who had positive attitudes towards PSD screening. The authors found that hospital staff used lengthy screening tests and complex procedures, with 81% of staff reporting that screening for PSD included questionnaires as well as either observations or interviews, while some staff used up to three screening tests combined with interviews and observations. Consistently, White et al. (2013) found time constraints to be the most common reported reason for infrequent screening for PSD in their semi-structured interviews with seven stroke clinicians in Australia. The consensus among these stroke clinicians was the perception that screening

for PSD is time consuming and difficult. These studies point to stroke clinicians' lack of knowledge or access to quick and simple PSD screening tools. The extensive number of recommended annual PSD screenings paired with evidence that time constraints are a primary barrier to adequate screening of PSD highlights the goal of increasing awareness of and access to validated PSD screening tools that are both simple and quick.

2) **Costly screening tools.** Some tools used to screen PSD can be accessed for free while other tools require initial and/or ongoing costs. Burton and Tyson (2015) considered only freely accessible tools as clinically feasible, noting that due to financial constraints, hospital staff opt for cheaper tools when selecting PSD screening methods. Lack of awareness of, or access to freely available appropriate tools may impede screening behaviour.
3) **Tools requiring training for delivery, scoring, or interpretation.** Schubert, Taylor, Lee, Mentari, and Tamaklo (1992) observed that the lack of PSD screening and under-diagnosis of PSD may be partially due to physicians' poor psychiatric training for such assessments. Other specialists, such as rehabilitation therapists, have been reported to lack education regarding recognizing patients with cognitive and affective disorders (Ruchinskas, 2002). Hart and Morris (2008) found that stroke staff reported lack of knowledge and training about PSD, and that low screening behaviour by staff was predicted by feelings of discomfort using available questionnaires. The perception of insufficient knowledge and training to screen for PSD is consistent with the fact that stroke staff tend to use lengthy and complex screening procedures. If staff possessed the awareness and access to valid PSD screening tools requiring no educational instruction, then lack of training may not be a barrier. In consultation with a group of occupational therapists and clinical psychologists working in stroke services in the UK, Burton and Tyson (2015) rated the clinical utility (i.e., feasibility) of tools used to screen for PSD. The authors considered that methods requiring specialist training to administer and score do not meet their criteria for good clinical utility. This research points the role of screening tools that require minimal to no training to improve screening practices.
4) **Difficulty selecting appropriate tools from the wide range available.** While more than 20 different tools are used to assess for PSD (Burton & Tyson, 2015; Roger & Johnson-Greene, 2009), in their questioning of hospital staff, Hart and Morris (2008) found no

single standard PSD screening tool. Interestingly, some of the most frequently used assessment tools have strikingly low correlations in stroke survivor populations (Roger & Johnson-Greene, 2009). Many of the tests used to screen for PSD have not been designed with stroke populations in mind or have not been adequately validated in stroke populations. Furthermore, stroke clinicians intending to screen for PSD need to select from a wide range of screening measures that vary in ease of access, the time and training required for administration, and validity in identifying PSD. Burton, Tyson and McGovern (2013) found that stroke clinicians reported a lack of knowledge and consensus for selecting appropriate stroke outcome measures as a key barrier to their general use. Alternatively, White et al. (2013) found that hospital clinicians used PSD screening tools despite their concerns about the tests' validity for PSD screening.

5) **Use of ineffective screening procedures**. Although professionals may attempt to assess or screen for PSD, use of ineffective methods can pose a barrier to identification and treatment. White et al. (2013) found that some stroke clinicians only identify severe depression with their existing screening methods, while Hart and Morris (2008) found that some professionals interpret patients' somatic mood symptoms as caused by stroke, and as unrelated to mood. This interpretation resulted in professionals not recognizing depressed mood in stroke patients. These findings point to the ineffective use of screening procedures as well as confusing screening procedures with assessment procedures, by placing too much emphasis on reducing false positives during so-called screening. This point is discussed above in the section entitled 'purpose of the two-step assessment process'. These concerns point to the need for simple, validated PSD screening tools with clear decision rules. Such tools should minimize the need for a screener to possess knowledge or training, particularly for teasing apart the complex etiology of reported symptoms.
6) **Concerns that screening will negatively impact patients.** Hart and Morris (2008) found that some stroke clinicians are concerned that PSD screening might trigger depression, and that these concerns act as a barrier to PSD screening behaviour. This concern may indicate the use of invasive questioning during screening and lack of the two-step assessment protocol in which all patients are first screened, and only patients screened as positive are given a diagnostic interview by a qualified practitioner. Indeed,

frequent use of lengthy or in-depth assessments on patients not picked up by screening protocols, as well as invasive questioning by unqualified professions, can be considered unethical for their needless use of patients' time and energy as well as for the emotional impact on the patients. These concerns can be avoided with the use of screening tools that are quick, simple, and cause little patient stress.

7) **Lack of awareness of PSD.** Walker et al. (2014) report that in medical consultations with stroke patients in Australia, depression is frequently unidentified because the relevant symptoms are normalized or are simply not discussed. Walker and colleagues and Seymour (2015) note that the tendency to not discuss PSD symptoms with patients is partly attributable to the priority being placed on medical assessments and the management of the medical condition. Salter et al. (2012) reported that 56% of the questioned clinicians in an Ontario rehabilitation facility were not aware of best practice recommendations. Hart and Morris (2008) found that 32% of their hospital staff respondents in the UK reported no awareness of the contents of their national guidelines for PSD screening.
8) **PSD screening not perceived as a priority.** There may be several reasons why stroke staff who are aware of PSD may not prioritize screening for PSD among their responsibilities, contributing to poorer screening behaviour. Notably, Hart and Morris (2008) reported no association between the profession of responding stroke staff and their intention to screen. The authors suggest that staff may have less intention to screen for PSD when: i) such screening is not included in their hospital protocol or unit policy for stroke care, ii) when there are unclear unit protocols regarding who screens and when, and iii) when staff perceive PSD screening not to be the norm amongst colleagues. Similarly, lack of inclusion of routine screening for PSD in other settings, such as rehabilitation units, would also be a barrier for routine PSD screening. Salter et al. (2012) highlight the importance of developing consistent clinical protocols and educating staff about standardized PSD screening in Ontario. Notably, staff attitudes towards PSD screening have been found to be relatively positive compared to actual screening behaviour (Hart & Morris; Salter et al., 2012). This finding emphasizes the role of other barriers, including lack of awareness and easy access to quick, simple, and free screening tools.

Considering the unique needs of stroke patients and their family members and caregivers. Patients, family members, and caregivers play an important role in PSD screening. As shown in the research outlined in the above section entitled 'actual PSD screening in relation to guidelines', actual PSD screening practices within hospital and rehabilitation facilities is not yet up to the standards outlined in best practice guidelines. Furthermore, once discharged into the community, patients' access to services where screening for PSD may take place becomes less standardized than before discharge. Stroke survivors may be dependent on themselves, their caregivers, or other specialists to identify potential PSD symptoms. With the goal of improving patient care, the CSBPR encourages direct inclusion of patients, families, and caregivers in various patient management and education processes (Lanctôt et al., 2019; Cameron, O'Connell, Foley, Salter, Booth, et al., 2016; Eskes et al., 2015; Lindsay et al., 2013). The CSBPR recommends delivery of education in all aspects of stroke care and recovery to patients, families, and caregivers so they may contribute effectively to stroke care and rehabilitation (Cameron et al.). For example, the CSBPR recommends providing patients, families, and caregivers with advocacy training for how to be active participants in care planning, decision making and in advocating on patients' behalf for PSD screening. The involvement of patients, families, and caregivers in home PSD screening processes would be consistent with the CSBPR's other recommendations for their inclusion, and may serve to further address gaps between best care and actual care.

According to interviews with stroke clinicians in Australia, clinicians already rely on insights from family members to inform their decisions pertaining to screening for PSD (White et al., 2013). However, opinions of family members may be less accurate than results of validated, self-report PSD screening tools designed to be used at home by patients, with assistance from family members and caregivers. Family members and caregivers are a support resource already existing in the process for identifying PSD, and may be more effectively empowered with appropriate home screening tools. Quick and simple PSD screening methods can be designed for administration and use at home with clear instructions for how to follow-up on screening test results.[8] Designing tools simple enough for untrained use is not a novel approach and it is consistent with the CSBPR emphasis on collaboration between diverse health-care practitioners, family members, caregivers, and patients in the continuous screening for PSD.

To improve screening practices, it is helpful that routinely used PSD screening tools be consistent with the needs of stroke patients, their families, and their caregivers. As discussed by Salter et al. (2007), depression screening tools designed for use in non-stroke populations often take more time to complete by stroke patients than the reported time requirements. Patients frequently need assistance in completing self-report questionnaires, and tests with multiple questions or with long or complex wording are time consuming and may confuse patients, their caregivers and family members. Many existing PSD screening tools that do not require training were also not designed to be administered, scored, or interpreted by patients, family members and caregivers. Such screening tools may lead to various problems when used by stroke patients, their families or care takers, including reduced use of screening tools, inaccurate results, unclear interpretations of results, and difficulty following appropriate actions for seeking further evaluation and treatment. Additionally, use of such tools needlessly increase the burden on patients, families, and caregivers.

Feasibility of Ultra-Short Screening Tools. Research described in the above section on the barriers to PSD screening shows the importance of accessible, validated tools that are highly feasible for routine use by busy screeners with minimal training. Consistently, the Stroke Improvement Programme[9] in the UK recommends that a simple and brief standardised measure be used to screen all stroke patients for mood disorders as step-one in a two-step assessment process followed by further evaluation (Gillham & Clark, 2011). Burton and Tyson (2015) developed a measure to rate the clinical utility of existing PSD screening tools based on hospital staff's priorities for selecting screening methods. These priorities were identified in consultation with a group of occupational therapists and clinical psychologists working in a range of stroke services in the UK. The identified priorities were for tools that were quick, easy, requiring minimal training, and freely available so long as accuracy ratings were similar to paid alternatives. These priorities reflected the need for screening tools that can be completed by any member of the multidisciplinary team in addition to their traditional workload. Tools that met the following criteria were identified as having acceptable clinical utility based on hospital staff's reported needs: a) time to administer and score of ≤ 5 minutes; b) access to freely available tests and required manuals; c) no additional costs per form; and d) no specialist training required. These specifications address several of the barriers to hospital staff compliance with screening standards and highlight the importance of developing feasible methods and protocols that address

the needs arising from the actual conditions in which these tools are used. Notably, the criteria by Burton and Tyson to rate tool clinical utility only seek to rate tools used by staff, and do not consider additional barriers to effective screening arising when tools are used by stroke patients, families, and caregivers. As discussed above in the section entitled 'considering the unique needs of stroke patients and their family members and caregivers,' these additional barriers, such as needing more time than average and increased difficulty with complexity, amplify the need for tools that may be simpler, shorter, quicker, and clearer than those outlined in criteria by Burton and Tyson.

Ultra-short verbal self-report screening tools have been defined as containing one to four items and requiring under two minutes to complete (Mitchell & Coyne, 2007). As discussed here, such screening tools are emphasized for their value in increasing the feasibility of depression screening by busy staff in other clinical settings. Ultra-short tools were developed to address findings that short questionnaires containing four to fifteen items and requiring two to five minutes to complete were still potentially too burdensome and were thus not being routinely used in clinical settings (reviewed by Mitchell & Coyne). To address barriers to routine screening of depression in clinical settings, the National Institute for Clinical Excellence (2004) recommended use of simple one- and two-item depression screening tools for use in primary and secondary care. Similarly, the National Clinical Guidelines for Stroke recommended a particular ultra-short screening tool, the Yale Question (Lachs et al., 1990; Mahoney et al., 1994), for screening depression in stroke patients (Intercollegiate Stroke Working Party, 2008). Mitchell and Coyne recommended use of two- or three-item depression screening, whereby persons screening positive are further evaluated in a second step, following their literature review of Standards for Reporting of Diagnostic Accuracy Studies (STARD)-compliant studies and analyses of the psychometric properties of ultra-short screening tools for depression in non-medical primary care samples.

An ultra-short PSD screening tool intended for use as step one in a two-step evaluation process is meant to yield a clear, binary decision: whether or not a patient is likely to have PSD and should be evaluated by a qualified professional. Such a tool is not meant to count clinically significant symptoms, establish the severity of a mood disorder, nor to provide a differential diagnosis. These are purposes of depression questionnaires that have been designed for other

uses, such as establishing specific diagnoses or outcomes, informing treatment, or measuring changes in symptoms.

Salter et al. (2007) identified the benefits of ultra-short screening tools to resolve the gap between best practice and actual practice in PSD screening. Simple, ultra-short self-report measures are expected to help address identified obstacles to routine screening as they require minimal costs in time and money, produce little patient burden, and minimize reliance on knowledge, training, or test interpretation skills. Routine implementation of ultra-short PSD screening tools in hospital protocols would also minimize the need for costlier interventions aimed at educating stroke staff to select their own screening tools and interpret complex symptoms.

Verbal self-report screening measures. Verbal self-report screening measures are the standard form of screening tools for PSD.[10] Verbal self-report tools are composed of language-based questions in which patients report their own mood, by completing a questionnaire or by responding verbally to questions asked or read orally. Notably, tests that rely on language, such as verbal self-report screening measures, may be difficult and inappropriate for use with stroke patients with impaired communication, for example, patients with moderate to severe aphasia.[11] To address this, tests involving visual analogue scales or pictures, as well as observer-rated measures, have been used with stroke patients unable to self-report their mood due to communication or cognitive impairments (Burton & Tyson, 2015).

Summary of screening tool specifications for effective screening. PSD screening tools that meet the following specifications are expected to contribute to improving feasibility of and compliance with PSD screening guidelines: 1) The PSD screening tool should be validated as sufficiently accurate in representative samples of appropriate stroke patients at specific time points or transition points following the stroke event. This point is discussed further below in the section entitled 'validation of PSD screening tools.' Screening test accuracy should be compared to a gold-standard reference test for PSD. The screening tests should 2) be easily accessible and free; 3) contain very brief and simple administration instructions; 4) be ultra-short including one or two simply worded questions; and 5) require minimal scoring to obtain results. Screening test results should 6) include a clear decision rule; 7) provide users with clear instructions on how to proceed (e.g., positive results indicate a need for evaluation by a qualified health-care professional); and 8) require minimal interpretation or analysis. It is expected that the

availability of tools meeting these specifications may serve to improve compliance with and accuracy of PSD screening when used with appropriate patients.[12]

Validation of PSD Screening Tools

Validating test accuracy. The validity of a PSD screening test is based on its accuracy in correctly identifying persons with and without PSD. To evaluate the accuracy of a PSD screening tool, it is compared to a gold-standard reference test for diagnosing PSD, and the tool under evaluation is referred as an index test. The gold-standard reference test is the best method for establishing the true status of the condition under assessment (e.g., Glasziou, Irwig, Bain, & Colditz, 2001; Whiting et al., 2004). Evaluating an index tool relative to a gold-standard reference test improves estimates of diagnostic precision of the index test and improves standardization of the operational definition of PSD. When evaluating an index test, all sample participants complete both the index test and the reference test to identify the accuracy with which the index test classifies participants. Some authors have suggested that minimal acceptable sensitivity should be Se ≥ 80% and specificity Sp ≥ 60% (Tyson, Burton, & McGovern, 2015; Bennett & Lincoln, 2006; Lincoln et al., 2003).

As discussed above, research indicates that the accuracy of PSD screening tools may vary with time post stroke (e.g., Berg et al., 2009). PSD may be associated with different symptoms and etiological mechanisms at different time points following stroke. Tools should, therefore, be validated for use at specific time points or transition points post stroke, rather than assuming the accuracy of a tool at one time post stroke, and generalizing to all other times post stroke.

Considering sources of bias in study design. An evaluation study lacks internal validity when it fails to measure what it purports to measure, due to bias and random error arising from the study design (Schmidt & Factor, 2013). In the context of a diagnostic accuracy study, *bias* occurs when the resulting estimates of the index test sensitivity and specificity are consistently overestimated or underestimated relative to the true accuracy parameters (Schmidt & Facto, 2013). Using a non-gold-standard test as a reference test for classifying depression may increase systematic-error in the study resulting from misclassification bias. For example, a non-gold-standard reference test may incorrectly miss cases of true depression. Furthermore, procedures for identifying and enrolling subjects can result in a sample that does not represent the target population and has a different association between the index test and reference test. For example, recruiting depressed and non-depressed participants through different processes can

result in confounding biases and in a disproportionate amount of diseased participants. Evaluation studies using prospective, consecutive recruitment as well as inclusion/exclusion criteria that reflect the target population help to reduce selection bias. Random error (or imprecision) results in the index test sensitivity and specificity randomly differing from the true accuracy parameters. A particularly important contributor to random error is sample size. Larger sample sizes reduce sampling error and increase the likelihood that the sample reflects the target population. Similarly, samples with higher frequency counts for true depression reduce sampling error and increase the likelihood that the depressed participants in the sample represent depressed people in the target population.

The external validity of an evaluation reflects the applicability of the study participants to the target population (Schmidt & Factor, 2013). It is important that PSD screening tests be evaluated with stroke samples that represent the target population, in this case, the population of stroke survivors who will be screened with the index test. Accuracy of index tests evaluated on non-stroke populations cannot be expected to generalize to stroke populations. For example, stroke patients are typically older and have more physical and cognitive impairments than non-stroke patients, which may impact what questions work best to distinguish depressed from non-depressed stroke survivors. Index tests evaluated on samples of stroke patients previously treated for depression, participating in private rehabilitation programs, or samples selected for their high cognitive or physical functioning, cannot be assumed to perform with equivalent accuracy in a population of stroke survivors in acute in-patient care. In addition, index tests evaluated in specific cultural settings may not function similarly in other cultural settings, where symptom profiles may be different and questions and symptoms may have different meanings to the tested individuals.

It is also important for screening tools to function well in subgroups of the target population. Several factors influence screening test accuracy and it cannot be assumed that the functioning and accuracy of questions is homogeneous across subsamples (Henkel et al., 2004). Stroke patients differ across several demographic characteristics and an ultra-short PSD screening tool should have acceptable functioning and accuracy with stroke survivors in disparate subgroups of the target population. Tests perform differently in population subgroups when people of different subgroups, with the same underlying levels of depression, respond differently to the same questions. Demographic characteristics, such as age, sex, language, and

level of education, often differentiate between cultural subgroups and may affect the functioning of index tests. In their study with primary care patients, Henkel and colleagues found that gender and age are important variables moderating the accuracy of depression screening measures. The authors describe research consistent with their results, suggesting that depressive symptoms present with different patterns in women compared to men, and in the young compared to the elderly, thus highlighting the importance of developing screening measures that perform well across subgroups of sex and age. In their development of a PSD measure, Cinamon, Finch, Miller, Higgins, and Mayo (2011) found that French and English versions were not similarly accurate, which the authors attributed to cultural differences associated with the respective languages. For example, French Canadians were significantly more likely to report getting upset over little things, enjoying getting up in the morning more, and finding it easier to make decisions, compared to English Canadians. Another factor to take into careful consideration is to ensure that tools function well across different levels of stroke severity, as the degree of impairment and disability may contribute to a variety of factors that influence screening test functioning, such as symptom profiles, perception of symptoms, and interpretation of test questions. Additionally, level of education may also moderate test functioning through cultural differences associated with obtaining different levels of education.

Accuracy of existing ultra-short verbal self-report screening tools for PSD. Burton and Tyson (2015) systematically reviewed accuracy studies evaluating mood screening tools with samples including stroke survivors, and reported the psychometric properties of each tool and their feasibility for routine use. Acceptable minimum accuracy was defined as Sensitivity $\geq$ 80% and Specificity $\geq$ 60% for identifying either major or any depression in at least one validated study. Feasibility, or clinical utility, was defined as $\leq$ 5 minutes to administer and score the test; the free availability of tests, manuals, and forms; and the requirement of no training to administer and interpret tests. The authors systematically identified studies written in English and published before June 2013, reporting accuracy of PSD screening tools relative to reference standard measurements of depression. The authors identified 15 verbal self-report index tests evaluated in PSD samples. Of these 15 tests, 10 met the accuracy criteria of the authors, while only four of the remaining 10 verbal tests met the authors' criteria for clinical utility. Two of the 15 evaluated verbal tests were ultra-short screening tools consisting of under five items and requiring under two minutes to complete. These ultra short tests include the two-item Patient Health

Questionnaire – PHQ-2 (Kroenke, Spitzer, & Williams, 2003) and the Yale–Brown single item screening question – The Yale (Lachs et al., 1990), which will be discussed in the following sections. The other two tools were short questionnaires, including the nine-item Patient Health Questionnaire – PHQ-9 (Spitzer, Kroenke, Williams, & Patient Health Questionnaire Primary Care Study Group, 1999), and the 15-item version of the Geriatric Depression Scale – GDS-15 (Sheikh & Yesavage, 1986). Table 1 lists the 15 identified verbal self-report tools evaluated for screening depression in stroke patients, and provides information on each test's length, training requirement, cost, and demonstrated minimum accuracy.

The two-item Patient Health Questionnaire – PHQ-2 (Kroenke et al., 2003) is an ultra-short self-report tool designed to be used in the initial screening for depression in primary care patients. The questionnaire asks patients the extent to which they have been bothered over the last two weeks by, 1) little interest and pleasure in doing things, and 2) feeling down, depressed, or hopeless. Patients provide their answer to each question on a four-point scale. Administration time is less than two minutes and scoring is simple. Response points are summed to give a score ranging from zero to six.

The Yale–Brown single item screening question (The Yale Question) is an ultra-short screening tool, developed by the Yale task force, to quickly screen for possible depression in elderly community patients (Lachs et al., 1990). The Yale Question asks patients to answer "Yes" or "No" to one question taken from the Yale-Brown obsessive-compulsive scale (Y-BOCS) "Do you often feel sad or depressed?" The Yale Question has been recommended for use as a screening tool as phase one, leading to a more detailed investigation in patients who respond in the affirmative (Watkins, Lightbody, Sutton, Holcroft, Jack et al., 2007). Administration is quick with no time required for scoring or interpretation. Due to its value as an ultra-short screening tool, the Yale Question is recommended for screening for depression in stroke patients by the United Kingdom's National Clinical Guidelines for Stroke (Intercollegiate Stroke Working Party, 2008).

Based on their analysis of results across reviewed studies, Burton and Tyson (2015) concluded there was insufficient data to support the accuracy of either the Yale question or a specific cut-off score for the PHQ-2 in identifying major depression or any depression post stroke. Burton and Tyson concluded that only the PHQ-9 and the GDS-15 met acceptable psychometrics with stroke patients at a specific cut-off score. Specifically, the GDS-15 shows

Table 1

Verbal self-report depression screening measures assessed by Burton and Tyson (2015)

Verbal self-report tool	Accurate ≥ one stroke sample	N items; completion time	Developed for what population?	Length	Training required (Y/N)	Freely available (Y/N)
BDI (Beck et al., 1961)	Yes	21 items, <10 min	To measure symptoms of depression in a psychiatric population	Long	N	N
BDI-II (Beck et al., 1996)	Yes	21 items, <10 min	To measure symptoms of depression in a psychiatric population	Long	N	N
BDI-FS (Beck et al., 2000)	No	7 items,	For medical patients	Short	Not reported	Not reported
CES-D (Radloff, 1977)	Yes	20 items, < 15 min	To measure symptoms of depression in the general population	Long	N	Y
GDS (Yesavage et al., 1982)	Yes	30 items, 8-10 min	Detecting depression in older adults	Long	N	Y
GDS-15 (Sheikh & Yesavage, 1986)	Yes	15 items, 5 min	Detecting depression in older adults	Short	N	Y
GHQ-28 (Goldberg & Williams, 1988)	Yes	28 items, <5 min	Screening for psychiatric disorders	Long	N	N
HADS-D or HADS-T (Zigmond & Snaith, 1983)	Yes	7-14 items, 2 to 6 min	HADS-T designed to measure emotional distress, HADS-D subscale designed to measure depression	Short to long	N	N
K10 (Kessler et al. 2003)	No	10 items,	Screening scale for psychological distress & level/ severity of mental disorder	Short	Not reported	Not reported
PHQ-2 (Kroenke et al., 2003)	Yes	2 items, <2 min	Screening for depression in general populations	Ultra-short	N	Y
PHQ-9 (Spitzer et al., 1999)	Yes	9 items, 3-5 min	Screening for depression in general populations	Short	N	Y

(continued)

Verbal self-report tool	Accurate ≥ one stroke sample	N items; completion time	Developed for what population?	Length	Training required (Y/N)	Freely available (Y/N)
SIDI (Rybarczyk et al., 1996)	No	30 items	Specifically for stroke patients in hospital settings to measure depression	Long	Not reported	Not reported
WDI (Snaith et al., 1971)	No	12 items	To measure severity of depression within a psychiatric population	Short	Not reported	Not reported
Yale (Lachs et al., 1990; Mahoney et al., 1994)	Yes	1 item, <5 min	Screening for depression in older patients	Short	N	Y
Zung SDS (Zung, 1965).	No	20-items	To identify clinically significant depression in a psychiatric population	Long	Not reported	Not reported

Note. Beck Depression Inventory (BDI); Beck Depression Inventory Second Edition (BDI-II); Beck Depression Inventory Fast Screen (BDI FS); Center for Epidemiological Studies-Depression Scale (CES-D), Geriatric Depression Scale (GDS); 15-item Geriatric Depression Scale (GDS-15); 28-item General Health Questionnaire (GHQ-28); Hospital Anxiety and Depression Scale Depression subscale (HADS-D); Hospital Anxiety and Depression Scale combined scales (HADS-T); Kessler-10 (K10); two-item Patient Health Questionnaire (PHQ-2); nine-item Patient Health Questionnaire (PHQ-9); Stroke Inpatient Depression Inventory (SIDI); Wakefield Depression Inventory (WDI); Yale–Brown single item screening question (Yale); Zung Self-Rating Depression Scale (Zung SDS). Reported completion times are often meant for general populations, not for stroke-population. It is expected that completion times will be longer in stroke populations.

accuracy for detecting any depressive disorder and the PHQ-9 for detecting major depressive disorder.

A limitation of the systematic review by Burton and Tyson (2015) regarding its ability to inform our understanding of PSD screening tools, is their broad inclusion criteria for evaluation studies. The authors included studies on samples not specific to stroke (e.g., brain injury samples containing stroke), studies evaluating index tests relative to non-gold-standard reference tests, which modifies the operational definitions of depression. Additionally, the authors include studies with outcomes not specific to PSD (e.g., a DSM-based diagnosis of depression or an adjustment disorder).

Notably, neither the PHQ-9 nor the GDS-15 are ultra-short screening tools ideal to be quickly and simply administered, scored, and interpreted by untrained staff, patients, caregivers, or family members. Completion time for the PHQ-9 and the GDS-15 are reported to be three to five minutes and five minutes, respectively, in primary care patients. As Burton and Tyson (2015) note, completion time is expected to be longer with stroke patients who often have communication and cognitive problems. Furthermore, additional time is required for the scoring and interpretation of both tests. With the PHQ-9, scoring instructions are somewhat complicated and potentially confusing for untrained patients and family members, resulting in increased error and inappropriate follow-up behaviour. Scoring of the GDS-15 requires some reverse coding, which may produce errors when used by patients and family members. The GDS-15 and the PHQ-9 take longer to administer, score and interpret when used by staff, rendering them less feasible for routine use compared with ultra-short screening tools. Taken together, Burton and Tyson's systematic review does not indicate a measure that meets the criteria of this research for a validated PSD screening tool that facilitates accurate screening.

Literature Review

Review objectives. This author sought to learn whether existing ultra-short verbal self-report depression screening tools have been shown to be accurate for detecting depression in stroke survivors, relative to a classification of any depression, mild depression, or major depressive disorder based on a gold-standard reference test, such as the DSM.

Literature review methods. Electronic databases, including Psychinfo, Medline, Psychology and Psychology and Behavioral Sciences Collection, American Doctoral Dissertations, Mental Measurements Yearbook with Tests in Print and Pubmed, as well as

Google Scholar were searched from their inception to the search date. The titles, abstracts, and full texts were screened by one reviewer. The goal was to identify studies evaluating and reporting the accuracy of ultra-short verbal self-report index tools for a depressive disorder relative to a gold-standard reference, in samples of stroke survivors. This research excluded conference papers and abstracts without published research articles. Also excluded were studies without reported accuracy statistics, studies where data from people with stroke could not be extracted, studies in which tools were tested specifically on samples of patients with aphasia or communication disorders, and studies with outcomes other than depression, or using non-gold-standard reference tests for depression.

The following key terms were used to search in each database: Stroke AND (depression OR depressive disorder OR mood disorder OR affective disorder) AND (detec* OR asses* OR evalua* OR scree* OR diagnos*). Term types were marked as "subject term" or "MeSH term" where available. No limits were applied to search criteria. The database searches occurred between November 18th and December 7th, 2016, inclusive. Through the above-mentioned database search method, this research identified reviews evaluating and reporting the accuracy of index tools for PSD. An initial examination of the titles and abstracts from this search published up until 2016 yielded five relevant reviews reporting on the accuracy of any verbal self-report depression screening measure evaluated in stroke populations. The identified reviews include Burton and Tyson (2015), Gilson (2012), Kitsos, Harris, Pollack and Hubbard (2011), Meader, Moe-Byrne, Llewellyn, and Mitchell (2014), and Salter et al. (2007). An examination of the reference lists of these five reviews was conducted to obtain all studies that met the inclusion-exclusion criteria. A particularly extensive systematic review, discussed in the previous section completed by Burton and Tyson, reported on articles written in the English language and published up until May 2013. Subsequently the electronic database search was updated to identify studies published in 2013 or later.

From research publications meeting inclusion and exclusion criteria, a range of data was extracted, including information on the tools evaluated, gold standard criteria, sample characteristics, time post stroke, selection criteria, and sensitivity and specificity. Information on each tool's availability, feasibility, and price was also gathered. There were not enough studies on each measure to compare measure functioning by time point and analyze for significant differences.

Literature review results. After reviewing studies included in previous literature reviews (Burton & Tyson 2015; Gilson, 2012; Kitsos et al., 2011; Meader et al., 2014; Salter et al., 2007) this author identified three studies that met criteria for inclusion in the literature review. The database search of articles published from 2013 to 2016, inclusive, yielded 21 articles identified as potentially relevant from the contents of their titles and abstracts. These articles included five found in Psychinfo, two additional articles from Medline, Psychology and Behavioral Sciences Collection, American Doctoral Dissertations, or Mental Measurements Yearbook with Tests in Print, one additional article from Pubmed, and 12 additional articles from Google Scholar. After reading the 21 articles, none met the inclusion-exclusion criteria reported above. Consistent with previous reviews, this literature review only identified candidate studies testing two ultra-short screening tools in stroke samples, namely the PHQ-2 and the Yale Question.

In total, only three original studies were identified that evaluated and reported on the accuracy of ultra-short verbal self-report screening measures for identifying a depressive disorder, relative to a gold-standard reference measure in stroke-specific samples. All three of these works assessed the PHQ-2. No published articles evaluating the Yale Question met the inclusion criteria for this research. Table 2 reports the authors, evaluated cut-off scores, sensitivity, specificity, gold-standard outcome measure, assessment time post stroke onset, sample characteristics, and exclusion criteria for these three articles.

The three studies in Table 2 only assessed the cut off-scores of '2' and/or '3'. A cut-off score of only '2' of 6 points is the most potentially sensitive PHQ-2 score assessed to date. However, de Man-van Ginkel et al. (2012) and Turner et al. (2012) found that even a cut of score of '2' was close but insufficiently sensitive (75 – 77%) for detecting major depression in a consecutive sample of patients at two months post stroke and in a sample of in and out patients ranging from three weeks to several years post stroke, respectively. Williams et al. (2005) reported that a cut-off score of '3' showed acceptable accuracy (Se=83%, Sp=84%) for detecting MDD, but not any depression (Se=78%, Sp=95%), in a clinical trial hospital sample within one to two months of stroke. However, a major limitation in the study by William et al. (2005) is the potential for verification bias (Whiting et al., 2004), as the reference test was only administered to patients who scored within the clinical range of either the PHQ-2 or the PHQ-9. This methodology can result in fewer false negatives and more true negatives than the true value,

Table 2

Studies assessing the accuracy of the PHQ-2 relative to a gold-standard measure of depression

Author	Tool (cut off)	Se/Sp %	Diagnosis, Gold-standard criteria	Time post stroke onset	Sample	Exclusion Criteria
de Man-van Ginkel et al. (2012)	PHQ-2 (2)	75/76	MDD, Structured CIDI	2 months	Prospective study, 164 consecutive hospital admissions, mean age=71 (SD =14) years	Severe cognitive or communication impairment, co-morbid or pre-morbid physical or psychiatric illness, non-English or Dutch speaking
Turner et al. (2012)	PHQ-2 (2)	77/63	MDD, SCID DSM-IV	Median=14 months (3 to 540 months)	72 patients in Australia, including 28% outpatients; 72% consecutive rehabilitation inpatients; mean age=67 yrs (SD =13) years	< 3 weeks post-stroke, non-English speaking, severe cognitive or physical impairment
	PHQ-2 (3)	69/83				
Williams et al. (2005)	PHQ-2 (3)	83/84*	MDD, SCID DSM-IV	1–2 months	316 case-matched depressed (n=145) and non-depressed (n=171) hospital patients mean age not reported.	Severe cognitive or communication impairment, not depressed
		78/95	MDD or other depression, SCID DSM-IV			

Note. Sensitivity (Se); Specificity (Sp); Composite International Diagnostic Interview (CIDI); transient ischemic attack (TIA); traumatic brain injury (TBI). *acceptable minimum accuracy

which can spuriously inflate sensitivity and specificity. In contrast, Turner et al. (2012) found a cut-off score of '3' was insufficiently sensitive (69%) for detecting MDD in stroke patients differing in time post stroke. Results assessed within this review show that the PHQ-2 with a cut-off score of '2' was close but below acceptable accuracy standards for a PSD screening tool. Results of these three identified studies do not provide support for use of the PHQ-2 or any other ultra-short tool for PSD screening with stroke survivors at any time since the stroke event.

Commenting on the Yale Question. While the Yale Question is an ultra-short depression screening tool recommended for screening depression in stroke patients by the United Kingdom's National Clinical Guidelines for Stroke (Intercollegiate Stroke Working Party, 2008), this research found no studies of the Yale Question that met our inclusion criteria. During the literature review, two studies were excluded (Watkins, Daniels, Jack, Dickinson, & van den Broek, 2001; Watkins et al., 2007) which evaluated the Yale Question as accurate relative to a test for depression that is not a gold gold-standard reference test and is inconsistent with DSM-based diagnostic criteria. One study (Turner-Stokes, Kalmus, Hirani, & Clegg, 2005) found the Yale Question insufficiently sensitive (Se = 64, Sp = 85) for detecting DSM-IV criteria for either minor or major depression on a sample of brain injury patients (67% stroke, 18% TBI, 15% other neurological conditions) in a rehabilitation in-patient unit for younger adults with severe complex disabilities. This study was excluded from the review as data from stroke patients could not be extracted. This author interprets existing literature to indicate that there is insufficient evidence to date to justify recommendations that the Yale Question be employed to screen for PSD. Additional studies are required to evaluate the Yale Question in stroke populations using gold-standard reference tests.

Limitations of research evaluating existing ultra-short tools for screening PSD. This literature review identified that only two ultra-short depression screening tools, the Yale and the PHQ-2, have been evaluated for screening depression in stroke patients. However, only three of the PHQ-2 were found to evaluate the ultra-short index test relative to a gold-standard measure of depression in a stroke-specific population. This author recommends that future studies provide all subjects with both the index and reference tests, avoid verification bias, assess screening tools of stroke-specific samples that represent the target population, and consider prospective consecutive designs to avoid selection bias.

Because PSD emerges at varied times following stroke and follows varied and dynamic trajectory paths, it should not be expected that screening tests will function identically at different time points following the stroke event. This point is discussed above in the section on the 'heterogeneous nature of PSD'. As PSD frequently emerges and should be assessed throughout the first year of stroke, PSD screening tools should be validated at several time points throughout the first year relative to gold-standard reference tests. Further research on existing ultra-short screening tools may support their accuracy with stroke patients at specific assessment times post stroke. There is little published information about the responsiveness of the measures to change over time in stroke populations. Longitudinal designs will be helpful for assessing tool accuracy at different assessment times post stroke. Studies on large samples varying in time since the stroke event should consider stratifying participants based on time elapsed since stroke.

Importantly, neither the PHQ-2 nor the Yale Question were developed for use in stroke patients. However, as stroke patients are older and have more disabilities than primary care patients, ultra-short tools may not generalize well to the stroke-specific population. This research has identified no studies evaluating whether questions or items other than those included in the PHQ-2 and the Yale, are better suited for screening depression in stroke patients. Such studies are needed to examine the specific questions that are most accurate for screening PSD in an ultra-short tool at different times after the stroke event.

Study Objectives

The principal objective of this diagnostic accuracy study was to determine what brief combinations of self-report mood questions can be used in ultra-short screening measures to accurately and feasibly screen for depression in stroke survivors, at different times post stroke, relative to a gold standard reference test for depression.

Proposing ultra-short PSD screening tests. Existing ultra-short screening tests demonstrate certain limitations for PSD screening, which this research addresses: tests have not been developed for stroke populations; are not sufficiently accurate for screening PSD; have not been assessed at different times following stroke; and may not contain the best items for screening PSD. An objective of this study was to contribute to developing ultra-short PSD screening tools that address these issues. The present tools are developed in a representative stroke population, and are designed to be administered, scored, and interpreted by untrained staff or at home by patients, family members, or caregivers. The intended use of successful index

tests is to replace other existing ultra-short screening tools used to screen for PSD as step one in a two-step assessment process. Proposed ultra-short screening tools may also replace existing screening measures when these existing tests are inappropriate for screening PSD, either because they have not been validated at specific time points with stroke patients, or because they are resource demanding or otherwise impractical. Screening tools developed in this study are designed to inform a binary clinical decision for whether the patient should be referred to a qualified health-care professional for further assessment of PSD, without the availability of additional information. Last, the developed screening tools are designed to be low cost for wide-spread distribution and use.

Development of a protocol to test, assess, and select candidate index tests. There is currently an insufficient research base on specific questions that should be used to identify PSD at various time points during the first-year post stroke. Therefore, an objective of this study was to examine the ability of a range of items and their combinations, obtained from validated mood questionnaires, for accurately discriminating PSD from no depression. A protocol was therefore developed in this study to systematically create, test, and rank candidate index tests based on their indicated validity at identifying depression, relative to a gold-standard reference test, at separate time points within the first year of stroke. The protocol was designed to select for tools that showed sufficient accuracy and functioning in sample subgroups, as well as in the whole sample. This study's protocol was designed to reduce the possibility that selected screening tests will bias or function differently in specific sub-groups of the target population, where subgroups pertain to sex, age, level of education, stroke severity, and language.

Addressing the argument that assessment time be considered when assessing tool validity. An objective in this study was to develop screening tools that are accurate at specific times post stroke, as well as to examine the extent to which ultra-short screening tool accuracy applies, or can be generalized, to different time points post stroke.

Hypotheses

This study will test the following two hypotheses:

H1: A single item or a brief combination of two items assessing mood state, can have acceptable accuracy for identifying cases of any depression and major depressive episodes relative to a gold-standard assessment of depression, at three months, six months, nine months, and 12 months following the stroke event.

H2: Different combinations of items assessing mood will be required to accurately identify depression at different time points. In other words, one or two-item models that show good accuracy at predicting depression at one time point, will not show sufficient accuracy at all time points.

Chapter 2: Methods

Ethics

This study was reviewed and approved by the Research Ethics Board of the McGill University Health Center, by the Office of Research at Concordia University, and by each participating hospital. Informed consent to participate in the study was given by patients and relatives. People unable to provide informed consent were excluded from the study. All stroke survivors were assured that participation was completely optional and that a decision not to participate would not result in any negative consequences. No compensation was provided to participants. Participants were referred to health practitioners when suspected of being at risk for adverse health events (e.g., depression or suicidality).

Participants

The sample is an inception cohort comprised of consenting persons consecutively admitted to one of three adult hospital sites of the McGill University Health Center (MUHC), located in Montreal, Canada, including the Royal Victoria Hospital, the Montreal General Hospital, and the Montreal Neurological Hospital. Stroke was verified with computed tomography or magnetic resonance imaging criteria. Eligible participants were admitted to a MUHC site within 10 days of stroke, spoke either English or French, and were to be discharged back into the community within the greater Montreal area, whether to their homes, to private residential settings, or to inpatient rehabilitation centers.

As is standard with stroke outcome studies, excluded from the sample were persons with death before discharge, severe illness expected to lead to death within one year of discharge,[13] severe cognitive impairment, sustained altered consciousness, and receptive aphasia, as these factors prevent capacity to provide informed consent and to complete questionnaires, and can introduce biases associated with assessment responses. Eligibility information, including discharge plans, was obtained in hospital medical charts. Severe cognitive impairment was defined as a score below 14/18 on the Brief Version Mini-Mental State Examination (Brief-MMSE; Paveza, Cohen, Blaser, & Hagopian, 1990; see measure in Appendix C). This criterion does not exclude participants with mild expressive aphasia who can answer questions adequately when provided with enough time and appropriate visual material.

Two additional exclusion criteria particular to this study, and included for convenience, were participants living further than 100 km from McGill University, and admission to a hospital more than 10 days post stroke onset, the date of the first assessment.

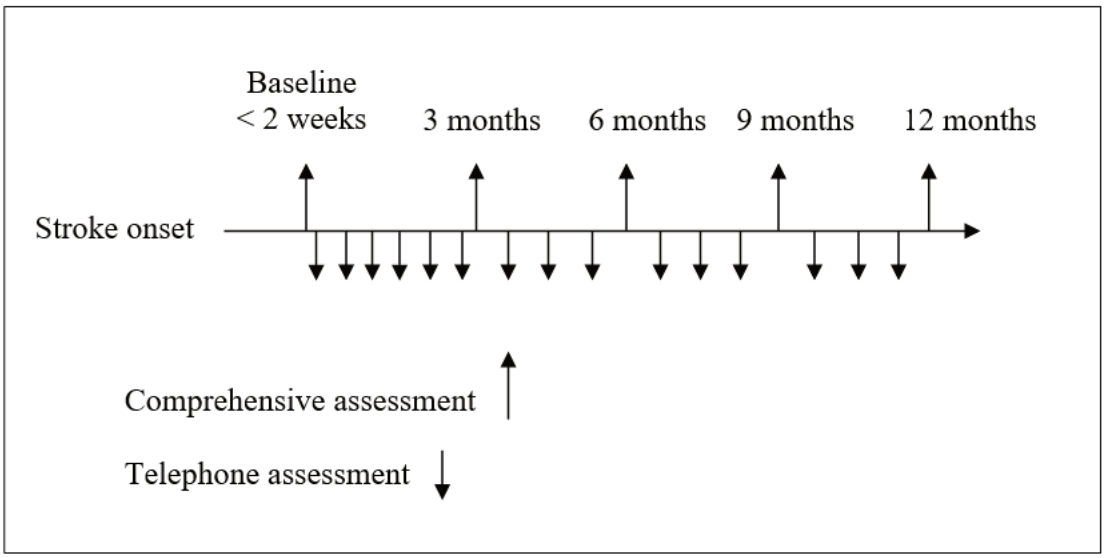
One hundred percent of the individuals admitted to a MUHC for stroke from September 2007 to March 2009 were screened for eligibility criteria. Eligible participants were approached, and informed consent was obtained or denied by participants and family members. After consent was obtained, demographic, clinical, and psychiatric data were obtained prospectively based on direct evaluation of patients, or by interviewing family members, according to the research protocol of this study. Additional retrospective data was extracted from medical charts.

Study Design

This study is a longitudinal, prospective, natural history study of an inception cohort, embedded within a larger study, lead by Dr. Mayo's research group at McGill University's clinical epidemiology laboratory,[14] aimed at tracking the dynamic pattern of depressive symptoms in patients in the year following stroke. This study design followed the most recent STARD checklist (Bossuyt et al., 2015).[15]

Consenting participants who met eligibility criteria received a comprehensive face to face assessment and administration of a battery of questions within the first 10 days post-stroke onset, whether in acute-care or discharged. The comprehensive assessment was then repeated at three, six, nine, and 12 months post-stroke. The battery consisted of numerous measures and questions, whereby only the measures relevant to this study are discussed here. Subjects were also interviewed by telephone, using a standardized script, every two weeks for the first three months and then monthly for the remainder of the study. Telephone interviews provided information about participants' reasons for refusing to participate in interviews. Other data collected during telephone interviews were not relevant to this study and are not discussed here. Table 10, in Appendix B shows the time post stroke when participants were administered each measure. The study design protocol is outlined in Figure 1, below.

Figure 1
Study Design



Materials and Measures

There are four types of variables included in this study. The first three are applicable to this study's analyses of substantive interest and include: (1) the reference standard measurements of depression; (2) candidate index tests developed in this study from items assessing mood state belonging to existing measures; and (3) covariates identified as population subgroups that may result in spectrum bias, including sex, age, language, level of education, and stroke severity. The fourth type of variable, (4) including stroke and clinical characteristics and socio-demographic information, was measured and analysed for the purpose of characterising the sample. All materials employed in this study have previously been used in stroke or comparable populations, including elderly patients, or patients with chronic health conditions. Validated English and French versions of measures were available. The four types of measures are described below, and copies of questionnaires described below can be found in Appendix C.

1. The reference standard measurement of depression. Depression was assessed at all five time points with criteria for current Major Depressive Episode (MDE) within the Mood Disorders Module of the Structured Clinical Interview for *Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision ([DSM-IV-TR]* American Psychiatric Association, 2000) *for Axis I Disorders, Non-Patient Edition ([SCID-I/NP]* First, Spitzer, Gibbon, & Williams, 2002). The SCID-I/NP is a semi-structured interview for making the major DSM-IV-TR Axis diagnoses, including diagnosis of a major depressive episode, for use in studies with subjects not identified as psychiatric patients (e.g., for research in primary care). Versions of the SCID are considered the 'gold standard' measure for the assessment of depression (e.g., Shear et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995), in accordance with DSM criteria and are accepted as the standard for diagnosing depression post-stroke (Robinson, 2003).

The mood module of the SCID-I/NP interview was administered at eight to 10 days post stroke and at each subsequent time point. It was administered face-to-face in the hospital or in the homes of participants by trained members of the research group and was supervised by a clinical psychologist. In cases where stroke-related disabilities prevented a patient from providing a self-report, the interview was completed in consultation with a proxy (i.e., the primary caregiver of a participant).

The study outcome is based on patient responses to criterion A and B[16] for a current MDE in the SCID-I/NP interview. In Criterion A, participants are interviewed to assess for nine core

symptoms that may be present during the same 2-week period which represent a change from previous functioning. The nine core, clinically significant MDE symptoms are as follows: 1) depressed mood; 2) loss of interest or pleasure; 3) weight loss or weight gain; 4) insomnia or hypomimia; 5) psychomotor agitation or retardation; 6) fatigue or loss of energy; 7) feelings of worthlessness or excessive/inappropriate guilt; 8) diminished ability to think, concentrate or indecisiveness; and 9) suicidal ideation. At least one of the symptoms must be either 1) depressed mood or 2) loss of interest of pleasure. Criterion B requires that the symptoms cause clinically significant distress or impairment. All participants were administered the first two items of criterion A, and only participants with either clinically significant depressed mood or loss of pleasure were administered the remaining sevens items. Items were scored as dichotomous variables: either the symptom was clinically significant, or it was not. Participants were classified at each time point, based on their SCID-I/NP responses, to one of the following four depressive outcome categories: [17]

a) **Major Depressive Episode (MDE):** Participants with five or more symptoms that represent a change in functioning within the same two-week period, where one symptom was either depressed mood or loss of interest or pleasure. The measured symptoms and the core diagnostic criteria for MDE are consistent across both the DSM-IV-TR and the most recent fifth edition of the manual, the DSM-5 (APA, 2013).
b) **Minor depression:** Participants with two to four symptoms that represent a change in functioning within the same two-week period, where one symptom was either depressed mood or loss of interest or pleasure. The term is listed in Appendix B of the DSM-IV-TR, and is classified more generally under the category Depressive Disorder Not Otherwise Specified (DD-NOS), which includes disorders with depressive features that do not meet full criteria for Major Depressive Disorder. In the DSM-5, the term is not listed, and persons showing the symptoms would be classified with Unspecified Depressive Disorder. The term 'minor depressive disorder', and its criteria have been frequently applied in the PSD literature and therefore the term is applied here.
c) **Depressive features:** Participants with only one symptom, either depressed mood or loss of interest or pleasure, were categorized in the 'depressive features' category.[18]
d) **No depression:** Participants with neither depressed mood nor loss of interest.

Any depression. The 'any depression' outcome category at each time point included all participants with either clinically significant depressed mood or loss of interest or pleasure for the same two-week period. This group including participants categorized with MDE, minor depression, or depressive features.

Several categorical outcome variables were generated based on participants' categorizations into the above-mentioned diagnostic categories. Participants were given scores on each of the following binary and ordinal variables at each time point: i) 'any depression' versus 'no depression', scored (1,0); and ii) MDE versus no MDE, scored (1,0). The 'no MDE' group included all participants not meeting criteria for an MDE, including participants categorized in the 'minor depression', 'depressive features', and 'no depression'. iii) A four-level ordinal depressive outcome variable ranked participant 'status and severity of depression' based on whether they met criteria for MDE, minor depression, depressive features, or no depression, and was scored (3, 2, 1, 0). iv) A three-level ordinal variable measuring only 'severity of depression' was based on diagnosis of MDE, minor depression, or depressive features, and was scored (3, 2, 1). Outcome variables were also generated to communicate a participant's most severe level of depression criteria met at all five time points and were used in analyses aimed at characterising the sample. Participants were scored (3, 2, 1, or 0) on this four-level ordinal outcome variable based on whether their most severe criteria at any time point in the year post stroke onset was '3' MDE; '2' 'minor depression'; '1' 'depressive features'; or '0' 'no depression'. Additional binary outcome variables were similar to the above-described classification, such as 'any depression' versus 'no depression'.

Participants' outcome variables were marked as missing for cases completely missing the SCID-I/NP at a given time point. For cases with missing items on an otherwise completed SCID-I/NP, outcome scores were generated when sufficient data was available. For example, when a participant endorsed either items one or two, but was missing responses to additional items, the binary outcome variables 'any depression' and 'depressive features' were scored '1', whereas values for 'MDE' and 'minor depression' outcome variables were marked as missing.

Notably, a history of depression was assessed at three months post stroke using the SCID-I/NP *past MDE* criterion within the mood module. This variable was used as a covariate.

2. Measures used in the development of index tests. This research examined functioning of one or two-item combinations, selected from longer questionnaires, for use as

ultra-short screening tools in predicting depressive outcomes according to the gold-standard reference test. All the items used in the index tests come from measures (i.e., questionnaires) previously validated for identifying problems with mood and depression. Items for index tests came from the following questionnaires:

The Stroke Specific Geriatric Depression Scale (SS-GDS; Cinamon et al., 2011) was developed by conducting a Rasch analysis from the 30-item GDS with a stroke population, resulting in removal of items that may be symptoms of the neurological incident rather than depression. Examples of removed items include "do you feel that you have more problems with your memory than most?" and "is your mind as clear as it used to be?" The SS-GDS is a self-rated questionnaire that is comprised of 17 simply worded questions about how participants feel (time duration is not specified) to which the participant answers "yes" or "no". Total scores for the SS-GDS are calculated on a continuous scale from "0" to "17", where greater scores indicate greater severity of depressive symptoms. In scoring the SS-GDS, all items are oriented so that a lower score represents better mental health. Notably, the original GDS is a screening measure for depressive symptoms in the elderly, which deemphasizes somatic symptoms, and has a reliability coefficient of 0.90 in a sample of geriatric stroke patients (Agrell & Dehlin, 1989). The SS-GDS has demonstrated evidence of reliability and validity (Cinamon et al., 2011). In this study, the SS-GDS was administered at eight to 10 days post stroke and at each subsequent time point.

The Mental Health Index (MHI-5) is a self-report questionnaire assessing general mental health[19] and is part of the Medical Outcomes Trust Study (MOS) 36-item *Short Form Questionnaire* (SF-36; Ware & Sherbourne, 1992), a larger questionnaire assessing participants perception of their health in eight conceptual domains. The SF-36 has been used in multiple patient populations and has been validated in stroke populations (e.g., Anderson, Laubscher, & Burns, 1996). The MHI-5 has been frequently used as a stand-alone measure of general mental health and has been shown to be a valid and reliable measure of mental health status (Ware & Gandek, 1998). In this study, the MHI-5 was administered at eight to 10 days post stroke and at each subsequent time point as a paper-pencil self-report questionnaire. The questionnaire asks participants to rate how they felt over the last four weeks with five questions, where responses are rated on a 6-point Likert-type scale ranging from 'all of the time' to 'none of the time'. The instrument contains the following question: 'How much of the time during the last month have you: (b) been a very nervous person?; (c) felt so down in the dumps that nothing could cheer you

up?; (d) felt calm and peaceful?; (f) felt downhearted and blue?; and (h) been a happy person? [20] Positively worded items are reverse coded. A participant's total MHI-5 score is calculated by summing up responses, and can range between 5 and 30, with higher scores indicating better mental health status. The MHI-5 scale can be directly transformed into a 0-100 scale, where all questions are given equal weight. As part of an international initiative, the SF-36 has been translated into Canadian French using a standard protocol, and French and English versions were available to participants.

The Stroke Impact Scale version 3.*0* (SIS V 3.0; Duncan et al., 1999) evaluates the impact of stroke in multiple domains, including emotional state. The SIS was developed as a comprehensive measure of health outcomes post-stroke and each of its eight domains can be scored independently. Domain three of the SIS V 3.0 asks participants nine questions (3a to 3i) in which they rated their emotional state in the last week on a five-point Likert-type scale, ranging from "none of the time" to "all of the time." The SIS V 3.0 has been validated for use with persons with hemorrhagic and ischemic stroke ranging in severity from mild to severe (Duncan et al., 1999; Duncan, Bode, Lai, Perera, & Glycine Antagonist in Neuroprotection Americas Investigators, 2003). The SIS was administered at time points two through five.

The EuroQOL five dimensions questionnaire (EQ-5D; The EuroQol Group, 1990) was developed as a self-completed questionnaire that describes health status in terms of five dimensions with five questions. Participants completed validated English (Canadian) or French (Canadian) versions of the paper-pencil EQ-5D index at time points two through five. For stroke populations, the EQ-5D has been found useful, with its ease of use reducing response errors in stroke populations when compared with the SF-36 (Dorman, Slattery, Farrell, Dennis, & Sandercock, 1998; Dorman, Waddell, Slattery, Dennis, & Sandercock, 1997a, 1997b; Fisher & Zorzitto, 1983). For the anxiety/depression dimension, participants are asked to choose one of three statements that best describes their degree of anxiety and depression felt on the day of the questionnaire. Participant responses to the depression/anxiety item were scored on a three-point scale.

A total of 32 items assessing mood state, including 17 SS-GDS items, five MHI-5 items, nine SIS items, and one EQ-5D item were included as possible contender screening items, although items from the SIS and the EQ-5D questionnaires were not administered at the first time point.

3. Covariates - population subgroups that may result in spectrum bias. Covariates included sex, age, language, level of education, and stroke severity. Sex, age, and level of education were obtained from a *Socio-Demographic and Health Related Questionnaire* developed by the research group of which this study is a part. This is a paper-pencil questionnaire completed by participants or a proxy at three days post stroke and at each subsequent time point. Sex was measured dichotomously. A participant's highest level of completed education was measured on a nine-point ordinal scale ranging from "completed grade six or less" to "completed graduate or professional school." Participants provided their date of birth, and age was calculated by subtracting the date of birth from the date of the stroke. A participant's spoken language was obtained from the recruiter/datasheet in medical charts and was measured as a multi-level nominal variable.

Stroke severity was assessed at baseline with the *Canadian Neurological Scale (CNS;* Côté et al., 1989). The CNS is a relatively quick and simple interviewer-administered tool used in the evaluation and monitoring of the neurological status of patients with stroke in the acute phase (Côté, Hachinski, Shurvell, Norris & Wolfson, 1986). The CNS demonstrates adequate reliability and validity (Côté et al., 1989), and evaluates 10 clinical domains of deficits due to stroke, including mentation (level of consciousness, orientation and speech) and motor function (face, arms and legs). Scores range from 1.5 to 11.5, with lower scores indicating greater neurological deficit. Stroke severity can be categorized as mild stroke (CNS $\geq$ 8), moderate stroke (score of 5 to 7), and severe stroke (score of 1 to 4). The CNS was completed at eight to 10 days post stroke and at discharge by a trained research assistant who observed participants in person and rated their ability to answer questions and perform activities.

Dichotomous covariates were created for categories of population subgroups viewed as potentially resulting in spectrum bias, including sex, age, language, level of education, and stroke severity. Non-dichotomous variables were converted into dichotomous variables to test the functioning of index tests across both levels. To allow for equal number of participants (n) in each group, participants were separated into groups by dividing the group at the median for age (young and old), stroke severity (high and low), and level of education (high and low). Participants were also separated into groups for language (speaks English and speaks no English) to ensure adequate functioning with participants who do not speak English.

4. Stroke characteristics, other socio-demographic, and clinical information.

Information characterizing a participant's stroke was obtained from medical charts. This information included the side of the hemiplegia (left, right or bilateral), the side of the lesion (left, right or bilateral), the type of stroke (ischemic or hemorrhagic), whether it was the first occurrence of a stroke incident or not, and the number of symptoms present at the time of stroke. Discharge destination and number of medications at discharge were also obtained by reviewing the discharge orders in medical charts. Medical condition co-morbidity was assessed with a checklist of co-existing conditions (Appendix C).

Participants were screened for severe cognitive impairment in the participant recruitment phase with the *interviewer-administered Brief version of the Mini-Mental State Examination* (Brief-MMSE; Paveza et al., 1990). The Brief-MMSE has four items measuring orientation (10 points), recall (three points), and spelling WORLD ("MONDE" in French) backwards (five points), for a possible total score ranging from zero to 18 points. These items have been shown to account for 98.8% of the variability of the full MMSE. Furthermore, a score ≤ 13 correctly identified cognitively impaired individuals with 95.5% sensitivity and 90.5% specificity (Koenig, 1996). Therefore, participants with scores below 14 on the MMSE were labeled with severe cognitive impairment and were not eligible to participate in this study.

A patient's function was assessed with the *Barthel Index* (BI; Mahoney & Barthel, 1965). The Barthel Index is widely used, interview administered measure, for the degree of functional independence and limitation in basic activities of daily living (Finch, Brooks, & Stratford, 2002). Total scores for the Barthel Index are calculated on an ordinal scale from '0', indicating functional dependence, to '100', indicating functional independence.[21] Granger and Hamilton (1990) reported that scores of 60 and above indicate independence, scores of 59 – 41 indicate marked dependence, score of 40 – 21 indicate severe dependence, and scores of 20 or below indicate total dependence. The Barthel Index has been shown to have high internal consistency (Wade et al., 1983) and is reported to be reliable for use in stroke rehabilitation with test-retest reliability of 0.89 and inter-rater reliability of 0.95 (Balu, 2009). The Barthel Index was completed at baseline and at discharge.

Participants completed the paper-pencil *OARS Social Resources Scale* at all five time points, beginning at three days after the stroke event. The OARS Social Resources Scale is one dimension of a larger questionnaire, the Multidimensional Functional Assessment Questionnaire

(OMFAQ) developed by the Duke OARS program (Older Americans Resources and Services, 1989). The Social Resources Scale measures the frequency of social interaction, availability of emotional support, and perceived quality of social support through a series of five self-report questions (Bushnik, 2011). Items analysed in this study include a participant's marital status and who they currently live with, answered through selecting best responses to multiple-choice questions. Responses to both variables were scored as multiple-level nominal variables.

In addition, at three days post stroke participants were asked to mark their current employment status as "employed" or "not employed" on the dichotomously worded question on brief paper-pencil *Socio-Demographic Questionnaire.*

Participants who refused interviews at a given time point were asked to complete *refuser interviews* over the phone with the interviewer. Refuser questionnaires included three MHI-5 items, including MHI-5 c, f, and h. The refuser questionnaire also asked participants "have you been depressed in the last 3 months?" to which participants were asked to provide a yes or no response.

Data entry and software. All data was entered into a Microsoft Access database, with data entry of all analyzed variables being verified by a second research assistant to minimize data-entry errors. Most statistical analyses were completed with Statistical Analysis Software (SAS), primarily SAS University Edition, alongside SAS 9.2 and SAS 9.3. Exact 95% confidence intervals for some accuracy measures (e.g., Se, Sp, and PPV) were calculated using Clopper-Pearson interval method for exact binomial confidence interval using the Medcalc online calculator.

Development and scoring of index tests (candidate models for PSD screening)

Dummy coding items. A total of 32 mood questions, including 17 SS-GDS items, five MHI-5 items, nine SIS items, and one EQ-5D item were included as possible contender screening items at time points two to five. Twenty-two mood questions were included at time one, as the SIS and the EQ-5D were not administered at that time point. Participants responses to all 32 (or 22) questions were recoded at each time point into binary (1,0) variables in order to separately assess the predictive accuracy of each response option (i.e., cut off point) of contender items. The already binary SS-GDS items were converted from yes/no to 1,0. Participant responses to the MHI-5, SIS, and EQ-5D were initially entered as multi-level ordinal items, which were each manually converted to dichotomous variables by representing each item value with a separate

dichotomous variable through "dummy coding". The number of created binary items per original question equaled *a*-1 generated items per question, where *a* is the number of response options on the Likert-type scale for that question. For example, if a participant's response was coded '4' on the MHI-b question with six possible response options, then five binary items were generated to reflect that score, namely 1, 1, 1, 0, 0. Dummy coding generated 42 items for T1 and 80 items for T2 to T5: 17 SS-GDS items, 25 MHI-5 items, 36 SIS items, and two EQ-5D items.

Combining items into candidate index tests. Dummy coded items were arranged in three ways to form three types of index test models. The first type of predictor model, called *univariate models*, involved testing each of the 80 dichotomous items independently as index tests. In univariate models, each item was assessed for its independent accuracy in predicting the reference standard. The second and third type of model involved combining pairs of dichotomous items. In the second type of index test model, referred to as *conjunctive pair models* (a term borrowed from logic), a participant was required to endorse both of two items for the conjunctive combination to be scored '1', whereas endorsing fewer than both items resulted in a score '0'. In the case where one response was missing and the other response was scored '0', the conjunctive pair was scored '0'. The conjunctive pair was scored as missing when responses to both items were missing or when responses to one item was missing and the other was scored '1'. In the third type of index test model, referred to as *alternate pair models*, a participant was required to endorse one or both of two items for the index test to be scored '1'. The index test was scored '0' when neither item in the model was endorsed. When one item response was missing and the other item response was scored '1', the index test was scored '1'. A participant's score on alternate pair models were considered missing when responses to both items were missing, or when one item response was missing while the other was scored '0'. All index tests were generated in SAS.

At T1, there were 42 dummy coded items to arrange into candidate index test models, including 42 univariate models (consisting of one variable), 903 conjunctive pair models and 903 as alternate pair models, for a total of 1848 models tested against the reference standard depression outcomes. At T2 to T5, there were 80 dummy coded items to arrange into candidate models, including 80 univariate models, 3240 conjunctive pair models, and 3240 as alternate pair models, for a total of 6560 models tested against reference standard outcomes at T2 through T5.

Handling of missing data. In this study, missing data occurred at each time point in the following three forms: a) a missing interview was defined as a participant missing all interviews and questionnaires at a given time point; b) missing questionnaires were defined as a participant missing all items in a given questionnaire; and c) missing items were defined as a participant omitting responses to specific questions within an otherwise attempted questionnaire. In this study, the participants who were present for interviews were considered as representative of the population of interest: participants willing and available to be screened, referred, and followed by healthcare professionals for PSD. Furthermore, good screening tests for depression should not contain questions that participants, and particularly depressed participants, routinely avoid answering. Therefore, when particular items or questionnaires showed greater levels of missing data than others, this missing data was interpreted as evidence of poor functioning of that item or questionnaire.

Pairwise deletion (or available case analysis) was the selected method to handle missing data. A consequence of pairwise deletion is variations in sample size across analyses. An unequal n, when comparing index tests, would not be disadvantageous in this study and instead preserves valuable information about potentially problematic items and index tests. Due to the sufficiently large sample size, the loss of statistical power from reduction in sample size was viewed as tolerable. Participants missing the SCID-I/NP, the gold-standard reference measure for depression, at any time point could not be included in any analyses testing candidate models for that time point. In addition, built directly into the six-step index test selection process, discussed later in the methods, were two steps developed to reduce the likelihood that missing values contributed bias to index test selection. Specifically, step three and step six identify and exclude candidate index test models, a) showing evidence of high levels of skipped items, and b) showing evidence of test and item missingness associated with the level of measured depression.

Statistical Tests and Analyses

Tests used to characterize the sample. This section describes the tests employed to complete analyses for descriptive statistics and for characterizing depression post stroke. To inform test selection and interpretation when assessing between-group differences, the Assumption of Normality was tested for all demographic and clinical variables with interval and ratio scales (see Table 11 in Appendix E). Significant Shapiro-Wilk (W) and Kolmogorov-Smirnov (D) tests as well as skewness or kurtosis greater or equal to an absolute value of two

were interpreted as evidence of non-normal distributions. As all tested variables showed some evidence of non-normal distributions, non-parametric tests were applied in all cases.[22]

To examine between-group differences in non-normally distributed variables, such as age, the non-parametric Wilcoxon Mann-Whiney U (WMW) was applied for two groups and the non-parametric Kruskal-Wallis (K-W) test was applied for two or more groups. Significant K-W tests on three or more groups were further analysed with post hoc K-W or WMW tests. All analyses were two-sided unless specified as one-sided. To examine between-group differences in percentage response to variables measured with nominal scales such as marital status, chi-squared tests of homogeneity were applied, and probability was calculated with two-sided Fisher's exact tests, unless otherwise specified. Probability was only estimated with χ^2 in cases where calculations were too computationally demanding for the Fisher's exact test.[23] Statistically significant findings with more than two groups or response options were further analyzed with post hoc chi-squared and Fisher's exact tests. The McNemar repeated measure design test was performed when comparing participant responses across time on two dichotomous variables using a two by two (2x2) contingency table (e.g., consistency in rates of depression across time points). Non-parametric Spearman correlations were applied to assess for correlations between two variables measured on ordinal or greater scales. Mean differences between groups were analysed with Satterthwaite t-tests. In all cases, an alpha of ≤ 0.05 was required for statistical significance and two-tailed tests were employed unless otherwise specified.

Accuracy statistics. Several accuracy statistics were employed to select the appropriate screening tests at each time point (classification performance). These measures are described below, with Appendix H available for additional information on the properties of the accuracy measures, their particular strengths and limitations in the context of PSD screening, as well as the rationale for their selection in this study. Methods for assessing accuracy of candidate index tests were completed through the cross-classification of the index tests with the reference test via 2x2 contingency tables. Participant responses to the ultra-short index tests were categorized as either positive or negative, and their results compared with the reference standard (depressed or not depressed), so that for each comparison of index to reference test, participants without missing data would fall into one of the following four categories: true positive (TP); false positive (FP); true negative (TN); and false negative (FN). The frequencies of TP, FN, FP, and

TN were used to calculate partial and global measures for classification performance and post-test risk.

Sensitivity (Se) or True positive rate (TPR), expressed as a percentage, defines the proportion of correctly identified depression cases of all true depression cases. Se provides the probability that the test result will be positive when the condition is present. *Specificity (Sp) or True Negative Rate (TNR),* expressed as a percentage, defines the proportion of correctly identified non-depressed cases of all non-depressed cases. Sp provides the probability that the test result will be negative when the condition is absent. Se and Sp are characteristics of the test's accuracy and are not influenced by prevalence. For the purposes of a screening test, high values for Se are prioritized over high values for Sp. It is suggested that good screening measures have a minimum Se ≥ 80% and Sp ≥ 60% (Tyson, Burton, & McGovern, 2015; Bennett & Lincoln, 2006; Lincoln et al., 2003). Higher point estimates of Se ≥ 85% and Sp ≥ 70% were required in this study so as to more accurately identify presence and absence of depressive symptoms for several reasons described in Appendix D.1. The following are the calculations for Se and Sp:

$$Se = TP / (TP+FN)$$

$$Sp = TN / (TN+FP)$$

Odds ratio (OR) is a measure of effect size of the association between the test result and the outcome (e.g., depression), and does not vary with prevalence. It provides a ratio for the odds of depression when the test is positive relative to the odds of depression when the test is negative and is equal to the cross-product ratio in a 2x2 contingency table. OR has a lower bound of 0 and no upper bound, whereby OR scores less than 1.0 correspond to probabilities below chance ($p < 0.5$), scores of 1.0 correspond to probabilities equal to chance, and scores above 1.0 correspond to probabilities above chance. An OR of 2.0 indicates that the odds of depression are twice as likely to occur in a positive versus a negative response on the binary index test, with each unit change of the index test. Confidence intervals for OR provide the reliability for the OR estimate. The following is the calculation for the OR:

$$OR = (TP/FP) / (FN/TN) \text{ or } (TP*TN)/(FP*FN)$$

C-statistic (also referred to as C-index or Concordance) is a global performance statistic that characterises the whole test. It is a goodness of fit statistic in logistic regression that indicates the extent to which the test predicts presence of a condition (e.g., depression) better

than chance. More specifically, for a randomly selected depressed participant and a randomly selected non-depressed participant, the C-statistic provides the probability that the two participants will be correctly scored on the index test (DeLong, DeLong, & Clarke-Pearson, 1988). The C-statistic is calculated from both Sp and 1-Se. In this study, C-statistics were calculated in SAS. The C-statistic ranges from 0.5 to 1.0, where 0.5 indicates a measure performing on par with chance and 1.0 indicates that the model perfectly discriminates both positive and negative outcomes. Fits of 0.7 indicate reasonable fit and values exceeding 0.8 indicate strong fit (Hosmer & Lemeshow, 2000). The C-statistic is equally influenced by sensitivity and specificity and is not affected by prevalence. A minimum cut-point of C-statistic ≥ 0.75 is required of index tests for identifying 'any depression,' while other primary measures of accuracy will be employed in the selection of appropriate index tests. Please see Appendix H for additional discussion on this topic. The following is the calculation for the C-statistic:

$$C = .5\ (1 + ((C-D)/C+D+T))$$

Six-Step Selection Process to Identify Best Performing Index Tests at Each Time point

A systematic, six-step protocol, presented below, was developed and followed to select the preferred index test for screening for depression from all candidate index test models at each time point. All candidate index tests that met criteria for retention at each given step were retained for assessment in the subsequent step. Candidate models that fell short of retention criteria at any step were eliminated from further consideration. The six-step protocol was repeated at each time point. Figure 2 shows a flow-chart illustrating the index test development and the six systematic and sequential rule-based steps for screening test selection. Please see Appendix D for an additional rationale for the factors included in the process for index test selection as well as an additional description of the methodology applied.

Step one: Retained index test models must show acceptable minimum accuracy:

a) Point estimate of Se must be ≥ 85% for identifying MDE criteria.[24]
b) Point estimate of Se must be ≥ 85% for identifying 'any depression'.
c) Point estimate of Sp must be ≥ 70% for identifying 'no depression'.
d) Point estimate of unweighted c-statistic must be ≥ 0.75 for discriminating 'any depression' from 'no depression'.

Step two: To minimize spectrum bias, retained index test models must show acceptable minimum accuracy and similar functioning in sample subgroups (old and young, men and women, English and French speaking, high and low education, and high and low stroke severity), according to the following criteria:

a) Within each subgroup, Se must be ≥70% and Sp must be ≥ 60% for discriminating 'any depression' from 'no depression,' when number of depressed participants included in the subgroup is ≥4. The Se criterion was waived when <4 depressed participants were included in analyses for a subgroup.
b) The model must show minimal difference in accuracy across both subgroups of each dichotomously stratified variable. Specifically, difference in Se and differences in Sp across subgroups, for predicting any depression, must each be <30%, when ≥5 depressed cases are included in a subgroup. When <5 any depression cases are included in a subgroup, <40% difference in Se was required.

Step three: Missingness of model index tests should not be predicted by level of true depression.

a) A model is dropped if frequency of missing values for an index test model, at a given time point, where frequency of missing values is due to skipping items within an otherwise completed questionnaire, is meaningfully associated with meeting criteria for 'any depression' at that time point. Chi-square tests were completed to assess the relationship between missingness of index tests at a given time point (missing v. not missing) and status of depression (any depression v. no depression) at that time point. A meaningful association was defined as an Odds Ratio ≥ 2.0.[25]

Step four: Where > one index test model at a given time point contain identical items, the one index test model with the preferred cut-off scores is retained based on the following criteria:

a) The accuracy of index test model composed of the same items were compared (Se for predicting MDE and Se, Sp, and C-statistic for predicting 'any depression'). Gains in Se for predicting MDE were prioritized over gains in Se for predicting any depression. In addition, gains in Se were prioritized over gains in Sp, unless loss in global accuracy (c-statistic) was substantial. Where Se or Se were equal, models with fewer depressed cases missing from the analysis were preferred. Attention was taken to ensure that gains in Se were not attributable to missing cases of true depression from the analysis.

b) Where no clear difference in accuracy existed between models (with the same items that differ only in cut-off score), the model containing the more inclusive (i.e., lower) cut-off score was retained, as it would likely identify more cases of true depression in the population.

Step five: Select two to three models with best overall accuracy at each time point:

a) This author consulted Se, Sp, C-statistic and PPV statistics and selected the two to three models clearly demonstrating superior accuracy. Gains in Se for predicting MDE were prioritized over gains in Se for predicting any depression. Furthermore, gains in Se were prioritized over gains in Sp, unless loss in global accuracy (c-statistic) was substantial. In cases where Se or Sp were equal, models with fewer depressed cases missing from the analysis were preferred. This author ensured that gains in Se were not attributable to missing cases of true depression from the analysis.

Step six: The best two to three models were compared at each time point, with the preferred index test model(s) being based on the following criteria:

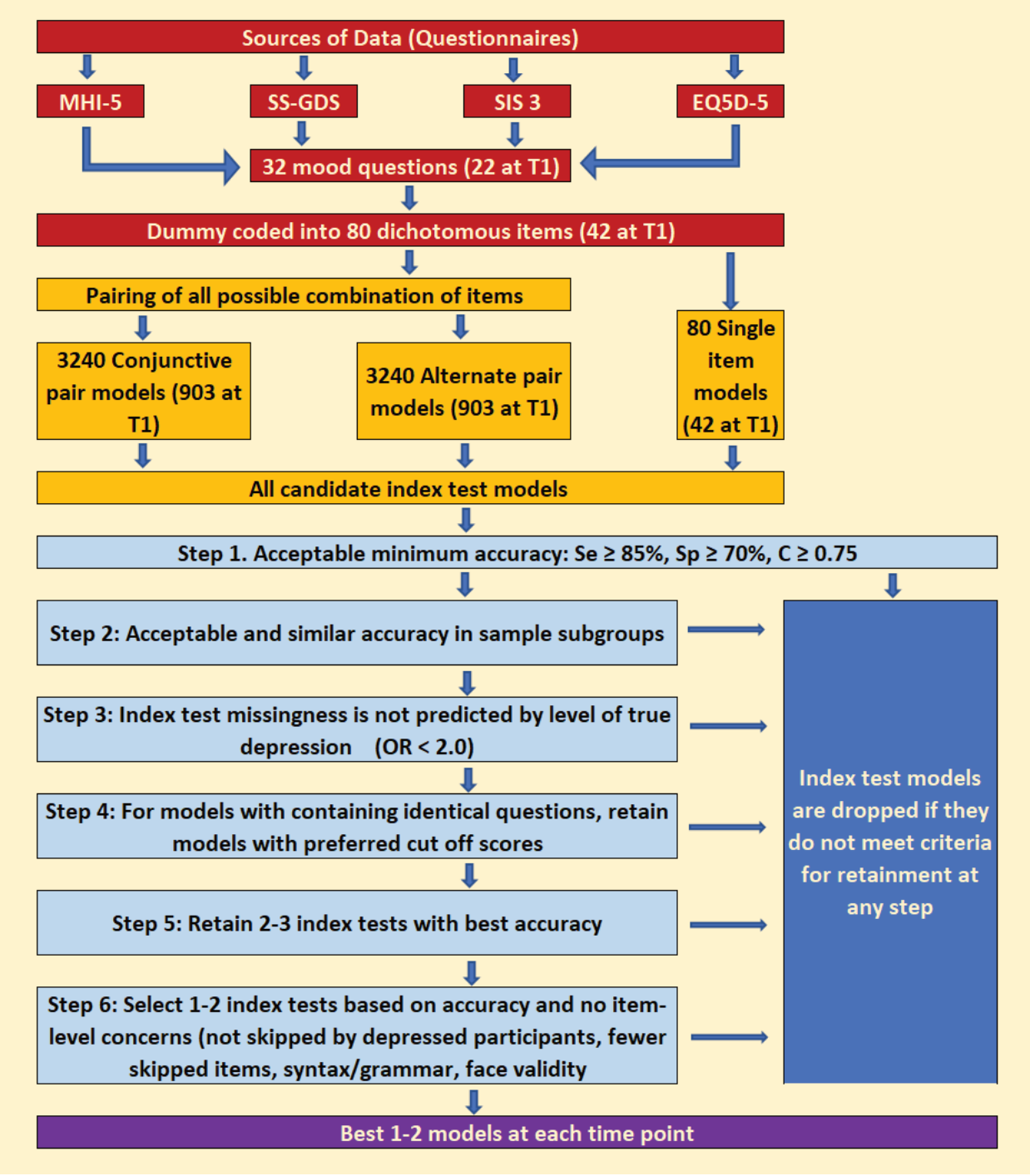
a) Models showing better accuracy statistics (Se, Sp, and C-statistic) were preferred (using the method specified in step five).
b) Where accuracy across models is equivalent, models with a greater number of depressed cases in the analysis are preferred.
c) Models showing no potential problematic items (higher levels of skipped items compared to other items in that questionnaire) were preferred.
d) Models containing questions with fewer concerns related to syntax and grammar were preferred. Item wording was examined via QUAID.[26]
e) Models containing questions without concerns related to face-validity were preferred. Face-validity was assessed through judgement of this study's author.

How Do the Best Index Tests at Each Time point Perform when Predicting Any Depression at Other Time points? Addressing Hypothesis Two

This study assessed the accuracy of the top two to three index test models at each time point for discriminating 'any depression' from no depression at all other time points. Specifically, Se and Sp point estimates were computed for all top models at all time points.

Figure 2

Flow-chart of index test development and selection process



Chapter 3: Results

Participants

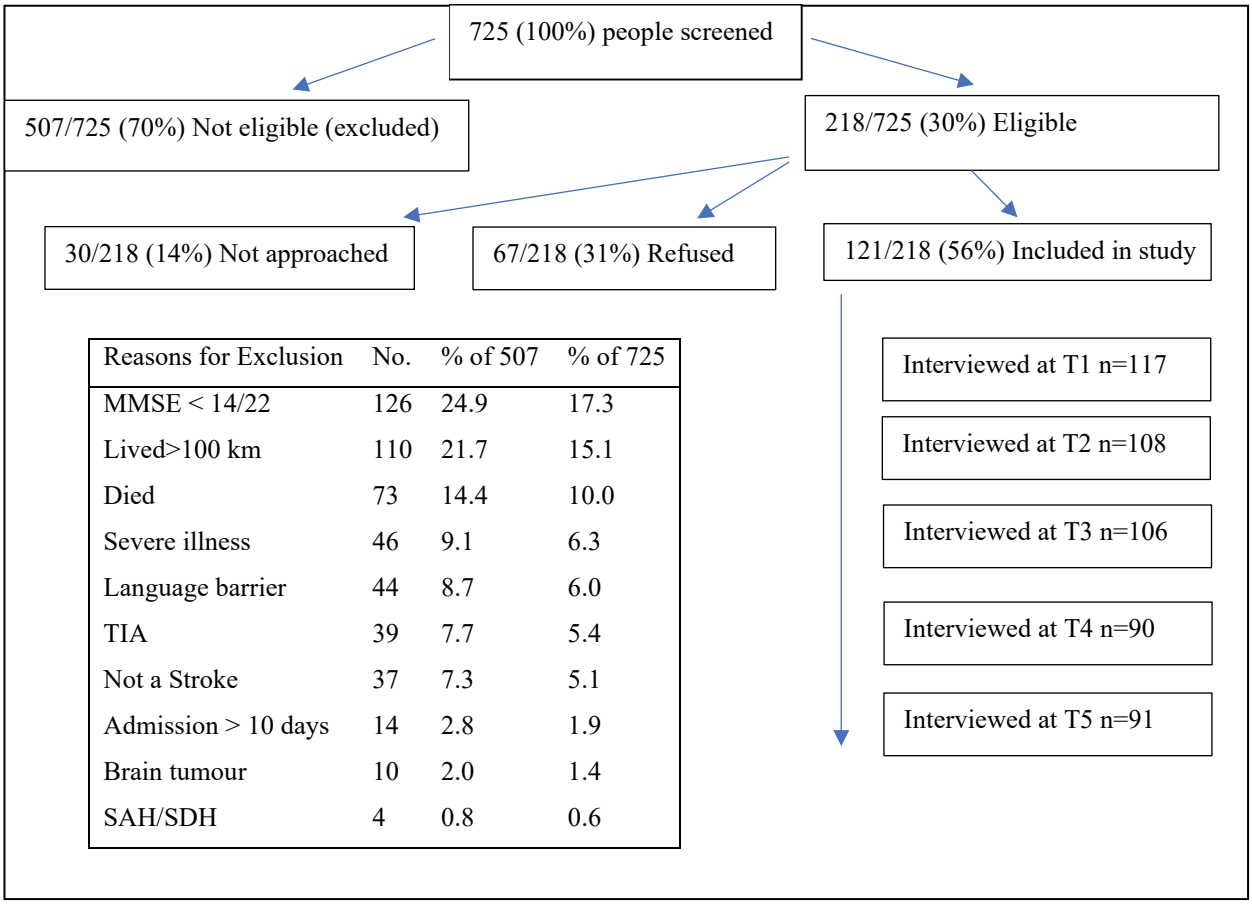
Participant recruitment. During the study period, 725 people were consecutively admitted within 10 days of stroke. Of admitted people, 70% were excluded for not meeting the inclusion/exclusion criteria. Of the 218 patients meeting inclusion criteria, 14% were not approached to participate in the study due to the lack of an available study coordinator.[27] Nine percent of approached patients refused to participate. Figure 3 provides a flow chart diagram of participant eligibility, exclusion, and inclusion in the study. Common reasons expressed for refusing to participate included "too much to cope with at the present time" (28%), "feeling too well" (15%), "too long a commitment" (11%) and "too depressed" (9%). The final sample consisted of 121 adults who agreed to participate, with an accrual rate of 17% (56% of eligible patients).

To investigate the generalizability of our sample, available descriptive and clinical characteristics of the present sample were compared with those of stroke survivors who refused to participate in this study. Two-tailed Satterthwaite t-tests compared mean and standard deviations of age, length of hospital stay (LOS), and Barthel Index at discharge.[28] A one-tailed t-test was employed to test if refusers had greater CNS scores. The chi-square test of homogeneity and Fisher's exact test for p values were employed to compare frequencies for distribution of sex and hospital attendance. Results are displayed in Table 3. Refusers had significantly better functioning at discharge, compared to study participants according to their Barthel Index Score. Consistently, stroke severity, as measured by CNS score, was significantly less severe in refusers compared to participants, assessed with a one-tailed t-test (but not a two-tailed t-test). No significant differences were found in age, sex, hospital, or length of hospital stay (LOS).

Participant characteristics. The following descriptive statistics and identified sex differences describe the sample of 121 at baseline, and results are presented in table 4. Tables 12 and 13 in Appendix E provide baseline demographic and clinical characterstics of 121 participants stratified and compared by sex. Date of admission to a hospital following stroke, retrieved from medical charts, ranged from September 2007 to March 2009. For most participants, the stroke was ischemic (92%) and their first reported stroke (84%), occurring in the right hemisphere (59%) with left side hemiplegia (55%). The mean and standard deviation (SD)

Figure 3

Flow chart of participant eligibility and inclusion in sample.



Reasons for Exclusion	No.	% of 507	% of 725
MMSE < 14/22	126	24.9	17.3
Lived>100 km	110	21.7	15.1
Died	73	14.4	10.0
Severe illness	46	9.1	6.3
Language barrier	44	8.7	6.0
TIA	39	7.7	5.4
Not a Stroke	37	7.3	5.1
Admission > 10 days	14	2.8	1.9
Brain tumour	10	2.0	1.4
SAH/SDH	4	0.8	0.6

Note. "Interviewed" is defined here as completing the SCID-I/NP interview at that particular time point.

Table 3

Comparing the characteristics of study participants with study refusers

Demographic and Clinical Variables at baseline	Participants (n=121)	Refusers (n=67)	Independent T- test with unequal variance: t(df), p; χ^2 (df, n), p
Age, mean (SD)	70.8 (13.1)	72 (14)	t (128) = -0.5758, p = 0.56
CNS score, mean (SD)	7.9 (2.6)	8.6 (2.5)	t (140) = -1.8125, p = 0.072
			For 1-tailed tests, p = 0.036*
LOS (days), mean (SD)	18.3 (15.7)	16 (13.0)	t (158) = 1.0771, p = 0.283
Barthel Index at discharge, mean (SD)	65.8 (29.8)	77 (30)	t (135) = -2.4574, p = 0.015*
Men, frequency (percentage)	69 (57%)	37 (55%)	χ^2 (1, N=188) = 0.057, p =0.878
Hospital, frequency (percentage)			χ^2 (2, N=188) = 0.380, p = 0.84
MGH	63 (52%)	33 (49)	
MNH	53 (44%)	30 (45)	
RVH	5 (4%)	4 (6)	

Note. Standard Deviation (SD); degrees of freedom (df); Canadian Neurological Stroke Scale (CNS); Length of hospital stay (LOS); Montreal General Hospital (MGH); Montreal Neurological Hospital (MNH); Royal Victoria Hospital (RVH). Mean differences between groups were analysed with the Satterthwaite t test, two-tailed tests, unless otherwise specified. Differences in frequencies between groups were analyzed with the Chi Square test and the reported significance is Fisher's exact test. *$p \leq 0.05$ significance level.

Table 4

Baseline demographic and clinical characteristics for all participants

Clinical variable of measure	Frequencies (Percent of Sample)
Age, *mean (SD)*	70.8 (13.1)
Men	69 (57%)
Spoken Language	121 (100%)
English, no French	69 (57.0%)
French, no English	33 (27.3%)
French and English	19 (15.7%)
Living with	
Spouse	55 (45.4%)
Alone	43 (35.5%)
Family	14 (11.6%)
Friends	2 (1.6%)
Other	7 (5.8%)
Highest Completed Education, mean rank (SD)	3.1 (1.9)
Less than grade 6 (0)	3 (2.4%)
Elementary school (1)	27 (22.3%)
High school (2)	30 (24.8%)
Some Cegep (3)	15 (12.4)
Some College or DEP (4)	7 (5.8%)
Graduated University (2-3 years) (5)	25 (20.7%)
Some graduate or professional School (6)	9 (7.4%)
Graduate or Professional School (7)	5 (4.1%)
Working before stroke	41 (33.9%)
Marital Status	120 (99%)
Single, never married	12 (9.9%)
Married or common law	58 (47.9%)
Widowed	34 (28.1%)
Divorced	16 (13.2%)
Separated	0
Type of Stroke	
Ischemic	111 (91.7%)
Haemorrhagic	10 (8.3%)

(continued)

Clinical variable of measure	Frequencies (Percent of Sample)
Lesion Side	
Left	46 (38.0%)
Right	71 (58.7%)
Bilateral	3 (2.5%)
Not noted	1 (0.8%)
Side of Hemiplegia	
Left	66 (54.6%)
Right	39 (32.2%)
Bilateral	0
None	16 (13.2%)
First stroke	102 (84.3%)
Any comorbid conditions	114 (96.6%)
Any dysphasia, recruiter sheet	19 (15.7%)
# of stroke symptoms, first three days post stroke, *mean (SD)*	3.6 (1.8)
CNS score, *mean (SD)*	7.9 (2.6)
CNS speech score (0, 0.5, 1), *mean (SD)*	0.65 (0.33)
Barthel at 3 days, (missing assigned 0), *mean (SD)*	48.8 (27.9)
Barthel at discharge, (missing assigned 0), *mean (SD)*	65.8 (29.8)
SF 36 Mental Health Index, *mean, (SD)*	70.7 (20.9)

Note. n=121. Values are frequencies with parentheses reflecting the percent unless otherwise specified. Canadian Neurological Scale (CNS); Degrees of freedom (DF); Standard deviation (SD).

of Canadian Neurological Scale (CNS) scores at eight to 10 days after the stroke event was 7.9 (SD = 2.6), which indicates mild to moderate stroke. At three days post stroke, 16% of participants had dysphasia. Participant Barthel Index scores at three days post stroke had a mean of 49 (SD = 28), indicating severe dependency, and a mean of 66 (SD = 30) by discharge, indicating moderate dependency. The mean SF-36 mental health index scores was 70.7 (SD = 20.9). At discharge, most participants (62%) went for rehabilitation and 31% went home. While the mode length of hospital stay (LOS) was eight days, LOS ranged from three to 102 days with 95% of participants discharged within 49 days.[29] Participants were prescribed a mean of 6.1 (SD = 4.2) different medications at discharge.

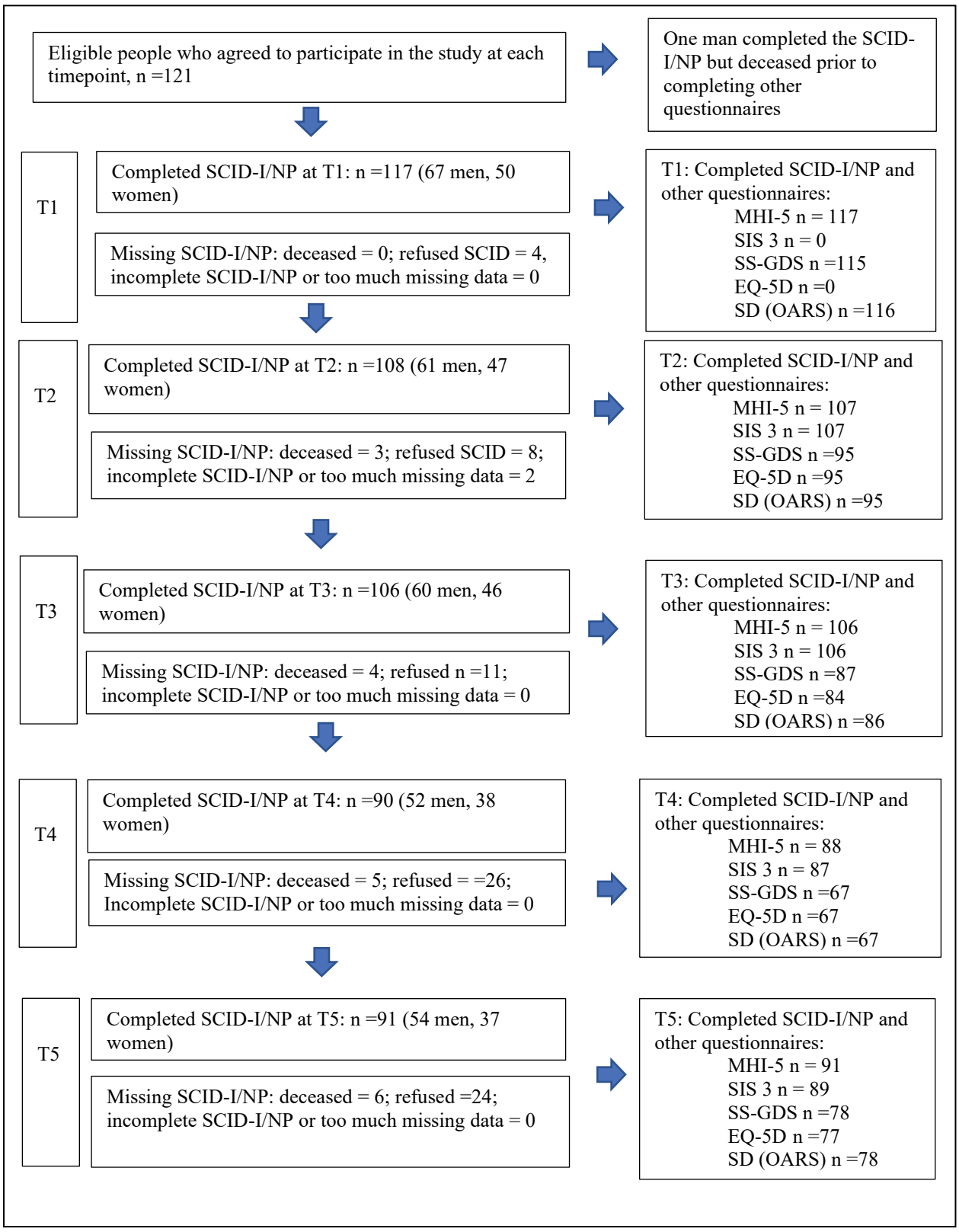
Fifty-seven percent of the cohort were men. Age ranged from 27 to 93 years, with a mean age of 70.8 years (SD = 13.1).[30] More than half of participants spoke English and no French. Level of education ranged from completing less than grade six to completing graduate or professional school, with half of the cohort reporting a minimum of at least some Cégep. Close to half of participants were married or in common law and were living with their spouce, while 35.5 percent of participants were living alone.

Sample Size and Missing Data

Frequencies and patterns of missing data. Figure 4 provides a flowchart of the 121 participants who completed the SCID-I/NP and mood-related questionnaires at each time point. Table 27 in Appendix G provides frequencies of participants missing complete interviews, missing specific questionnaires, and missing specific items at each time point. The rate of missing data increased with each subsequent time point, except for T5 for which the rate of missing data decreased. Thirty-five participants (30%) were missing a complete interview at least once during the study, and 44 participants (36%) were missing at least one SCID-I/NP interview during the study. The frequencies of participants without completed SCID interviews were four at T1, 13 at T2, 15 at T3, 31 at T4, and 30 at T5. Of the SCID-I/NP completers, the MHI-5 was the most frequently answered questionnaire at each time point. While Figure 4 shows that the frequency of participants attempting the SIS 3 questionnaire appears similarly high to the MHI-5, in fact many participants who completed the first SIS 3 item, SIS 3a, skipped all remaining SIS 3 items, resulting in missing item responses more similar in frequency to the SS-GDS. This data can be seen in Table 27 of Appendix I.

Figure 4

Flowchart of frequency of responders to measures at each time point



Reasons for missing data. Of the 44 participants missing at least one complete interview, 6 (14%) passed away during the study. Deceased participants were 50% men, 83% above the median age (72 years), 82% below the median level of education, 67% English speaking, and 67% above the median CNS score at baseline. None of the six deceased participants met criteria for any depression at any point during the study. To pre-emptively measure and control for biases associated with missing data, interview refusers were asked to participate in refuser telephone interviews, which asked reasons for refusal. Table 24 in Appendix G lists the reported reasons by participants for refusing to the complete the SCID-I/NP interview at each time point. Qualitatively speaking, death, sickness, feeling good, and busyness, were among the reported reasons for refusal. However, most participants provided no reasons for missing an interview. Due to the high percentage of missing data in refuser phone questionnaires (less than 24 % completion rate of MHI-5 items within a month of the missed interview), relationships between missing interviews and mood within the sample could not be estimated.

Dichotomous Covariates

To reduce the likelihood of spectrum bias in the developed screening tools, index tests were tested, in the six step selection protocol, for equal and effective functioning across both levels of each of the five dichomous covariates for sex, age, education, language, and CNS score. Participants were categorized into dichotomous subgroups for each covariate. In dichotomizing covariates, cut-off points for covariate subgroups were selected to allow for equal n in subgroups of each covariate, with the exception of sex. Please see Table 14 in Appendix E for information on the covariates, including definitions, cut off scores, dichotomization process, and group n.

Characterising Depression Post Stroke

Prevalence of each severity level of depression in the first-year post stroke onset, as well as incidence at each time point, were assessed. Table 5 lists the frequencies and percentages of the complete case analysis (n = 121) of participants meeting criteria for all severities of depression at any time in the study and at each of five time points. The incidence rate of post stroke depression varied with the severity level or diagnostic criteria used to define depression. Overall, within the first year after the stroke event, 52 participants (43% of the full sample) met criteria for any depression, defined as clinically significant endorsement of either depressed mood or loss of interest and pleasure, at least once in the first year. Of these 52 participants, 17

Table 5

Distribution of the total sample by most severe depression criteria met at any time during the study, and by criteria met at each time point

	Most severe criteria met at any time: n (% of 89)	T1: n (% of 117)	T2: n (% of 108)	T3: n (% of 106)	T4: n (% of 90)	T5: n (% of 91)
No depression	37 (42%)	109 (93.2%)	85 (78.7%)	82 (77.4%)	70 (77.8%)	74 (81.3%)
Any depression	52 (58%)	8 (6.8%)	23 (21.3%)	24 (22.6%)	20 (22.2%)	17 (18.7%)
MDE	17 (19%)	2 (1.7%)	10 (9.3%)	7 (6.6%)	2 (2.2%)	5 (5.5%)
Minor depression	18 (20%)	4 (3.4%)	5 (4.6%)	9 (8.5%)	9 (10.0%)	8 (8.8%)
Depressive features	17 (19%)	2 (1.7%)	8 (7.4%)	8 (7.5%)	9 (10.0%)	4 (4.4%)
Depressive features, with 1+ missing item	4 (4%)	1 (0.9%)	3 (2.8%)	2 (1.9%)	4 (4.4%)	1 (1.1%)
Depressive features, no missing items	13 (15%)	1 (0.9%)	5 (4.6%)	6 (5.7%)	5 (5.6%)	3 (3.3%)
Unknown, n (% of 121)	32 (26.4%)	4 (3.3%)	13 (10.7%)	15 (12.4%)	31 (25.6%)	30 (24.8%)

Note. Standard deviation (SD), Major depressive episode (MDE). Cell values refer to frequency and percentage of eligible participants. For each cell, percentage is calculated with a consistent column denominator, which represents available data. For example, for the 'most severe criteria met at any time' column, percentage is calculated from a denominator of 89 participants, for whom status of depression within the first year is known. As an exception to this rule, the cell percentages listed in the 'unknown' row are calculated from a denominator of all participants (n=121).

participants (14% of 121) met DSM-5 criteria for a major depressive episode (MDE) at least once during the study, 18 participants (15%) met DSM-IV-TR criteria for minor depression but not an MDE, and 17 participants (14%) met the criteria for depressive features, but not a more severe diagnosis. Notably, four of the 17 participants who met criteria for depressive features had at least one SCID-I/NP item response missing, and therefore may have been experiencing more severe depressive symptoms than was indicated on their SCID-I/NP interview. In all analyses, these four participants are included within the 'depressive features,' unless otherwise specified. Thirty-seven participants (30%) consistently showed no depression throughout the first year following stroke. The depression status of 32 participants (26%) could not be determined due to missing data.[31] As displayed in Table 5, the highest rates of any depression amongst respondents occurred from three to nine months post stroke (21 – 23%).

Time points when criteria for depression (all diagnostic severities) were first met are presented in Table 17 of Appendix F. As expected, the time point when depressive symptoms were first reported was heterogeneous across participants, and ranged from 10 days to 12 months post stroke.[32] Three months post stroke was the mode time for first reporting symptoms that met criteria for both any depression (n =18, 35% of 52) and for a major depressive episode (n =8, 47% of 17). Appendix F provides additional analyses and discussion on differences across time points in diagnosis and severity of depression, as well as relationships between demographic and clinical characteristics and depression onset time and severity.

Results of Six-Step Selection Process for Identifying the Best Index Tests at Each Time Point

Overview of results and tables. Table 6 provides a summary of requirements for retention at each step, as well as the number of candidate models retained at each step, at each time point. As shown in Table 6, the number of models demonstrating acceptable performance at each step varied across time points, suggesting that items function differently across time. At T1, T3, and T5, several models showed adequate accuracy and were retained at Step 1, whereas at T2 and T4, few models met criteria for this step. Table 7 summarizes quantitative and qualitative information used in the selection process of step 6 for all items retained in step 5. Table 28 in Appendix I presents a detailed examination of qualitative concerns related to individual items that inform model selection in step six. Some of these concerns include item wording, such as inclusion of terminology that is potentially unfamiliar to the participant, as well as imprecise,

Table 6

Number of models retained in each step of the six step selection process, at each time point

Models retained in each step (n)	T1 (3 to 10 days)	T2 (3 months)	T3 (6 months)	T4 (9 months)	T5 (12 months)
Step 1: Acceptable minimal accuracy	18 conjunctive pair models, 33 alternate pair models 2 univariate	3 alternate pair models	36 alternate pair models	5 alternate pair models	3 conjunctive pair models, 16 alternate pair models
Step 2: Acceptable model functioning across subgroups.	53 of 53 models	2 of 3 models	23 of 36 models	3 of 5 models	11 of 19 models
Step 3: Avoid biased results due to missingness associated with depression.	53 of 53 models	2 of 2 models	16 of 23 models	3 of 3 models	8 of 11 models
Step 4: Among tests composed of same items, select model with best cut-off scores.	34 of 53 models	2 of 2 models	9 of 16 models	3 of 3 models	6 of 8 models
Step 5: Select 2-3 best models, based on accuracy.	3 of 34 models	2 of 2 models	3 of 9 models	3 of 3 models	2 of 6 models
Step 6: Authors suggested best model(s) based on test accuracy and concerns at the item level.	2 selected	1 selected.	1 selected.	2 selected	1 selected

ambiguous, or confusing language; concerns about face validity of items in stroke populations; and concerns following relatively higher levels of skipped items. Table 8 presents the accuracy statistics of the final one to two models retained in step six criteria for each time point. The following are the results of the selection process presented by time point.

T1 results

Results of T1 selection process. Fifty-three models were retained in step one, 52 of which correctly identified seven of eight cases of any depression (Se=88%) and two of two MDE cases (Se=100%). One model correctly identified eight of eight (Se=100%) cases of any depression. Fifty-one models contained a combination of two items, whereas two models contained only one item. All 53 candidate T1 models passed step two criteria for acceptable model functioning across subgroups. As no depressed participants missed any T1 models due to skipping items, all models were retained in step three. Nineteen models were dropped in step four while the 34 models with the best cut-off scores were retained. Three models were retained in step five, of which, all three models showed superior C-Statistics, two models demonstrated superior specificity, and one model had superior Se. The latter model was not selected in step six as the incremental increase in Se was not considered to be worth the substantial loss in specificity associated with considerably higher FP rates, relative to the other two models. No substantial item level concerns were noted for any model.

Final T1 screening models. This methodology identified two models considered the best for ultra-short PSD screening tools for use at three to 10 days post stroke. Both models included the item MHI-5 H with cut off score of '5' (have you been a happy person? - A little of the time or less often) in combination with a second item. In model one, stroke survivors were required to endorse both MHI-5 H with cut off score of '5' and MHI-5 F with cut off score of '4' (feeling downhearted and blue some of the time or more often). In model two, stroke survivors were required to endorse both MHI-5 H with cut off score '5' and SS-GDS 2 (often getting bored). Both models appear to perform equally well at three to 10 days post stroke. The selected models for the ultra-short screening tool at three to 10 days post stroke correctly identified seven of eight cases of any depression, including two of two cases of MDE, three of four cases of minor depression, and two of two cases of depressive features. In identifying any depression, the selected models had Se=88% (95% CI= 47% to 100%), Sp≥ 90 (95% CI = 83% to 95%), C ≥ 89%, positive predictive value (PPV) ≥ 39% (95% CI = 26% to 54%), and Relative Risk (RR) ≥

Table 7

Selection criteria and rationale for model selection in step six

Timepoint	Model	TP	FP	TN	FN	Se % Any dep (95% CI)	Se % MDE (95% CI)	Sp % Any dep (95% CI)	C	Concerns at item level for a given timepoint?	Reason for author's selection
T1	SS-GDS 17 (NO) OR MHI-5 H (5)	8	28	79	0	100 (63.06 to 100)	100 (15.81 to 100.00)	74 (64.4 to 81.85)	0.869	-	Rejected for lesser accuracy. Incremental increase in Se, not worth loss in overall accuracy
	** MHI-5 F (4) AND MHI-5 H (5)	7	9	100	1	88 (47.35 to 99.68)	100 (15.81 to 100.00)	92 (84.90 to 96.15	0.896	-	Selected for best accuracy.
	** SS-GDS 2 (Yes) AND MHI-5 H (5)	7	11	98	1	88 (47.35 to 99.68)	100 (15.81 to 100.00)	90 (82.66 to 94.85)	0.887	-	Selected for best accuracy
T2	SS-GDS 14 (No) OR MHI-5 C (5)	16	24	55	2	89 (65.29 to 98.62)	90 (55.50 to 99.75)	70 (58.25 to 79.47)	0.793	GDS14, higher rate of skipped items by depressed pts	Not selected due to concerns at the item level associated with skipped items
	** SS-GDS 10 (Yes) OR MHI-5 C (5)	16	23	56	2	89 (65.29 to 98.62)	90 (55.50 to 99.75)	71 (59.58 to 80.57)	0.799	-	Selected due to minimal concerns at the item level.
T3	SIS 3 C (2) OR MHI-5 F (4)	19	20	53	2	90 (69.62 to 98.83)	100 (59.04 to 100.00)	73 (60.91 to 82.39)	0.815	Face validity concerns with measuring degree of 'feeling like a burden' in stroke population.	Not selected due to concerns at the item level associated with face-validity.

(Table 7 continued)

	** SIS 3 G (1) OR MHI-5 F (4)	19	20	53	2	90 (69.62 to 98.83)	100 (59.04 to 100.00)	73 (60.91 to 82.39)	0.815	-	Selected based on accuracy measures and fewest concerns at the item level.
	GDS 8 (Yes) OR MHI-5 B (3)	17	15	57	2	89 (66.86 to 98.70)	86 (42.13 to 99.64)	79 (67.98 to 87.84)	0.843	-	Not selected due to poorer accuracy with MDE cases and fewer depressed included in analyses.
T4	SIS 3 F (4) OR MHI-5 F (3)	13	14	43	2	87 (59.54 to 98.34)	100 (15.81 to 100.00)	75 (62.24 to 85.87)	0.811	More concerns associated with item wording.	Not selected for concerns regarding wording at the item level.
	** SS-GDS 8 (Yes) OR MHI-5 D (5)	12	13	42	2	86 (57,19 to 98.22)	100 (15.81 to 100.00)	76 (62.98 to 86.77)	0.81	-	Not selected because one fewer depressed case in the model
	** SS-GDS 8 (Yes) OR MHI-5 F (3)	13	15	41	2	87 (59.54 to 98.34)	100 (15.81 to 100.00)	73 (59.70 to 84.17)	0.799	-	Selected for relatively equal accuracy and fewer concerns at the item level.
T5	** EQ-5D (2) AND SS-GDS 1 (No)	11	13	51	2	85 (54.55 to 98.08)	100 (47.82 to 100.00)	80 (67.77 to 88.72)	0.822	-	Selected due to fewer item-level concerns.
	GDS 1 (No) AND MHI-5 D (2)	13	20	46	2	87 (59.54 to 98.34)	100 (47.82 to 100.00)	70 (57.15 to 80.41)	0.782	Concerns with face validity at the item level.	Not selected due to concerns at the item level.

Note. The authors recommended model at each timepoint. TP, TN, FP, FN refer to discrimination of any depression from no depression. Sensitivity, specificity, positive and negative predictive value are expressed as percentages. Confidence intervals for sensitivity and specificity are "exact" Clopper-Pearson confidence intervals. Wordings for items and responses are provided in table 8.

Table 8

Performance of the final one to two models retained in step six at each timepoint

Timepoint	Model		Se % Any depression (95% CI)	Se % MDE (95% CI)	Sp % Any depression (95% CI)	C Any depression	PPV % Any depression (95% CI)	RR Any depression (95% CI)
T1	MHI-5 F (4) AND MHI-5 H (5)	Q: Have you felt downhearted and blue? A: Some of the time or more often) AND Q: Have you been a happy person? A:a little of the time or less often	88 (47.35 to 99.68)	100 (15.81 to 100.00)	92 (84.90 to 96.15	0.896	43.75 (28.30 to 60.52)	44.19 (5.82 to 335.73)
	SS-GDS 2 (Yes) AND MHI-5 H (5)	Q: Do you often get bored? A: yes AND Q: Have you been a happy person? A: a little of the time or less often	88 (47.35 to 99.68)	100 (15.81 to 100.00)	90 (82.66 to 94.85)	0.887	38.89 (25.53 to 54.15)	38.50 (5.04 to 294.39)
T2	SS-GDS 10 (Yes) OR MHI-5 C (5)	Q: Do you feel that your life is empty? A: Yes OR Q: Have you felt so down in the dumps that nothing could cheer you up? A: A little of the time or more often	89 (65.29 to 98.62)	90 (55.50 to 99.75)	71 (59.58 to 80.57)	0.799	41.03 (32.22 to 50.45)	11.90 (2.90 to 48.87)
T3	SIS 3 G (1) OR MHI-5 F (4)	Feel quite nervous? A: All of the time OR Q: Have you felt downhearted and blue? A: Some of the time or more often	0.90 (69.62 to 98.83)	100 (59.04 to 100.00)	0.73 (60.91 to 82.39)	0.815	48.72 (38.94 to 58.59)	13.40 (3.31 to 54.22)

(Table 8 continued)

4	GDS 8 (Yes) OR MHI-5 D (5)	Q: Are you bothered by thoughts you can't get out of your head? A: Yes OR Q: Have you felt calm and peaceful? (a little of the time (5) or less often.	0.86 (57.19 to 98.22)	100 (15.81 to 100.00)	0.76 (62.98 to 86.77)	0.81	48.00 (35.41 to 60.85)	10.56 (2.57 to 43.44)
	GDS 8 (Yes) OR MHI-5 F (3)	Q: Are you bothered by thoughts you can't get out of your head? A: Yes OR Q: Have you felt downhearted and blue? A: A good bit of the time or more often.	0.87 (59.54 to 98.34)	100 (15.81 to 100.00)	0.73 (59.70 to 84.17)	0.799	46.43 (34.99 to 58.25)	9.98 (2.44 to 40.91)
5	EQ-5D (2) AND SS-GDS 1 (No)	Q: My health state is A: moderately anxious or depressed or worse AND Q: feel full of energy? A: No	0.85 (54.55 to 98.08)	100 (47.82 to 100.00)	0.80 (67.77 to 88.72)	0.822	45.83 (33.07 to 59.16)	12.14 (2.91 to 50.62)

te. TP, TN, FP, FN refer to discrimination of any depression from no depression. PPV is positive predictive value, RR is ative risk, CI is confidence intervals, Q is Question, A is Answer. Sensitivity, specificity, positive and negative predictive lue are expressed as percentages. Confidence intervals for sensitivity and specificity are "exact" Clopper-Pearson nfidence intervals.

38 (95% CI=5.04 to 294.39). Participants who tested positive in the final two models were approximately 39 times and 44 times, respectively, more likely to have any depression compared to participants who tested negative in the final models.

T2 results

Results of T2 selection process. Only three candidate index test models were retained in step one. Two models were retained in step two, while one model was not retained as it had low sensitivity among French participants. Both remaining models were retained in steps three, four, and five. The final two models showed a similar accuracy. In step six, one model was not selected due to concerns at the item level. Specifically, it contained item SS-GDS 14 which was skipped by at least two depressed participants who had completed the SS-GDS at that time point, suggesting problems with item functioning amongst depressed participants.

Final T2 screening models. The selected model for an ultra-short PSD screening test at three months post stroke identifies stroke survivors as depressed when they endorse either SS-GDS 10 (Do you feel that your life is empty? – Yes) or MHI-5 C with cut off score of '5' (Have you felt so down in the dumps that nothing could cheer you up? A little of the time or more often). This ultra-short combination of items identified 16 of 18 cases of any depression, which included nine of 10 cases of MDE, three of three cases of minor depression, and four of five cases of depressive features. In identifying cases of any depression this model had Se=89% (95% CI= 65% to 99%), Sp≥ 71 (95% CI = 60% to 80%), C = 80%, PPV = 41% (95% CI = 32% to 50%) and RR ≥ 12 (95% CI= 2.9 to 49). Participants who tested positive in the final model were approximately 12 times (RR) more likely to have any depression compared to participants who tested negative.

T3 results

Results of T3 selection process. Thirty-six models (all alternate pair models) were retained in step one. Thirteen of these models were discarded in step two for showing problematic functioning within one or more covariate subgroups for language, age, sex, and CNS score. Seven models were dropped in step three and an additional seven models were dropped in step four. Of the nine best models retained in step four, three models showed clearly superior accuracy statistics and were retained in step five. In step six, one model was not selected for face validity concerns at the item level as it identified stroke patients as depressed when they reported feeling like a burden to others, all of the time. This author interpreted that this item may be

measuring patient perceptions of the impact of their stroke on others, rather than a symptom of depression. A final model was selected as more depressed participants were included in the analyses for the model and it showed more favorable accuracy amongst participants with MDE.

Final T3 screening models. In the selected model for use as an ultra-short PSD screening tool at six months post stroke, stroke patients are identified as depressed when they either report a cut off score of '1' on item SIS 3G (Feel quite nervous? All of the time) or when patients report a cut off score of '4' on item MHI-5 F (have you felt downhearted and blue? Some of the time or more often). This candidate index test identified 19 of 21 cases of any depression, which included seven of seven cases of MDE, eight of nine cases of minor depression, and four of five cases of depressive features. In identifying any depression, this model had Se=90% (95% CI= 70% to 99%), Sp= 73 % (95% CI = 61% to 82%), C = 82%, PPV = 49% (95% CI = 39% to 58%) and RR ≥ 13 (95% CI= 3.3 to 54). Participants who tested positive in the final model were approximately 13 times more likely to have any depression compared to participants who tested negative on the final model.

T4 results

Results of T4 selection process. Five models met step one criteria. In step two, two models were dropped for demonstrating poor functioning with French participants. All three models were retained in step three. However, one of the three models was flagged for showing sub-threshold functioning issues. Specifically, a diagnoses of any depression was associated with missing this model due to skipping items included in the model, despite otherwise completing the questionnaires containing those items (OR=1.9, NS). Three of these three models were retained in steps four and five. In step six, one model was not selected due to concerns regarding potential ambiguity of item wording as well as for the sub-threshold functioning issues flagged in step three. The two final models in step six showed similar accuracy and no concerns at the item level.

Final T4 screening models. Both final models required participants to endorse either SS-GDS 8 (Are you bothered by thoughts you can't get out of your head? - Yes) or another item. In the first model, stroke patients are identified as depressed if they endorse either SS-GDS 8 or MHI-5 D with cut-off score of '5' (Have you felt calm and peaceful? A little of the time or less often). This candidate index test identified 12 of 14 cases of any depression, which included two of two cases of MDE, six of seven cases of minor depression, and four of five cases of

depressive features. In identifying any depression, this model had Se=86% (95% CI= 57% to 98%), Sp= 76 % (95% CI = 63% to 87%), C = 81%, PPV = 48% (95% CI = 35% to 61%) and RR ≥ 10 (95% CI= 2.6 to 43).

In the second final model, stroke patients are identified as depressed if they endorse either SS-GDS 8 or MHI-5 F with cut off score of '3' (Have you felt downhearted and blue? A good bit of the time or more often.) This candidate index test identified 13 of 15 cases of any depression, which included two of two cases of MDE, seven of nine cases of minor depression, and four of four cases of depressive features. In identifying any depression this model had Se=87% (95% CI= 60% to 98%), Sp= 73 % (95% CI = 60% to 84%), C = 80%, PPV = 46% (95% CI = 35% to 58%), and RR ≥ 10 (95% CI= 2.4% to 41%). Participants who tested positive in either of the final two models were approximately 10 times more likely to have any depression compared to participants who tested negative on the final model.

T5 results

Results of T5 selection process. Nineteen models met step one criteria. Of these models, eight models were dropped in step two for demonstrating poor accuracy in with women, French participants, and in subgroups of education and stroke severity. Three models were dropped in step three, two were dropped in step four, and the two best models were retained in steps four and five. In step six, one model was selected because it showed fewer potential concerns related to face validity at the item level.

Final T5 screening models. In the final ultra-short PSD screening test model at 12 months post stroke, patients are identified as depressed when they report '2' or more on item EQ-5D and endorse item 1 of the SS-GDS. The model rates a stroke patient as depressed when the patient rates both themselves at '2' or more on item EQ-5D (My health state is moderately anxious or depressed (2) or worse), and endorses SS-GDS 1 (Do you feel full of energy? - No). This final screening model identified 11 of 13 cases of any depression, which included five of five cases of MDE, five of six cases of minor depression, and one of two cases of depressive features. In identifying any depression this model had Se=85% (95% CI= 54% to 98%), Sp= 80 % (95% CI = 68% to 89%), C = 82%, PPV = 46% (95% CI = 33% to 59%) and RR ≥ 12 (95% CI= 2.9 to 51). Participants who tested positive in the final model were approximately 12 times more likely to have any depression compared to participants who tested negative on the final model.

How Do the Best Index Test Models at Each Time Point Fair when Predicting Any Depression at Other Time Points?

The accuracy of the best performing two to three index test models at each time point were assessed for accurately discriminating 'any depression' from no depression at all other time points. Results of the Se and Sp values for each analysis are presented in Table 9. Results show that no index test was able to predict any depression with the minimum Se and Sp defined in step one criteria for retention. These results support the idea that the effectiveness of ultra-short index tests is time dependent, and are consistent with the hypothesis that different screening questions are required at each time point.

Addressing Hypotheses

Results from this study support the two hypotheses. H1: Ultra-short screening tools containing two items, but not one item, showed acceptable minimum accuracy in identifying PSD relative to a gold-standard measure of depression. H2: The ultra-short screening tools that were best at identifying PSD at any one time point post stroke were not acceptably accurate at any other time points following stroke. In other words, different ultra-short screening tools were required for each of five time points to identify PSD with acceptable accuracy.

Table 9

Testing the accuracy with which top index tests screen for any depression at each of the other time points

Best Index Test Models	Predicting any depression T1		Predicting any depression T2		Predicting any depression T3		Predicting any depression T4		Predicting any depression T5	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
T1 best models										
Q: Are you hopeful about the future? A: No *OR* Q: Have you been a happy person? A: A Little of the time (or less often)	1.00	0.74**	0.53	0.83	0.59	0.80	0.64	0.82	0.57	0.81
^^Q: Have you felt downhearted and blue? A: Some of the time (or more often) *AND* Q: Have you been a happy person? A: A little of the time (or less often)	0.88	0.92**	0.39	0.98	0.33	0.95	0.15	0.99	0.33	0.92
^^Q: Do you often get bored? A: Yes *AND* Q: Have you been a happy person? A: A little of the time (or less often)	0.88	0.90**	0.29	0.96	0.09	0.95	0.25	0.97	0.27	0.96
T2 best models:										
^^Q: Do you feel that your life is empty? A: Yes *OR* Q: Have you felt so down in the dumps that nothing could cheer you up? A: A little of the time (or more often)	0.88	0.59	0.89	0.71**	0.95	0.66*	0.71	0.53	0.80	0.65*
Q: Do you think it is wonderful to be alive now? A: No *OR* Q: Have you felt so down in the dumps that nothing could cheer you up? A: A little of the time (or more often)	0.75	0.62	0.89	0.70**	0.95	0.66*	0.76	0.62	0.80	0.66*
T3 best models:										
Q: Feel that you are a burden to others? A: All of the time *OR* Q: Have you felt downhearted and blue? A: Some of the time (or more often)	-	-	0.74	0.75	0.90	0.73**	0.78	0.65	0.64	0.72
^^Q: Feel quite nervous? A: All of the time *OR* Q: Have you felt downhearted and blue? A: Some of the time or more often	-	-	0.74	0.76	0.90	0.73**	0.72	0.67	0.64	0.72

(Table 9 continued)

Best Index Test Models	Predicting any depression T1		Predicting any depression T2		Predicting any depression T3		Predicting any depression T4		Predicting any depression T5	
	Se	Sp	Se	Sp	Se	Sp	Se	Sp	Se	Sp
re you bothered by thoughts you can't get out of your head? A: *OR* Q: Have you been a very nervous person? A: A good bit of the time (or more often)	0.8	0.66*	0.56	0.68	0.89	0.79**	0.71	0.74	0.71	0.70
T4 best models:										
joy things as much as ever? A: Some of the time (or less often) : Have you felt downhearted and blue? A: Most of the time (or often)	-	-	0.68	0.74	0.75	0.68	0.87	0.75**	0.69	0.73
Are you bothered by thoughts you can't get out of your head? s *OR* Q: Have you felt calm and peaceful? A: A little of the or less often)	0.63	0.64	0.53	0.73	0.81	0.72*	0.86	0.76**	0.71	0.66
Are you bothered by thoughts you can't get out of your head? s *OR* Q: Have you felt downhearted and blue? A: A good bit time (or more often)	0.75	0.69	0.63	0.74	0.81	0.76*	0.87	0.73**	0.71	0.73
T5 best models:										
My health state is A: moderately anxious or depressed (or) *AND* Q: Feel full of energy? A: No	-	-	0.57	0.85	0.54	0.82	0.45	0.82	0.85	0.80**
you feel full of energy? A: No *AND* Q: Have you felt calm eaceful? A: Most of the time (or less often)	1.00	0.55	0.72	0.56	0.93	0.56	0.64	0.59	0.87	0.70**

No model met step one retention criteria when predicting any depression at other timepoints. **models meeting the ition of acceptable minimum accuracy at that timepoint (Se≥0.85, Sp≥0.70). *models meeting acceptable minimum acy at that timepoint (Se≥0.80, Sp≥0.60). ^^Models selected in step six

Chapter 4: Discussion

Addressing Hypotheses

This study tested the hypothesis that one or two-item screening tools would have acceptable accuracy for identifying cases of any depression and major depressive episodes relative to a gold-standard assessment of depression. At each of the five tested time points, two-item models showed very good accuracy for identifying PSD. These results support the conclusion that ultra-short PSD screening tools comprised of two items can be used to accurately screen for PSD at various time points within the first year following a stroke event.

Notably, there was less support for use of single-item models for identifying cases of any depression and major depressive episodes relative to a gold-standard assessment of depression. At three months, six months, nine months, and 12 months following the stroke event, this study tested 80 single item (i.e., univariate) models derived from every cut off score of 32 items assessing mood state obtained from four validated questionnaires. Results showed that of the 80 tested models, no single-item at any cut-off score was effective for use as an ultra-short PSD screening tool at three months, six months, nine months, or 12 months post stroke. At three to 10 days post stroke, this study tested 42 single-item models derived from 22 items assessing mood state obtained from two questionnaires. Results showed that only two of the 42 single-item models, involving items MHI-5 H and MHI-5 F, met the criteria for acceptable minimum accuracy for identifying depression in stroke patients. These two single-item models were, however, always outperformed by two-item models. Results did not support the hypothesis that single-item tools are accurate at identifying PSD.

This study also tested the hypothesis that different combinations of items assessing mood are required to accurately identify depression at different time points. Consistent with this hypothesis, this study found that the best models for ultra-short screening tests at any one specific time point did not meet the defined minimum accuracy requirements at any other time point and was always outperformed by a different combination of items. These results support the conclusion that different ultra-short combinations of items may be required to effectively screen for PSD at different times post stroke. The results suggest that PSD may be best assessed with use of time-sensitive ultra-short screening tools developed and validated for use at specific time points post stroke.

This study tested the accuracy of two-item models using two methods for combining pairs of dummy coded items into ultra-short screening tools. Tests with alternate pair models identified patients as depressed when the patients endorsed at least one of two items. Tests with conjunctive pair models identified patients as depressed when the patient endorsed both of two items. Results showed that alternative pair models showed good accuracy and functioning at all five time points. In fact, at three months, six months, and nine months post stroke, several alternate pair models (but no conjunctive pair models) met minimum accuracy requirements for identifying any depression. The success of alternative pair models, relative to conjunctive pair models between three to nine months post stroke, suggests the benefits of asking patients if they have either one symptom or another symptom, and highlights the heterogeneous presentation of the condition across stroke survivors and the insufficient sensitivity of assessing only one symptom. In other words, patients with PSD may experience the condition differently and report different symptoms at specific time points. In contrast, at three to 10 days and at 12 months post stroke, some conjunctive pair models as well as alternate pair models met sufficient accuracy requirements. Furthermore, at these two time points, conjunctive pair models were selected as they were considered best for use in ultra-short screening measures. The finding that conjunctive pair models were selected at T1 and T5 highlights that at these time points, test specificity was improved when two symptoms were required. For example, loss of energy (a somatic symptom of depression) was only discriminating at 12 months and only in combination with another symptom. It is possible that this finding may be explained by an increased presence of somatic symptoms in stroke populations, making it less discriminating of depression.

A qualitative comparison of results from steps one through six of the selection process suggest some potential differences in results across time points. While potential patterns are purely speculative, they may highlight areas for investigation in future research.

At three to 10 days post stroke, the item MHI-5 H with cut off score of 5 (Have you been a happy person? - A little of the time (5) or less often) was particularly important in identifying PSD. As a univariate model, this item and severity rating met minimum accuracy requirements for identifying any depression and it was included in each of the top three models at T1. Model sensitivity was better, however, when the depressed participants were asked to either endorse infrequent happiness or loss of hope. Results show that the best two models at this time identified patients as depressed when patients reported infrequent happiness as well as a second

symptom - downhearted and blue in model one, and often feeling bored in model two. While infrequent happiness correctly identified depressed patients, specificity was improved by also requiring either low mood or boredom. Importantly, neither infrequent happiness nor the top three models at T1 showed acceptable minimum accuracy at identifying PSD at any later time point.

At three months post stroke, only two models met requirements for minimum accuracy and functioning for identifying any depression, and neither of these models met the minimum accuracy requirements at any other time point. However, at six and 12 months, both models met Lincoln and colleagues' (2003) previous definition of minimum accuracy (Se ≥ 80%, Sp ≥ 60%). Both final models identified patients as depressed when patients either reported feeling so down in the dumps that nothing could cheer them up at least a little of the time or more often, or an alternative symptom. In the final model, the alternative symptom was feeling that life is empty.

At six months and nine months, feeling downhearted and blue, nervousness, and being bothered by thoughts were repeatedly included in the top three models at each time point, which were all alternate pair models. These findings suggest that nervousness and being bothered by thoughts may become important indicators of depression later on after stroke for some patients. For patients six months post stroke, the final selected model identified patients as depressed if they reported feeling downhearted and blue at least some of the time, or if they reported feeling nervous all of the time. At nine months post stroke, the final two models identified patients as depressed if they reported being bothered by thoughts or a second symptom, either feeling frequently downhearted or blue or infrequently calm and peaceful.

The level of a patient's energy appeared to become important for identifying depression at 12 months post stroke. In the two final models, patients were identified with PSD if they reported both not feeling full of energy and extreme anxiety or depression, and if they reported both not feeling full of energy and feeling calm and peaceful at any level less than all of the time. The final two models were conjunctive pairs and indicate that while sensitivities of the individual items were high, specificity was improved by inclusion of another item.

Contributions of Results

This study contributes to literature aimed at improving screening for Post Stroke Depression. This study is unique in that it investigates and contributes to our understanding of which ultra-short combination of items assessing mood state work best for screening depression,

relative to a gold-standard assessment, in stroke populations at three to 10 days, three, six, nine, and 12 months. Existing ultra-short depression screening tools were developed for use in primary care and a review of the literature indicates a lack of evidence supporting their accuracy for use in stroke populations. According to the literature review in this study, no previous studies have attempted to develop ultra-short screening tools for depression in stroke populations.

This appears to be the first study to examine what brief combination of items assessing mood state from validated mood questionnaires are optimal for PSD screening at specific time points within the first year post stroke. This study involved systematically generating and testing 80 single-item screening tool models and 6560 two-item screening tool models for accuracy at screening depression in stroke survivors at five separate time points within the first year of stroke. Using a systematic statistical approach, this study provides support for the most effective data-driven screening tool models for use at each time point.

This study provides data on the sensitivity and specificity of proposed ultra-short time-sensitive PSD screening tools, developed in a representative stroke sample in a longitudinal study. Results of this study suggest a time-specific ultra-short screening tool for use at separate time points following stroke, including at three to 10 days, three months, six months, nine months, and 12 months post stroke. Each ultra-short screening model contained two items that work best to identify the presence of mild and severe forms of PSD. Each model showed acceptable accuracy with stroke survivors and acceptable functioning across levels of sex, age, education, stroke severity, and the spoken languages of French and English.

As hypothesized, the ultra-short combination of items required to best identify PSD were different at each time point post stroke. When each of the top two to three models at each time point were tested at other time points, no model met our criteria for acceptable minimum accuracy at any other time point. Instead, different combinations of items worked best for identifying PSD at each specific time point. Results suggest that PSD may be best assessed with use of time-sensitive ultra-short screening tools developed and validated for use at specific time points post stroke. These findings are consistent with previous research indicating that assessment time post stroke may be associated with different PSD symptom profiles (Bush, 1998; Paradiso et al., 1997; Tateno et al., 2002) and that depression screening tests and their items perform differently at different times post stroke onset (Berg et al., 2009; Jackson & Alshekhlee, 2015).

While ultra-short screening tools have been recommended for efficiently screening PSD, the results contribute evidence that two-item screening tools would function better than one item screening tools at all time points within the first year of stroke. At every time point, two-item models outperformed one-item models at accurately identifying PSD.

This study also proposes a methodology for testing and considering model functioning across sample subgroups and validity as well as minimum accuracy. This methodology involved assessing candidate index test performance through a series of six steps to identify the best performing two to three ultra-short screening tool models at each time point, and proposing a preferred model based on an analysis of indicators for each model's accuracy and functioning. The screening tools were required to have a minimum sensitivity of 85% and a minimum specificity of 70%, which is higher than previous studies which required a lower accuracy of 80% sensitivity and 60% specificity. As the intent of this study was to develop the best time-specific PSD screening measures, identifying candidate index models with the highest accuracy was an important goal associated with this process.

In this study, the incidence rate of some severity of PSD within the first year was high (43%). Only 30% of participants consistently showed no depression throughout the first year and the depression status of 26% of participants could not be determined due to missing data. While three months post stroke was the mode PSD onset time, most cases presented onset at other time points between a few days to 12 months post stroke. These findings emphasize the importance of screening for PSD at multiple times post stroke.

Bridging the Gap Between Actual Practice and Best Practice

Results of this research contribute to bridging the gap between actual practice and best practice. It is intended that results from this study will help to engender widespread use of ultra-short screening tools that are both validated for accurate detection of post stroke depression at specific times post stroke, and that are quick and simple enough to support frequent use. Burton and Tyson (2015) state that clinical utility is rarely considered in tool development although it is key to uptake in clinical practice and research.

Tool use. Ultra-short PSD screening tools can be distributed to health-care professionals working with stroke survivors at various stages following stroke. The tools developed in this study are very short and simple and can be used in under a minute by untrained persons to screen for positive PSD results. The suggested screening tools have a specific decision rule to reduce

the need for untrained practitioners to score or interpret results. All patients screening positive should be referred to qualified practitioners for a thorough assessment of depression, prior to treatment. The tools are designed to identify both mild and severe forms of PSD, which allows all patients who may benefit from mental health care to be followed and to receive treatment.

Canadian Stroke Best Practice Recommendations (Lindsay, Gubitz., Bayley, & Phillips, 2013) emphasizes the value of empowering patients and families as an important component to working together towards reducing the gap between best practice and actual practice. Consistent with this value, families, caregivers, and patients can be given basic information about PSD upon discharge that includes the brief screening tools. Access to relevant information at home will help patients, families and caregivers advocate for their needs, for example, by requesting assessment for PSD by an appropriate professional.

Developing protocols and awareness. An important goal will be to promote awareness and compliance with best practice recommendations for PSD screening amongst both health care professionals as well as family members, patients and caregivers. The CSBPR suggests that PSD screening occur through all stages and settings following a stroke, including in acute care, rehabilitation, prevention clinics and outpatient community settings (including primary care, home care and long-term care). The CSBPR recommends that PSD screening and assessment practices be integrated into existing stroke protocols in various stages of stroke care, including during rehabilitation, return to the community, and during ongoing primary care and prevention clinic surveillance of patients. In their investigation of barriers to compliance with PSD screening requirement, Morris, Jones, Wilcox, and Cole (2012) emphasize the importance of developing systematic protocols to encourage widespread awareness of, access to, and use of feasible and validated PSD screening tools. This author agrees with Morris and colleagues, and highlights the importance developing protocols, and enhancing awareness of said protocols, which involve use of ultra-short, valid PSD screening tools, in order to continue closing the gap between actual practice and best practice. Morris et al. (2012) also note the importance of addressing knowledge-related barriers to professionals' compliance with existing screening protocols, as protocols are not effective if they are not followed. Increasing an awareness of PSD amongst varied professionals, its frequency of occurrence, and the value of screening may contribute to improving screening practices. Initiatives to address knowledge-related barriers to screening may include knowledge dissemination and translation initiatives to health-care

providers, patients, caregivers, and family members regarding the value of routine screening for PSD.

Knowledge dissemination initiatives may include access to free, simple, resource-effective, and validated PSD screening tools. Example initiatives may include distribution of pamphlets to all stroke patients, caregivers, and family members during acute care or at discharge. Stroke care pamphlets may provide key messages to improve stroke rehabilitation, including a brief description of PSD, ultra-short PSD screening questions to be asked at key times post stroke, with clear guidelines for when and how to obtain consultation about PSD with appropriate professionals. Karamchandani et al. (2015) report providing a packet of PSD education material to all stroke patients with minimal depression consisting of a PSD pamphlet and factsheet, as well as information on other post-stroke mood disorders.

Study Limitations and Directions for Future Research

Results of this study on a representative stroke sample are not expected to generalize to non-stroke populations, to stroke populations that differ importantly in demographic or clinical profiles, or to stroke survivors with conditions excluded from this study's sample.

Results from this study are considered exploratory. It will be necessary that the best functioning time-specific PSD screening tools developed and proposed here be assessed relative to a gold-standard reference test for PSD on a separate, representative stroke sample.

While all index tests proposed in this study showed good functioning at each level of the six assessed covariates, not all potential covariates were assessed. It is possible that proposed index tests function better in subgroups of the sample that were not measured, contributing to spectrum bias.

Methodology of this study required patients to complete multiple questionnaires and items. It is possible that the questions identified for use in ultra-short PSD screening tools may elicit different responses when completed outside of the context of large questionnaires and batteries. This possibility should be considered in future research validating the time-specific PSD screening tools developed in this study. For example, it is recommended that measures be administered in the format for the final screening tools offered in Appendix J. It is also recommended that tools be assessed for effective use within the contexts where they may be used.

In this study, stroke patients were not asked and did not directly contribute their perspective on the PSD screening tools proposed, as this objective was outside the scope of this study. As stroke survivors may contribute information to the understanding of the content validity of proposed PSD screening tools, a future research objective may include directly asking stroke survivors for their thoughts on ultra-short PSD screening tools.

Proxies' completion of self-report may have introduced some bias into results as ratings by caregivers are not identical to ratings by stroke patients themselves. Demonstrating evidence of the bias introduced by caregivers, Berg et al. (2009) showed that proxies tend to rate patient depression as more severe than patients rate their own depression. In this study, too few proxies responded to directly test the functioning of index tests completed by proxies during the test selection process. Further research could directly assess the accuracy and functioning of the index tests proposed here when tests are completed by proxies.

The first time point of the SCID-I/NP administration occurred at 10 days post stroke. In this study, the DSM-5 two-week criterion for an MDE was waived for T1, and symptoms were considered endorsed if present since the time of stroke (10 days). This approach was taken to prevent the loss of important information about episode severity. In particular, this study was designed to distinguish between stroke patients with one symptom, two to four symptoms, and five or more symptoms, so as to ensure that candidate index tests did not miss the most severe cases of depression. A limitation of this approach was that some patients may have been miscategorized with an MDE if their symptoms resolved between 11 and 14 days post stroke. A potential benefit of testing at 10 days rather than 14 days post stroke is that early screening can allow more stroke patients with mood difficulties to be flagged before they are discharged from the hospital. Early detection and treatment are encouraged by the CSBPR as it is viewed as helpful to prognosis.

Future research can assess the impact of integration of ultra-short screening tools into protocols used in hospital and rehabilitation centers, to track for improvements in the screening rate of PSD. The ultra-short screening tools proposed in this study are designed to address the barriers to PSD screening by health care practitioners. The tools can also be sent home with patients and families to improve ability to identify PSD at home. This study proposes an example of how ultra-short screening tools can be packaged for distribution, with the goals of improving knowledge translation and screening rates (Appendix J).

One typical limitation of this study is the impact of missing data on study results. In this research, measures were taken to assess, measure, and control for variables thought to be associated with missingness. As discussed by Graham (2009), attrition-related missingness is ubiquitous to longitudinal studies, can be expected to be partially missing not at random (MNAR), and therefore should be treated as MNAR. In this study, it was expected that missing values for reference standard outcomes of depression were at least partially dependent on these unmeasured levels of depression. The possibility of biased parameter estimates due to data MNAR cannot be completely eliminated and is therefore a limitation in this study. Study responders were interpreted as reflective of the actual population of interest. Missing interviews were interpreted as reflective of a meaningful subgroup of stroke survivors likely to miss or refuse mental health assessment. Patterns of frequently missing items and questionnaires in otherwise completed interviews were interpreted as evidence of poor functioning of that questionnaire or item. Therefore, pairwise deletion was selected for handling missing data as it was viewed as the best method for limiting biased parameter estimates and biased conclusions that do not generalize to the population of interest.

The primary limitation of pairwise deletion in this study is reductions in sample size and loss of some statistical power. Due to the number of participants and cases of true depression at each time point, there was a limit to the analytical investigations that were completed with sufficient power. As a result of lower power, the lower bound confidence intervals for sensitivity of the best performing screening tools relative to the gold-standard were as low as 47% in T1 and ranged from 47% to 70% across time points. These results suggest that chance may account for the excellent sensitivity of the identified screening tools. In contrast, all of the screening tools proposed at each time point in this study had lower bound confidence intervals for specificity at 60% or greater. These results suggest that the tools are likely to meet specificity requirements in similar populations. Future research testing the proposed screening tools on larger samples sizes will be required to validate the tools' accuracy with improved confidence intervals for sensitivity at the time points identified in this study.

There was insufficient power to assess for significance in the change of the selected models' sensitivity and specify across time points. Future studies may measure the effect size and statistical significance of change in PSD screening tool accuracy across time. Results of such a study may further inform the understanding of how PSD screening tools function

differently at different times post stroke and may contribute further evidence supporting the need to validate PSD screening tools at specific times post stroke.

Chapter 5: Conclusion

Results of this study indicate that two-item screening tools containing items assessing mood state, can accurately screen for depressive disorders in stroke patients and that the accuracy of ultra-short index tests for PSD are time dependent. Results are consistent with previous studies suggesting that different screening questions are required at each time point following stroke. Based on this study's representative stroke sample, this research proposes the best combinations of questions at five time points within the first year of stroke that were most effective for identifying mild and severe forms of depression, relative to the DSM-5 reference test. Availability of feasible and accurate PSD screening tools are expected to improve screening rates thereby increasing rates of early identification and treatment of PSD. Improving rates of early detection and treatment of PSD is expected to reduce negative outcomes associated with stroke, improve stroke rehabilitation outcomes, and reduce costs and burden on patients, families, caregivers, medical services, and society at large. Continued development of PSD screening tools can minimize the substantial negative impacts and costs of PSD through addressing barriers to effective screening practices, and thereby contribute to closing the gap between actual practice and best practice in PSD screening.

References

Aben, I., Lodder, J., Honig, A., Lousberg, R., Boreas, A., & Verhey, F. (2006). Focal or generalized vascular brain damage and vulnerability to depression after stroke: A 1-year prospective follow-up study. *International Psychogeriatrics*, *18*(1), 19–35. doi:10.1017/S104161020500270X

Aben, I., Verhey, F., Strik, J., Lousberg, R., Lodder, J., & Honig, A. (2003). A comparative study into the one-year cumulative incidence of depression after stroke and myocardial infarction. *Journal of Neurology, Neurosurgery & Psychiatry*, *74*(5), 581–585. doi:10.1136/jnnp.74.5.581

Adams, G. F. & Hurwitz, L. J. (1963). Mental barriers to recovery from strokes. *Lancet, 282*(7307)*,* 533–537. doi:10.1016/S0140-6736(63)92636-9

Agrell, B., & Dehlin, O. (1989). Comparison of six depression rating scales in geriatric stroke patients. *Stroke*, *20*(9), 1190–1194. doi:10.1161/01.STR.20.9.1190

Alexopoulos, G. S., Wilkins, V. M., Marino, P., Kanellopoulos, D., Reding, M., Sirey, J. A., . . . & Kiosses, D. N. (2012). Ecosystem focused therapy in poststroke depression: A preliminary study. *International Journal of Geriatric Psychiatry*, *27*(10), 1053–1060. doi:10.1002/gps.2822

Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, *112*(4), 545–557. doi:10.1037/0021-843X.112.4.545

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (Rev. 3rd ed.). Washington, DC: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (Rev. 4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author. doi:10.1176/appi.books.9780890425596

Andersen, G., Vestergaard, K., & Lauritzen, L. (1994). Effective treatment of poststroke depression with the selective serotonin reuptake inhibitor citalopram. *Stroke*, *25*(6), 1099–1104. doi:10.1161/01.STR.25.6.1099

Andersen, G., Vestergaard, K., Ingemann-Nielsen, M., & Lauritzen, L. (1995). Risk factors for post-stroke depression. *Acta Psychiatrica Scandinavica*, *92*(3), 193–198. doi:10.1111/j.1600-0447.1995.tb09567.x

Anderson, C., Laubscher, S., & Burns, R. (1996). Validation of the Short Form 36 (SF-36) health survey questionnaire among stroke patients. *Stroke*, *27*(10), 1812–1816. doi:10.1161/01.STR.27.10.1812

Ayerbe, L., Ayis, S., Crichton, S., Wolfe, C. D., & Rudd, A. G. (2013). The natural history of depression up to 15 years after stroke: The South London stroke register. *Stroke*, *44*(4), 1105–1110. doi:10.1161/STROKEAHA.111.679340

Ayerbe, L., Ayis, S., Wolfe, C. D., & Rudd, A. G. (2013). Natural history, predictors and outcomes of depression after stroke: systematic review and meta-analysis. *The British Journal of Psychiatry*, *202*(1), 14-21. doi:10.1192/bjp.bp.111.107664

Ayerbe, L., Ayis, S., Rudd, A. G., Heuschmann, P. U., & Wolfe, C. D. (2011). Natural history, predictors, and associations of depression 5 years after stroke: The South London stroke register. *Stroke*, *42*(7), 1907–1911. doi:10.1161/STROKEAHA.110.605808

Balu, S. (2009). Differences in psychometric properties, cut-off scores, and outcomes between the Barthel Index and Modified Rankin Scale in pharmacotherapy-based stroke trials: systematic literature review. *Current Medical Research and Opinion*, *25*(6), 1329–1341. doi:10.1185/03007990902875877

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *BDI-II, Beck depression inventory: Manual*. San Antonio, TX: Psychological Corporation. doi:10.1037/t00742-000

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). *Manual for the Beck depression inventory - Fast screen for medical patients*. San Antonio TX: Psychological Corporation.

Beck, A. T., Ward, C., Mendelson, M., Mock, J., & Erbaugh, J. (1961). Beck depression inventory (BDI). *Archives of General Psychiatry*, *4*(6), 561–571. doi:10.1001/archpsyc.1961.01710120031004

Beekman, A. T. F., Penninx, B. W. J. H., Deeg, D. J. H., Ormel, J., Smit, J. H., Braam, A. W., & Van Tilburg, W. (1998). Depression in survivors of stroke: a community-based study of prevalence, risk factors and consequences. *Social Psychiatry and Psychiatric Epidemiology*, *33*(10), 463-470. doi:10.1007/s001270050080

Bennett, H. E., & Lincoln, N. B. (2006). Potential screening measures for depression and anxiety after stroke. *International Journal of Therapy and Rehabilitation*, *13*(9), 401–406. doi:10.12968/ijtr.2006.13.9.21784

Berg, A., Lönnqvist, J., Palomäki, H., & Kaste, M. (2009). Assessment of depression after stroke: A comparison of different screening instruments. *Stroke*, *40*(2), 523–529. doi:10.1161/STROKEAHA.108.527705

Berg, A., Palomäki, H., Lehtihalmes, M., Lönnqvist, J., & Kaste, M. (2003). Poststroke depression: An 18-month follow-up. *Stroke*, *34*(1), 138–143. doi:10.1161/01.STR.0000048149.84268.07

Bhogal, S. K., Teasell, R., Foley, N., & Speechley, M. (2004). Lesion location and poststroke depression: Systematic review of the methodological limitations in the literature. *Stroke*, *35*(3), 794–802. doi:10.1161/01.STR.0000117237.98749.26

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L., . . . & Kressel, H. Y. (2015). STARD 2015: An updated list of essential items for reporting diagnostic accuracy studies. *Radiology*, *277*(3), 826-832. doi:10.1148/radiol.2015151516

Bowen, A., Knapp, P., Hoffman, A., & Lowe, D. (2005). Psychological services for people with stroke: Compliance with the UK National Clinical Guidelines. *Clinical Rehabilitation*, *19*(3), 323–330. doi:10.1191/0269215505cr799oa

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. doi: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. doi:10.3109/09638288.2012.709305

Bush, B. A. (1998). Longitudinal assessment of poststroke depression: The stability of the factor structure of the CES-D in the first year poststroke. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *58*(12-B), p. 6802.

Bushnik, T. (2011). Older Americans' resources and services multidimensional functional assessment questionnaire. In J. S. Kreutzer, J. DeLuca, & V. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 132–203). New York, NY: Springer. doi:10.1007/978-0-387-79948-3

Cameron, J. I., O'Connell, C., Foley, N., Salter, K., Booth, R., Boyle, R., . . . & Lindsay, P. (2016). Canadian stroke best practice recommendations: Managing transitions of care following stroke, guidelines update 2016. *International Journal of Stroke*, *11*(7), 807–822. doi:10.1177/1747493016660102

Canadian Stroke Network. (2011). *The quality of stroke care in Canada*. Ottawa, ON: Author.

Carod-Artal, J., Egido, J. A., González, J. L., & Varela de Seijas, E. (2000). Quality of life among stroke survivors evaluated 1 year after stroke: Experience of a stroke unit. *Stroke*, *31*(12), 2995-3000. doi:10.1161/01.STR.31.12.2995

Carter, L. T., Oliveira, D. O., Duponte, J., & Lynch, S. V. (1988). The relationship of cognitive skills performance to activities of daily living in stroke patients. *The American Journal of Occupational Therapy*, *42*(7), 449–455. doi:10.5014/ajot.42.7.449

Chemerinski, E., Robinson, R. G., Arndt, S., & Kosier, J. T. (2001). The effect of remission of poststroke depression on activities of daily living in a double-blind randomized treatment study. *The Journal of Nervous and Mental Disease*, *189*(7), 421–425. doi:10.1097/00005053-200107000-00002

Chen, Y., Guo, J. J., Zhan, S., & Patel, N. C. (2006). Treatment effects of antidepressants in patients with post-stroke depression: A meta-analysis. *Annals of Pharmacotherapy*, *40*(12), 2115–2122. doi:10.1345/aph.1H389

Cinamon, J. S., Finch, L., Miller, S., Higgins, J., & Mayo, N. (2011). Preliminary evidence for the development of a stroke specific geriatric depression scale. *International Journal of Geriatric Psychiatry*, *26*(2), 188–198. doi:10.1002/gps.2513

Clark, M. S., & Smith, D. S. (1998). The effects of depression and abnormal illness behaviour on outcome following rehabilitation from stroke. *Clinical Rehabilitation*, *12*(1), 73–80. doi:10.1191/026921598669567216

Collin, S. J., Tinson, D., & Lincoln, N. B. (1987). Depression after stroke. *Clinical Rehabilitation*, *1*(1), 27–32. doi:10.1177/026921558700100107

Côté, R., Battista, R. N., Wolfson, C., Boucher, J., Adam, J., & Hachinski, V. (1989). The Canadian Neurological Scale: Validation and reliability assessment. *Neurology*, *39*(5), 638–643. doi:10.1212/WNL.39.5.638

Côté, R., Hachinski, V. C., Shurvell, B. L., Norris, J. W., & Wolfson, C. (1986). The Canadian Neurological Scale: A preliminary study in acute stroke. *Stroke*, *17*(4), 731–737. doi:10.1161/01.STR.17.4.731

Coulter, D. M., & Edwards, I. R. (1990). Mienserin and agranulocytosis in New Zealand. *Lancet*, *336*(8718), 785–787. doi:10.1016/0140-6736(90)93248-N

Coupland, C. A. C., Dhiman, P., Barton, G., Morriss, R., Arthur, A., Sach, T., & Hippisley-Cox, J. (2011). A study of the safety and harms of antidepressant drugs for older people: A cohort study using a large primary care database. *Health Technology Assessment*, *15*(28), 1–202. doi:10.3310/hta15280

de Abajo, F. J., Rodríguez, L. A. G., & Montero, D. (1999). Association between selective serotonin reuptake inhibitors and upper gastrointestinal bleeding: Population based case-control study. *British Medical Journal*, *319*(7217), 1106–1109. doi:10.1136/bmj.319.7217.1106

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*, *43*(3), 854–856. doi:10.1161/STROKEAHA.111.640276

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, *44*(3), 837–845. doi:10.2307/2531595

Demirtas, H., Freels, S. A., & Yucel, R. M. (2008). Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continuous outcomes: A simulation assessment. *Journal of Statistical Computation and Simulation*, *78*(1), 69–84. doi:10.1080/10629360600903866

Diller, L., & Bishop, D. S. (1995). Depression and stroke. *Topics in Stroke Rehabilitation*, *2*(2), 44–55. doi:10.1080/10749357.1995.11754069

Dorman, P., Slattery, J., Farrell, B., Dennis, M., & Sandercock, P. (1998). Qualitative comparison of the reliability of health status assessments with the EuroQol and SF-36 questionnaires after stroke. *Stroke*, *29*(1), 63–68. doi:10.1161/01.STR.29.1.63

Dorman, P. J., Waddell, F., Slattery, J., Dennis, M., & Sandercock, P. (1997a). Is the EuroQol a valid measure of health-related quality of life after stroke? *Stroke*, *28*(10), 1876–1882. doi:10.1161/01.STR.28.10.1876

Dorman, P. J., Waddell, F., Slattery, J., Dennis, M., & Sandercock, P. (1997b). Are proxy assessments of health status after stroke with the EuroQol questionnaire feasible, accurate, and unbiased? *Stroke*, *28*(10), 1883–1887. doi:10.1161/01.STR.28.10.1883

Duncan, P. W., Bode, R. K., Lai, S. M., Perera, S., & Glycine Antagonist in Neuroprotection Americas Investigators. (2003). Rasch analysis of a new stroke-specific outcome scale: The Stroke Impact Scale. *Archives of Physical Medicine and Rehabilitation*, *84*(7), 950–963. doi:10.1016/S0003-9993(03)00035-2

Duncan, P. W., Lai, S. M., Wallace, D., Embretson, S., Johnson, D., & Studenski, S. (1999). *Stroke impact scale version 3.0*. Kansas City, MO: University of Kansas Medical Center. doi:10.1037/t05210-000

Eskes, G. A., Lanctôt, K. L., Herrmann, N., Lindsay, P., Bayley, M., Bouvier, L., ... & Gubitz, G. (2015). Canadian stroke best practice recommendations: Mood, cognition and fatigue following stroke practice guidelines, update 2015. *International Journal of Stroke*, *10*(7), 1130-1140. doi:10.1111/ijs.12557

Eskes, G., & Salter, K. (2013). Mood and cognition in patients following stroke. In M. Lindsay, G. Gubitz, M. Bayley, & S. Phillips (Eds.) *Canadian best practice recommendations for stroke care* (4th ed.) Chapter 7 (pp. 7-36). Ottawa, ON: Canadian Stroke Network.

Fang, J., & Cheng, Q. (2009). Etiological mechanisms of post-stroke depression: A review. *Neurological Research*, *31*(9), 904–909. doi:10.1179/174313209X385752

Finch, E., Brooks, D., & Stratford, P. W. (2002). *Physical rehabilitation outcome measures: A guide to enhanced clinical decision making* (2nd ed.). Hamilton, ON: BC Decker. doi:10.2310/6640.2003.35271

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. (2002). *Structured clinical interview for DSM-IV-TR axis I disorders – Non-Patient edition* (SCID-I/NP). New York, NY: State Psychiatric Institute.

Fisher, R. H., & Zorzitto, M. L. (1983). Placement problem: Diagnosis, disease or term of denigration? *Canadian Medical Association Journal*, *129*(4), 331–334.

Fonda, S. J., Wallace, R. B., & Herzog, A. R. (2001). Changes in driving patterns and worsening depressive symptoms among older adults. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *56*(6), S343–S351. doi:10.1093/geronb/56.6.S343

Gainotti, G., Antonucci, G., Marra, C., & Paolucci, S. (2001). Relation between depression after stroke, antidepressant therapy, and functional recovery. *Journal of Neurology, Neurosurgery & Psychiatry*, *71*(2), 258–261. doi:10.1136/jnnp.71.2.258

Gibson, L., MacLennan, W. J., Gray, C., & Pentland, B. (1991). Evaluation of a comprehensive assessment battery for stroke patients. *International Journal of Rehabilitation Research*, *14*(2), 93–100. doi:10.1097/00004356-199106000-00001

Gillen, R., Tennen, H., McKee, T. E., Gernert-Dott, P., & Affleck, G. (2001). Depressive symptoms and history of depression predict rehabilitation efficiency in stroke patients. *Archives of Physical Medicine and Rehabilitation*, *82*(12), 1645–1649. doi:10.1053/apmr.2001.26249

Gillham, S., & Clark, L. (2011). *Psychological care after stroke: Improving stroke services for people with cognitive and mood disorders*. Retrieved from https://www.nice.org.uk/media/default/sharedlearning/531_strokepsychologicalsupportfinal.pdf

Gilson, R. (2012). *Validation of the distress thermometer among stroke survivors* (Doctoral dissertation, University of Southampton, England). Retrieved from https://eprints.soton.ac.uk/359648/

Glasziou, P., Irwig, L., Bain, C., & Colditz, G. (2001). *Systematic reviews in health care: A practical guide*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511543500

Goldberg, D. P., & Williams, P. (1988). *User's guide to the general health questionnaire*. Basingstoke, England: NFER-Nelson.

Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576. doi:10.1146/annurev.psych.58.110405.085530

Granger, C. V., & Hamilton, B. B. (1990). Measurement of stroke rehabilitation outcome in the 1980s. *Stroke*, *21*(9 Suppl), II46–II47.

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. doi:10.1111/ijs.12357

Hackett, M. L., & Anderson, C. S. (2005). Treatment options for post-stroke depression in the elderly. *Aging Health*, *1*(1), 95-105. doi:10.2217/1745509X.1.1.95

Hackett, M. L., Anderson, C. S., House, A., & Xia, J. (2008). Interventions for treating depression after stroke. *Cochrane Database of Systematic Reviews*, (4), CD003437. doi:10.1002/14651858.CD003437.pub3

Harel, O., & Zhou, X. H. (2007). Multiple imputation: Review of theory, implementation and software. *Statistics in Medicine*, *26*(16), 3057–3077. doi:10.1002/sim.2787

Hart, S., & Morris, R. (2008). Screening for depression after stroke: An exploration of professionals' compliance with guidelines. *Clinical Rehabilitation*, *22*(1), 60–70. doi:10.1177/0269215507079841

Haun, J., Rittman, M., & Sberna, M. (2008). The continuum of connectedness and social isolation during post stroke recovery. *Journal of Aging Studies*, *22*(1), 54–64. doi:10.1016/j.jaging.2007.03.001

Hays, R. D., Sherbourne, C. D., & Mazel, R. M. (1993). The rand 36-item health survey 1.0. *Health Economics*, *2*(3), 217–227. doi:10.1002/hec.4730020305

Healey, A. K., Kneebone, I. I., Carroll, M., & Anderson, S. J. (2008). A preliminary investigation of the reliability and validity of the brief assessment schedule depression cards and the beck depression inventory-fast screen to screen for depression in older stroke survivors. *International Journal of Geriatric Psychiatry*, *23*(5), 531–536. doi:10.1002/gps.1933

Henkel, V., Mergl, R., Kohnen, R., Allgaier, A. K., Möller, H. J., & Hegerl, U. (2004). Use of brief depression screening tools in primary care: Consideration of heterogeneity in performance in different patient groups. *General Hospital Psychiatry*, *26*(3), 190–198. doi:10.1016/j.genhosppsych.2004.02.003

Herrmann, N., Black, S. E., Lawrence, J., Szekely, C., & Szalai, J. P. (1998). The Sunnybrook stroke study: A prospective study of depressive symptoms and functional outcome. *Stroke*, *29*(3), 618–624. doi:10.1161/01.STR.29.3.618

Hinkle, J. L. (1998). Biological and behavioral correlates of stroke and depression. *Journal of Neuroscience Nursing*, *30*(1), 25–31. doi:10.1097/01376517-199802000-00004

Hochstenbach, J., Prigatano, G., & Mulder, T. (2005). Patients' and relatives' reports of disturbances 9 months after stroke: Subjective changes in physical functioning, cognition, emotion, and behavior. *Archives of Physical Medicine and Rehabilitation*, *86*(8), 1587–1593. doi:10.1016/j.apmr.2004.11.050

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York, NY: John Wiley & Sons. doi:10.1002/0471722146

House, A., Knapp, P., Bamford, J., & Vail, A. (2001). Mortality at 12 and 24 months after stroke may be associated with depressive symptoms at 1 month. *Stroke*, *32*(3), 696–701. doi:10.1161/01.STR.32.3.696

Huang, Y. (2010). Cost-benefit analysis of a community –based stroke prevention program in Bao Shan District, Shanghai, China. *International Journal of Collaborative Research on Internal Medicine and Public Health, 2*(9), 307-316. Retrieved from https://www.researchgate.net/publication/266456233_Cost-Benefit_Analysis_of_a_Community-Based_Stroke_Prevention_Program_in_Bao_Shan_District_Shanghai_China

Iacoboni, M., Padovani, A., Di Piero, V., & Lenzi, G. L. (1995). Post-stroke depression: Relationships with morphological damage and cognition over time. *The Italian Journal of Neurological Sciences*, *16*(3), 209–216. doi:10.1007/BF02282991

Intercollegiate Stroke Working Party. (2002). *Concise report on the National sentinel stroke audit 2001/2*. London, England: Royal College of Physicians.

Intercollegiate Stroke Working Party. (2004). *Concise report on the National sentinel stroke audit 2004*. London, England: Royal College of Physicians.

Intercollegiate Stroke Working Party. (2008). *National clinical guidelines for stroke* (2nd ed.). London, England: Royal College of Physicians.

Intercollegiate Stroke Working Party. (2016) *National clinical guidelines for stroke, 5th edition*. London: Royal College of Physicians.

Jackson, N., & Alshekhlee, A. (2015). Abstract T P310: PHQ-2 Screens for early depression in acute stroke. *Stroke*, *46*(Suppl_1), ATP310.

Johnson, J. L., Minarik, P. A., Nyström, K. V., Bautista, C., & Gorman, M. J. (2006). Poststroke depression incidence and risk factors: An integrative literature review. *Journal of Neuroscience Nursing*, *38*(4 Suppl), S316-S327. doi:10.1097/01376517-200609000-00008

Jorge, R. E., Robinson, R. G., Arndt, S., & Starkstein, S. (2003). Mortality and poststroke depression: A placebo-controlled trial of antidepressants. *American Journal of Psychiatry*, *160*(10), 1823–1829. doi:10.1176/appi.ajp.160.10.1823

Jørgensen, L., Engstad, T., & Jacobsen, B. K. (2002). Higher incidence of falls in long-term stroke survivors than in population controls: Depressive symptoms predict falls after stroke. *Stroke*, *33*(2), 542–547. doi:10.1161/hs0202.102375

Jun, E.M., Roh, Y.H., & Kim, M.J. (2013). The effect of music movement therapy on physical and psychological states of stroke patients. *Journal of Clinical Nursing, 22*(1-2) 22-31. doi:10.1111/j.1365-2702.2012.04243.x. Epub 2012 Sep 17

Karamchandani, R. R., Vahidy, F., Bajgur, S., Vu, K. Y. T., Choi, H. A., Hamilton, R. K., . . . & Savitz, S. I. (2015). Early depression screening is feasible in hospitalized stroke patients. *PloS One*, *10*(6), e0128246. doi:10.1371/journal.pone.0128246

Kauhanen, M. L., Korpelainen, J. T., Hiltunen, P., Brusin, E., Mononen, H., Määttä, R., . . . & Myllylä, V. V. (1999). Poststroke depression correlates with cognitive impairment and neurological deficits. *Stroke*, *30*(9), 1875–1880. doi:10.1161/01.STR.30.9.1875

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., . . . & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry*, *60*(2), 184–189. doi:10.1001/archpsyc.60.2.184

Kim, P., Warren, S., Madill, H., & Hadley, M. (1999). Quality of life of stroke survivors. *Quality of Life Research*, *8*(4), 293–301. doi:10.1023/A:100892743

King, R. B. (1996). Quality of life after stroke. *Stroke*, *27*(9), 1467–1472. doi:10.1161/01.STR.27.9.1467

Kitsos, G., Harris, D., Pollack, M., & Hubbard, I. J. (2011). Assessments in Australian stroke rehabilitation units: A systematic review of the post-stroke validity of the most frequently used. *Disability and Rehabilitation*, *33*(25-26), 2620–2632. doi:10.3109/09638288.2011.575526

Koenig, H. G. (1996). An abbreviated mini-mental state exam for medically ill older adults. *Journal of the American Geriatrics Society*, *44*(2), 215–216. doi:10.1111/j.1532-5415.1996.tb02449.x

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. doi:10.1046/j.1468-1331.1999.630309.x

Kotila, M., Waltimo, O., Niemi, M. L., Laaksonen, R., & Lempinen, M. (1984). The profile of recovery from stroke and factors influencing outcome. *Stroke*, *15*(6), 1039–1044. doi:10.1161/01.STR.15.6.1039

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2003). The patient health questionnaire-2: Validity of a two-item depression screener. *Medical care*, *41*(11), 1284–1292. doi:10.1097/01.MLR.0000093487.78664.3C

Krueger, H., Koot, J., Hall, R. E., O'Callaghan, C., Bayley, M., & Corbett, D. (2015). Prevalence of individuals experiencing the effects of stroke in Canada: Trends and projections. *Stroke*, *46*(8), 2226–2231. doi:10.1161/STROKEAHA.115.009616

Krueger, H., Lindsay, P., Cote, R., Kapral, M. K., Kaczorowski, J., & Hill, M. D. (2012). Cost avoidance associated with optimal stroke care in Canada. *Stroke*, *43*(8), 2198–2206. doi:10.1161/STROKEAHA.111.646091

Kutlubaev, M. A., & Hackett, M. L. (2014). Part II: Predictors of depression after stroke and impact of depression on stroke outcome: An updated systematic review of observational studies. *International Journal of Stroke*, *9*(8), 1026-1036. doi:10.1111/ijs.12356

Lachs, M. S., Feinstein, A. R., Cooney, L. M., Drickamer, M. A., Marottoli, R. A., Pannill, F. C., & Tinetti, M. E. (1990). A simple procedure for general screening for functional disability in elderly patients. *Annals of Internal Medicine*, *112*(9), 699–706. doi:10.7326/0003-4819-112-9-699

Lanctôt, K. L., Lindsay, M. P., Smith, E. E., Sahlas, D. J., Foley, N., Gubitz, G., . . . & Herrmann, N. (2019). Canadian stroke best practice recommendations: Mood, cognition and fatigue following stroke, update 2019. *International Journal of Stroke*, *10*(7), 1130-1140. doi:10.1177/1747493019847334

Lightbody, C. E., Auton, M., Baldwin, R., Gibbon, B., Hamer, S., Leathley, M. J., . . . & Watkins, C. L. (2007). The use of nurses' and carers' observations in the identification of

poststroke depression. *Journal of Advanced Nursing*, *60*(6), 595–604. doi:10.1111/j.1365-2648.2007.04508.x

Lincoln, N. B., & Flannaghan, T. (2003). Cognitive behavioral psychotherapy for depression following stroke: A randomized controlled trial. *Stroke*, *34*(1), 111–115. doi:10.1161/01.STR.0000044167.44670.55

Lincoln, N. B., Nicholl, C. R., Flannaghan, T., Leonard, M., & Van der Gucht, E. (2003). The validity of questionnaire measures for assessing depression after stroke. *Clinical Rehabilitation*, *17*(8), 840–846. doi:10.1191/0269215503cr687oa

Lindsay, M. P, Gubitz, G., Bayley, M., & Phillips, S. (Editors). (2013). *Chapter 7: Mood and Cognition in Patients Following Stroke*. In Lindsay, M. P., Gubitz, G., Bayley, M., Hill, M. D., Davies-Schinkel, C., Singh, S., and Phillips, S. (Editors) On behalf of the Canadian Stroke Strategy Best Practices and Standards Writing Group. *Canadian Best Practice Recommendations for Stroke Care (4th ed.) Ottawa*, Ontario, Canada: Canadian Stroke Network, 2013. Retrieved from https://pdfs.semanticscholar.org/ba28/b9ed17db2e0c2c9c2a779e5517b249204f79.pdf

Liu, B., Anderson, G., Mittmann, N., To, T., Axcell, T., & Shear, N. (1998). Use of selective serotonin-reuptake inhibitors or tricyclic antidepressants and risk of hip fractures in elderly people. *Lancet*, *351*(9112), 1303–1307. doi:10.1016/S0140-6736(97)09528-7

Mahoney, F. I., & Barthel, D. W. (1965). Functional evaluation: The Barthel Index: A simple index of independence useful in scoring improvement in the rehabilitation of the chronically ill. *Maryland State Medical Journal, 14,* 61–65.

Mahoney, J., Drinka, T. J., Abler, R., Gunter-Hunt, G., Matthews, C., Gravenstein, S., & Carnes, M. (1994). Screening for depression: Single question versus GDS. *Journal of the American Geriatrics Society*, *42*(9), 1006–1008. doi:10.1111/j.1532-5415.1994.tb06597.x

Management of Stroke Rehabilitation Working Group (2010). VA/DOD clinical practice guideline for the management of stroke rehabilitation. *Journal of Rehabilitation Research and Development*, *47*(9), 1.

Meader, N., Moe-Byrne, T., Llewellyn, A., & Mitchell, A. J. (2014). Screening for poststroke major depression: A meta-analysis of diagnostic validity studies. *Journal of Neurology, Neurosurgery & Psychiatry, 85*(2), 198–206. doi:10.1136/jnnp-2012-304194

Mitchell, A. J., & Coyne, J. C. (2007). Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *British Journal of General Practice*, *57*(535), 144–151.

Mitchell, P. H., Veith, R. C., Becker, K. J., Buzaitis, A., Cain, K. C., Fruin, M., . . . & Teri, L. (2009). Brief psychosocial–behavioral intervention with antidepressant reduces poststroke depression significantly more than usual care with antidepressant: Living well with stroke: Randomized, controlled trial. *Stroke*, *40*(9), 3073–3078. doi:10.1161/STROKEAHA.109.549808

Morris, P. L., Robinson, R. G., Andrzejewski, P., Samuels, J., & Price, T. R. (1993). Association of depression with 10-year poststroke mortality. *The American Journal of Psychiatry*, *150*(1), 124–129. doi:10.1176/ajp.150.1.124

Morris, R., Jones, J., Wilcox, J., & Cole, S. (2012). Depression and anxiety screening after stroke: Adherence to guidelines and future directions. *Disability and Rehabilitation*, *34*(9), 733–739. doi:10.3109/09638288.2011.619623

Morrison, V., Pollard, B., Johnston, M., & MacWalter, R. (2005). Anxiety and depression 3 years following stroke: demographic, clinical, and psychological predictors. *Journal of Psychosomatic Research*, *59*(4), 209-213. doi:10.1016/j.jpsychores.2005.02.019

Mulherin, S. A., & Miller, W. C. (2002). Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Annals of Internal Medicine, 137*(7), 598–602. doi:10.7326/0003-4819-137-7-200210010-00011

Narushima, K., Chan, K. L., Kosier, J. T., & Robinson, R. G. (2003). Does cognitive recovery after treatment of poststroke depression last? A 2-year follow-up of cognitive function associated with poststroke depression. *American Journal of Psychiatry*, *160*(6), 1157–1162. doi:10.1176/appi.ajp.160.6.1157

National Audit Office (2010). *Tackling inequalities in life expectancy in areas with the worst health deprivation.* London, England: National Audit Office.

National Institute for Clinical Excellence *(*2004*). Depression: Management of depression in primary and secondary care*. London, England: Author. Retrieved from http://www.scamfyc.org/documentos/depresion%20NICE.pdf

National Stroke Foundation. (2010). *Clinical guidelines for stroke management 2010*.

Melbourne, Australia: Author. Retrieved from https://www.pedro.org.au/wp-content/uploads/CPG_stroke.pdf

National Stroke Foundation. (2017). *Clinical guidelines for stroke management 2017*. Melbourne, Australia: Author. Retrieved from https://informme.org.au/en/Guidelines/Clinical-Guidelines-for-Stroke-Management

Ostir, G. V., Berges, I. M., Ottenbacher, A., & Ottenbacher, K. J. (2011). Patterns of change in depression after stroke. *Journal of the American Geriatrics Society*, *59*(2), 314–320. doi:10.1111/j.1532-5415.2010.03266.x

Ouimet, M. A., Primeau, F., & Cole, M. G. (2001). Psychosocial risk factors in poststroke depression: A systematic review. *The Canadian Journal of Psychiatry*, *46*(9), 819–828. doi:10.1177/070674370104600905

Paolucci, S., Antonucci, G., Grasso, M. G., Morelli, D., Troisi, E., Coiro, P., . . . & Bragoni, M. (2001). Post-stroke depression, antidepressant treatment and rehabilitation results. *Cerebrovascular Diseases*, *12*(3), 264–271. doi:10.1159/000047714

Paolucci, S., Antonucci, G., Pratesi, L., Traballesi, M., Grasso, M. G., & Lubich, S. (1999). Poststroke depression and its role in rehabilitation of inpatients. *Archives of Physical Medicine and Rehabilitation*, *80*(9), 985–990. doi:10.1016/S0003-9993(99)90048-5

Paolucci, S., Gandolfo, C., Provinciali, L., Torta, R., Sommacal, S., DESTRO Study Group, & Toso, V. (2005). Quantification of the risk of poststroke depression: The Italian multicenter observational study DESTRO. *Acta Psychiatrica Scandinavica*, *112*(4), 272–278. doi:10.1111/j.1600-0447.2005.00590.x

Paradiso, S., Ohkubo, T., & Robinson, R. G. (1997). Vegetative and psychological symptoms associated with depressed mood over the first two years after stroke. *The International Journal of Psychiatry in Medicine*, *27*(2), 137–157. doi:10.2190/BWJA-KQP3-7VUY-D06T

Parikh, R. M., Eden, D. T., Price, T. R., & Robinson, R. G. (1989). The sensitivity and specificity of the Center for epidemiologic studies depression scale in screening for post-stroke depression. *The International Journal of Psychiatry in Medicine*, *18*(2), 169–181. doi:10.2190/BH75-EUYA-4FM1-J7QA

Parikh, R. M., Robinson, R. G., Lipsey, J. R., Starkstein, S. E., Fedoroff, J. P., & Price, T. R. (1990). The impact of poststroke depression on recovery in activities of daily living over

a 2-year follow-up. *Archives of Neurology*, *47*(7), 785–789. doi:10.1001/archneur.1990.00530070083014

Paveza, G. J., Cohen, D., Blaser, C. J., & Hagopian, M. (1990). A brief form of the mini-mental state examination for use in community care settings. *Behavior, Health, & Aging*, *1*(2), 133–139.

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. doi:10.1046/j.1468-1331.2001.00182.x

Powers, W. J., Rabinstein, A. A., Ackerson, T., Adeoye, O. M., Bambakidis, N. C., Becker, K., ... & Jauch, E. C. (2018). 2018 guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, *49*(3). doi:10.1161/STR.0000000000000163

Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401. doi:10.1177/014662167700100306

Ramasubbu, R., Robinson, R. G., Flint, A. J., Kosier, T., & Price, T. R. (1998). Functional impairment associated with acute poststroke depression: The Stroke data bank study. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *10*(1), 26–33. doi:10.1176/jnp.10.1.26

Reynolds, C. F. (1992). Treatment of depression in special populations. *The Journal of Clinical Psychiatry*, *53*(9, Suppl), 45–53.

Robinson, R. G. (2003). Poststroke depression: Prevalence, diagnosis, treatment, and disease progression. *Biological Psychiatry*, *54*(3), 376–387. doi:10.1016/S0006-3223(03)00423-2

Robinson, R. G., & Jorge, R. E. (2015). Post-stroke depression: a review. *American Journal of Psychiatry*, *173*(3), 221-231. doi:10.1176/appi.ajp.2015.15030363

Robinson, R. G., Schultz, S. K., Castillo, C., Kopel, T., Kosier, J. T., Newman, R. M., . . . & Starkstein, S. E. (2000). Nortriptyline versus fluoxetine in the treatment of depression and in short-term recovery after stroke: A placebo-controlled, double-blind study. *American Journal of Psychiatry*, *157*(3), 351–359. doi:10.1176/appi.ajp.157.3.351

Robinson-Smith, G., Johnston, M. V., & Allen, J. (2000). Self-care self-efficacy, quality of life, and depression after stroke. *Archives of Physical Medicine and Rehabilitation*, *81*(4), 460–464. doi:10.1053/mr.2000.3863

Roger, P. R., & Johnson-Greene, D. (2009). Comparison of assessment measures for post-stroke depression. *The Clinical Neuropsychologist*, *23*(5), 780–793. doi:10.1080/13854040802691135

Ruchinskas, R. (2002). Rehabilitation therapists' recognition of cognitive and mood disorders in geriatric patients. *Archives of Physical Medicine and Rehabilitation*, *83*(5), 609–612. doi:10.1053/apmr.2002.32496

Rybarczyk, B., Winemiller, D. R., Lazarus, L. W., Haut, A., & Hartman, C. (1996). Validation of a depression screening measure for stroke inpatients. *The American Journal of Geriatric Psychiatry*, *4*(2), 131–139. doi:10.1097/00019442-199621420-00005

Salter, K., Bhogal, S. K., Foley, N., Jutai, J., & Teasell, R. (2007). The assessment of poststroke depression. *Topics in Stroke Rehabilitation*, *14*(3), 1–24. doi:10.1310/tsr1403-1

Salter, K., McClure, J., Mahon, H., Foley, N., & Teasell, R. (2012). Adherence to Canadian best practice recommendations for stroke care: Assessment and management of poststroke depression in an Ontario rehabilitation facility. *Topics in Stroke Rehabilitation*, *19*(2), 132–140. doi:10.1310/tsr1902-132

Salter, K. L., Foley, N. C., Zhu, L., Jutai, J. W., & Teasell, R. W. (2013). Prevention of poststroke depression: Does prophylactic pharmacotherapy work? *Journal of Stroke and Cerebrovascular Diseases*, *22*(8), 1243-1251. doi:10.1016/j.jstrokecerebrovasdis.2012.03.013

Särkämö, T., Tervaniemi, M., Laitinen, S., Forsblom, A., Soinila, S., Mikkonen, M., ... & Peretz, I. (2008). Music listening enhances cognitive recovery and mood after middle cerebral artery stroke. *Brain*, *131*(3), 866-876. doi: 10.1093/brain/awn013

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology & Laboratory Medicine*, *137*(4), 558–565. doi:10.5858/arpa.2012-0198-RA

Schubert, D. S., Taylor, C., Lee, S., Mentari, A., & Tamaklo, W. (1992). Detection of depression in the stroke patient. *Psychosomatics*, *33*(3), 290–294. doi:10.1016/S0033-3182(92)71967-7

Schulz, R., Beach, S. R., Ives, D. G., Martire, L. M., Ariyo, A. A., & Kop, W. J. (2000). Association between depression and mortality in older adults: The Cardiovascular health study. *Archives of Internal Medicine*, *160*(12), 1761–1768. doi:10.1001/archinte.160.12.1761

Schwartz, J. A., Speed, N. M., Brunberg, J. A., Brewer, T. L., Brown, M., & Greden, J. F. (1993). Depression in stroke rehabilitation. *Biological Psychiatry*, *33*(10), 694–699. doi:10.1016/0006-3223(93)90118-W

Scottish Intercollegiate Guidelines Network (SIGN) (2010). *Management of patients with stroke: Rehabilitation, prevention and management of complications, and discharge planning: A national clinical guideline.* Edinburgh, Scotland.

Seymour, A. M. (2015). *Depression screening for patients with stroke in an acute care setting* (Master thesis, Simon Fraser University, Burnaby, Canada). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.877.4938&rep=rep1&type=pdf

Shah, S., Vanclay, F., & Cooper, B. (1989). Improving the sensitivity of the Barthel Index for stroke rehabilitation. *Journal of Clinical Epidemiology*, *42*(8), 703–709. doi:10.1016/0895-4356(89)90065-6

Shear, M. K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., & Hanekamp, M. (2000). Diagnosis of nonpsychotic patients in community clinics. *The American Journal of Psychiatry*, *157*(4), 581–587. doi:10.1176/appi.ajp.157.4.581

Sheikh, J. I., & Yesavage, J. A. (1986). Geriatric depression scale (GDS): Recent evidence and development of a shorter version. *Clinical Gerontologist: The Journal of Aging and Mental Health,* 5(1-2), 165–173. doi:10.1300/J018v05n01_09

Shen, X., Liu, M., Cheng, Y., Jia, C., Pan, X., Gou, Q., . . . & Zhang, L. (2017). Repetitive transcranial magnetic stimulation for the treatment of post-stroke depression: A systematic review and meta-analysis of randomized controlled clinical trials. *Journal of Affective Disorders*, *211*, 65-74. doi:10.1016/j.jad.2016.12.058

Sinyor, D., Amato, P., Kaloupek, D. G., Becker, R., Goldenberg, M., & Coopersmith, H. (1986). Post-stroke depression: Relationships to functional impairment, coping strategies, and rehabilitation outcome. *Stroke*, *17*(6), 1102–1107. doi:10.1161/01.STR.17.6.1102

Sivrioglu, E. Y., Sivrioglu, K., Ertan, T., Ertan, F. S., Cankurtaran, E., Aki, O., . . . & Kirli, S. (2009). Reliability and validity of the geriatric depression scale in detection of poststroke

minor depression. *Journal of Clinical and Experimental Neuropsychology*, *31*(8), 999-1006. doi:10.1080/13803390902776878

Snaith, R. P., Ahmed, S. N., Mehta, S., & Hamilton, M. (1971). Assessment of the severity of primary depressive illness: Wakefield self-assessment depression inventory. *Psychological Medicine*, *1*(2), 143–149. doi:10.1017/S0033291700000064

Søndergaard, M. P., Jarden, J. O., Martiny, K., Andersen, G., & Bech, P. (2006). Dose response to adjunctive light therapy in citalopram-treated patients with post-stroke depression. *Psychotherapy and Psychosomatics*, *75*(4), 244–248. doi:10.1159/000092895

Spalletta, G., & Robinson, R. G. (2010). How should depression be diagnosed in patients with stroke? *Acta Psychiatrica Scandinavica*, *121*(6), 401–403. doi:10.1111/j.1600-0447.2010.01569.x

Spalletta, G., Bossu, P., Ciaramella, A., Bria, P., Caltagirone, C., & Robinson, R. G. (2006). The etiology of poststroke depression: A review of the literature and a new hypothesis involving inflammatory cytokines. *Molecular Psychiatry*, *11*(11), 984–991. doi:10.1038/sj.mp.4001879

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Patient Health Questionnaire Primary Care Study Group. (1999). Validation and utility of a self-report version of PRIME-MD. *Journal of the American Medical Association, 282*(18), 1737–1744. doi:10.1001/jama.282.18.1737

Spratt, M., Carpenter, J., Sterne, J. A., Carlin, J. B., Heron, J., Henderson, J., & Tilling, K. (2010). Strategies for multiple imputation in longitudinal studies. *American Journal of Epidemiology*, *172*(4), 478–487. doi:10.1093/aje/kwq137

Stein, P. N., Sliwinski, M. J., Gordon, W. A., & Hibbard, M. R. (1996). Discriminative properties of somatic and nonsomatic symptoms for post stroke depression. *The Clinical Neuropsychologist*, *10*(2), 141–148. doi:10.1080/13854049608406674

Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the Structured Clinical Interview for DSM-III—R and clinical diagnoses. *Journal of Nervous and Mental Disease*, *183*(6), 365–369. doi:10.1097/00005053-199506000-00003

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for

some traditional and novel measures. *Epidemiology*, *21*(1), 128–138. doi:10.1097/EDE.0b013e3181c30fb2

Stroke Mood and Cognition Best Practices Writing Group. (2013). Mood and cognition in patients following stroke. In M. P. Lindsay, G. Gubitz, M. Bayley, & S. Phillips (Eds.), Canadian best practice recommendations for stroke care (4th ed.). Ottawa, ON: Canadian Stroke Network. Retrieved from https://pdfs.semanticscholar.org/ba28/b9ed17db2e0c2c9c2a779e5517b249204f79.pdf

Tateno, A., Kimura, M., & Robinson, R. G. (2002). Phenomenological characteristics of poststroke depression: Early-versus late-onset. *The American Journal of Geriatric Psychiatry*, *10*(5), 575-582. doi:10.1097/00019442-200209000-00011

Tham, W. Y., Kwan, Y., & Ang, A. (2012). Performance of the patient health questionnaire-2 (PHQ-2) as a screening tool for depression in stroke patients. *Biological Psychiatry*, *71*(8), 141S. doi:10.1016/j.biopsych.2012.02.013

The EuroQol Group (1990). EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy*, *16*(3), 199–208. doi:10.1016/0168-8510(90)90421-9

Tsouna-Hadjis, E., Vemmos, K. N., Zakopoulos, N., & Stamatelopoulos, S. (2000). First-stroke recovery process: The role of family social support. *Archives of Physical Medicine and Rehabilitation*, *81*(7), 881–887. doi:10.1053/apmr.2000.4435

Turnbull, O. H., Evans, C. E., & Owen, V. (2005). Negative emotions and anosognosia. *Cortex*, *41*(1), 67–75. doi:10.1016/S0010-9452(08)70179-5

Turner, A., Hambridge, J., White, J., Carter, G., Clover, K., Nelson, L., & Hackett, M. (2012). Depression screening in stroke: A comparison of alternative measures with the structured diagnostic interview for the Diagnostic and statistical manual of mental disorders, 4th ed. (major depressive episode) as criterion standard. *Stroke*, *43*(4), 1000–1005. doi:10.1161/STROKEAHA.111.643296

Turner-Stokes, L., Kalmus, M., Hirani, D., & Clegg, F. (2005). The depression intensity scale circles (DISCs): A first evaluation of a simple assessment tool for depression in the context of brain injury. *Journal of Neurology, Neurosurgery & Psychiatry*, *76*(9), 1273–1278. doi:10.1136/jnnp.2004.050096

Tyson, S. F., Burton, L., & McGovern, A. (2015). The impact of a toolkit on use of standardised measurement tools in stroke rehabilitation. *Clinical Rehabilitation*, *29*(9), 926–934. doi:10.1177/0269215514562590

van de Weg, F. B., Kuik, D. J., & Lankhorst, G. J. (1999). Post-stroke depression and functional outcome: A cohort study investigating the influence of depression on functional recovery from stroke. *Clinical Rehabilitation*, *13*(3), 268–272. doi:10.1191/026921599672495022

van Dijk, M. J., de Man-van Ginkel, J. M., Hafsteinsdóttir, T. B., & Schuurmans, M. J. (2016). Identifying depression post-stroke in patients with aphasia: A systematic review of the reliability, validity and feasibility of available instruments. *Clinical Rehabilitation*, *30*(8), 795–810. doi:10.1177/0269215515599665

Wade, D. T., Skilbeck, C. E., & Hewer, R. L. (1983). Predicting Barthel ADL score at 6 months after an acute stroke. *Archives of Physical Medicine and Rehabilitation*, *64*(1), 24–28.

Walker, S., Walker, J., Richardson, G., Palmer, S., Wu, Q., Gilbody, S., . . . & Sculpher, M. (2014). Cost-effectiveness of combining systematic identification and treatment of co-morbid major depression for people with chronic diseases: The example of cancer. *Psychological Medicine*, *44*(7), 1451–1460. doi:10.1017/S0033291713002079

Ware, J. E., & Gandek, B. (1998). Overview of the SF-36 health survey and the international quality of life assessment (IQOLA) project. *Journal of Clinical Epidemiology*, *51*(11), 903–912. doi:10.1016/S0895-4356(98)00081-X

Ware, J. E., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection. *Medical Care*, *30*(6), 473–483. doi:10.1097/00005650-199206000-00002

Watkins, C., Daniels, L., Jack, C., Dickinson, H., & van den Broek, M. (2001). Accuracy of a single question in screening for depression in a cohort of patients after stroke: Comparative study. *British Medical Journal, 323*(7322), 1159. doi:10.1136/bmj.323.7322.1159

Watkins, C. L., Lightbody, C. E., Sutton, C. J., Holcroft, L., Jack, C. I., Dickinson, H. A., . . . & Leathley, M. J. (2007). Evaluation of a single-item screening tool for depression after stroke: A cohort study. *Clinical Rehabilitation*, *21*(9), 846–852. doi:10.1177/0269215507079846

Wei, N., Yong, W., Li, X., Zhou, Y., Deng, M., Zhu, H., & Jin, H. (2015). Post-stroke depression and lesion location: A systematic review. *Journal of Neurology*, *262*(1), 81–90. doi:10.1007/s00415-014-7534-1

White, J. H., Towers, S. E., Turner, A., & Hambridge, J. (2013). Electronic screening and decision support for poststroke depression: An exploration of doctors' and patients' perceptions of acceptability. *Archives of Physical Medicine and Rehabilitation*, *94*(4), 788–790. doi:10.1016/j.apmr.2012.09.009

Whiting, P., Rutjes, A. W. S., Dinnes, J., Reitsma, J. B., Bossuyt, P. M. M., & Kleijnen, J. (2004). Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technology Assessment*, *8*(25), 1–234. doi:10.3310/hta8250

Whyte, E. M., & Mulsant, B. H. (2002). Post stroke depression: Epidemiology, pathophysiology, and biological treatment. *Biological Psychiatry*, *52*(3), 253–264. doi:10.1016/S0006-3223(02)01424-5

Williams, L. S., Brizendine, E. J., Plue, L., Bakas, T., Tu, W., Hendrie, H., & Kroenke, K. (2005). Performance of the PHQ-9 as a screening tool for depression after stroke. *Stroke*, *36*(3), 635–638. doi:10.1161/01.STR.0000155688.18207.33

Williams, L. S., Ghose, S. S., & Swindle, R. W. (2004). Depression and other mental health diagnoses increase mortality risk after ischemic stroke. *American Journal of Psychiatry*, *161*(6), 1090–1095. doi:10.1176/appi.ajp.161.6.1090

Winstein, C. J., Stein, J., Arena, R., Bates, B., Cherney, L. R., Cramer, S. C., ... & Lang, C. E. (2016). Guidelines for adult stroke rehabilitation and recovery: a guideline for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, *47*(6).

Word Health Organization. (2002). *The world health report 2002: reducing risks, promoting healthy life*. World Health Organization.

Wu, C. S., Wang, S. C., Cheng, Y. C., Gau, S. S. (2011). Association of cerebrovascular events with antidepressant use: A case-crossover study. *American Journal of Psychiatry, 168*(5), 511-521. doi:10.1176/appi.ajp.2010.10071064

Xu, X. M., Zou, D. Z., Shen, L. Y., Liu, Y., Zhou, X. Y., Pu, J. C., . . . & Wei, Y. D. (2016). Efficacy and feasibility of antidepressant treatment in patients with post-stroke depression. *Medicine*, *95*(45). doi: 10.1097/MD.0000000000005349

Yesavage, J. A., Brink, T. L., Rose, T. L., Lum, O., Huang, V., Adey, M., & Leirer, V. O. (1982). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research*, *17*(1), 37–49. doi:10.1016/0022-3956(82)90033-4

Yi, Z. M., Liu, F., & Zhai, S. D. (2010). Fluoxetine for the prophylaxis of poststroke depression in patients with stroke: A meta-analysis. *International Journal of Clinical Practice*, *64*(9), 1310-1317. doi: 10.1111/j.1742-1241.2010.02437.x

Zawadzka, E., & Domańska, Ł. (2014). Assessment of select dimensions of patients' emotional functioning at different time periods after stroke. *Applied Neuropsychology: Adult*, *21*(2), 87–93. doi:10.1080/09084282.2012.747959

Zhang, Z. J., Chen, H. Y., Yip, K. C., Ng, R., & Wong, V. T. (2010). The effectiveness and safety of acupuncture therapy in depressive disorders: systematic review and meta-analysis. *Journal of Affective Disorders*, *124*(1-2), 9-21. doi:10.1016/j.jad.2009.07.005

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, *67*(6), 361–370. doi:10.1111/j.1600-0447.1983.tb09716.x

Zung, W. W. K. (1965). A self-rating depression scale. *Archives of General Psychiatry*, *12*(1), 63–70. doi:10.1001/archpsyc.1965.01720310065008

Appendix A: Mood and stroke best practice recommendations

Mood and stroke best practice recommendations 2015 (relevant sections)

- **1.0 All patients with stroke should be considered to be at high risk for PSD, which can occur at any stage of recovery (Evidence Level A).**

- **1.1 Screening for PSD** Side note: Common risk factors associated with PSD include increasing stroke severity, functional dependence, presence of cognitive impairment, and history of previous depression. Increased functional dependence (e.g. requiring help with activities of daily living) and having a history of prestroke depression may be the two most salient risk factors for the development of PSD. Communication deficits and social isolation may also be considered as possible risk factors for depression. Refer to Transitions of Care Module (available at www.strokebestpractices.ca) for information on depression in family and informal caregivers of people with stroke.

    1. All patients with stroke should be screened for depressive symptoms, given the high prevalence of depression poststroke, the need for screening to detect depression, and the strong evidence for treating symptomatic depression poststroke (Evidence Level B).

    2. Screening should be undertaken using a validated tool to maximize detection of depression (Evidence Level B); *table 1A – a summary of suggested validated tools – is available at* www.strokebestpractices.ca.

    3. Stroke patient assessments should include evaluation of risk factors for depression, particularly a history of depression (Evidence Level C).

    4. For patients who experience some degree of communication challenge or deficits following stroke, appropriate strategies for screening of possible PSD should be implemented to ensure adequate assessment and access to appropriate treatment (Evidence Level C).

- **1.2 Timing of screening for PSD**

    1. Screening for PSD may take place at various stages throughout the continuum of stroke care, particularly at transition points (Evidence Level C). Repeated screening may be required since the ideal timing for screening for PSD is unclear.

    2. Screening for depressive symptoms could be considered during acute care stay in patients at high risk for depression, particularly if evidence of depression or mood changes is noted. Stroke patients who are identified as at risk could be screened before discharge from acute care (Evidence Level C).

    3. Screening for depressive symptoms should be considered during transition points in care, such as from an inpatient acute setting to an inpatient rehabilitation setting, and or before return to the community (Evidence Level C).

4. Screening for depressive symptoms should be considered following discharge to the community, at stroke prevention clinic assessments, during follow-up appointments, and during periodic health assessments with primary care practitioners and consulting specialists (Evidence Level C).

- **1.3 Assessment for PSD**

    1. Patients identified with a high probability of clinically significant PSD during screening should be assessed in a timely manner by a healthcare professional with expertise in diagnosis, management, and follow-up of depression in patients following stroke (Evidence Level C).

- **1.5 Pharmacotherapy for PSD***Note: Watchful waiting is defined as a period of time when the patient who displays mild depressive symptoms is monitored closely without additional therapeutic interventions to determine whether the mild depressive symptoms will improve. The timeframe for watchful waiting varies in the literature somewhere between two and four-weeks. It is often described as including suggestions to the patient for self-help strategies and participation in physical exercise and other strategies noted in Section 1.4 above.

    1. Patients with *mild* depressive symptoms or those diagnosed with minor depression may initially be managed by 'watchful waiting'* (Evidence Level B).

    4. Following initial treatment for PSD, patients should continue to be monitored for recurrence of depressive symptoms, as part of ongoing comprehensive stroke management (Evidence Level C). The involvement and feedback of patients, family, and caregivers can be an important component of ongoing monitoring.

- **1.8 Ongoing monitoring, support, and education** Note: Additional materials available on the Stroke Best Practices website (www.strokebestpractices.ca/depression) include a table summarizing the psychometric properties of a selected set of screening and assessment tools that have been validated for use with stroke patients, or frequently reported in the stroke literature, and a table summarizing the pharmacotherapeutic properties, side effects, drug interactions, and other important information on selected classes of medications available for use in Canada and more commonly recommended for PSD.

    a. Patients and families should be given information and education about the potential impact of stroke on their mood and that of family and caregivers; patients and families should be provided with the opportunity to talk about the impact of stroke on their lives at all stages of care (Evidence level C).

Note. Source: Eskes et al., 2015

Mood and stroke best practice recommendations 2019 update (relevant sections)

Section 1: Post-stroke depression update 2019

Definitions and descriptions

Depression following stroke: *Within this module, we consider depression following stroke*. The DSM5 category that applies is *mood disorders due to another medical condition such as stroke with depressive features, major depressive-like episode, or mixed-mood features*. It is often associated with large vessel infarction.

- A patient who is a candidate for this diagnosis would present with depressed mood or loss of interest or pleasure along with four other symptoms of depression (e.g. weight loss, insomnia, psychomotor agitation, fatigue, feelings of worthlessness, diminished concentration, suicidal ideation) lasting two or more weeks.

- Several mechanisms, including biological, behavioral, and social factors, are involved in its pathogenesis.

- Symptoms usually occur within the first three months after stroke (early onset depression following stroke); however, may occur at any time (late onset depression following stroke). Symptoms resemble those of depression triggered by other causes, although there are some differences; people who have experienced a stroke with depression following stroke experience more sleep disturbances, vegetative symptoms, and social withdrawal.

Recommendations

1.0 All people who have experienced a stroke should be considered at risk for post-stroke depression, which can occur at any stage of recovery [Evidence Level A].

i. People who have experienced a stroke and families should be given information and education about the potential impact of stroke on their mood [Evidence level C].

ii. People who have experienced a stroke and families should be provided with the opportunity to talk about the impact of stroke on their lives at all stages of care [Evidence level C]. *Refer to the CSBPR Transitions of Care Module for further information on Patient and Family Education, and Community Follow-up.*

1.1. Screening for post-stroke depression

i. All people who have experienced a stroke should be screened for post-stroke depression if deemed medically appropriate, given the high prevalence of post-stroke depression and the evidence for treating symptomatic depression post stroke [Evidence Level B]. *Note: "Medically appropriate" excludes people who have experienced a stroke who are unresponsive or who have deficits that interfere with screening for*

mood disorders. Any pre-stroke mental health or cognitive diagnoses should be taken into consideration during the screening process.

ii. Screening should be undertaken by trained professionals using a validated screening tool to maximize detection of depression [Evidence Level B]. *Summary of suggested validated screening tools at www.strokebestpractices.ca.*

iii. Stroke assessments should include evaluation of risk factors for depression, particularly a history of depression [Evidence Level C]. *Refer to note below for list of risk factors.*

iv. For people who experience some degree of communication challenge or deficits following stroke, appropriate strategies that do not rely on verbal communication should be implemented for screening of possible post-stroke depression to ensure adequate detection and assessment, and access to appropriate treatment [Evidence Level C]. *Refer to the CSBPR Stroke Rehabilitation Module for further information on communication deficits.*

Note: Common risk factors associated with post-stroke depression include increasing stroke severity, functional dependence, presence of cognitive impairment, and history of previous depression. Increased functional dependence (e.g. requiring help with activities of daily living) and having a history of pre-stroke depression may be the two most salient risk factors for the development of post-stroke depression. Communication deficits and social isolation may also be considered as possible risk factors for depression. Refer to CSBPR Transitions of Care and Participation Module for information on depression in family and informal caregivers of people with stroke.

1.2 Assessment for post-stroke depression

i. People who have experienced a stroke whose screening indicates a high risk for depression should be assessed in a timely manner by a healthcare professional with expertise in diagnosis, management, and follow-up of depression [Evidence Level C].

Clinical considerations 1.2: Timing of screening for post-stroke depression (new in 2019)

i. Screening for post-stroke depression may take place at various stages throughout the continuum of stroke care, especially at transition points, as time of onset for post-stroke depression can vary and include:

a. At transfer from an inpatient acute setting to an inpatient rehabilitation setting;

b. From an inpatient rehabilitation setting before return to the community;

c. During secondary prevention clinic visits;

d. Following discharge to the community, during follow-up appointments with consulting specialists, and during periodic health assessments with primary care practitioners.

ii. Screening for depressive symptoms could be considered during the initial acute care stay, if deemed medically appropriate, particularly if evidence of depression or mood changes is noted or if risk factors for depression are present, as outlined in Section 1.1, iii.

iii. Repeated screening may be required since the ideal timing for screening for post-stroke depression is unclear.

1.4. Pharmacotherapy for post-stroke depression

i. People who have experienced a stroke with mild depressive symptoms or those diagnosed with minor depression may initially be managed by "watchful waiting"* (Evidence Level B]. *See note below for definition of watchful waiting**.

a. Pharmacological treatment should be considered and started if the depression is persistent or worsens and interferes with clinical goals [Evidence Level B].

**Note: Watchful waiting is defined as a period when the patient who displays mild depressive symptoms is monitored closely without additional therapeutic interventions to determine whether the mild depressive symptoms will improve. The timeframe for watchful waiting varies in the literature, typically between 2 and 4 weeks. It is often described as including suggestions to the patient for self-help strategies and participation in physical exercise.*

Clinical considerations 1.4

i. The involvement and feedback of people who have experienced a stroke, family, and caregivers is an important component of ongoing monitoring for post-stroke mood changes and conditions.

ii. Counseling and education should include information about potential relapse or recurrence of symptoms, signs to be aware of, the importance of adherence with prescribed medication regime, and contacting their primary care physician or mental health expert should those signs reappear.

1.7 Ongoing monitoring, support, and education

i. People who have experienced a stroke and families should continue to be given information and education about the potential impact of stroke on mood [Evidence level C].

Note. Source: Lanctôt et al., 2019

Appendix B: Questionnaires administered at each time point

Table 10

Questionnaires administered at each time point

Measure	T1 (3-10 days)	T2 (3 months)	T3 (6 months)	T4 (9 months)	T5 (12 months)	Discharge
SCID-I/NP current MDE criterion	✔ Day 8-10	✔	✔	✔	✔	
The Stroke Specific Geriatric Depression Scale (SS-GDS)	✔ Day 8-10	✔	✔	✔	✔	
Mental Health Index (MHI-5)	✔ Day 8-10	✔	✔	✔	✔	
Stroke Impact Scale (SIS)		✔	✔	✔	✔	
The EuroQOL five dimensions questionnaire (EQ-5D)		✔	✔	✔	✔	
Socio-Demographic and Health Related Questionnaire (full of brief version)	✔ ✔ Day 3	✔	✔	✔	✔	
Canadian Neurological Scale (CNS)	✔ 8-10 days					✔
OARS Social Resources Scale	✔ Day 3	✔	✔	✔	✔	
SCID-I/NP past MDE criterion		✔				
Barthel Index (BI)	✔					✔
Interviewer-administered Brief version of the Mini-Mental State Examination (Brief-MMSE)	✔ Recruitment phase					
Phone interview for refusers (within a month of missed interview)	✔	✔	✔	✔	✔	✔

Note. A checkmark indicates that a measure was administered at that time point.

Appendix C: Copies of Interviews and Questionnaires Included in this Study

DSM-5 Criteria for Major Depressive Episode

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest of pleasure.
Note: Do not include symptoms that are clearly attributable to another general medical condition.

(1) Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful).

(2) Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).

(3) Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day.

(4) Insomnia or hypersomnia nearly every day.

(5) Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).

(6) Fatigue or loss of energy nearly every day.

(7) Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).

(8) Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).

(9) Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The episode is not attributable to the physiological effects of a substance or another medical condition.

Note: Criteria A–C represent a major depressive episode.

Note. Depression was assessed at all five time points with criterion A for a current Major Depressive Episode (MDE) within the Mood Disorders Module of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision – DSM-IV-TR (American Psychiatric Association, 2000) Axis I Disorders, Non-Patient Edition – SCID-I/NP (First, Spitzer, Gibbon, & Williams, 2002). Due to copyright restrictions, the SCID-

I/NP interview could not be printed here. Above, please find the most recent DSM-5 (APA, 2013) criteria for Major Depressive Episode. In this study, a major depressive episode is met when criteria A and B are met, while criteria C was not assessed. The criteria for an MDE are essentially equivalent in the DSM-5 and the DSM-IV-TR for the purposes of this study. In this study, a Major Depressive Episode is diagnosed when at least five of the core symptoms in criterion A are met ([MDE], APA, 2000, p. 375), a Minor Depressive Disorder ([MIND], APA, p. 777) is diagnosed when at least two but less than five of the core symptoms in criterion A are met, and a label of depressive features is given when either core symptom (1) or (2) of criterion A is met. For both MDE and MIND, at least one item must be either MDE criterion (1) or (2). MIND is classified in the Depressive Disorder Not Otherwise Specified (NOS) section of the DSM-IV-TR.

The Stroke Specific Geriatric Depression Scale (SS-GDS)

Please circle yes or no for each question.

1. Do you feel full of energy?	YES / NO
2. Do you often get bored?	YES / NO
3. Do you think that most people are better off than you are?	YES / NO
4. Do you frequently get upset over little things?	YES / NO
5. Do you have trouble concentrating?	YES / NO
6. Do you often get restless and fidgety?	YES / NO
7. Are you basically satisfied with your life?	YES / NO
8. Are you bothered by thoughts you can't get out of your head?	YES / NO
9. Do you frequently feel like crying?	YES / NO
10. Do you feel that your life is empty?	YES / NO
11. Are you afraid that something bad is going to happen to you?	YES / NO
12. Do you enjoy getting up in the morning?	YES / NO
13. Do you often feel helpless?	YES / NO
14. Do you think it is wonderful to be alive now?	YES / NO
15. Do you worry a lot about the past?	YES / NO
16. Are you in good spirits most of the time?	YES / NO
17. Are you hopeful about the future?	YES / NO

SF-36 Mental Health Index (MHI-5)

These questions are about how you feel and how things have been with you during the past 4 weeks. For each question, please give the one answer that comes closest to the way you have been feeling. How much of the time during the past 4 weeks (circle one number on each line)?

	All of the time	Most of the time	A Good Bit of the time	Some of the time	A Little of the time	None of the time
b. Have you been a very nervous person?	1	2	3	4	5	6
c. Have you felt so down in the dumps that nothing could cheer you up?	1	2	3	4	5	6
d. Have you felt calm and peaceful?	1	2	3	4	5	6
f. Have you felt downhearted and blue?	1	2	3	4	5	6
h. Have you been a happy person?	1	2	3	4	5	6

The Stroke Impact Scale version 3.0 (SIS V 3.0) items (3a to 3i)

The purpose of this questionnaire is to evaluate how stroke has impacted your health and life. We want to know from YOUR POINT OF VIEW how stroke has affected you. We will ask you questions about impairments and disabilities caused by your stroke, as well as how stroke has affected your quality of life. Finally, we will ask you to rate how much you think you have recovered from your stroke.

These questions are about how you feel, about changes in your mood and about your ability to control your emotions since your stroke.

3. In the past week, how often did you...	None of the time	A little of the time	Some of the time	Most of the time	All of the time
a. Feel sad?	5	4	3	2	1
b. Feel that there is nobody you are close to?	5	4	3	2	1
c. Feel that you are a burden to others?	5	4	3	2	1
d. Feel that you have nothing to look forward to?	5	4	3	2	1
e. Blame yourself for mistakes that you made?	5	4	3	2	1
f. Enjoy things as much as ever?	5	4	3	2	1
g. Feel quite nervous?	5	4	3	2	1
h. Feel that life is worth living?	5	4	3	2	1
i. Smile and laugh at least once a day?	5	4	3	2	1

The EuroQOL five dimensions questionnaire (EQ-5D)

Please indicate which statement best describes your own health state today. Do not tick more than one box in each group.

Question 5

Anxiety / Depression

I am not anxious or depressed	☐
I am moderately anxious or depressed	☐
I am extremely anxious or depresses	☐

Socio-Demographic and Health Related Questionnaire

The following are questions on activities and work you have done in the last month and Socio-Demographic Information**.**

Please circle the number that corresponds with the appropriate answer.

1. Are you F (1) M (0)
2. What is your date of birth? _____ / _____ / _____
   Day Month Year
3. What is the highest level of education you have completed?
   (0) Grade 6 or less
   (1) Grade7 to 12 (not graduating from high school)
   (2) Graduated high school or equivalent
   (3) Part College or CEGEP
   (4) Graduated 2 years college or DEC
   (5) Graduated 2-3 years university
   (6) Part graduate/professional school
   (7) Completed graduate or professional school
   (8) Completed graduate or professional school
4. If failed to complete a program they were enrolled in why did they not complete it?

   _____

   _____

   _____

   _____
5. In which Country were you born? _____
6. What was the first language you learned as a child? _____
7. Do you have children ? Yes (1) No (2)
   If yes, how many? (1) one (2) two (3) three (4)four (5) >five
   How old are they ? _____
8. Are you currently employed ? Yes (1) No (0)
   If yes, -how many hours per week do you work ? _____
   -what kind of work do you do? _____

How long have you worked there?_____

If less than 6 months ask Why did you leave your last job?_____

b. Do you volunteer or help out in the community? (1) Yes No (0)

If yes, a) For how many hours per week ? _____

b) What do you do? _____

c) How long have you been doing this_____

9. Do you have a family doctor ? Yes (1) No (0)
10. Are you currently taking any medications ? Yes (1) No (0)
11. Have you fallen during the past month? Yes (1) No (0)

If yes, did you fall:

(1) at home Yes (1) No (0)

(2) outside the home Yes (1) No (0)

(3) did you injure yourself? Yes (1) No (0)

(4) What type of injury? _____

12. During the last month how many times have you……

been hospitalised	0	1	2	3	__	reasons _____
been to an emergency room	0	1	2	3	__	reasons _____
seen a doctor	0	1	2	3	__	reasons _____
seen a physiotherapist	0	1	2	3	__	reasons _____
seen an occupational therapist	0	1	2	3	__	reasons _____
seen a speech therapist	0	1	2	3	__	reasons _____
seen a dietician	0	1	2	3	__	reasons _____
seen other therapists	0	1	2	3	__	reasons _____
received someone from the CLSC	0	1	2	3	__	reasons _____
received any other services	0	1	2	3	__	reasons _____
(like meals on wheels, homecare, housekeeper etc)	0	1	2	3	__	reasons _____

13. Which of the following categories best describe your family income? (before income tax)

(1) Less than $20,000 (4) $40,000 to $49,999

(2) $20,000 to $29,999 (5) $50,000 to $69,999

(3) $30,000 to $39,999 (6) Over $70,000

(7) Refused to answer

14. In general, how do your family finances work out at the end of the month? Do you usually have:

(1) some money left over

(2) just enough to make ends meet

(3) not enough to make ends meet

(4) refused to answer

15. During the past 6 months, have you had any accidents causing an injury that led to a restriction of activities? (1) Yes (0) No

If yes, a) What type of accident? _____

b) What type of injury? _____

16. Do you currently smoke? (0) No ⇒ go to Question 14

(1) Yes, on a regular basis

(2) Yes, on occasions

17. Currently, approximately how many cigarettes do you smoke a day? _____

18. Which statement best describes your experience with cigarette smoking:

(0) I have never smoked cigarettes

(1) I occasionally smoke cigarettes

(2) I have smoked cigarettes on a daily basis in the past

19. Have you ever consumed beer, wine, liquor or other alcoholic drinks? (0.5% beer is not considered as alcohol)

(1) Yes (0) No ⇒ go to the end

Over the past 6 months, have you ever consumed beer, wine, liquor or other alcoholic drinks?

(1) Yes (0) No ⇒ go to the end

20. How often have you consumed alcoholic drinks during the past 6 months?

Did you drink:

(1) Everyday (4) Once a week

(2) 4 to 6 times per week (5) Once or twice a month

(3) 2 to 3 times per week (6) Less than once a month

Past periods of Psychopathology

21. Have you ever seen anyone for emotional or psychiatric problems? (1) Yes (0) No

22. Was there ever a time someone else thought you should see someone for emotional or psychiatric problems? (1) Yes (0) No

If yes how many times_____

23. Have you ever seen anyone for drug or alcohol problems? (1) Yes (0) No

If yes how many times_____

24. Thinking back over your whole life when were you the most upset?_____

Why?_____

25. Thinking back over your whole life when were you the feeling the best?

_____

_____

Checklist of Co-existing conditions

Please indicate whether you are presently treated for each of the identified conditions.

(1)	Hypertension	(11)	Diabetes
(2)	Heart attack (MI)	(12)	Glaucoma
(3)	Angina	(13)	Cataracts
(4)	Stroke	(14)	Impaired hearing
(5)	Hemiplegia	(15)	Parkinson`s disease
(6)	Bronchitis	(16)	Thyroid problem
(7)	Emphesema	(17)	Cancer (specify type, Location)
(8)	Asthma		_____
(9)	Arthritis	(18)	Liver disease
(10)	Ulcer disease	(19)	Other, specify, _____

Was this questionnaire filled out by the subject (1) or the caregiver/proxy (0)? _____

OARS Social Resources Scale

HEIGHT_____**WEIGHT**_____

Now, I would like to ask you some questions about your family and friends.

1. Are you single, married, widowed, divorced or separated?
   (1) single never married
   (2) married/common law
   (3) widowed
   (4) divorced
   (5) separated
   (9) not answered

2. Who lives with you?
   No = 0 Yes = 1

a)	0	1	no one
b)	0	1	husband or wife
c)	0	1	children
d)	0	1	grandchildren
e)	0	1	parents
f)	0	1	grandparents
g)	0	1	brothers and/or sisters
h)	0	1	other relatives
i)	0	1	friends
j)	0	1	non-related paid (includes free room) helper
k)	0	1	others (specify) _____

3. How many people do you know well enough to visit within their homes?
   (3) five or more
   (2) three to four
   (1) one to two
   (0) none
   (9) not answered

4. About how many times did you talk to someone (friends, relatives or others) on the telephone in the past week? (either you called them or they called you)
   If the subject has no phone, the question still applies.
   (3) once a day or more
   (2) 2 – 6 times
   (1) once
   (0) not at all
   (9) not answered

5. How many times during the past week did you spend some time with someone who does not live with you, that is you went to see them or they came to visit you or you went out to do things together?
   (3) once a day or more
   (2) 2 – 6 times
   (1) once
   (0) not at all
   (9) not answered

6. Do you have someone you trust and can confide in?
   (2) yes
   (0) no
   (9) not answered

7. Do you find yourself feeling lonely quite often, sometimes or almost never?
   (0) quite often
   (1) sometimes
   (2) almost never
   (0) not answered

8. Do you see your relatives and friends as often as you want to or are you somewhat unhappy about how little you see them?
   (1) as often as wants to
   (2) somewhat unhappy about how little
   (9) not answered

9. Is there someone who would give you any help at all if you were sick or disabled, for example, your husband / wife, a member of your family or a friend?
   (1) yes
   (0) no one willing and able
   (9) not answered

   If yes, ask a) and b)

   a) Is there someone who would take care of you as long as you needed, or only for a short time, or only someone who would help you now and then (for example, taking you to the doctor or fixing lunch occasionally, etc.)
      (1) Someone who would take care of the subject indefinitely (as long as needed)
      (2) Someone who would take care of the subject for a short time (a few weeks to six months)
      (3) Someone who would help the subject now and then (taking him to the doctor, fixing lunch, etc.)
      (9) not answered

   b) Who is this person?
   Relationship: _____

Canadian Neurological Stroke Scale

Patient Name: _____

Rater Name: _____

Date: _____

Mentation			Score
Level Consciousness		Alert	3.0
		Drowsy	1.5
Orientation		Oriented	1.0
		Disoriented/NA	0.0
Speech	Normal	1.0	
		Expressive Deficit	0.5
		Receptive Deficit	0.0

TOTAL: _____

Section A1	Motor Functions	Weakness	Score
NO COMPREHENSION DEFICIT			
	Face	None	0.5
		Present	0.0
	Arm: Proximal	None	1.5
		Mild	1.0
		Significant	0.5
		Total	0
	Arm: Distal	None	1.5
		Mild	1.0
		Significant	0.5
		Total	0
	Leg: Proximal	None	1.5
		Mild	1.0
		Significant	0.5
		Total	0
	Leg: Distal	None	1.5
		Mild	1.0
		Significant	0.5
		Total	0

TOTAL: _____

Section A2	Motor Functions	Weakness	Score
COMPREHENSION DEFICIT			
	Face	Symmetrical	0.5
		Asymmetrical	0.0
	Arms	Equal	1.5
		Unequal	0.0
	Legs	Equal	1.5
		Unequal	0.0

TOTAL: _____

The interviewer-administered Brief version of the Mini-Mental State Examination

Sample Items

Orientation to Time "What is the date?" Registration "Listen carefully. I am going to say three words. You say them back after I stop. Ready? Here they are... HOUSE (pause), CAR (pause), LAKE (pause). Now repeat those words back to me." [Repeat up to 5 times, but score only the first trial.] Naming "What is this?" [Point to a pencil or pen.]

Barthel Index (BI)

Patient Name: _____
Rater Name: _____
Date: _____

Activity	Score
FEEDING 0 = unable 5 = needs help cutting, spreading butter, etc., or requires modified diet 10 = independent	_____
BATHING 0 = dependent 5 = independent (or in shower)	_____
GROOMING 0 = needs to help with personal care 5 = independent face/hair/teeth/shaving (implements provided)	_____
DRESSING 0 = dependent 5 = needs help but can do about half unaided 10 = independent (including buttons, zips, laces, etc.)	_____
BOWELS 0 = incontinent (or needs to be given enemas) 5 = occasional accident 10 = continent	_____
BLADDER 0 = incontinent, or catheterized and unable to manage alone 5 = occasional accident 10 = continent	_____
TOILET USE 0 = dependent 5 = needs some help, but can do something alone 10 = independent (on and off, dressing, wiping)	_____
TRANSFERS (BED TO CHAIR AND BACK) 0 = unable, no sitting balance 5 = major help (one or two people, physical), can sit 10 = minor help (verbal or physical) 15 = independent	_____

MOBILITY (ON LEVEL SURFACES)
0 = immobile or < 50 yards
5 = wheelchair independent, including corners, > 50 yards
10 = walks with help of one person (verbal or physical) > 50 yards
15 = independent (but may use any aid; for example, stick) > 50 yards _____

STAIRS
0 = unable 5 = needs help (verbal, physical, carrying aid)
10 = independent _____

TOTAL (0–100): _____

Depression-Related Questions on Telephone Interview for Refusers

Questions from SF-36 MHI-5, rated on a 6 point likert-type scale, ranging from all of the time to none of the time:

c. Have you felt so down in the dumps that nothing could cheer you up?

f. Have you felt downhearted or blue

h. Have you been a happy person?

Appendix D: Additional Description and Rationale for Factors Included in the Index Test Selection Process

D.1. Supplementary Information on Use of Accuracy Statistics in the Test Selection Process

Selected tests were required to show good Se for both MDE and any depression to reflect the clinical position that stroke survivors with depressed mood or anhedonia warrant assessment and follow up by medical professionals.

When selecting between multiple candidate index tests which all met defined minimum criteria for Se, Sp, and C as well as additional criteria defined below, tests showing additional gains in accuracy were preferred. In these cases, additional gains in Se were prioritized over gains in Sp, unless loss of global accuracy (C) was substantial. Additionally, gains in Se in identifying MDE were prioritized over gains in Se for identifying any depression because it was more important that index tests do well at screening for the most severe depression.

Exact 95% confidence intervals (CI) were calculated for Se and Sp point estimates using Clopper-Pearson interval method for exact binomial CI. CIs were not directly used to inform test selection due to insufficient power associated with few depressed cases in the analyses. Instead, CIs were presented to inform reliability of the final selected index tests.

D.2. Minimizing bias in assessment tools due to spectrum effect

The selection process was designed to minimize the possibility that screening tools show spectrum bias or spectrum effect, which occurs when a test does not adequately represent all population subgroups (Mulherin & Miller, 2002). To minimize bias associated with spectrum effects, selected screening tests were required to demonstrate adequate minimum accuracy (Se ≥70% and Sp ≥ 60%) for identifying depression in all sample subgroups stratified dichotomously by sex, age, language, level of education, and stroke severity. The number of subgroups for each covariate was limited to two groups with equal n, to maintain sufficient sample size and power. The criteria that Se must be ≥70% permitted one FN classification when four depressed cases were included in the analysis. The Se criterion of ≥70% was waived when <4 depressed participants were included in analyses for a subgroup. Additionally, tests were required to perform with reasonably similar accuracy across both levels of each binary subgroup. Reasonably similar accuracy was defined in this study as <30% difference in both Se and Sp. In the case where <5 depressed were included in the analyses, the cut off for acceptable difference in Sp remained at <30% and while the cut off for acceptable difference in Se was increased to

<40%. For the purposes of these analyses, depression was defined broadly as 'any depression,' due to limitations in sample size. Candidate index tests not meeting these criteria were interpreted as showing evidence of spectrum bias and were not retained in the test selection process.

D.3. Minimizing Bias Due to Missing Data in Assessment Tools

As noted in the section on handling missing data, skipped as well as missing values for index tests were interpreted as meaningful. Step three of the selection process addressed the goal of minimizing the possibility that selected index tests include items that are more likely to be skipped by depressed participants, resulting in spuriously inflated accuracy statistics. In step three, Chi-square tests were completed to assess the relationship between missingness of index tests at a given time point (missing versus not missing) and status of depression (any depression versus no depression) at that time point. A meaningful association (i.e., sufficient effect size) was defined as an Odds Ratio ≥ 2.0, and index tests with ORs ≥ 2.0 were dropped in step three of the test selection process. The index tests were dropped because such an association was indicative of poor test functioning amongst depressed participants. Eliminated from these analyses were cases missing the entire questionnaire or interview, as this type of missing data was not viewed as suggestive of poor item functioning. Cases with missing test values remaining in these analyses were specifically due to skipping an item in question within an otherwise completed questionnaire. The association also provided evidence of missingness associated with an underlying level of depression that would bias resulting index test parameters. Statistical significance with 95% confidence was not required because reducing type II error was prioritized over reducing type I error for this calculation.

In step six, the best two to three models at each time point were compared across a number of criteria including test accuracy, number of depressed cases in the analyses, and concerns at the item level within each test. This study set no definitive hierarchy for these criteria. Instead, the weighting of each criteria was considered on a case-by-case basis, with the rationale for test selection clearly explained in Table 7 of the results section. Several types of item-level problems were considered, including evidence of poor item functioning due to patterns of missingness as well as concerns with syntax, grammar and face validity. Preference was placed on index tests containing questions with wording that were simple, easily understood, and had questions that appeared valid for assessment of depression in stroke survivors. Tests

containing items that had a concerning pattern of missingness were identified as indicative of poor item functioning. A concerning pattern of item missingness was defined as: a) items missed with a frequency of ≥4 more than the most frequently answered items on that questionnaire at a given time point; b) items missed with a frequency of ≥ 7 more than the most frequently answered items on that questionnaire summed over all time points; c) items missed by depressed participants with a frequency of ≥ 2 more than the most frequently answered items on that questionnaire at a given time point; and d) items missed by depressed participants with a frequency of ≥ 5 more than the most frequently answered items on that questionnaire summed over all time points.

Appendix E: Supplementary Tables from the Results Section

Table 11

Testing the Assumption of Normality in descriptive statistic variables measured on interval scales, complete case analysis

Variable of measure	Mean (SD)	Median	Mode	Range	Tests of Normality	Skewness	Kurtosis
Age**	70.8 (13.1)	72.1	-	27.0 – 93.5	All ≤0.036	-0.78	0.83
Number of symptoms in three days post stroke	3.6 (1.8)	3.0	2.0	0 - 8	All < 0.01	0.33	-0.55
CNS score	7.9 (2.6)						
Barthel at 3 days, *(missing assigned 0)*	48.8 (27.9)	50.0	30.0	0 - 100	All ≤ 0.0147	0.025	-0.99
Barthel at discharge, (missing assigned 0)	65.8 (29.8)	75.0	100.0	0 - 100	All < 0.01	-0.57	-0.85
Length of Hospital Stay (LOS)	18.3 (15.7)	14.0	8.0	3 - 102	All < 0.01	2.64	8.82
SS-GDS Index	4.3 (2.9)	4.0	2.0	0 - 13	All < 0.01	0.76	0.22
MHI-5 Index	22.7 (5.2)	24.0	26.0	6 - 30	All < 0.01	-0.92	0.56
Number of medications at discharge	6.1 (4.2)	6.0	6.0	0 - 22	All < 0.01	0.77	0.94

Note. Tests of Normality include: Shapiro-Wilk (W), Kolmogorov-Smirnov (D), Cramer-von Mises (W-Sq), Anderson-Darling (A-Sq), and each of their respective p values.

Table 12

Demographics at baseline for participants stratified and compared by sex

Variable of Measure	Means (SD) or Frequencies (Percent of Sample)	Comparing for Sex Differences	Men n =69 (57%)	Direction of Effect	Women n =52 (43%)
Age, *mean (SD)***	70.8 (13.1)	χ^2 (1, 121) = 6.69, p = 0.0097**	68.3 (12.8)	<	74.2 (12.8)
Spoken Language	121 (100%)	χ^2 (2, 121) = 4.16, p = 0.13			
English, no French	69 (57.0%)				
French, no English	33 (27.3%)				
French and English	19 (15.7%)				
Living with whom***		χ^2 (4, 121) = 21.68, p <0.0001***	69 (100%)	≠	52 (100%)
Spouse***	55 (45.4%)	χ^2 (1, 121) = 15.39, p = 0.0001***	42 (60.9%)	>	13 (25.0%)
Alone	43 (35.5%)	χ^2 (1, 121) = 3.01, p = 0.089	20 (29.0%)		23 (44.2%)
Family*	14 (11.6%)	χ^2 (1, 121) = 5.23, p = 0.041*	4 (5.8%)	<	10 (19.2%)
Friends	2 (1.6%)	χ^2 (1, 121) = 1.53, p = 0.50	2 (2.9%)		0
Other*	7 (5.8%)	χ^2 (1, 121) = 5.54, p = 0.042*	1 (1.4%)	<	6 (11.5%)
Highest Completed Education, mean rank (SD)*	3.1 (1.9)	χ^2 (1, 121) = 3.14, p = 0.076, p = 0.038* with 1-sided Wilcoxon	3.4 (2.0)	>	2.7 (1.7)
				Men cumulative	
Less than grade 6 (0)	3 (2.4%)		1 (1.4%)	2 (3.8%)	2 (3.8%)
Elementary school (1)	27 (22.3%)		15 (21.7%)	12 (23.1%)	12 (23.1%)
High school (2)	30 (24.8%)		14 (20.3%)	16 (30.8%)	16 (30.8%)

(continued)

Variable of Measure	Means (SD) or Frequencies (Percent of Sample)	Comparing for Sex Differences	Men n =69 (57%)	Direction of Effect	Women n =52 (43%)
Some Cegep (3)	15 (12.4)		8 (11.6%)	7 (13.5%)	7 (13.5%)
Some College or DEP (4)	7 (5.8%)		4 (5.8%)	3 (5.8%)	3 (5.8%)
Graduated University (2-3 years) (5)	25 (20.7%)		16 (23.2%)	9 (17.3%)	9 (17.3%)
Some graduate or professional School (6)	9 (7.4%)		6 (8.7%)	3 (5.8%)	3 (5.8%)
Graduate or Professional School (7)	5 (4.1%)		5 (7.2%)	0	0
Working before stroke**	41 (33.9%)	χ^2 (1, 121) = 6.6, p = 0.012**	30 (43.5%)	>	11 (21.2%)
Marital Status***	120 (99%)	χ^2 (3, 120) = 25.64, p < 0.0001***			
Single, never married	12 (9.9%)	χ^2 (1, 120) = 0.003, p =1.0	7 (10.1%)		5 (9.6%)
Married or common law***	58 (47.9%)	χ^2 (1, 120) = 18.53, p < 0.0001***	45 (65.2%)	>	13 (25%)
Widowed***	34 (28.1%)	χ^2 (1, 120) = 22.4, p < 0.0001***	8 (11.6%)	<	26 (50%)
Divorced	16 (13.2%)	χ^2 (1, 120) = 0.01, p = 1.0	9 (13.0%)		7 (13.5%)
Separated	0		0		0

Note. n=121. Values are frequencies with parentheses reflecting the percentage unless otherwise specified. Sex difference variables measured on ordinal or greater scales were calculated with non-parametric Kruskal-Wallis (K-W) test. A one-sided exact Wilcoxon Mann-Whiney U (WMW) test was employed to test for sex differences in level of education. Differences in categorical variables were calculated with chi-square tests and p values were calculated with Fisher's two-sided exact test, unless otherwise specified. Significant difference between men and women *$p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. In the direction of effect column, '<' indicates that the value for men was significantly lower than for women (e.g., men were significantly younger than women), '>' indicates that the value for men was significantly greater than for women (e.g., compared to women, men were significantly more likely to live with a spouse than not live with a spouse), and '≠' indicates that there was a significant difference between at least two of the measured groups. There were identified sex differences in descriptive statistics for the sample of 121 at baseline. Women were older than men. Women's age range was from 44 to 93

years and men's age range was from 27 to 87 years. The level of education was higher in men than women. Compared to women, men were more likely to be married or living in common law, living with their spouse, and working outside the home. Women were more likely than men to be widowed, living with a family member, or living in an "other" living arrangement. Results are presented in Tables 12 and 13.

Table 13

Clinical characteristics at baseline for participants stratified and compared by sex

Clinical variable of measure	Means (SD) or Frequencies (Percent of Sample)	Comparing for Sex differences
Type of Stroke		χ^2 (1, 121) = 0.75, p = 0.51
Ischemic	111 (91.7%)	
Haemorrhagic	10 (8.3%)	
Lesion Side		χ^2 (3,121) = 1.44, p = 0.90
Left	46 (38.0%)	
Right	71 (58.7%)	
Bilateral	3 (2.5%)	
Not noted	1 (0.8%)	
Side of Hemiplegia		χ^2 (2, 121) = 3.22, p = 0.19
Left	66 (54.6%)	
Right	39 (32.2%)	
Bilateral	0	
None	16 (13.2%)	
First stroke	102 (84.3%)	χ^2 (1, 121) = 1.19, p = 0.32
Any comorbid conditions	114 (96.6%)	χ^2 (1, 118) = 0.078, p =1.0
Any dysphasia, recruiter sheet	19 (15.7%)	χ^2 (1, 121) = 0.86, p = 0.45
# of stroke symptoms, first three days post stroke, mean (SD)	3.6 (1.8)	z = 0.13, p = 0.90
CNS score, mean (SD)	7.9 (2.6)	z = -1.59, p = 0.110
CNS speech score (0, 0.5, 1), mean (SD)	0.65 (0.33)	z=-1.46, p=0.14
Barthel at 3 days, (missing assigned 0), mean (SD)	48.8 (27.9)	z = -1.32, p = 0.19
Barthel at discharge, (missing assigned 0), mean (SD)	65.8 (29.8)	z = -1.64, p = 0.10
SF 36 Mental Health Index, Mean, (SD)	70.7 (20.9)	z = -0.44, p = 0.66

(continued)

Clinical variable of measure	Means (SD) or Frequencies (Percent of Sample)	Comparing for Sex differences
Length of Hospital Stay (LOS), mean (SD)	18.3 (15.7)	$z = 0.88$, $p = 0.38$
Number of medications at discharge, *mean (SD)*	6.1 (4.2)	$z = 1.49$, $p = 0.14$
Discharge destination		χ^2 (4, 121) = 4.3, p = 0.40
Deceased before discharge	3 (2.5%)	
Rehabilitation	76 (62.8%)	
Home	37 (30.6%)	
Transfer to another acute care facility	4 (3.3%)	
Long Term Care Facility	1 (0.8%)	
History of any Depression	36 (29.8%)	χ^2 (1, 94) = 0.52, p = 0.52
History of MDE	17 (14.0%)	χ^2 (1, 94) = 2.25, p = 0.18

Note. Canadian Neurological Scale (CNS); Degrees of freedom (DF); Standard deviation (SD). Values are frequencies (percent) unless otherwise specified. Sex differences variables measured on ordinal or greater scales, calculated with non-parametric Kruskal-Wallis (K-W) tests for three or more groups and Wilcoxon Mann-Whiney U (WMW) tests for two groups. Differences in categorical variables, calculated with chi-square tests and p values, calculated with Fisher's two-sided exact test, unless otherwise specified. *$p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$.

Table 14

Characteristics of participants stratified by covariate subgroups

Dichotomized Covariates	Subgroups	Definition	n (% of sample)
Sex	Women	Reported by participant	52 (43%)
	Men	Reported by participant	69 (57%)
Age	Young	< 72 years at baseline	60 (50%)
	Old	≥72 years at baseline	61 (50%)
Education	Low	Level 0-2: highest level of completed education is ('Grade 6 ≥' to 'Graduated highschool or equivalent')	60 (50%)
	High	Level 3-7: highest level of completed education is ('Part of College or CEGEP' to 'Completed graduate or profectional school')	61 (50%)
CNS score	Low	CNS total score <8.5 (of 11.5 scale)	56 (46%)
	High	CNS total score >8.4 (of 11.5 scale)	65 (54%)
Language	English	Speak English (+ any number of other languages)	88 (73%)
	French	Speaks French and no English	33 (27%)

Note. n=121. CNS = Canadian Neurological Stroke Scale. Proxies were consulted for four participants at T1, for three participants at T2, and for one participant at T3, T4, and T5. Proxy support was too infrequent to be considered a meaningful confound in this study.

Appendix F: Analyses and Discussion Aimed at Characterizing Depression Post Stroke

Table 15 shows the frequency of participants who met depression criteria, as well as the frequency of participants who endorsed each of the nine core criteria from the SCID-I/NP interview for depression at baseline. Table 16 shows frequencies of participants meeting SCID-I/NP-based criterion for depression at each time point as well as when depression was first met during the course of the study.

Table 17 shows results of non-parametric Spearman correlations between demographic and clinical characteristics. Women were older than men and participants living alone were more likely to be female and older. All scores on measures for depression and mental health status at baseline were significantly correlated. Number of medications at discharge was correlated with length of hospital stay as well as increased depression and lower mental health status at baseline. A history of meeting criteria for any depression in the past was associated with lower mental health status at baseline on the MHI-5.

Table 15

Sample frequency for meeting SCID-I/NP criteria for depression and for endorsing core MDE symptoms at baseline

Depression	Frequency Endorsed			Percent of depressed (n=8)
	Men (n =4)	Women (n =4)	All (n =8)	
Binary Outcomes:				
Major Depressive disorder	1	1	2	25
Minor Depression or more severe	2	2	4	50
Any depression	4	4	8	100
DSM-5 Symptoms				
1. Depressed or sad mood	3	4	7	
2. Diminished Interest/pleasure	3	4	7	
3. Significant weight loss or weight gain	1	1	2	
4. Insomnia or hypersomnia	1	1	2	
5. Psychomotor agitation or retardation	1	3	4	
6. Fatigue or loss of energy	2	2	4	
7. Worthlessness or guilt	1	0	1	
8. Diminished thinking/concentration/indecisiveness	0	0	0	
9. Thoughts of death or suicidal ideation	0	0	0	

Note: One of eight participants were not asked about symptoms 3-9. DSM 5 (diagnostic and statistical manual 5th ed.).

Table 16

Sample frequencies for participants meeting diagnostic criteria for depression by time point, and for time point when depression criteria were first met

Time points	Frequency meeting MDE criteria at each time point (% of 17 meeting MDE criteria at any point in first year)	Frequency first reporting MDE criteria at each time point (of n =17)	Frequency meeting any depression criteria at each time point (% of 52 participants meeting criteria at any time during first year)	Frequency first reporting any depression criteria at each time point (of n =52)
T1	2 (12%)	2 (12%)	8 (15%)	8 (15%)
T2	10 (62%)	8 (47%)	23 (45%)	18 (35%)
T3	7 (47%)	4 (24%)	24 (47%)	14 (27%)
T4	2 (15%)	0 (0%)	20 (44%)	4 (8%)
T5	5 (33%)	3 (18%)	17 (40%)	8 (15%)

Note. MDE (Major Depressive Episode)

Table 17

Spearman`s Rank Correlation Coefficients between depression and other clinical and demographic characteristics at baseline, complete case analysis

	1	2	3	4	5	6	7	8	9	10	11
1. Sex	-										
2. Age	.24**	-									
3. Any depression (SCID-I/NP) (T1)	.04	-.11	-								
4. n depressive symptoms (SCID-I/NP) (at T1)	.04	-.10	n/a	-							
5. severity of depression (SCID-I/NP) (T1)	.04	-.11	n/a	n/a	-						
6. SS-GDS total score (T1)	.05	-0.11	0.27**	.26**	.27**	-					
7. MHI-5 total score (T1)	-.04	.05	-.30***	-.30**	-.30***	-.56***	-				
8. N medications at discharge	.14	.18	.13	.12	.12	.23*	-.22*	-			
9. Length of Hospital Stay	.08	-.03	-.03	-.03	-.03	.12	-.13	.20*	-		
10. Live alone	.32***	.27**	.12	.12	.12	.03	-.02	.15	-.03	-	
11. Past history of Depression	.07	-.11	.11	.11	.11	.18	-.26*	.10	.04	.07	-

Note. Numbers 1-11 in the column titles denote the variables listed in rows with the corresponding numbers. SS-GDS = Geriatric Depression Scale - Stroke Specific; Mental Health Inventory - 5; SCID-I/NP= Structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders. p<.05; ** p<.01; ***p<.001.

Comparing the Demographic and Clinical Characteristics of Depressed and Non-Depressed Participants

To better understand the relationship between PSD and demographic and clinical characteristics, this author examined whether participants, grouped by most severe level of depression identified during the course of the study, differed significantly in demographic and clinical characteristics. Severity of depression was categorized in three methods shown in subsections of Table 18: a) depression v. no depression; b) major depressive episode (MDE) v. no MDE; and c) no depression, depressive features, minor depression, and MDE. Omitted from all analyses are participants who were missing data such that it could not be determined whether the participants met depression criteria during the course of this study. In Table 18, age, length of hospital stay, number of medications at discharge, neurological status, number of depressive symptoms, and mental health status scores at baseline, all measured on ordinal or interval scales, were compared across participants grouped by most severe level of depression in the first year. Table 19 presents results of post hoc analyses on significant results from section 'c' of Table 18.

As expected, participants meeting criteria for any depression during the study had significantly worse mental health or depression status on the MHI-5 and the SS-GDS at baseline, compared to participants who did not meet criteria for any depression during the course of the study. Similarly, participants meeting criteria for MDE during the study had significantly worse mental health and depression on these measures at baseline compared to participants never meeting criteria for an MDE during the study. Post hoc analyses in Table 19 reveal that participants whose most severe episode during the study was a minor depressive episode had significantly worse scores on the MHI-5 and SS-GDS at baseline compared to the no depression group, and had worse SS-GDS scores at baseline compared to the depressive features group. Similarly, participants meeting criteria for a an MDE during the study had significantly worse MHI-5 and SS-GDS scores at baseline compared to participants not meeting criteria for an MDE and compared to participants meeting criteria for depressive features. Interestingly, participants whose most severe level of depression was 'depressive features' had significantly higher neurological status (CNS score) at baseline compared to each of the other groups. Participants categorized by most severe status and severity of depression did not differ significantly in measures of age, length of hospital stay, or number of medications at discharge. No other differences were significant.

Table 18

Comparing demographic and clinical characteristics in participants grouped by most severe depression criteria met in the study

		K-W Test (2 levels) Any v. no depression	Median Complete sample	Any depression	No depression
A	Age (years)	X^2 (1, 89) = 0.02, p = 0.90	72.1		
	Days in Hospital	X^2 (1, 89) = 0.44, p = 0.50	14.0		
	CNS score	X^2 (1, 89) = 0.01, p = 0.91	8.5		
	SS-GDS score	X^2 (1, 86) = 10.54, p = 0.0012***	4.0	5.0	3.0
	MHI-5 score	X^2 (1, 89) = 10.10, p = 0.0015***	24.0	21.5	25.0
	n meds at discharge	X^2 (1, 86) = 0.55, p = 0.46	6.0		
		K-W Test (2 levels) MDE v. NO MDE	Median Complete sample	MDE	No MDE
B	Age (years)	X^2 (1, 68) = 0.04, p = 0.85	72.1		
	Days in Hospital	X^2 (1, 68) = 1.21, p = 0.27	14.0		
	CNS score	X^2 (1, 68) = 1.74, p = 0.19	8.5		
	SS-GDS score	X^2 (1, 68) = 5.40, p = 0.02*	4.0	6.0	3.0
	MHI-5 score	X^2 (1, 68) = 12.16, p = 0.0005***	24.0	20.0	25.0
	n meds at discharge	X^2 (1, 67) = 2.47, p = 0.12	6.0		

		K-W Test (4 levels) All 4 levels of depression	Median No depression	Depressive features	Minor Depression	MDE
C	Age (years)	X^2 (3, 89) = 0.35, p = 0.95				
	Length of hospital stay (days)	X^2 (3, 89) = 0.30, p = 0.38				
	CNS score	X^2 (3, 89) = 11.92, p = 0.0076**	8.5	9.5	7.25	7.5
	SS-GDS score	X^2 (3, 86) = 17.20, p = 0.0006***	3.0	4.0	6.5	6.0
	MHI-5 score	X^2 (3, 89) = 16.44, p = 0.0009***	25.0	25.0	21.0	20.0
	n meds at discharge	X^2 (3, 86) = 3.31, p = 0.35				

Note. Analyses are completed with the Kruskal-Wallis (K-W) test, a non-parametric test comparing medians. MHI-5: Higher scores indicate better health status. SS-GDS: Higher scores indicate more depressive symptoms. CNS (Canadian Neurological Stroke scale): Higher scores represent improved neurological status (decreasing stroke severity). *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0.001.

Table 19

Post hoc tests comparing the CNS scores of participants grouped by most severe level of depression met during the study

		No depression, n = 37	Depressive features, n = 17	Minor episode, median, n = 18
CNS score	No depression, median = 8.5			
	Depressive features, median = 9.5	X^2 (1, 54) = 5.64, p = 0.018*		
	Minor depressive episode, median = 7.2	X^2 (1, 55) = 1.66, p = 0.20	X^2 (1, 35) = 8.60, p = 0.003**	
	Major depressive episode, median = 7.5	X^2 (1, 55) = 1.68, p = 0.19	X^2 (1, 34) = 7.76, p = 0.005**	X^2 (1, 35) = 0.12, p = 0.73
SS-GDS score	No depression, median = 3.0			
	Depressive features, median = 4.0	X^2 (1, 53) = 0.68, p = 0.41		
	Minor depressive episode, median = 6.5	X^2 (1, 53) = 12.11, p = 0.0005***	X^2 (1, 32) = 6.67, p = 0.01**	
	Major depressive episode, median = 6.0	X^2 (1, 54) = 7.86, p = 0.0050**	X^2 (1, 33) = 4.05, p = 0.044*	X^2 (1, 33) = 0.72, p = 0.39
MHI-5 score	No depression, median = 25.0			
	Depressive features, median = 25.0	X^2 (1, 54) = 0.60, p = 0.44		
	Minor depressive episode, median = 21.0	X^2 (1, 55) = 7.20, p = 0.007**	X^2 (1, 35) = 2.93, p = 0.087	
	Major depressive episode, median = 20.0	X^2 (1, 54) = 12.58, p = 0.0004***	X^2 (1, 34) = 6.71, p = 0.01**	X^2 (1, 35) = 0.38 p = 0.54

Note. Analyses are completed with the Kruskal-Wallis (K-W) test. MHI-5: Higher scores indicate better health status; SS-GDS: Higher scores indicate more depressive symptoms. CNS (Canadian Neurological Stroke scale): Higher scores represent improved neurological status (decreasing stroke severity). *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

The next set of analyses examined whether participants grouped by the most severe level of depression during the study (same groupings as in Tables 18 and 19) differed across demographics and clinical characteristics measured on nominal scales. Assessed nominal characteristics include the dichotomous covariates listed in Table 14, including sex (male v. female), age (young v. old), education level (high v. low), CNS score (high v. low), and language (English v. only French). Additional variables in analyses included work status before the stroke (working v. not working), stroke type (ischemic v. haemorrhagic), history of depression (any depression v. none), history of a major depressive episode (MDE v. no MDE), living with whom at the time of the stroke (alone v. with someone), martial status (married/common-law v. unmarried), and lesion side (left, right, or bilateral). Results presented in Table 20 and Table 21 show post hoc chi square analyses for significant 4x2 contingency tables.

Severity of depression was significantly predicted by level of education, CNS score, work status, and history of an MDE. Specifically, compared to participant not meeting MDE criteria during the course of the study, participants meeting criteria for an MDE during the study had lower education and were less likely to be working at the time of the stroke. Compared to participants not meeting criteria for any depression during the study, participants meeting criteria for any depression during the study were more likely to have reported a history of a previous major depressive episode. Participants only meeting criteria for depressive features during the study were more likely to have reported a history of MDE, compared to the no depression group and compared to the minor depressive episode group. Severity of depression within the first year post stroke was not significantly associated with nominally measured sex, age, language, type of stroke, and history of any depression.

Table 20

Comparing demographic and clinical characteristics in participants grouped by most severe depression criteria met in the study, results of chi-squared analyses

	A	b	c
Demographic/Clinical correlates, dichotomized, measured at baseline	Any depression v. none	MDE v. less than MDE	All levels of depression (None, depressive features, minor, MDE)
Men v. women	χ^2 (1, 89) = 1.08, p = 0. 38	χ^2 (1, 68) = 0.75, p = 0.40	χ^2 (3, 89) = 1.80, p = 0.62
Age (old v. young)	χ^2 (1, 89) = 0.0004, p =1.00	χ^2 (1, 68) = 9.08, p = 0.78	χ^2 (3, 89) = 0.70, p = 0.89
Education (high v. low)*	χ^2 (1, 89) = 0.40, p = 0.67	χ^2 (1, 68) = 5.04, p = 0.048*	χ^2 (3, 89) = 6.91, p = 0.077
CNS score (high v. low)**	χ^2 (1, 89) = 0.17, p = 0.83	χ^2 (1, 68) = 1.96, p = 0.26	χ^2 (3, 89) = 11.18, p = 0.0083**
Language (English v. French only)	χ^2 (1, 89) = 1.58, p = 0.23	χ^2 (1, 68) = 2.52, p = 0.12	χ^2 (3, 89) = 3.09, p = 0.38
Working (yes v. no)*	χ^2 (1, 89) = 2.74, p =0.12	χ^2 (1, 68) = 4.068, p = 0.050*	χ^2 (3, 89) = 4.14, p = 0.26
Stroke type (ischemic v. haemorrhagic)	χ^2 (1, 89) = 0.0052, p = 1.00	χ^2 (1, 68) = 0.24, p = 1.00	χ^2 (3, 89) = 0.45, p = 1.00
past depression (any v. none)	χ^2 (1, 75) = 2.07, p = 0.17	χ^2 (1, 63) = 3.94, p = 0.073	χ^2 (3, 75) = 7.73, p = 0.059
Past MDE (MDE v. no MDE)***	χ^2 (1, 75) = 6.70, p = 0.015*	χ^2 (1, 63) = 1.16, p = 0.44	χ^2 (3, 75) = 15.92, p = 0.0012***
Living alone (yes v. no)	χ^2 (1, 89) = 0.18, p = 0.83	χ^2 (1, 68) = 0.51, p = 0.57	χ^2 (3, 89) = 2.71, p = 0.45
Lesion Side (left, right, bilateral)	χ^2 (2, 88) = 1.58, p = 0.43	χ^2 (2, 68) = 2.53, p = 0.38	χ^2 (6, 88) = 6.02, p = 0.36

Note. Analyses are completed with chi square tests of homogeneity. Probability is calculated by the two-sided Fisher's exact test. MDE: Major depressive episode; MHI-5: Higher scores indicate better health status; SS-GDS: Higher scores indicate more depressive symptoms; CNS (Canadian Neurological Stroke scale): Higher scores represent improved neurological status (decreasing stroke severity). *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

Table 21

Post hoc chi square analyses for significant differences in Table 20

		No depression	Depressive features	Minor depressive episode
CNS score	No depression			
(high v. low)	Depressive features	χ^2 (1, 54) = 6.80, p = 0.014**		
	Minor depressive episode	χ^2 (1, 55) =0.23, p = 0.78	χ^2 (1, 35) = 7.44, p = 0.012**	
	Major depressive episode	χ^2 (1, 54) =1.21, p = 0.38	χ^2 (1, 34) = 10.09, p = 0.0039**	χ^2 (1, 35) = 0.30, p = 0.73
Past MDE	No depression			
(MDE v. no MDE)	Depressive features	χ^2 (1, 47) = 13.97, p = 0.0007***		
	Minor depressive episode	χ^2 (1, 47) = 0.052, p = 1.00	χ^2 (1, 26) = 6.50, p = 0.030*	
	Major depressive episode	χ^2 (1, 49) = 4.18, p = 0.062	χ^2 (1, 28) = 2.16, p = 0.24	χ^2 (1, 28) = 1.71, p = 0.33

Note. Analyses are completed with chi square tests of homogeneity. Probability is calculated by the two-sided Fisher's exact test. MDE: Major depressive episode; CNS score (Canadian Neurological Stroke Scale score) – Higher scores represent improved neurological status (decreasing stroke severity). *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

Assessing for Relationships in Demographic and Clinical Characteristics and Time of Depression Onset Post Stroke

This author was interested in examining whether clinical and demographic variables were associated with length of time between the stroke event and the onset of depression. Separate analyses were conducted to assess the relationship with onset time of a) any depression, b) depressive features or minor depression, and c) an MDE, and are presented in subsections of Table 22. As shown in Table 22, MHI-5 score at baseline was positively correlated with onset time of any depression, such that improved health status at baseline was associated with later onset of any depression following a stroke event. In addition, number of medications at discharge was negatively correlated with onset time of any depression, such that more medications at discharge was associated with earlier onset of any depression following a stroke event. No other correlations were significant.

Table 23 shows results of non-parametric Kruskal-Wallis (K-W) tests examining whether nominally classified demographic and clinical characteristics were associated with onset time of depression after the stroke. The same nominal variables were assessed here as were shown in Tables 18 and 19. Consistent with Table 22, separate analyses were conducted to assess the relationship with onset time of a) any depression, b) depressive features or minor depression, and c) an MDE. Results, presented in Table 23, show that compared to English speaking participants, participants who spoke French and no English had a significantly earlier onset of either depressive features or minor depression. In addition, compared to non-working participants, participants working at the time of their stroke had earlier onset of either depressive features or minor depression. Results of all other analyses were non-significant.

Table 22

Differences characteristics of participants grouped by first reported onset time of each severity of depression criteria met with the SCID-I/NP interviews

Demographic and Clinical correlates, measured at baseline	Time point when first reported:		
	a. Any depression	b. Depressive features or minor depression	c. MDE
Age (years)	rs (52) = -0.067, p = 0.64	rs (24) = 0.22, p = 0.30	rs (17) = -0.042, p = 0.87
Length of hospital stay (days)	rs (52) = 0.079, p = 0.58	rs (24) = -0.11, p = 0.61	rs (17) = -0.098, p = 0.71
CNS score	rs (52) = 0.14, p = 0.33	rs (24) = 0.14, p = 0.52	rs (17) = 0.39, p = 0.12
SS-GDS score	rs (49) = -0.23, p = 0.12	rs (21) = 0.032, p = 0.89	rs (17) = -0.17, p = 0.51
MHI-5 score**	rs (52) = 0.37, p = 0.0065**	rs (24) = 0.29, p = 0.17	rs (17) = 0.24, p = 0.36
n meds at discharge*	rs (50) = -0.34, p = 0.015*	rs (23) = -0.30, p = 0.16	rs (17) = -0.39, p = 0.12

Note. Analyses were completed with Spearman rank-order correlations, a nonparametric measure of association based on the ranks of the data values. MDE: Major depressive episode; CNS score (Canadian Neurological Stroke Scale score) – Higher scores represent improved neurological status (decreasing stroke severity). *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

Table 23

Differences in nominally measured characteristics of participants, grouped by first reported onset time of each severity of depression criteria met with the SCID-I/NP interviews

Demographic and Clinical correlates, measured at baseline	Time point when first reported:		
	a. Any depression	b. Depressive features or minor depression	c. MDE
Men v. women	χ^2 (1, 52) = 0.001, p = 0.97	χ^2 (1, 24) = 0.29, p = 0.59	χ^2 (1, 17) = 0.95, p = 0.33
Age (old v. young)	χ^2 (1, 52) = 0.001, p = 0.97	χ^2 (1, 24) = 0.86, p = 0.35	χ^2 (1, 17) = 0.17, p = 0.68
Education (high v. low	χ^2 (1, 52) = 0.013, p = 0.91	χ^2 (1, 24) = 0.073, p = 0.79	χ^2 (1, 17) = 0.20, p = 0.65
CNS score (high v. low)	χ^2 (1, 52) = 2.76, p = 0.097	χ^2 (1, 24) = 0.11, p = 0.74	χ^2 (1, 17) = 2.94, p = 0.086
Language (English v. French only)*	χ^2 (1, 52) = 0.083, p = 0.77	χ^2 (1, 24) = 4.11, p = 0.043*	χ^2 (1, 17) = 1.13, p = 0.72
Working (yes v. no)	χ^2 (1, 52) = 0.67, p = 0.41	χ^2 (1, 24) = 3.75, p = 0.052*	χ^2 (1, 17) = 1.16, p = 0.28
Stroke type (ischemic v. haemorrhagic)	χ^2 (1, 52) = 0.51, p = 0.48	χ^2 (1, 24) = 0.050, p = 0.82	χ^2 (1, 17) = 0.30, p = 0.59
past depression (any depression v. none)	χ^2 (1, 41) = 0.57, p = 0.45	χ^2 (1, 17) = 0.36, p = 0.55	χ^2 (1, 15) = 0.00, p = 1.00
Past MDE (MDE v. no MDE)	χ^2 (1, 41) = 0.011, p = 0.92	χ^2 (1, 17) = 0.010, p = 0.92	χ^2 (1, 15) = 0.41, p = 0.52
Living alone (yes v. no)	χ^2 (1, 52) = 0.19, p = 0.66	χ^2 (1, 24) = 2.35, p = 0.12	χ^2 (1, 17) = 0.95, p = 0.33
Lesion Side (left, right, bilateral)	χ^2 (2, 51) = 0.46, p = 0.79	χ^2 (2, 23) = 2.47, p = 0.29	χ^2 (2, 17) = 0.38, p = 0.54

Note. Analyses were completed with the Kruskal-Wallis (K-W) test, a non-parametric test comparing medians. MDE: Major depressive episode; CNS score (Canadian Neurological Stroke Scale score) – Higher scores represent improved neurological status (decreasing stroke severity). *$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$.

Appendix G: Supplementary Missing Data Section

A goal in this study was to select methods for handling missing data that would mitigate the possibility of developing index tests based on biased parameter estimates and drawing biased conclusions that do not generalize to the population of interest. The sections below describe the steps taken to accomplish this goal as well as the results of these steps.

Rationale and Methods. Missing data can influence the results of statistical analyses leading to biased conclusions about the population of interest. First introduced by Rubin (1976), missing data is frequently categorized as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). The distinction between the mechanisms of missing data (i.e., MAR, MCAR, and MNAR) inform the degree to which analyses on a sample can bias results and limit the ability to draw conclusions about a population. Data is considered MCAR when missingness does not depend on observed or unobserved variables. The cause of the missingness is due to other variables that are irrelevant. Results of analyses conducted on variables with data MCAR, by definition, do not bias conclusions about the population of interest. Data that is MAR (or ignorable missingness) occurs when missingness has nothing to do with the missing values themselves but is due to other variables that were measured. Once these causal variables are controlled for in the model, missingness no longer results in bias (Graham, 2009). Finally, data that is MNAR (or non-ignorable missingness) occurs when the probability of missing values on given variable is related directly to the underlying levels of the missing variable (Harel & Zhou, 2007).

In this study's analyses of substantive interest, missed interviews, questions, and items are MNAR when missingness is directly related to underlying unmeasured levels of depression. Data that is MNAR is difficult to identify because, by definition, it cannot be predicted from other measured variables, including previous scores for the same variable (Graham, 2009). Therefore, data MAR is a non-testable assumption. In the review, 'Missing data analysis, making it work in the real word', Graham argues that in the real world, missing data does not fall neatly into one of these categories but falls on a continuum somewhere between MAR and MNAR. In handling missing data, Graham argues that attrition-related missingness is ubiquitous to longitudinal studies and recommends attrition related missingness and missingness in general be treated as MNAR (i.e., not purely MAR). Consistent with Graham, it was expected that missing values for reference standard outcomes of depression were at least partially dependent

on these unmeasured levels of depression. For example, it was anticipated in this study that some participants feeling too depressed or too good to want to participate in interviews or questionnaires (MNAR).

In line with Graham (2009), this author interpreted that participants missing complete interviews, for example, due to death, level of interest, or inability, as mirroring the normal tendency of some stroke patients to refuse mental health assessments and are reflective of a meaningful subgroup of the population. Participants present for interviews were considered reflective of the population of interest: participants willing and interested to be screened, referred, and followed by healthcare professionals for PSD. As well, when particular items or questionnaires showed greater levels of missing data than others, the missing data was interpreted as evidence of poor functioning of that item or questionnaire. Data missing at the item level resulting in missing values for candidate index tests was interpreted as meaningful and indicative of problems with the index test questions that could generalize to use of the index tests in the population of interest. Good index tests for depression should not contain questions that participants, and particularly depressed participants, routinely avoid answering.

In keeping with this theoretical position, no missing data was statistically imputed. Responders were interpreted as reflecting our actual population of interest and imputation would result in inclusion of non-responders in analyses. Multiple Imputation (MI) is often viewed as the preferred method for imputation so additional reasons for not using it are discussed here. Most importantly, MI is only valid under the assumption that data is MAR (Spratt et al., 2010) and results is biased parameter estimates when data is MNAR (Graham, 2009). This is a problem since, as noted above, it is assumed that some data is MNAR (i.e., that unmeasured level of depression is directly related to some of the missingness). Furthermore, participants missing complete interviews or multiple interviews within the study are likely to also be missing sufficient information to produce an accurate MI model auxiliary that generates unbiased parameter estimates[33] under the assumption that data is missing MAR. Imputing values using inaccurate models would add further bias to results. Additional reasons to not use MI is that MI assumes normal distribution of data, is often not the case in this study.[34] Last, MI works best when missingness is spread across variables, which it was not in this study.

Instead, pairwise deletion was the selected method to handle missing data in this study. This choice is further discussed in the methods section of this dissertation.

Measuring Missing Data. The extent of missing data was described in the results section of this dissertation.

Examining for Patterns in Missing Data: Methods, Results, and Interpretation.

Methods were included to attempt to understand reasons for missingness, to measure variables related to missingness, and to reduce bias associated with missingness in the analyses of substantive interest.

First, to pre-emptively assess and understand the causes of missingness, interview refusers were asked to participate in refuser interviews by telephone, which asked reasons for refusal. Notably, many refusers also refused participation in refuser questionnaires for reasons expected to be partially explained by underlying levels of depression. Table 24 lists reported reasons by participants for refusing to the complete the SCID-I/NP interview at each time point. Because many refusers provided no reason for missing an interview, no patterns could be meaningfully identified for underlying reasons for missingness. Qualitatively, death, sickness, feeling good, and busyness, were among the reported reasons for refusal.

Second, interview refusers were also asked to respond to selected MHI-5 mood items as part of the refuser telephone questionnaire. As shown in Table 25, less than 24% of refusers at all time point completed MHI-5 items from the refuser phone questionnaire within a month of the missed interview. Higher percentages of responses at each time point (0% to 89% response rate) were provided to the question 'have you been depressed in the last 3 months?' also asked on the refuser questionnaire. The percentage of responders who reported 'yes' ranged from 12% to 40% across time points. Due to the high percentage of missing data in the refuser questionnaires, results from phone interviews were not regarded as sufficiently reliable to infer relationships between missing interviews and mood within the sample.

Third, to examine for evidence of data that was MAR, analyses were completed to determine whether covariates predict missing at least one SCID-I/NP interview. As the SCID-I/NP is the reference test, its completion was necessary for inclusion in all analyses testing candidate index tests. As shown in Table 26 below, the frequency of participants without a completed SCID interview at T1 was four, at T2 was 13, at T3 was 15, at T4 was 31, and at T5 was 30. Forty-four of 121 participants (36%) were missing at least one SCID-I/NP interview throughout the course of the study. Table 26 also presents results of chi-square tests of homogeneity assessing for significant relationships between missing SCID-I/NP interviews and

dichotomized covariates for sex, level of education, language, age, and CNS score, listed in Table 14. Results of analyses showed that the dichotomous covariates were not significantly associated with missing a SCID interview at any time or at specific time points.

Forth, methods for minimizing bias associated with missing data were included in step three of the six-step section for index tests at each time point. This step was implemented to reduce the likelihood that selected index tests contain items that are likely to be skipped by people who are depressed. In step three, candidate models were dropped when there was evidence that the reasons for their missing items was related to underlying levels of depression. Specifically, candidate models were dropped when their frequency of missing values for an index test model at a given time point was meaningfully associated with meeting criteria for 'any depression' at that time point. Included in the missing value count were instances in which items were skipped within an otherwise attempted questionnaire. Chi-square tests were completed to assess the relationship between missingness of index tests at a given time point (missing v. not missing) and status of depression (any depression v. no depression) at that time point. A meaningful association was defined as an Odds Ratio ≥ 2.0. In this step, seven and three candidate models were dropped at T3 and T5, respectively.

Fifth, methods for minimizing bias associated with missing data were included in step six of the six-step section for index tests at each time point. In step six, models were preferred when they contained items that did not show problems related to higher levels of skipped items, compared to other items in that questionnaire. This measure was taken to reduce the likelihood that selected index tests contain items that have increased likelihood of being skipped by stroke survivors. Response frequency at each time point for all items and questionnaires contributing to candidate index tests are presented in Table 27, Appendix I. Table 27 provides the frequencies of questionnaire and item responses by time point categorized by (A) all participants and (B) participants who were diagnosed with any depression at that time point. Concerning patterns of missingness indicative of a problematic items are defined in the notes under Table 27 below. Results of these examinations revealed that two SS-GDS items at T2 and one SS-GDS item at T5 were skipped more frequently than other SS-GDS items by participants categorized with any depression, and were thus flagged as potentially problematic. In step six of the six-step selection process, candidate index tests containing these items were considered potentially problematic. How these results influenced index test selection in step 6 is described in Appendix I.

Table 24

Participants' reported reasons for skipping interviews and questionnaires

Time point	Frequency of refusal of SCID-I/NP interview	Reported Reasons (frequency) for refusing SCID-I/NP
T1	4	▪ 4x no provided reason
T2	13	▪ 3x deceased ▪ 1x feeling too good ▪ 1x on vacation ▪ 2x sick in hospital / too sick to respond ▪ 2x too little data ▪ 4x no provided reason
T3	15	▪ 4x deceased ▪ 1x too busy moving ▪ 10x no provided reason
T4	31	▪ 5x deceased ▪ 1x too busy ▪ 25x no provided reason
T5	30	▪ 6x deceased ▪ 24x no provided reason

Note. SCID-I/NP (Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision Axis I Disorders, Non-Patient Edition).

Table 25

Frequency of interview refusers at each time point, and frequency of refusers with available mood items

Time point	Completely missing interviews by time point, not due to death	n refusers with responses to MHI-5 items	n refusers with responses to other mood items	n (%) endorsing depression item
T1	n = 2 missed	n = 0 (0%) within 1 month of refused T1 interview	n = 0, (0%)	n/a
T2	n = 8 missed	n = 2 (12%) within 1 month of refused T2 interview	n = 5, (62%)	n = 2 (40%)
T3	n = 9 missed	n = 1 (11%) within 1 month of refused T3 interview	n = 8, (89%)	n = 1 (12%)
T4	n = 26 missed	n = 6 (23%) within 1 month of refused T4 interview	n = 12, (46%)	n = 3 (25%)
T5	n = 19 missed	n = 2 (10%) within 1 month of refused T5 interview	n = 8, (42%)	n = 3 (38%)

Note. MHI-5 (Mental Health Index)

Table 26

Examining whether missing SCID-I/NP interviews are predicted by dichotomous covariates

	Missing SCID at any time (n = 44, 36%)	Missing SCID at T1 (n = 4, 3%)	Missing SCID at T2 (n = 13, 11%)	Missing SCID at T3 (n = 15, 12%)	Missing SCID at T4 (n = 31, 26%)	Missing SCID at T5 (n = 30, 25%)
Men v. women	χ^2 (1,121) = 0.64, p = 0.45	χ^2 (1,121) = 0.08, p = 1.0	χ^2 (1,121) = 0.12, p = 0.78	χ^2 (1,121) = 0.06, p = 1.0	χ^2 (1,121) = 0.08, p = 0.83	χ^2 (1,121) = 0.80, p = 0.40
Level of education (High v. Low)	χ^2 (1,121) = 1.45, p =0.26	χ^2 (1,121) = 1.07, p =0.36	χ^2 (1,121) = 0.10, p =0.78	χ^2 (1,121) = 0.74, p =0.42	χ^2 (1,121) = 1.20, p =0.30	χ^2 (1,121) = 1.73, p =0.21
English v. French only	χ^2 (1,121) = 1.62, p = 0.21	χ^2 (1,121) = 0.01, p = 1.0	χ^2 (1,121) = 2.82, p = 0.11	χ^2 (1,121) = 3.25, p = 0.11	χ^2 (1,121) = 1.42, p = 0.25	χ^2 (1,121) = 1.77, p = 0.24
Young v. old (>71 y.o.)	χ^2 (1,121) = 0.47, p = 0.57	χ^2 (1,121) = 0.0003, p = 1.0	χ^2 (1,121) = 4.09, p = 0.075	χ^2 (1,121) = 1.81, p = 0.27	χ^2 (1,121) = 0.98, p = 0.40	χ^2 (1,121) = 2.66, p = 0.14
CNS score (high v. low)	χ^2 (1,121) = 0.38, p = 0.57	χ^2 (1,121) = 0.023, p = 1.0	χ^2 (1,121) = 0.34, p = 0.57	χ^2 (1,121) = 0.001, p = 1.0	χ^2 (1,121) = 0.48, p = 0.54	χ^2 (1,121) = 0.022, p = 0.68

Note. Probabilities calculated with Fisher's two-sided exact test. SCID is short for Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision Axis I Disorders, Non-Patient Edition (SCID-I/NP) Missing SCID-I/NP includes those missing for any reason including refusal, mistake, death. *p ≤ 0.05, **p ≤ 0.01, ***p ≤ 0. 001.

Appendix H: Supplementary Section on Accuracy Measures Employed in this Study

Supplementary information is presented here on the properties of the accuracy measures, their particular strengths and limitations in the context of PSD screening, as well as rationale for their selection. In consideration of the strengths and limitations of the various accuracy measures for the purpose of screening test selection, Se and Sp were preferred as the primary point estimates as they provide more flexibility than weighted or unweighted global statistics like C-statistic. Positive and Negative Predictive Values and Relative Risks were reported for final selected index tests. The relationship between accuracy statistics employed in this study are shown in Figure 5 of this appendix.

Sensitivity and Specificity. Se and Sp are not influenced by prevalence. As a result, when comparing index tests and the functioning of index tests across time, the variation in percentage of depressed participants included in each analysis (attributable to varying incidence of PSD across time and patterns of missing data) do not impact the Se and Sp values, allowing for a more clear comparison of accuracy. Additionally, Se and Sp scores are expected to remain consistent across samples with varying prevalence, in conditions where the assumption of consistent spectrum effects is not violated. Due to their accuracy in measurement and flexibility for use in selecting best models, Se and Sp were the primary accuracy measures employed in the selection process. For step one of the selection process, requiring Se $\geq$ 85%, if fewer than seven depressed cases were included in the analysis, no number of false negative cases were permitted by the criteria. In the case of seven depressed cases in the analysis, one false negative was permitted.

The C-Statistic. The C-statistic offers a useful and frequently employed index for comparing tests where both sensitivity and specificity are considered. This statistic is generally regarded as important and a score of 0.7 is generally viewed as acceptable. Therefore, an acceptable minimum value of C is selected in this study, although due to the problems inherent to C-statistics described here, more importance was placed on desired Se and Sp values. A problem with relying exclusively on C statistic is its weighting of Se and Sp. In selecting a model index test based on the highest C-statistic, the implicit assumptions are that: a) sensitivity and specificity are valued equally; and b) that the relative costs of each false negative and false positive case are inversely proportional to the prevalence. These assumptions regarding the equal weighting of FN and FP cases do not hold in this study. Notably, relying on the weighted

Youden index (a weighted global accuracy index) is also not optimal because prevalence of PSD changes across samples, which then impacts weighting. For example, when prevalence is 15%, in general, FN cases are weighted over five times that of FP cases. However, when prevalence is 33%, FN cases are weighted over two times more than FP cases. This may be fine when these assumptions are acceptable. The benefit of this system is that C-statistic would remain constant across samples with changing prevalence where sample characteristics impacting spectrum effect remain constant. In binary models like in this study, the C-statistic is equal to the area under the Receiver Operating Characteristic (ROC) curve. To calculate C, all possible combinations of depressed (R+) and non-depressed (R-) cases are paired. For each predictive model, C is the number of pairs for which R+ and R- were accurately discriminated by the index test (where R+ = T+ and R- =T-), D is the number of pairs incorrectly discriminated (where R+ = T- and R- =T+), and T is the number of pair given the same value on the index test (where R+ = T+ and R- =T+, or where R+ = T- and R- =T-). Somer's D is the C-D divided by the number of pairs. The C-statistic then is 0.5 (1+Somer's D) (Allison, 2003).

Positive Predictive Values. Positive predictive value (PPV), expressed as a percentage, provides a partial estimate of post-test risk and is the probability that subjects with a positive screening test result truly have the condition. PPV is calculated by determining the proportion of truly depressed cases of all cases with positive index tests results. PPV increases as prevalence increases and its calculation requires a known prevalence. In addition, spectrum effects influence PPV. Because prevalence of PSD varies substantially across tested samples, reported PPVs for index tests would not be reliable for use in the population of interest. Therefore, PPV values will be reported for the selected index tests for informative purposes within the present sample, but cannot be expected to generalize to other samples. The following is the formula for PPV:

$$PPV=TP / (TP + FP)$$

Relative Risk. Relative Risk (RR) provides an estimate of how many times more likely the risk of depression will be given the test result. RR is calculated by the probability of depression with a positive test divided by the probability of depression with a negative test. Where RR = 1, there is no difference in risk of depression between positive and negative test results. RR estimates > 1 and < 1 indicate that the risk of depression increases with positive test results and negative test results, respectively. Because RR is calculated from Negative Predictive

Value (NPV) and PPV, it is influenced proportionally by prevalence.[35] For RR to be useful and reliable in the population of interest, prevalence must be known and stable. As with the PPV, given substantial variance in prevalence across PSD samples, calculated RR values in this study will have limited meaning outside of the present sample and should not be generalized to the population of interest. Therefore, RR values will be reported for the selected index tests for informative purposes within the present sample. Confidence intervals for RR provide the reliability for the RR estimate, where significant RRs have both lower and upper bounds above zero (or below zero). The following is the formula for RR:

$$RR = PPV / (1-NPV)$$

Figure 5

Relationship of accuracy statistics, with example.

N_T=76		**Patients' Depression diagnosis (as confirmed on the SCID-I/NP)**			
		Depression Present (Positive)N_{R+}=11	Depression Absent (Negative) N_{R-}=65		
Exposure: Results of ultra-short screening tool	Positive Test N_{I+}=25	**True Positive** (TP) = 10	**False Positive (Type I error)** (FP) = 15	**PPV** = TP / (TP + FP) = 10 / (10 + 15) = 40%	**RR** = *PPV / (1-NPV)* =0.4/0.02 = 20
	Negative Test N_{L-}=51	**False Negative (Type II error)** (FN) = 1	**True Negative** (TN) = 50	**NPV** = TN / (FN + TN) = 50 / (1 + 50) = 98%	**OR** = (TP/FP) / (FN/TN) =(10/15) / (1/50) =33
		Se = TP / (TP + FN) = 10 / (10 + 1) =91% **Prevalence** =(TP+FN)/Total N =11/76 = 14%	**Sp** = TN / (FP + TN) = 50 / (15 + 50) = 77%		
		C- Statistic = *.5 (1+ ((C-D)/C+D+T))*			

Appendix I: Supplementary Tables and Figures on Results of the Six Step Selection Process

Table 27

Frequencies of missing responses for complete interviews, specific questionnaires, and specific items at each time point for (A) all participants and (B) participants who were diagnosed with any depression at that time point.

Missing responses	T1 (A)	T1 (B)	T2 (A)	T2 (B)	T3 (A)	T3 (B)	T4 (A)	T4 (B)	T5 (A)	T5 (B)	Total N	
Missing full interview (SCID-I/NP, MHI-5, SS-GDS, & SIS3) for any reason	2		11		13		31		25		35	
Deceased (at time point)	1		3		4		5		6		6	
SCID-I/NP (missing & cannot score)	4		13		15		31		30		44	
SCID-I/NP item 1	4		11		15		31		29		43	
SCID-I/NP item 2	4		13		16		31		30		45	
Missing Questionnaires											Total (A)	Total (B)
MHI-5	2	0	12	0	13	0	31	0	25	0	81	0
SIS 3	N/A	N/A	25	5	33	9	52	9	40	5	150	28
GDS	4	0	26	5	32	9	52	9	38	3	148	26
Missing items												
MHI-5 b	2	0	12	0	13	0	31	0	25	0	81	0
MHI-5 c	3	0	12	0	15	1	32	0	25	0	84	1
MHI-5 d	2	0	12	0	13	0	31	0	25	0	81	0
MHI-5 f	2	0	12	0	13	0	31	0	26	1	82	1
MHI-5 h	2	0	12	0	13	0	31	0	26	1	82	1
EQ-5D	N/A	N/A	25	6	34	4	52	9	39	4	150	23
SIS 3a	N/A	N/A	12	0	13	0	33	0	27	1	85	1
SIS 3b	N/A	N/A	25	5	33	9	52	9	40	5	150	28
SIS 3c	N/A	N/A	26	5	33	9	53	9	40	5	152	28
SIS 3d	N/A	N/A	26	5	33	9	53	9	40	5	152	28
SIS 3e	N/A	N/A	25	5	33	9	53	9	40	5	151	28
SIS 3f	N/A	N/A	26	5	33	9	52	9	40	5	151	28
SIS 3g	N/A	N/A	26	6	35	9	54	9	40	5	155	29
SIS 3h	N/A	N/A	26	5	33	9	53	9	40	5	152	28
SIS 3i	N/A	N/A	25	5	33	9	53	9	41	6	152	29
SS-GDS 1	5	0	26	5	32	9	52	9	38	3	148	26

(continued)

Missing responses	T1 (A)	T1 (B)	T2 (A)	T2 (B)	T3 (A)	T3 (B)	T4 (A)	T4 (B)	T5 (A)	T5 (B)	Total N	
SS-GDS 2	5	0	26	5	32	9	52	9	39	3	149	26
SS-GDS 3	6	0	26	5	32	9	52	9	40	4	150	27
SS-GDS 4	5	0	27	5	32	9	52	9	40	4	151	27
SS-GDS 5	5	0	26	5	32	9	52	9	39	4	149	27
SS-GDS 6	4	0	26	5	32	9	52	9	41	5**	151	28
SS-GDS 7	5	0	26	5	33	9	52	9	38	3	149	26
SS-GDS 8	5	0	26	5	33	9	53	9	38	3	150	26
SS-GDS 9	5	0	29	8**	32	9	53	10	38	3	152	30
SS-GDS 10	5	0	27	6	32	9	52	9	39	3	150	27
SS-GDS 11	6	0	26	5	32	9	52	9	39	4	149	27
SS-GDS 12	5	0	27	6	33	9	53	9	39	3	152	27
SS-GDS 13	5	0	26	5	32	9	52	9	38	3	148	26
SS-GDS 14	5	0	29	8**	32	9	52	9	38	3	151	29
SS-GDS 15	4	0	26	5	32	9	52	9	38	3	148	26
SS-GDS 16	4	0	26	5	32	9	52	9	38	3	148	26
SS-GDS 17	4	0	27	6	33	9	53	9	40	4	153	28

Note. Frequency of missing items was the same for all items across all cut off points. Sum of Missing (A) and (B) include only T2-5, to allow for comparisons of total missing responses across questionnaires, as some questionnaires were not administered at T1. N/A= Not administered at that time point. **Signals problematic items. Concerning patterns of missingness indicative of a problematic items was defined as: a) items missed with a frequency of ≥ 4 more than the most frequently answered items on that questionnaire at a given time point; b) items missed with a frequency of ≥7 more than the most frequently answered items on that questionnaire summed over all time points; c) items missed by depressed participants with a frequency of ≥2 more than the most frequently answered items on that questionnaire at a given time point; and d) items missed by depressed participants with a frequency of ≥ 5 more than the most frequently answered items on that questionnaire summed over all time points. SIS item response frequency were not compared to SIS 3a response frequency because the frequency of missing responses after item SIS 3a was substantially elevated and appears to represent a problem with the questionnaire as a whole in the context of the entire battery rather than any problematic item.

le 28

mining step five criteria for potential concerns at item level in the top two to three models

ne nt	Q.1 wording (required response with cut off score)	Con-junction	Q.2 wording (required response with cut off score)	Syntactical Concerns with Q.1 and 2 (QUAID)	Theoretical and Grammatical Concerns with Q.1 and 2 or B	Problems evidenced skipping these items?
	SS-GDS 17: Are you Hopeful for future - No	OR	MHI-5 H: have you been a happy person - A little of the time (5) or less often.	SS-GDS 17: Unfamiliar technical terms: future. MHI-5 H: no indicated terminology problems.	"Been" Do they mean before the stroke or following the stroke?	No
	MHI-5 F: Have you felt downhearted and blue? - Some of the time (4) or more often	AND	MHI-5 H: Have you been a happy person - A little of the time (5) or less often	MHI-5 F: Unfamiliar technical terms: downhearted. MHI-5 H: no indicated terminology problems.	"Been" Do they mean before the stroke or following the stroke?	No
	SS-GDS 2: Do you often get bored? - Yes	AND	MHI-5 H: Have you been a happy person? - A little of the time (5) or less often	SS-GDS 2: Vague or imprecise relative terms: Frequency ambiguity: often The point or value on the implicit underlying scale is vague or imprecise. MHI-5 H: no indicated terminology problems.	"Been" Do they mean before the stroke or following the stroke?	No

Table 28 continued)

T2	SS-GDS 14: Do you think it is wonderful to be alive now? - No	OR	MHI-5 C (5): Have you felt so down in the dumps that nothing could cheer you up? - A little of the time (5) or more often	SS-GDS 14: Vague or imprecise relative terms: Temporal ambiguity: now; Deictic term: now MHI-5 C: no reported terminology problems.	MHI-5 C contains 2 conditions which may be confusing; Difficulty with translation of "down in the dumps".	SS-GDS 14 skipped by 3 patients who completed questionnaire (all who missed it were depressed)!
	SS-GDS 10: Do you feel that your life is empty?* - Yes	OR	MHI-5 C (5): Have you felt so down in the dumps that nothing could cheer you up? - A little of the time (5) or more often	SS-GDS 10: No reported terminology problems. MHI-5 C: no reported terminology problems.	MHI-5 C contains 2 conditions which may be confusing; Difficulty with translation of "down in the dumps". Emptiness may be a confusing term.	No
T3	SIS 3 C: Feel that you are a burden to others? - All of the time (1)	OR	MHI-5 F: Have you felt downhearted and blue? Some of the time (4) or more often.	SIS 3 C: Unfamiliar technical terms and value or ambiguous noun: others MHI-5 F: Unfamiliar technical terms: downhearted.	Is feeling like a burden in the case of stroke problematic? Is endorsement confounded by whether you are a burden?	No
	SIS 3 G (1): Feel quite nervous? - All of the time (1)	OR	MHI-5 F (4): Have you felt downhearted and blue? - Some of the time (4) or more often.	SIS 3 G: Question looks good to me. MHI-5 F: Unfamiliar technical terms: downhearted	How much is quite?	No

(Table 28 continued)

	SS-GDS 8: Are you bothered by thoughts you can't get out of your head? - Yes	OR	MHI-5 B (3): Have you been a very nervous person? - A good bit of the time (3) or more often	SS-GDS 8: Unfamiliar technical term/vague or ambiguous noun: head. MHI-5 B: no indicated terminology problems	"Bothered" may be ambiguous.	No
T4	SIS 3 F: Enjoy things as much as ever? Some of the time (3) or less often	OR	MHI-5 F: Have you felt downhearted and blue? – Most of the time (2) or more often	SIS 3 F: Vague or imprecise relative terms: Quantification ambiguity: much; Frequency ambiguity: ever. Vague or ambiguous noun-phrases: Abstract noun: things. MHI-5 F: Unfamiliar technical terms: downhearted.	This item measures anhedonia. As much as ever may be confusing. Will they compare current state to pre-stroke, to their average pre-stroke state, to their average post stroke state, or to their happiest time.	Sub-threshold problems (noted in step 3)
	SS-GDS 8: Are you bothered by thoughts you can't get out of your head? - Yes	OR	MHI-5 D (5): Have you felt calm and peaceful? - A little of the time (5) or less often	SS-GDS 8: Unfamiliar technical terms / and vague or ambiguous term: head. MHI-5 D: No indicated terminology problems	Bothered may be ambiguous. Otherwise, good grammar and face validity.	No
	SS-GDS 8: Are you bothered by thoughts you can't get out of your head? - Yes	OR	MHI-5 F (3): Have you felt downhearted and blue? - A good bit of the time (3) or more often	SS-GDS 8: Unfamiliar technical terms/ and vague or ambiguous noun: head. MHI-5 F: Unfamiliar technical terms: downhearted.	"Bothered" may be ambiguous. Good simple grammar/structure.	No

(Table 28 continued)

T5	EQ-5D (2): My health state is moderately anxious or depressed or worse	AND	SS-GDS 1: Do you feel full of energy? - No	EQ-5D: Unfamiliar technical terms: moderately. SS-GDS 1: "Question looks good to me. Answer looks good to me."	Good face validity. Simple grammar.	No
	SS-GDS 1: Do you feel full of energy? - No	AND	MHI-5 D: Have you felt calm and peaceful? - Most of the time (2) or less often	SS-GDS 1: "Question looks good to me. Answer looks good to me." MHI-5 D: no indicated terminology problems	Looks good grammatically. It is not intuitive that feeling calm and peaceful most of the time or less would predict depression -> question has poor face validity.	No

Note. The authors recommended model at each timepoint. Accuracy statistics refer to classification of 'any depression' unless otherwise specified.

Appendix J: Final PSD Screening Tools at each Time point

Figure 6

PSD screening questionnaire for use between three and 10 days after a stroke (example 1)

10 days after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. In the past 4 weeks, have you felt downhearted and blue?	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
2. In the past 4 weeks, have you been a happy person	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time

Answers

At 10 days after stroke, if answers to EITHER question 1 OR to question 2 fall inside a white box, this patient likely does not have depression currently. Please continue to screen for depression throughout this patient's recovery from stroke.

If this patient's answers to BOTH question 1 AND question 2 fall inside a grey box, this patient is at high risk for depression. Please refer this patient to a qualified mental health professional for a diagnostic assessment for depression.

Figure 7

PSD screening questionnaire for use between three and 10 days after a stroke (example 2)

10 days after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. Do you often get bored?					YES	NO
2. In the past 4 weeks, have you been a happy person	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time

Answers

At 10 days after stroke, if answers to EITHER question 1 OR to question 2 fall inside a white box, this patient likely does not have depression currently. Please continue to screen for depression throughout this patient's recovery from stroke.

If this patient's answers to BOTH question 1 AND question 2 fall inside a grey box, this patient is at high risk for depression. Please refer this patient to a qualified mental health professional for a diagnostic assessment for depression.

Figure 8

PSD screening questionnaire for use at three months after a stroke

3 months after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. Do you feel that your life is empty?				YES		NO
2. In the past 4 weeks, have you felt so down in the dumps that nothing could cheer you up?	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time

Answers

At 3 months after stroke, if a patient's answers to BOTH question 1 AND question 2 fall inside a white box, this patient likely does not have depression currently. Please continue to screen for depression throughout the patient's recovery from stroke.

At 3 months after stroke, if a patient's answers to EITHER question 1 OR to question 2 fall inside in a grey box, this patient is at high risk for depression and should be referred to a qualified mental health professional for a diagnostic assessment for depression.

Figure 9

PSD screening questionnaire for use at six months after a stroke

6 months after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. In the past 4 weeks, have you felt downhearted and blue?	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time
2. In the past week, how often did you feel quite nervous?	All of the time	Most of the time	Some of the time	A little of the time	None of the time	

Answers

At 6 months after stroke, if a patient's answers to BOTH question 1 AND question 2 are in a white box, this patient likely does not have depression currently. Please continue to screen for depression throughout the patient's recovery from stroke.

At 6 months after stroke, if a patient's answer to EITHER question 1 OR question 2 is in a grey box, this patient is at high risk for depression and should be referred to a qualified mental health professional for a diagnostic assessment for depression.

Figure 10

PSD screening questionnaire for use at nine months after a stroke (example 1)

9 months after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. Are you bothered by thoughts you can't get out of your head?					YES	NO
2. In the past 4 weeks, have you felt calm and peaceful?	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time

Answers

At 9 months after stroke, if a patient's answer to BOTH question 1 AND to question 2 are in a white box, this patient likely does not meet criteria for depression currently. Please continue to screen for depression throughout the patient's stroke recovery.

At 9 months after stroke, if a patient's answer to EITHER question 1 OR to question 2 is in a grey box, this patient is at high risk for depression and should be referred to a qualified mental health professional for a diagnostic assessment for depression.

Figure 11

PSD screening questionnaire for use at 9 months after a stroke (example 2)

9 months after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. Are you bothered by thoughts you can't get out of your head?					YES	NO
2. In the past 4 weeks, have you felt downhearted and blue?	All of the time	Most of the time	A good bit of the time	Some of the time	A little of the time	None of the time

Answers

At 9 months after stroke, if a patient's answer to BOTH question 1 AND to question 2 are in a white box, this patient likely does not meet criteria for depression currently. Please continue to screen for depression throughout the patient's stroke recovery.

At 9 months after stroke, if a patient's answer to EITHER question 1 OR to question 2 is in a grey box, this patient is at high risk for depression and should be referred to a qualified mental health professional for a diagnostic assessment for depression.

Figure 12

PSD screening questionnaire for use at 12 months after a stroke

12 months after stroke, please answer the following questions:

For each question, please give the one answer that comes closest to the way you have been feeling:

1. Do you feel full of energy?	YES	NO	
2. Please indicate which statement best describes your own health state today.	I am not anxious or depressed	I am moderately anxious or depressed	I am extremely anxious or depressed

Answers

At 12 months after stroke, if a patient's answer EITHER question 1 OR question 2 is in a white box, this patient likely does not have depression currently.

At 12 months after stroke, if a patient's answer to BOTH question 1 AND to question 2 are in a grey box, this patient is at high risk for depression and should be referred to a qualified mental health professional to be assessed for a diagnosis of depression.

Endnotes

[1] Hacket & Pickles (2014) report that the 31% rate for PSD was likely an under-estimation of the true frequency, which they attribute to methodological limitations.

[2] Selective serotonin reuptake inhibitors (SSRIs), commonly prescribed for the treatment of PSD, have been associated with increased risk for mortality, stroke / transient ischemic attack and other risks in older adults (Coupland et al., 2011; Wu, Wang, Cheng, & Gau, 2011). In stroke survivors, pharmacological treatment has been associated with bladder, bowel, and cardiac problems (Hinkle, 1998), leucopenia and agranulocytosis (Coulter & Edwards, 1990), increased bleeding (DeAbajo, Garcia, & Montero, 1999), and increased risk of hip fractures (Liu et al., 1998) for which stroke patients are vulnerable due to their increased incidence of falls (Carod-Artal, Egido, González, & Varela de Seijas, 2000).

[3] A stroke is a sudden loss of brain function caused most frequently by the interruption of flow of blood to the brain (ischemic stroke) and less frequently by the rupture of blood vessels in the brain (haemorrhagic stroke).

[4] The classification of DSM-5 disorders is harmonized with the ICD classifications, so that disorders listed in the DSM-5 include codes ICD-9-CM and ICD-10-CM codes (APA, 2013).

[5] The Canadian Stroke Network has over 100 researchers at 24 universities across Canada. http://canadianstrokenetwork.ca/en/

[6] The CSBPR guidelines are aimed at increasing capacity for service delivery, promoting up-to-date evidence-based care, reducing variations in care, and to reducing the gap between knowledge and clinical practice.

[7] The most recent update (Lanctôt, Lindsay, Smith, Sahlas, Foley, et al., 2019) was published at the time of submission of this dissertation. References to the new update have been retrospectively included throughout the document where possible.

[8] To contribute to initiatives aimed at empowering patients in their own care, the epidemiology laboratory at McGill University has already designed simple screening tools for home use by patients with Multiple Sclerosis.

[9] Stroke Improvement Programme website: https://improvement.nhs.uk/resources/transforming-stroke-care-pathway/

[10] Burton and Tyson (2015) systematically reviewed studies investigating accuracy of screening tools for depression in stroke patients. The authors identified 27 tools, including 15 verbal self-report tools, seven tools incorporating visual aids, and five observational tools.

[11] van Dijk, de Man-van Ginkel, Hafsteinsdóttir, and Schuurmans (2016) report examples of methods used in studies to get around this language problem. The authors recommend methods involving informants completing the interview on the patient's behalf (e.g., spouse of children) as well as use of non-language-based external validators.

[12] Appropriate patients include patients represented by the sample and do not include patients that meet the exclusion criteria.

[13] Such as organ failure or malignancy.

[14] Study investigators/collaborators included Dr. Miller PhD; Dr. Leslie Fellows MD FRCP; Dr. Lisa Koski PhD; Dr. Robert Cote MD FRCP; Dr. Ronald Ludman MD; and Rosa Sourial MSc (A)

[15] The STARD checklist provides a measure for assessing a study's "Standards for Reporting Diagnostic accuracy studies".

[16] In this study, depression was assessed with criterion A and B for current MDE in the SCID-I/NP interview. These criteria correspond with criterion A and B from the DSM-5 (APA, 2013). SCID-I/NP criteria C and D for a current MDE were not assessed. Criterion C of the SCID-1/NP corresponds with criterion C of the DSM-5 and rules out the possibility that the mood episode can be better explained by the direct physiological effects of substance (e.g., drug of abuse, medication) or a general medical condition (e.g., stroke). Criterion D of the SCID-I/NP rules out the possibility that symptoms be better explained by bereavement. This criterion has been removed from the DSM-5. This author argues that screening tools for PSD should identify all patients with clinically significant mood disturbances. It is not the job of a screening tool to tease apart complex etiological factors (e.g., criterion C and D of the SCID-I/NP). This can be the job of professional assessment in step two of the assessment process.

[17] The DSM-5 MDE criteria specify that symptoms must be present nearly every day for a two-week period. Unless symptoms precipitated the stroke, this criterion cannot be met for patients who are less than two weeks post stroke. In this study, in the acute phase post stroke (at 10 days) the 14-day criterion was waived. This choice is further discussed in the discussion of this dissertation.

[18] DSM-5 diagnostic criteria for depressive disorder due to another medical condition, such as stroke, with depressive features as the specifier require clinically significant depressed mood or loss of interest and pleasure, and does not require other symptoms. This author agrees that in the context of stroke, presence of depressed mood or loss of interest or pleasure are sufficient to warrant attention of a medical professional. This diagnosis also requires evidence that symptoms are due to the direct pathophysiological consequence of stroke. This study did not identify whether depressed symptoms were directly caused by stroke and describes existing research which questions the feasibility and value of making such an etiological inference.

[19] This measure is also referred to as the RAND 36-item Health Survey 1.0 subscale for mental health (Hays, Sherbourne, & Mazel, 1993).

[20] Items one and two are thought to tap anxiety while items three through five are thought to tap depression.

[21] The interviewer rates patients' actual, not potential, functioning over the preceding 24 to 48-hour period, on 10 items, each representing one of 10 activities of daily living: feeding, bathing, grooming, dressing, bowel and bladder continence, toileting, transfers, mobility, and stair use.

Total scores for the Barthel Index are calculated on an ordinal scale from '0', indicating functional dependence, to '100', indicating functional independence. Each self-care item is rated on a three-point scale indicating if the patient can perform the activity independently, with assistance or supervision, or not at all as well as the amount of assistance and time required. Items carry variable weights (0, 5, 10 or 15). It also has been found to correlate highly with other indices of ADL in stroke populations, supporting concurrent validity (Balu, 2009). The Barthel Index has been frequently used with stroke patients (e.g., Gibson, MacLennan, Gray, & Pentland 1991; Shah, Vanclay, & Cooper, 1989; Carter, Oliveira, Duponte, & Lynch, 1988; Wade, Skilbeck, & Hewer, 1983) and is used to indicate the amount of nursing care needed (Granger & Hamilton, 1990).

[22] Non-parametric tests were also preferred as they can be considered more appropriate when assumptions of homogeneity of variance are violated.

[23] Fisher's exact test is valid for all sample sizes. Fisher's exact test was applied as a default because significance of the deviation from a null hypothesis can be calculated precisely, rather than relying on an approximation that increases in accuracy as sample size increases, as with the chi-square test. While traditionally it is considered necessary to use Fisher's exact test when assumptions required for chi-squared are not met (more than 20% of cells having frequencies below five), with improved technology, Fisher's exact test can be efficiently applied with larger sample sizes.

[24] Sp for MDD was not examined because it does not make sense to correctly classify people with less severe depression as not depressed.

[25] Statistical significance with 95% confidence was not required because reducing type II error was prioritized over reducing type I error.

[26] QUAID (question-understanding aid) is a software tool that assists survey methodologists, social scientists, and designers of questionnaires in improving the wording, syntax, and semantics of questions. Website: http://quaid.cohmetrix.com/

[27] Missingness due to lack of a study coordinator occurred at random (non-systematically), and resulting missing data is considered missing completely at random, thereby not biasing results.

[28] The Satterthwaite t-test provides better estimates for group differences where variances between groups are not equivalent, and it provides equivalent answers as pooled t-tests when the variances in each group are equivalent.

[29] Mean (SD) length of hospital stay (LOS) for all participants was 18 (15.7) days. The distribution of LOS was positively skewed and non-normal with significant univariate outliers.

[30] Younger participants were significant outliers (skewness= -0.78, kurtosis = 0.83).

[31] Depression status of participants could not be determined due to missing data if the participant was missing at least one SCID-I/NP interview and never met criteria for any depression at any time point.

[32] Time one of the SCID-I/NP administration occurred at 10 days post stroke. The DSM-5 two-week criterion for symptoms was waived for this timepoint and symptoms were considered endorsed if present since the time of stroke (10 days).

[33] The model used to generate values must also be correct and must match up with the model used in analyses (Allison, 2003).

[34] Using multiple imputation with non-normal data has been found to produces accurate parameter estimates and standard errors with relatively large sample sizes (i.e., 400+) (Demirtas, Freels, & Yucel, 2008).

[35] Negative Predictive Value (NPV), expressed as a percentage, provides a partial estimate of post-test risk and is the probability that subjects with a negative screening test result truly does not have the condition. It is included with PPV in the calculation for RR. NPV is calculated as follows: NPV = TN / (FN + TN)