Table 1A: Selected Validated Screening and Assessment Tools for Post-Stroke Depression

This table provides a summary of 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. This list is not exhaustive, rather it highlights the more commonly used and validated tools. It is recommended that these tools be considered as first line options for all stroke services. (Table completed by Katherine Salter, PhD candidate with thesis research in Post-Stroke Depression).

Assessment Tool and Link	# of Items	Response Format	Total Score	Stroke-specific reliability/validity	Interpretation of Scores*	Sensitivity/Specificity for PSD
Recommended First Line Tools						
Geriatric Depression Scale (GDS) http://web.stanford.edu/~yesavage/GDS.html	30	Self-report Yes/No responses	0-30	**Reliability**: Though thoroughly evaluated in populations of elderly individuals, relatively little has been done specific to individuals with stroke. Agrell and Dehlin (1989) reported high internal consistency ($\alpha=0.90$) as did Sivrioglu et al. (2009) ($\alpha=0.88$). **Concurrent Validity**: Agrell and Dehlin (1989) reported good correlations between GDS scores and scores on self-report and observational depression assessment scales. **Discriminative Validity**: Sivrioglu et al. (2009) demonstrated significant differences in GDS scores between groups of depressed vs. non-depressed participants ($p<0.001$).	Normal = 0 – 10, scores ≥11 indicate presence of depression; 11-20 = mild depression, 21-30 = moderate to severe depression (McDowell et al. 1996)	Many studies have examined the relative sensitivity and specificity of the GDS – most have reported sensitivity and specificity values > 80% (Stiles and McGarrahan (1998). Within the stroke population, Johnson et al. (1995) using a cut-off of 10/11, Johnson et al. (1995) reported sensitivity = 85%, specificity = 66% and a misclassification rate of 29%. More recently, using DSM-IV-TR as the criterion for diagnosis, Sivrioglu et al. (2009) reported sensitivity = 69% & specificity = 75% for using a cutoff point of 10/11, and sensitivity = 66% and specificity = 79% for a cut off of 11/12 .
Hospital Anxiety and Depression Scale (HADS) http://www.strokengine.ca/assess/hads/	14 (2 x 7-item sub-scales)	Self-report Multiple choice response options graded on a 4 pt scale	0-42 (0-21 for each subscale)	**Reliability**: Visser et al (1995) reported test retest reliability (0.87); reported internal consistency reliability for the depression portion of the HADS has been >0.70 (Johnston et al. 2000, Aben et al. 2002); most recently Sagen et al (2009) reported $\alpha=0.83$. **Construct validity**: Reported satisfactory on confirmatory factor analysis (Johnston et al. 2000). **Discriminative validity**: HADS-D and HADS-A scores obtained by stroke	Scale authors recommended either 8/9 (high sensitivity) or 10/11 (high specificity) be used to identify the presence of depression using the depression subscale of the HADS (Zigmond and Snaith 1983).	Aben et al. (2002) reported sensitivity of 72.5% and specificity of 78.9% for the HADS-D, using a cut-off score of ≥7. For the total scale, using a cut-off of ≥11, sensitivity and specificity were 86.8% and 69.9% respectively. Johnson et al. (1995) used a cut-off of 4/5 for the HADS-D and demonstrated a sensitivity of 93% and specificity of 44% while O-Rourke et al. (1998) reported sensitivity of 80% and specificity of 79% using the same cut-off point as Aben et al. More recently, Sagen et al. (2009) reported

Assessment Tool and Link	# of Items	Response Format	Total Score	Stroke-specific reliability/validity	Interpretation of Scores*	Sensitivity/Specificity for PSD
				patients differed significantly from controls (p<0.001) (Visser et al. 1995).	Alternate cut-off points have been evaluated for the post stroke population.	sensitivity and specificity for the HADS-total (relative to the DSM-IV) of 90% and 83% (cut off of ≥11), 79% and 85% (cut off of ≥12) respectively. For the HADS-D, sensitivity = 79% and specificity = 82% (cut off of ≥5). AUC for HADS-D was 0.87 (95% CI 0.78-0.96) and for HADS-total 0.91 (95% CI 0.85-0.97) (Sagen et al. 2009)
Patient Health Questionnaire -9 (PHQ-9) http://strokengine.ca/assess/module_phq9_intro-en.html http://www.phqscreeners.com/	9	Multiple choice response options, 4pt scale	0-27	**Reliability**: Inter-rater reliability = 0.98, test re-test = 0.75 and internal consistency = 0.79 (de Man-van Ginkel et al. 2012). **Concurrent validity**: PHQ-9 was significantly correlated with GDS-15 scores (r=0.8, p<0.01) (de Man-van Ginkel et al. 2012).	Scores ≥10 (sensitivity=80%, specificity=78%) for identification of PSD 6-8 weeks post stroke (deMan van Ginkel et al. 2012)	A single study evaluated the sensitivity and specificity of the PHQ-9 for both major depression and any depression against a structured clinical interview in a subgroup of outpatients with stroke who endorsed either 2 or more symptoms on the PHQ-9 or either of the PHQ-2 items at study baseline (Williams et al. 2005). The authors reported sensitivity of 91% and specificity of 89% for major depression as well as sensitivity of 78% and specificity of 96% for any depression associated with a cut-off score ≥10. These numbers may, however, have been influenced by the pre-screening (using items from the PHQ-9) and formal assessment of selected individuals only. De Man-vanGinkel et al. (2012) also reported the results of a validation study that evaluated the PHQ-9 against the results of a composite international diagnostic interview for the DSM-IV conducted with 164 individuals with stroke (outpatients approximately 6-8 weeks post stroke). Similar to Williams et al., the authors reported that the accuracy of the PHQ-9 was best using a cutoff of ≥10 with a sensitivity of 80% and specificity of 78%. Using the PHQ-9 in patients pre-screened with the PHQ-2 increased the accuracy of identification (sensitivity = 87%) (de man-van Ginkel et al. 2012).
Additional Tools for Consideration						
Beck Depression Inventory (BDI-II) http://strokengine.ca/assess/module_bdi_intro-en.html	21	Self-report Multiple-choice response set graded for severity	0-63	**Reliability**: Aben et al. (2002) confirmed high internal consistency reliability of the BDI in a population of individuals with stroke. Outside of the stroke population estimates of internal consistency tend to exceed 0.80	Threshold for presence of depression = 10; 10 – 18 = mild depression, 19 – 29 = moderate	ROC analysis completed by Lincoln et al. (2003) suggests that the accepted cut-off point indicative of presence of depression might be too low – recommends 15/16 to optimize sensitivity; however specificity is reduced relative to the DSM-III-R. Aben et

Assessment Tool and Link	# of Items	Response Format	Total Score	Stroke-specific reliability/validity	Interpretation of Scores*	Sensitivity/Specificity for PSD
http://www.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=015-8018-370				(Beck et al. 1988) **Predictive validity**: BDI scores are predictive of functional recovery and need for institutional care following stroke (Kotila et al. 1999, Desrosiers et al. 2002).	depression, 30 – 63 = severe depression (Beck et al. 1988)	al. (2002) reported the standard cut-off points to be acceptable for used for individuals with stroke.
Center for Epidemiological Studies Depression Scale (CES-D) http://cesd-r.com/	20	Self-report 4-pt scale	0-60	**Reliability**: Internal consistency reliability has been reported ranging from 0.64-0.86 (Agrell & Dehlin 1989, Toedter et al. 1995). Reported item-to-total correlations ranged from 0.39-0.75 (Shinar et al. 1986). **Concurrent validity**: Results of the CES-D used to assess individuals with stroke have correlated significantly with results of other standardized self-report and observational depression assessment tools (Agrell and Dehlin 1989, Shinar et al. 1986, Parikh et al. 1988)	Presence of depression = ≥16 (McDowell et al. 1996)	Using the suggested cut-off score, Shinar et al. and Parikh et al. reported sensitivity of 73% and 86%, and specificity of 100% and 90% respectively (relative to the DSM-III-R).
Tools to Consider for Aphasic Patients						
Stroke Aphasic Depression Questionnaire-10 (SADQ-10) http://strokengine.ca/assess/module_sadq_intro-en.html http://www.nottingham.ac.uk/medicine/about/rehabilitationageing/publishedassessments.aspx	10	Observer rating of observed behaviour 4-point scale	30	**Reliability**: Using carers of individuals with aphasia to complete follow-up assessments, 4-week test-retest reliability was reported to be 0.69 for the SADQ-10 (Sutcliffe and Lincoln 1998). Internal consistency has been reported as α = 0.80 (Sutcliffe and Lincoln 1998, Lincoln and Sutcliffe 2000). **Construct validity**: Results of factor analysis suggested that the SADQ-10 items may be unidimensional (Sutcliffe and Lincoln 1998) **Concurrent validity**: SADQ-10 scores have been positively associated with scores on the HADS-D, HADS-A, Wakefield Depression Inventory (Sutcliffe and Lincoln 1998), and the GDS-15 (Leeds et al. 2004), though correlations with healthcare professional ratings have varied (Lincoln and Sutcliffe 2000).	Scores ≥ 15 may represent presence of depression (Leeds et al. 2004).	Using the suggest cut-off score of ≥15, Leeds et al. (2004) reported sensitivity = 70% and specificity = 77% in a group of stroke rehabilitation inpatients. Based around cut-offs used for the HADS, Bennett et al. (2006) identified a cut-off of 17/18 on the SADQ-H (sensitivity= 100% and specificity=81%), and an optimum cut-off of 5/6 on the SADQ-H 10 (sensitivity = 100% and specificity = 78%).

Assessment Tool and Link	# of Items	Response Format	Total Score	Stroke-specific reliability/validity	Interpretation of Scores*	Sensitivity/Specificity for PSD
Aphasia Depression Rating Scale (ADRS) http://strokengine.ca/assess/module_adrs_intro-en.html	9	Observer rating based on interview & observation Rating scale varies per item	0-32	**Reliability**: Test retest reported to be 0.89 by scale authors. Interobserver reliability = 0.89 (Benaim et al. 2004). **Concurrent validity**: ADRS scores were correlated with CAS ratings and with results of HRSD (Benaim et al. 2004).	Scores of ≥ 9 are used to indicate the presence of depression (Benaim et al. 2004).	Using the cut-off indicated as appropriate by the scale author, sensitivity of 83% and specificity of 71% were reported (relative to a psychiatric diagnosis) (Benaim et al. 2004).
Tools for Consideration in Children						
Children's Depression Inventory (CDI) http://www.mhs.com/product.aspx?gr=edu&id=overview&prod=cdi2#description (The CDI 2 has been recently released but test details are not available free of charge)	27	Self-report 3 pt scale	0-54	The psychometric properties of this scale have not been investigated within a stroke-specific population.	Scores of ≥ 19 have been identified as representing the 90th percentile within a general population of children in grades 3-9 (Smucker et al. 1986).	n/a
Kidscreen 52 (Generic HRQL measure) http://www.kidscreen.org/	52	Self-report 5 pt scale	Scores for each dimension are calculated as T-values (mean=50; SD=10).	The psychometric properties of this scale have not been investigated within a stroke-specific population.	Higher scores indicate higher Health-Related Quality of Life and well-being.	n/a

- It should be emphasized that a score indicating depression on a screening tool is not equivalent to a diagnosis of depression. Rather, a positive score indicates the need for further follow-up and assessment
- **more detailed review of these screening tools may be obtained via the ebrsr.com, strokengine.com or in Salter et al. (2007)