Outcome Measure	**Hamilton Rating Scale for Depression (Ham-D)**
Sensitivity to Change	No
Population	Adult
Domain	Psychological Status
Type of Measure	Clinician-rated scale
ICF-Code/s	d1
Description	The Hamilton Rating Scale for Depression (HDRS, also known as the Ham-D) is reportedly the most widely used clinician-administered depression assessment scale. The original version contains 17 items (HDRS17) which asks about symptoms of depression experienced over the past week. Although the scale was designed for completion after an unstructured clinical interview, there are now semi-structured interview guides available. The HDRS was originally developed for hospital inpatients, thus the emphasis on melancholic and physical symptoms of depression. A later 21-item version (HDRS21) included 4 items intended to subtype the depression, but which are sometimes, incorrectly, used to rate severity. Numerous versions with varying lengths include the HDRS17, HDRS21, HDRS29, HDRS8, HDRS6, HDRS24, and HDRS7. For the HDRS17, a score of 0–7 is generally accepted to be within the normal range (or in clinical remission), while a score of 20 or higher (indicating at least moderate severity) is usually required for entry into a clinical trial. For each item, clinicians are asked to identify the one "cue" which best characterizes the patient. The response options vary from item to item. Some items are on a scale from 0 to 4 (e.g., Item 1 – Depression Mood, has the response options from *0 = Absent* to *4 = Patient reports virtually only these feeling states in his/her spontaneous verbal and non-verbal communication*). Others are on scale from 0 to 2 (e.g., Item 4 – Insomnia, has the response options *0 = No difficulty falling asleep* to *2 = Complains of nightly difficulty falling asleep*). Administration time is 20-30 minutes for the 17-item measure. Exert from Bagby et al (2004): Although Hamilton (1) explained the rationale for the inclusion of both 3-point and 5-point items, the argument was not made on the grounds of differential weighting. Hamilton believed that certain items would be difficult to anchor dimensionally and therefore assigned them fewer response options. The end result is that certain items contribute more to the total score than others. Contrasting psychomotor retardation and psychomotor agitation, for example, reveals that a severe manifestation of the former contributes 4 points, whereas an equally severe manifestation of the latter contributes 2 points. Similarly, someone who weeps all the time can contribute 3 or 4 points on depressed mood, whereas someone who feels tired all the time can contribute only 2 points

	on the general somatic symptoms item.
Properties	There is a plethora of information in the psychometric properties of this scale, very few in TBI. The psychometric properties of the scale are summarized by Bagby et al. (2004) in their meta-review. Except where indicated, their summaries are noted below. Construct validity: Unidimensionality of the scale has been examined using both IRT and Rasch analysis. These studies indicate that several items on the scale are not consistent with a unidimensional construct. Gibbons et al. (1993) in their IRT analysis used a subset of Ham-D items to create a unidimensional scale. This subset included depressed mood, which was sensitive at low levels; work/interests, psychic anxiety, and loss of libido, which were sensitive at mild levels; somatic anxiety, psychomotor agitation, and guilt, which were sensitive at moderate levels; and suicide, which was sensitive at severe levels. Similarly, Bech et al. (1981) used Rasch analysis to create a 6-item shorter scale composed of items for depressed mood, guilt, work/interests, psychomotor retardation, anxiety psychic, and general somatic symptoms. This shortened scale has been confirmed in two subsequent Rasch analysis studies (Maier & Philipp 1988; Maier, Heuser, et al. 1988). Internal Consistency: Cronbach's alpha estimates for the overall scale have ranged from 0.46 to 0.97, and 10 of 13 studies reported estimates above .70. Inter-rater reliability: Pearson's *r* estimates have ranged from .65 to .98. Nine of 11 studies have reported inter-rater reliability to be above .82. Intraclass correlation coefficients have ranged from .46 to .99. Ten of 17 studies had interclass *r* coeffients above .80. Test-retest reliability: Estimates have ranged from .81 to .98 (4 studies). Convergent validity: Correlations with the BDI range from .27 to .85. Correlations with the Brief Psychiatric Rating Scale range from .56 to .89. Concurrent/Predictive validity: A recent Brazilian study in TBI individuals found that the Ham-D discriminates well between those individuals with and without depression (Schwarzbold et al., 2014). Sensitivity and specificity: In the Brazilian TBI study, score of 7 or more led to a sensitivity of 92.9% and specificity of 78.1% (Schwarzbold et al., 2014).
Advantages	• The Hamilton depression scale's internal reliability is adequate • Convergent validity and discriminant validity are adequate
Disadvantages	• Scale was not developed specifically for TBI. Few studies in TBI samples, but has been used in one recent intervention study for depression after TBI (Fann et al., 2015), without showing positive

	treatment effects. • Many scale items are poor contributors to the measurement of depression severity; others have poor interrater and retest reliability. • For many items, the format for response options is not optimal. • Content validity is poor. • The factor structure of the Hamilton depression scale is multidimensional but with poor replication across samples. • Bagby et al. (2004) summarises his findings by suggesting that there is evidence suggests that the Hamilton Depression Scale is psychometrically and conceptually flawed.
Additional Information	
Reviewers	Jennie Ponsford

References

Bagby, R.M., Ryder, A.G., Schuller, D.R., Marshall, M.B (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight. *American Journal of Psychiatry, 161 (12),* 2163-2177

Fann, J.R., Bombardier, C.H., Vannoy, S., (2015). Telephone and in-person cognitive behavioural therapy for major depression after traumatic brain injury: a randomized controlled trial. *Journal of Neurotrauma, 32*, 45-57

Bech, P. (1981). Rating scales for affective disorders: their validity and consistency. *Acta Psychiatr Scand Suppl*, 295, 1–101.

Gibbons, R. D., Clark, D. C., Kupfer, D. J. Exactly what does the Hamilton Depression Rating Scale measure? *J Psychiatr Res, 27*, 259–273

Hamilton, M. (1960). A rating scale for depression*. J Neurol Neurosurg Psychiatry, 23*, 56–62

Maier, W., Philipp, M., Heuser, I., Schlegel, S., Buller, R., Wetzel, H.(1988). Improving depression severity assessment, I: reliability, internal validity and sensitivity to change of three observer depression scales*. J Psychiatr Res, 22*, 3–12.

Maier, W., Heuser, I., Philipp, M., Frommberger, U., Demuth, W. (1988). Improving depression severity assessment, II: content, concurrent and external validity of three observer depression scales. *J Psychiatr Res, 22*, 13–19.

Schwarzbold, M. L., Diaz, A. P., Nunes, J. C., Sousa, D. S., Hohl, A., Guarnieri, R., . . . Walz, R. (2014). Validity and screening properties of three depression rating scales in a prospective sample of patients with severe traumatic brain injury. *Revista brasileira de psiquiatria*(AHEAD), 000-000.