# The Depression Inventory Development Scale: Assessment of Psychometric Properties Using Classical and Modern Measurement Theory

A. Vaccarino[1], A. Kalali[2], S. Kennedy[3], R. Lam[4], R. Milev[5], R. Uher[6], P. Blier[7], F. Farzan[8], J. Foster[9], B. Frey[9], S. Gilbert Evans[1], F. Leri[10], G. MacQueen[11], D. Mueller[12], S. Parikh[13], S. Rotzinger[14], C. Soares[5], S. Strother[15], G. Turecki[16], K. Evans[1]

[1]Indoc Research, Toronto, Canada, [2]Dept Psychiatry, University California, [3]Dept Psychiatry, University Toronto, [4]Dept Psychiatry, University of British Columbia, [5]Dept Psychiatry, Queen's University, [6]Dept Psychiatry, Dalhousie University, [7]Dept Psychiatry, University Ottawa, [8]Simon Frazer University, [9]St Joseph's Healthcare, McMaster University, [10]Dept Psychology, Guelph University, [11]Hotchkiss Brain Institute, University of Calgary, [12]CAMH, [13]University of Michigan, [14]Dept Psychiatry, University Health Network, [15]Rotman Research Institute, Baycrest Centre, [16]Dept Psychiatry, McGill University

Introduction

The Depression Inventory Development (DID) initiative[1] aims to develop a comprehensive rating scale for MDD with more favourable psychometrically properties than current "gold standards." Using an iterative process between field testing and psychometric analysis, an empirically-driven and collaborative protocol for item development was established. We report here results of latest iteration of DID items to assess symptoms related to *anhedonia, cognition, fatigue, general malaise, motivation, anxiety, negative thinking, pain, sleep and appetite.*
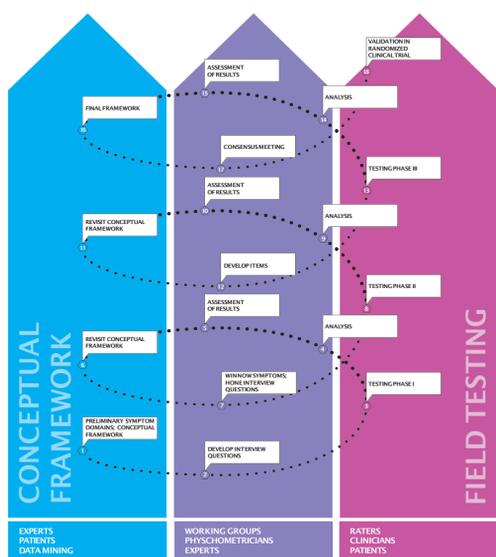
Methods

The present study was part of a larger multi-site, open label study conducted by the Canadian Biomarker Integration Network in Depression (CAN-BIND).[2] Participants were administered 32 DID items using a semi-structured interview at each of two visits (MDD: Baseline, n = 211 and Week 8, n = 177; Healthy comparators: Baseline, n = 112 and Week 8, n = 104). Classical test theory, item response theory and Rasch measurement theory were applied to assess the psychometric properties of the DID items and determine which should be removed, modified or advanced. Participants were also administered the MADRS and QIDS-SR that allowed DID items to be evaluated against existing "benchmark" items. Data were managed on the Brain-CODE platform[3]

Results

Baseline demographic and clinical characteristics

	MDD (n=211)	Healthy Controls (n=112)
Female, n (%)	133 (63.0%)	71 (63.4)
Age in yrs, mean $\pm$ SD (range)	35.30 + 12.65 (18-61)	33.04+10.74 (18-60)
MADRS $\pm$ SD (range)	29.85+5.61 (19-47)	0.76+1.67 (0-10)
QIDS-SR $\pm$ SD (range)	15.76+4.11 (3-26)	2.28+1.98 (0-12)

Classical Test Theory

The 32 DID items were examined for data quality using the following CTT acceptability criteria:

- Missing data <10%
- Endorsement of zero or maximum option score <80%
- Endorsement of any single option <50%
- Aggregate frequency endorsement of adjacent options >10%
- Skewness < |2|
- Inter-item correlations/redundancy r<0.75

Based on CTT analyses, 8 items were removed from the item bank because of low variability in the range of response options, floor effects and/or skewedness: *Increased Appetite, Decreased Eating, Increased Eating, Headaches, Health-Bothered, Health-Intensity, Prolonged Sleep and Weakness.*

References

1. Vaccarino *et al.* Innov Clin Neurosci (2016) 13:20
2. Lam *et al.* BMC Psychiatry (2016) 16:105
3. Vaccarino *et al.* Front Neuroinformatics (2018) 12:28
4. Samejima. Psychometrika Monographs (1969) 34 (Suppl. 4)

Item Response Theory

IRT was used to assess an item's ability (at the individual option level) to discriminate levels of depressive severity. Individual DID items were analysed separately against the ten MADRS items using Samejima's Graded Response Model[4] in IRTPro Version 3.0. Option Characteristic Curves (OCCs) were generated to display the probability of selecting a particular option of a DID item as a function of overall depressive severity (MADRS). Discriminatory/slope (denoted as "a") provided quantitative measures of item discrimination, with slope parameters of $a < 0.65$ used as an indication of poor discrimination.

OCCs (smooth lines, left axis) and IICs (dotted lines, right axis) for DID items (A) showing good, (B) moderate, and (C) poor discriminative properties.
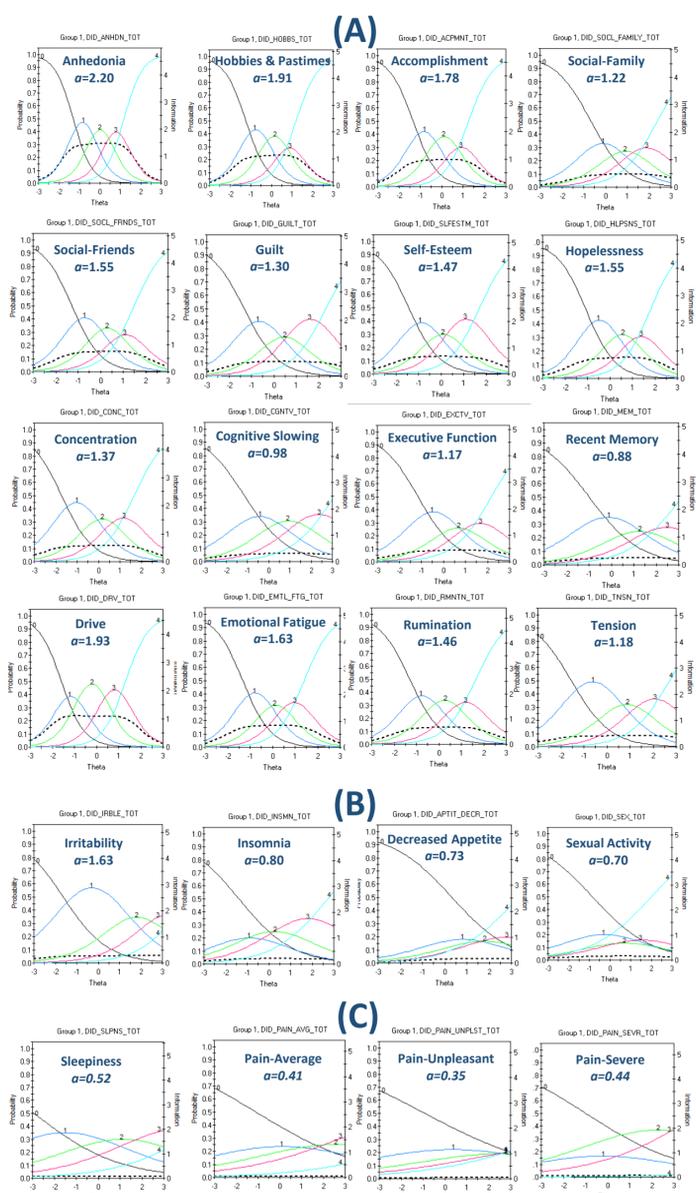
Rasch Measurement Theory

RMT (in WinSteps 4.2) was used to assess performance of the remaining items together as a scale. Person-item maps were generated that plot persons (individual participants) and items on a single continuum to compare the range and position of the person measure distribution to that of the item measure.

Person-item location for DID, MADRS and QIDS-SR items
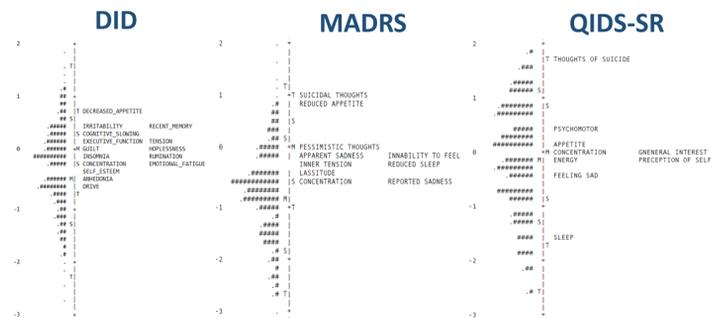
Additional DID Analyses

- Good internal consistency (α=0.95)
- Item-total correlations ranging from r = 0.49 (Decreased Appetite) to r = 0.84 (Drive)
- All items showed sensitivity to anti-depressant treatment, with effects sizes ranging from $n_p^2 = 0.10$ (Decreased Appetite) to $n_p^2 = 0.48$ (Anhedonia)

Conclusions

CTT, IRT and RMT analyses were used to assess the performance of the DID items and identify which would be advanced, modified or removed. Fifteen DID items are recommended for inclusion in the penultimate scale, with modifications as required.

Acknowledgements