

Development and Preliminary Psychometric Evaluation of the Motivation and Energy Inventory – Short Form (MEI-SF)

S.E. Fehnel and L.D. McLeod, RTI Health Solutions, RTI International
 H.M. Edin and S.L. Hogue, Global Health Outcomes North America, GlaxoSmithKline

Introduction

Because depressed patients commonly experience reductions in motivation and energy, both as symptoms of their disorder and as side effects of pharmacotherapy, it is important to identify treatments that can restore their vitality. The Motivation and Energy Inventory (MEI), a 27-item questionnaire utilizing a 4-week reference period, was developed to facilitate such efforts. The constructs addressed by the MEI were identified through a combination of literature review, consultation with experts, and three patient focus groups. Item wording and response scales were then tested and modified through four iterative sets of cognitive interviews conducted with 32 additional patients.

After initial development of the MEI, the data from two clinical trials, as well as a separate test-retest study conducted by RTI-HS, were utilized to evaluate the psychometric properties of the MEI. Consistent with its measurement model, the psychometric results confirmed that the 27-item instrument has three factors generally addressing physical energy, mental energy, and social motivation. Furthermore, these results provided strong evidence for the internal consistency, test-retest reliability, validity, and responsiveness of all three MEI subscales. While the psychometric soundness and utility of the 27-item MEI have been demonstrated, it seemed that the availability of an alternative form, perhaps with fewer items and a shorter reference period, would facilitate the use of this instrument in future studies.

The objectives of the current project, therefore, were to create and conduct a preliminary psychometric evaluation of an acute, short form of the MEI, capable of both reducing patient burden and maximizing responsiveness to treatment effects.

Methods

Item Reduction

To identify candidate items for deletion from the questionnaire, three steps were taken, utilizing data from the two aforementioned clinical trials.

- Pairwise t-tests were conducted, comparing changes on each MEI item for responders and nonresponders, with the goal of identifying items that were not optimally sensitive to differences between these two groups. Responders were subjects whose Hamilton Rating Scale for Depression (HAM-D; Hamilton, 1960) scores decreased by 50% or more.

- Changes in item-level scores were correlated with changes in HAM-D scores for both groups (responders and nonresponders), as well as the entire sample, to identify any items only weakly related to changes in the severity of depressive symptoms.

- Inter-item correlations were examined to locate item pairs that were so highly correlated they could be considered redundant.

Preliminary Psychometric Evaluation

After the final 18-item set was identified, a preliminary investigation of its reliability, validity, and responsiveness (as administered within the context of the original MEI) was then conducted utilizing existing data from the two clinical trials and the test-retest study.

- Factor Structure:** Iterative exploratory factor analyses were conducted to determine the optimal subscale structure using the baseline data from both clinical trials.

- Internal Consistency:** Cronbach's alpha was computed for the summed score of the 18 items using both baseline and end-of-study data from all three studies.

Test-Retest Reliability:

- Intra-class correlations (ICCs) and Pearson correlations (*r*) were computed using the data from the test-retest study.
- For comparison, reliability estimates for the Center for Epidemiologic Studies – Depression Scale (CES-D; Radloff, 1977) and the Medical Outcomes Study Short-Form 36 (SF-36) Vitality Subscale (Ware et al., 1993) were also computed.

Construct Validity:

- Using data from both clinical trials, correlations among baseline scores on the MEI-SF, Quality of Life in Depression Scale (QLDS; Hunt & McKenna, 1992), HAM-D, Hamilton Rating Scale for Anxiety (HAM-A; Hamilton, 1959), several work productivity measures, and Clinical Global Impressions – Severity (CGI-S; Guy, 1976) were examined.

Methods (continued)

- Using the data from the test-retest study, summed scores for these 18 items were correlated with scores on the CES-D and SF-36 Vitality Subscale at baseline (test) and the end of the study (retest).

Responsiveness:

- Changes in subscale scores for responders were compared to those for nonresponders using data from both clinical trials.
- A variant of Guyatt's responsiveness statistic (below) was used to compare changes in the MEI-SF scores between baseline and follow-up for responders and nonresponders (Guyatt, Walter, & Norman, 1987).

$$\frac{(\text{mean change score for responders}) - (\text{mean change score for nonresponders})}{\text{standard deviation of the change in scores of nonresponders}}$$

Results

Item Reduction

Taken together, the results of the pairwise t-tests examining the ability of each MEI item to distinguish between responders and nonresponders and the correlations between changes in MEI item-level scores and HAM-D scores suggested that the four items addressing the frequency of social contacts, social activities, recreational activities, and exercise were among the least sensitive to change in both trials. This finding is not surprising, because these four items address behavioral changes that tend to manifest themselves more gradually than mental or emotional changes, which are addressed in the remaining items. These items were subsequently removed from consideration for inclusion on the MEI-SF.

Baseline inter-item correlations for five pairs of items suggested at least moderate redundancy, ranging from .58 to .73. A decision was made to include only one item from each of these pairs on the MEI-SF. The "winning" items were chosen based on their performance in other analyses, such as those examining responsiveness. Ultimately, 18 items were chosen for inclusion on the MEI-SF. The content of these items is summarized in Table 1.

Table 1. Content of the MEI-SF

1. Felt enthusiastic when beginning your day
2. Satisfied with what you accomplished during the day
3. Had trouble getting out of bed because you didn't want to face the day
4. Had trouble finishing things because you lost interest
5. Felt overwhelmed, even by small tasks
6. Procrastinated or put things off until another day
7. Had problems concentrating
8. Had trouble making decisions
9. Avoided social conversations with others
10. Preferred to be alone
11. Had trouble keeping things organized
12. Felt physically tired during the day
13. Felt energetic
14. Felt motivated
15. Was interested in taking on additional tasks or projects
16. Was interested in meeting new people
17. Was interested in talking with others
18. Was interested in social activities like visiting friends, going out to dinner, parties

Note: While the reference period for the MEI is the preceding 4 weeks, the MEI-SF utilizes a 7-day reference period. All items use either a 5- or 7-point Likert-type response scale.

Results (continued)

Preliminary Psychometric Evaluation

Factor Structure

After deciding on the final 18 items, an iterative series of exploratory factor analyses were conducted to determine the optimal subscale structure. As with the full MEI, a three-factor solution appeared the most interpretable among the multidimensional options. The 18 items, however, did not produce a clean, consistent multifactor pattern. For example, the factors on which several items loaded differed between the two trials and another item loaded highly on two factors. On the other hand, every item loaded adequately (from .38 to .66) under the one-factor model. This is not surprising, because there is a high degree of relatedness among many of the individual items, as well as the three original MEI subscales. Based on these results, only a single MEI-SF score was computed for all remaining analyses.

Internal Consistency

Alpha coefficients for the 18 items chosen for the MEI-SF ranged from .86 to .94 at baseline and from .95 to .96 at the end of the three studies, providing strong preliminary evidence for the internal consistency of the MEI-SF.

Test-Retest Reliability

Table 2 displays estimates of test-retest reliability for the MEI-SF items. These preliminary results strongly support the stability of the MEI-SF total score and suggest that it is similar to that of the CES-D and the SF-36 Vitality Subscale.

Table 2. Test-Retest Reliability for the MEI-SF, the CES-D, and the SF-36 Vitality Subscale (N = 143)

Scale	ICC	r
MEI-SF	.73	.80
CES-D	.70	.80
SF-36 Vitality	.71	.79

Construct Validity

The correlations presented in Tables 3 and 4 indicate stronger relations between the MEI-SF and measures addressing similar constructs as compared to measures addressing more disparate constructs, providing preliminary evidence for the construct validity of the MEI-SF. For example, the relations between the MEI-SF and the measures addressing depression-specific quality of life (assessed by the QLDS), patient-reported depression severity (assessed by the CES-D) and vitality (assessed by the SF-36 Vitality Subscale) are the strongest among all those examined, both because of the similarity of constructs addressed and the fact that the persons who responded to all these measures were the patients themselves.

Table 3. Pearson Correlations Between the MEI-SF and Related Measures at Baseline in the Two Clinical Trials

Scale	Trial 1	Trial 2
QLDS Total	-.71	-.62
HAM-D	-.21	-.25
HAM-A	-.27	-.18
Work Effectiveness (1 item)	.37	.30
Work Productivity (3-item composite)	-.32	-.31
CGI Severity	-.21	-.13

Table 4. Pearson Correlations Between the MEI-SF and Related Measures in the Test-Retest Study

Scale	Test	Retest
CES-D	-0.80	-0.78
SF-36 Vitality Subscale	0.83	0.85

Results (continued)

Responsiveness

Table 5 displays descriptive statistics, as well as Guyatt's responsiveness statistic, an effect size comparing changes in the MEI-SF scores between baseline and follow-up for responders and nonresponders, for both clinical trials. Differences in summed scores for these 18 items are statistically significant ($p < .0001$) between the two groups and the responsiveness statistics are substantially higher than 0.80, the effect size Cohen (1977) suggested as representative of a large effect. Furthermore, the correlations between changes on the sum of the MEI-SF items and HAM-D scores in the trials were -.61 and -.62, respectively, providing further suggestion that the MEI-SF will be very sensitive to changes in depression severity.

Table 5. Changes in MEI-SF Scores for Responders vs. Nonresponders.

	Responders			Nonresponders			t-Value*	Effect Size
	N	Mean	SD	N	Mean	SD		
Trial 1	187	30.50	16.94	82	6.76	15.19	10.91	1.56
Trial 2	152	31.36	17.84	77	8.57	14.86	9.64	1.53

*Both t-tests significant at $p < .0001$.

Conclusions and Future Directions

While the psychometric properties of the MEI-SF as a stand-alone instrument, utilizing a one-week reference period, have yet to be established, the methods used to choose its items and the results of preliminary analyses suggest that this instrument may be of even greater utility than the original MEI, particularly for trials of short duration. A full psychometric evaluation of this short, acute form of the MEI will be conducted later this year.

References

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences* (p. 8). New York: Academic Press.

Guy, W. (1976). *ECDEU assessment manual for psychopharmacology* (United States Department of Health, Education, and Welfare Publication No. 76-338). Rockville, MD: National Institute of Mental Health, 218-222.

Guyatt, G. H., Walter, S. D., & Norman, G. (1987). Measuring change over time: Assessing the usefulness of evaluative instruments. *Journal of Chronic Diseases*, 40, 171-178.

Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32, 50-55.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry*, 23, 56-62.

Hunt, S. M., & McKenna, S. P. (1992). The QLDS: A scale for measurement of quality of life in depression. *Health Policy*, 22, 307-319.

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

Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey: Manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Center.

Presenter

Sheri Fehnel, PhD
 Global Head, Patient-Reported Outcomes
 RTI Health Solutions, RTI International
 3040 Cornwallis Road, PO Box 12194,
 Research Triangle Park, NC 27709-2194
 Phone: 919.541-7454 Fax: 919.541-7222
 E-mail: sfehnel@rti.org www.rtihs.org
 Presented at: ISOQOL 2004 Symposium
 June 2004 Boston, MA