Journal of Behavioral Medicine, Vol. 20, No. 2, 1997

Validating the MPI-DLV Using Experience Sampling Data

R. Lousberg,[1,4] A. J. M. Schmidt,[2] N. H. Groenman,[1] L. Vendrig,[2] and C. I. M. Dijkman-Caes[3]

Accepted for publication: July 17, 1996

The present study investigates the relationship between scale scores on the Dutch version of the Multidimensional Pain Inventory (MPI-DLV) and data derived from comparable items from an experience sampling procedure. Fifty-seven chronic pain patients participated in the study, which lasted 6 consecutive days. Special attention was given to the relationship between the MPI-DLV pain intensity score and the mean experience sampling pain intensity score. Significant correlations were found between the MPI-DLV scales Pain Severity, Interference, Solicitous, Punishing and Distracting Responses, and Household Chores and their experience sampling analogues. A marginally significant correlation was found with regard to the MPI-DLV Life Control scale. The General Activity and Affective Distress scales had no relationship with the analogous experience sampling items. The significant correlations were regarded as further validation of the MPI-DLV. A regression analysis revealed that 58% of the variance of the experience sampling pain intensity score could be explained by the MPI-DLV present pain intensity item score.

KEY WORDS: chronic pain; experience sampling; Multidimensional Pain Inventory, Dutch version; (MPI-DLV); validity assessment.

[1]Center for Pain Expertise, Academic Hospital Maastricht, Maastricht, The Netherlands.
[2]Department of Medical Psychology, University of Limburg, Maastricht, The Netherlands.
[3]Education Department, Academic Hospital Maastricht, Maastricht, The Netherlands.
[4]To whom correspondence should be addressed at Center for Pain Expertise, Academic Hospital Maastricht, P. Debeyelaan 25, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.

0160-7715/97/0400-0195$12.50/0

INTRODUCTION

The use of self-report questionnaires in multidisciplinary pain assessment has become increasingly popular, as evidenced by the rapidly growing number of self-report questionnaires measuring one or more pain related aspects and the fact that scores derived from questionnaires often form a major part of the effect variables in treatment evaluation studies.

One of the questionnaires often used in pain research is the Multidimensional Pain Inventory (MPI) (Kerns *et al.,* 1985). This list is designed to assess the psychosocial and behavioral aspects of pain. The psychometric properties (reliability and validity) are good. The MPI has recently been translated into German (MPI-D) (Flor *et al.,* 1990) and into Dutch (MPI-DLV) (Lousberg, 1994). Apart from some small differences, the factor structure of the American version could be replicated in both studies. Also, the reliability and validity analyses of the MPI-D and MPI-DLV have yielded satisfactory results. In addition, evidence is reported that these three MPI versions are sensitive to treatment effects (Flor *et al.,* 1990; Kerns *et al.,* 1986; Lousberg, 1994).

Although there are some items that ask for a judgment at that particular moment (e.g., "Rate the level of pain *at the present moment*"), the MPI can be considered to be a list that measures relatively stable behavioral and emotional aspects of pain (from the patient's perspective). For example, the MPI scale "Interference" measures the degree to which the pain generally interferes with the patient's daily life. In clinical practice, one of the objectives in the assessment phase of a chronic pain problem is, indeed, to get an idea of how a patient functions, thinks, and behaves in daily life. However, since the MPI is usually completed in a clinical setting, and since it is more or less stable/"trait" information that is obtained, inferences regarding the patient's behavior, pain experience, etc., in daily life cannot really be drawn. It is therefore important, from a clinical point of view, to acquire greater certainty as to the generalizability, ecological validity, of the MPI scale scores.

The results of the German study mentioned above (Flor *et al.,* 1990), have shed more light on this issue. Making use of diary data, it was found that the MPI-D Pain Intensity scale had a significantly positive correlation (0.61) with the mean diary pain intensity score (based on one week's data). A significant positive correlation (0.48) was also found between the MPI-D Interference scale and an analogous question in the diary. Correlations between the MPI-D activity subscales and the corresponding diary activity scales did not, however, reach significance; nor did the MPI-D General Activity scale correlate significantly with the diary general activity scale ($r = -.27$; n.s.). Apart from the MPI-D General Activity (sub)-scale(s), these

results can be seen as a further, ecological, validation of the MPI. This study has prompted at least two other research issues. First, it would be of value to know the connection between daily records and several other MPI scales. Second, the relationship between the MPI-D Pain Severity scale and the mean diary score was expressed as a correlation, indicating the strength of the relationship. It would be useful to obtain a better insight into the nature of this relationship.

It was decided to employ an experience sampling method (ESM) to investigate these two issues. ESM is a technique for obtaining a representative sample of moments in a person's daily life. The reliability and validity of this technique have already been demonstrated by Czikszentmihalyi and Larson (1987), and as argued by DeVries (1987), this method avoids the problem of general and retrospective recall (see also Clark, 1988). Another advantage of ESM is that anticipation at moments of measurement is prevented.

In the present study a watch was used that indicated to the patient (in a random time schedule) when to answer some questions in a booklet. Key items were created to represent/mirror the content of the MPI-DLV scales. It was predicted that the mean score of the experience sampling items would have a positive correlation with the corresponding MPI-DLV scale scores. Further, a regression analysis was carried out to investigate the nature of the relationship between the present pain estimate item on the MPI-DLV and the mean experience sampling pain intensity rating.

METHOD

Subjects

All patients were referred to the pain center at the Academic Hospital Maastricht. After a patient was seen by the neurologist he/she was asked to participate in the study if the following criteria were met: pain duration longer than 6 months, pain is not caused by cancer, age between 18 and 65, no alcohol or other drug dependence, having a significant other, absence of severe psychopathology, and no major difficulties with reading or writing.

Patients were told the purpose of the experience sampling experiment, and their informed consent was obtained. They were told that the experience sampling experiment would start the next day and would last 6 full days. They were then given instructions on how to handle the watch and what to do with the booklets, etc. They were told that they could turn off the watch's signal function (if, for example, they were in church) but that

it was important that the booklets were filled in completely. They were also assured that, should they have any questions or problems, they could phone the interviewer. The patients then completed the MPI-DLV. Finally, they were given six booklets (each for 1 day), the watch, and a set of instructions on both the experience sampling questions and the watch. An appointment was made for a debriefing session in which the watch and booklets should be returned. The briefing period took about 30 min.

During the debriefing session the booklets were checked for missing data. If there were any problems, these were discussed. Patients were also asked whether, during the 6 days of the experience sampling procedure, anything had happened that might have had a major influence on the experiment.

Fifty-seven chronic pain patients (26 males and 31 females) satisfied the inclusion criteria and gave their informed consent. The mean age was 42.33 (range, 21–64; SD = 10.0). The mean duration of pain was 7.1 years (range, 1–33 years; SD = 7.5 years). Table I presents the IASP primary pain-locus distribution.

Sampling Method

In most cases, the experience sampling experiment started on a Tuesday morning and lasted until Sunday night. A Seiko RC-1000 watch was used to indicate to the subjects when they should fill in/answer the items/questions. The signals (beeps) sounded eight times a day between 0830 and 2230, on a random time schedule. In addition, patients had to answer one question about pain severity at the moment they woke up and just before going to sleep. At the latter moment they had also to answer some questions about partner responses with regard to their pain behavior

Table I. Primary IASP Pain Locus

	N	%
Head, face, and mouth	6	10.5
Cervical region	2	3.5
Upper shoulder and upper limbs	5	8.8
Thoracal region	4	7.0
Abdominal region	2	3.5
Lower back, lumbar spine, sacrum, and coccyx	18	31.6
Lower limbs	5	8.8
Pelvic region	1	1.8
Anal, perianal, and genital region	1	1.8
More than three major sites	13	22.9

that very day (see also footnote 5). Responses made more than 10 min after a beep were considered invalid.

Experience Sampling Booklets

The first five scales on the MPI-DLV are Pain Severity, Interference (by the pain with the patient's daily life), Life Control, Affective Distress, and Support. The next three scales measure Punishing, Solicitous and Distracting Responses (by the partner to the patient's pain behavior). The last scale, General Activity, measures the frequency of some common daily activities. The General Activity scale is divided into three subscales: Household Chores, Outdoor Work, and Social Activities/Activities Away From Home.

Table II shows the ESM items used to mirror the MPI-DLV scales (with the exception of the Support scale). All the items were scored on 7-point Likert scales (0–6). The Likert scales of the first three ESM items ranged from "no pain/interference/control at all" (0) to "a great deal of pain/interference/control" (6). The Likert scales of ESM items representing the partner response scales ranged from "not at all" (0) to "very often" (6). The physical activity item ranged from "resting, lying down, doing nothing" (0) to "hard physical work" (6). The change in mood item ranged from "normal mood" (0) to "a sharp change" (6). If a patient indicated a change (i.e., a score from 1 to 6), he/she had to fill in whether this change was perceived as pleasant or unpleasant. The item "change in mood" was chosen to represent the MPI-DLV Affective Distress scale, rather than the more simple and straightforward item "mood." This was due to a specific interest in mood changes (reported elsewhere). In order to carry out the planned correlation analysis, the ESM mood change score was transformed as follows: a 0 score (normal mood) was recoded 3. Mood change scores in the

Table II. MPI-DLV Scales with Their Corresponding ESM Items

MPI-DLV scale	ESM item
Pain Severity	Pain severity
Interference	Pain interference
Life Control	Control of situation
Affective Distress	Change in mood
Punishing Responses	Irritation
Solicitous Responses	Takes over/take off work
Distracting Responses	Attention distraction
General Activity	Physical activity

"pleasant" direction were recoded 2.5, 2, 1.5, 1, 0.5, or 0. Mood change scores in the "unpleasant" direction were recoded 3.5 to 6 (at 0.5 intervals). In other words, the transformed mood change scale ranged from 0 (a very pleasant change) via 3 (normal mood) to 6 (a sharp change in the unpleasant direction). It was predicted that this transformed scale would show a positive correlation with the MPI-DLV Affective Distress scale.

With respect to the MPI-DLV scale Punishing Responses, which includes also frustration and anger, it was decided to choose "Irritation" as the ESM mirror-item because the concept of irritation is generally easily understood (some people may have problems with the word frustration). Anger was not chosen since real anger as a response to pain behavior may not happen as often as irritation, therefore probably resulting in a very skewed distribution.

Since the MPI-DLV General Activity scale contains three kinds of activities (Household Chores, Outdoor Work, and Social Activities/Activities Away From Home), it could scarcely be represented by one key item. Based on a previous finding (the positively significant correlation between bicycle ergometer performance and the MPI-DLV General Activity scale), it was hypothesized that a measure of actual physical activity level would correlate positively with the MPI-DLV General Activity scale. In order to get more specific information about daily activities, three open questions were added that had to be answered at each signal: What are you doing? Where are you? and Who are you with? These three daily life activity subcategories were coded. Previous research using the same coding system had shown that the interrater reliability was high [κ's > 0.90 (Delespaul, 1994)]. On the basis of the coded data, a "household activity frequency score" was computed for each patient. This score was correlated with the MPI-DLV General Activity subscale Household Activities.

Analyses

Analyses were carried out as follows: for each ESM variable a mean score was computed by adding up all nonmissing scores and dividing this total by the number of valid observations.[5] Correlation analyses were car-

[5]The maximum number of valid observations for the analogues of the MPI-DLV (sub)scales Interference, Life Control, Affective Distress, General Activity, and Household Chores was 48 (8 times a day multiplied by 6 days). The Pain Severity item had a maximum of 60: 10 ratings per day (8 beeps, 1 score when the patient woke up, and 1 just before going to sleep). For the analogues of the Punishing, Solicitous, and Distracting Responses scales, the maximum number of observations was six (one rating per day, and made at the end of that day). One rating per day was decided upon for these three variables because in many cases there is a structural absence of the patient's partner (at work).

ried out to assess the strength of the relationship between the MPI-DLV scales and the corresponding experience sampling items. The ESM irritating response variable showed a very positively skewed distribution. A logarithmic transformation was carried out on this variable. (The correlation between the transformed and the nontransformed variable was 0.97.) A regression analysis was carried out, taking the MPI-DLV item "Rate the level of pain *at the present moment*" as the independent variable and the mean score on the related ESM item as the dependent variable.

RESULTS

Compliance

All but one of the patients who gave their informed consent completed the study without serious difficulties. The one who did not gave up after a day and a half because he felt the experiment was too stressful. This patient was excluded from the analyses.

A further measure of compliance was derived from the amount of missing data. There were two kinds of missing data: the "real" missing data (no answer to a question) and the invalid answers (given more than 10 min after a signal). Missing data are reported here as a percentage (number of missing observations on a variable divided by the maximum number of possible observations on this variable). Table III shows these percentages for each experience sampling variable. As can be seen from this table, the total amount of missing data on the ESM variables is relatively large (except for on the partner response scales). The nonresponse to the item "physical activity" was very high. A closer inspection of the missing data in the ESM booklets indicated that, in almost all cases, there was at least one long period (about half a day or longer) during which the watch's signal function was turned off. Another explanation given by some patients for missing data was that in noisy situations they simply could not hear the signal. The average number of nonmissing data (valid observations) varied from 32, of 48, (physical activity) to 50, of 50, (pain intensity), and this was considered sufficient for obtaining reasonable estimates of the ESM variable means. With regard to the partner response items, for almost all patients the mean value could be based on the maximum of six observations.

Table IV presents the correlations between the MPI-DLV scales and the related ESM items, together with their significance. As can be seen, there are (relatively) strong positive relationships between the MPI-DLV scales Pain Severity, Interference, Solicitous Responses, Punishing Responses, Distracting Responses, and Household Chores and their

Table III. Percentage of Missing Data on the ESM Variables[a]

	PS	I	LC	AD	GA	HC	SR	PR	DR
Not filled in	11	13	13	24	27	12	6	5	6
Invalid	5	6	6	2	6	6	0	0	0
Total	16	19	19	26	33	18	6	5	6

[a]PS, Pain Severity; I, Interference; LC, Life Control; AD, Affective Distress; GA, General Activity; HC, Household Chores; SR, Solicitous Responses; PR, Punishing Responses; DR, Distracting Responses.

Table IV. The Relationship Between MPI-DLV Scales and Corresponding ESM Items

	Pearson correlation	*p* value
Pain Severity	0.75	<.001
Interference	0.60	<.001
Life Control	0.25	.06
Affective Distress	0.20	.14
General Activity	0.16	.22
Household Chores	0.40	.002
Solicitous Responses	0.52	<.001
Punishing Responses	0.23	.04
Distracting Responses	0.31	.02

corresponding ESM variables. The correlation between the MPI-DLV Life Control scale and its corresponding ESM variable reached only marginal significance. The ESM items change in mood and physical activity did not correlate with the corresponding MPI-DLV scales.

Frequency distributions showed an approximately normal distribution for both variables, allowing a regression analysis to be carried out. The resulting regression equation was ESM = 1.19 + 0.70 * MPI-DLV. The F statistic of the regression model was highly significant [$F(1,55) = 52.6, p < 0.0001$], as were the t values of the regression coefficient ($t = 7.30, p < .0001$) and constant ($t = 3.40, p = .001$). Forty-nine percent of the variance could be explained by the model. Residual analyses, as well as a visual inspection of the scatter plot, showed that there was one "clear" outlier (MPI-DLV value, 0; ESM value, 4.12; standardized residual, 2.9). Disregarding this case resulted in a remarkably improved data fit [$F(1,54) = 73.9, p < .0001$, Multiple $R = .76, R^2 = .58$]. The new regression equation was: ESM = 0.78 + 0.81 * MPI-DLV. The constant in the equation was significantly different from 0 ($t = 2.32, p = .02$). Figure 1 contains a scatter plot of the standardized scores on the item "rate the level of pain at the present moment" on the MPI-DLV Pain Severity scale and its ESM equivalent.

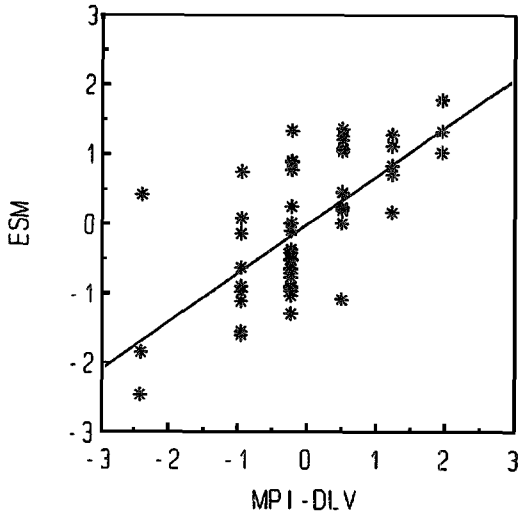

Fig. 1. Pain severity standardized scores: ESM versus MPI-DLV.

Finally, the difference between the mean pain severity score of the MPI-DLV and the related ESM item was tested. The mean MPI-DLV pain severity score was 3.52; the mean ESM score was 3.31. This difference was not significant ($t = 1.02, p = .314$).

DISCUSSION

In principle, a validation process is never-ending, in that all information on new validity facets merely adds to a more complete understanding of the properties of the instrument under consideration. As stated above, validation research on the MPI has already been successful. One of the purposes of the present investigation was to obtain additional information on the ecological validity of the MPI-DLV, in particular, on the relationship between relatively stable/trait information from the MPI-DLV scales and analogous items (measuring state information) in an experience sampling procedure.

Six MPI-DLV scales showed significant correlations with the corresponding ESM items: Pain Severity, Interference, Solicitous Responses, Punishing Responses, Distracting Responses, and Household Chores. The correlations between the MPI-DLV Pain Severity and Interference scales

and their ESM analogues are comparable to those reported by Flor and co-workers (1990). The ESM situation control item correlated marginally significantly with the MPI-DLV Life Control scale. The results relating to these MPI-DLV scales seem to show that valid information can be obtained about a patient's behavior and thoughts in daily life, using a comparatively short instrument (the MPI-DLV).

No relationship was found between the MPI-DLV scales Affective Distress and General Activity and their related ESM items. As already mentioned, the problem relating to the General Activity scale had to do with the lack of an adequate analogous item for the ESM. Nevertheless, based on the results of a working-to-tolerance bicycle ergometer test (Lousberg, 1994), where a positive correlation had been found between the patients' performance levels and the General Activity scales, expectations were that the ESM "physical activity" item would also show a positive correlation with this scale. As a matter of fact, a moderate positive correlation (0.16) was found but it failed to reach significance. It is interesting, however, that the MPI-DLV General Activity subscale Household Chores did show a fairly high correlation with a (reliably) coded household activities score. As previously mentioned, Flor and co-workers (1990) also failed to find a significant correlation between the diary and the MPI-D General Activity scale. These findings concur with Linton's conclusion that there is a discrepancy between the patient's perception of activities and the objective activity level (Linton, 1985). All in all, it is not clear at the moment whether the MPI-DLV General Activity scale adequately represents patient's physical activity level.

With respect to the ESM "mood change" item, the lack of significance can most likely be attributed to the content of the item (the emphasis on change instead of on mood alone). That is, the ESM "mood change" item may have been limited by its failure to incorporate various forms of affective distress including irritability, tension, and anxiety, which are all measured by the MPI-DLV. In conclusion, to put it even more strongly, the nonsignificant relationships between MPI-DLV and ESM on the two scales are not necessarily an indication that the MPI-DLV needs further validation but are likely to be caused by the limitation of ESM in sampling these phenomena.

Another aim of this study was to investigate the nature of the relationship between the MPI-DLV and the ESM with regard to present pain intensity. Examination of the data revealed one outlier. The problem of what to do with outliers is still unsolved (Cohen and Cohen, 1983). There were no reasons—such as a failure to carry out an instruction, a low intelligence level, etc.—which could account for the large discrepancy and, therefore, provide a reasonable argument for disregarding the case. Nevertheless, be-

cause of the relatively small number of subjects and the strong influence of this patient's score on the regression model, the authors still consider the second regression model (with the outlier omitted) to be a better estimation of the relationship. The parameters of the second regression equation convey useful information. A low score on the MPI-DLV pain intensity item is probably an underestimation of patient's mean daily pain intensity level. On the other hand, a high score on the MPI-DLV pain item should probably be interpreted as a slight overestimation of the mean daily pain level. As for the small discrepancies in the low and high regions, it can be argued that this is a phenomenon of "regression to the mean." It is very likely that a patient with a maximum score (6) on the MPI-DLV pain item will report at least some lower scores during the experience sampling period. Similarly, a minimum MPI-DLV score of 0 will very likely be transformed into some higher scores measured in daily life (if not, the term chronic pain patient becomes questionable). In general, though, it can be stated that the MPI-DLV score on pain intensity reflects the patient's mean pain experience level in daily life.

A major point of concern in the present study was the amount of missing data. Although the planned analyses could be carried, in future research studies, measures will have to be taken to minimize the possibility of missing data, especially when fluctuations in pain are under investigation (Affleck *et al.*, 1991; Jamison and Brown, 1991).

In conclusion, the present results lend further support to the validity of six of the MPI-DLV scales. Given the scores on these MPI-DLV scales, valid conclusions can be drawn about some aspects of a patient's behavior/thoughts in daily life. In addition, the finding that data from a short questionnaire strongly resemble those yielded by extensive and patient-loaded information-acquisition methods is very important. As far as the MPI-DLV scales Affective Distress and General Activity are concerned, further research is needed before conclusions can be drawn about their generalizability to daily life situations.

REFERENCES

Affleck, G., Tennen, H., Urrows, S., and Higgins, P. (1991). Individual differences in the day to day experience of chronic pain: A prospective daily study of rheumatoid arthritis patients. *Health Psychol.* 10: 419-426.

Clark, D. A. (1988). The validity of measures of cognition: A review of the literature. *Cog. Ther. Res.* 12: 1-20.

Cohen, J., and Cohen, P. (1983). *Applied Multiple Regression/Correlation Analyses for the Behavioral Sciences*, Erlbaum, Hillsdale, NJ.

Czikzentmihalyi, M., and Larson, R. (1987). Validity and reliability of the experience-sampling method. *J. Nerv. Ment. Dis.* 175: 526-536.

Delespaul, P. (1994). *Personal communication.*
DeVries, M. (1987). Investigating the mental disorders in their natural settings. *J. Nerv. Ment. Dis.* 175: 509-513.
Flor, H., Rudy, T. E., Birbaumer, N., and Schugens, M. M. (1990). Zur Anwendbarkeit des West Haven-Yale Multidimensional Pain Inventory im deutschen Sprachraum. *Der Schmerz* 4: 82-87.
Jamison, R. N., and Brown, G. K. (1991). Validation of hourly pain intensity profiles with chronic pain patients. *Pain* 45: 123-128.
Kerns, R. D., Turk, D. C., and Rudy, T. E. (1985). The West Haven Yale Multidimensional Pain Inventory (WHYMPI). *Pain* 20: 345-356.
Kerns, R. D., Turk, D. C., Holzman, A. D., and Rudy, T. E. (1986). Comparison of cognitive-behavioral and behavioral approaches to the outpatient treatment of chronic pain. *Clin. J. Pain* 1: 195-203.
Linton, S. J. (1985). The relationship between activity and chronic back pain. *Pain* 21: 289-294.
Lousberg, R. (1994). *Chronic Pain, Multiaxial Diagnostics and Behavioral Mechanisms,* Academic dissertation, Datawyse University Press, Maastricht.