RUNNING HEAD: ASSESSMENT OF OPIOID MISUSE IN CHRONIC PAIN

DEVELOPMENT OF A SELF-REPORT SCREENING INSTRUMENT FOR
ASSESSING POTENTIAL OPIOID MEDICATION MISUSE
IN CHRONIC PAIN PATIENTS*

Laura L. Adams, Ph.D., Robert J. Gatchel, Ph.D., Richard C. Robinson, Ph.D.,
Peter Polatin, M.D., Noor Gajraj, Ph.D., Martin Deschner, Ph.D., and Carl Noe, M.D.

The Eugene McDermott Center for Pain Management
The University of Texas Southwestern Medical Center at Dallas

*Supported in part by Grants 2R01-MH46452, 2R01-DE10713 and 2K02 MH01107 from
the National Institutes of Health, and a grant from the Sid Richardson Foundation.
ABSTRACT

This study comprised the first step in the psychometric development of a self-report screening instrument for risk of opioid medication misuse among chronic pain patients. A 26-item instrument, the Pain Medication Questionnaire (PMQ), was constructed based on suspected behavioral correlates of opioid medication misuse, which heretofore have received limited empirical investigation. The PMQ was administered to 184 patients at an interdisciplinary pain treatment center. Reliability coefficients for the PMQ were found to be of moderate but acceptable strength. Construct and concurrent validity were examined through correlation of PMQ scores to measures of substance abuse, physical and psychological functioning, and physicians’ risk assessments. To explore high and low cutoff points for misuse risk, subgroups were formed according to the upper and lower thirds of PMQ scores and compared on validity measures. Higher PMQ scores were associated with history of substance abuse, higher levels of psychosocial distress, and poorer functioning. Future psychometric analyses will consider predictive validity and examine shortened versions of the instrument.

KEY WORDS: chronic pain; interdisciplinary treatment; opioid medication misuse; Pain Medication Questionnaire; psychometric properties.
INTRODUCTION

While the use of opioid medication has long been accepted as the standard of care for managing acute and cancer-related pain [1], their use in treating non-cancer pain syndromes has been surrounded by controversy. Many physicians express uncertainty regarding the long-term effectiveness of opioids, as well as their potential for eliciting addiction. The specter of legal liability has compounded these fears, as patient-initiated litigation has emerged against physicians for allegedly causing opioid addiction [2]. Federal regulations require a physician to demonstrate a legitimate medical purpose for prescribing controlled substances and to assess for addiction potential prior to prescribing opioid medications [3].

Confusion surrounding a specific operational definition of opioid misuse among pain patients has complicated the process of effectively assessing and predicting its occurrence [1, 4, 5, 6]. The phenomena of “physical dependency” and “tolerance” are sometimes misconstrued as indices of opioid misuse, despite being typical, physiologically based outcomes of ongoing usage of opioid medications. Several investigators thus argue that the traditional model for assessing disordered substance use, set forth in the Diagnostic and Statistical Manual of Mental Disorders – 4th Edition (DSM-IV) [7], is not wholly appropriate for identifying opioid misuse in pain patients [6, 8]. While some DSM-IV behavioral criteria might be relevant (such as continued drug use despite negative physical, psychological or social effects), the criteria involving physical dependence and withdrawal are not applicable [9].

To address uncertainties in defining aberrant opioid use among chronic pain patients, three relevant organizations (American Society of Addiction Medicine, American Academy of Pain Medicine, and American Pain Society) developed a set of guidelines to help clinicians with
this assessment [9]. These organizations define opioid addiction as “a primary, chronic, neurobiological disease, with genetic, psychosocial, and environmental factors influencing its development and manifestations” [9]. This definition includes several categories of behavior, including impaired control over drug use, compulsive use, continued use despite harm, and craving.

Some investigators have explicated this definition further by enumerating specific behaviors that are assumed, in clinical practice, to reflect problematic opioid use. Portenoy [1] compiled a list of “aberrant drug-related behaviors,” which were divided into two categories of risk. Among the “probably more predictive” behaviors identified were: 1) forging prescriptions; 2) stealing or borrowing drugs from others; 3) frequently losing prescriptions; and 4) resisting changes to pain treatment, despite adverse side effects. The “probably less predictive” behaviors included: 1) aggressive complaining about the need for more drugs; 2) drug hoarding; and 3) unsanctioned dose escalation or other form of noncompliance. With a similar list, Savage [10] suggested that opioid addiction might be revealed through such behaviors as: 1) unwillingness to taper opioids to try alternative pain treatments; 2) decreased levels of function despite apparent analgesia; and 3) frequent requests for medication before renewal is due. While a single incident of these behaviors would not necessarily indicate a problem, a pattern of multiple incidents would be more suggestive. In addition, some correlates of opioid misuse go beyond specific drug-seeking behaviors, to more general domains of life functioning. For example, individuals with addictive diseases tend to have difficulties fulfilling their usual work and domestic roles, and they often have dysfunctional relationships [11].

While no widely used or thoroughly validated screening instrument currently exists for assessing risk of opioid misuse in chronic pain patients, several investigators have begun to
develop and validate relevant criteria [12, 13]. Chabal and colleagues [12] developed a “prescription abuse checklist” consisting of five criteria, including: 1) excessive focus on opiate issues during clinic visits; 2) a pattern of early refills or dose escalation in the absence of clinical change; 3) multiple phone calls or visits about opiate prescriptions; 4) a pattern of prescription problems; and 5) supplemental sources of opiates. Pain patients meeting three out of five of these criteria are presumably classified as opioid abusers. Compton and colleagues [13] have made inroads by developing a 42 item, interview-based screening questionnaire for assessing opioid abuse. Among other findings, their study showed that three items were particularly useful in identifying misusers of opioids: 1) showing a tendency to increase opioid dose; 2) having a preference for route of administration; and 3) considering oneself addicted. More recently, Passik and colleagues [14] piloted a questionnaire among a small group of cancer and HIV patients, evaluating medication use, present and past drug abuse, patients’ beliefs about addiction risk, and aberrant drug-taking attitudes and behaviors. A more thorough review of early initiatives to assess opioid abuse can be found in Robinson et al. [9].

Scope of the Present Investigation

The primary goal of the present study was to initiate the first step in an iterative process that will lead to the psychometric development of a self-report screening instrument, the Pain Medication Questionnaire (PMQ), to assess risk for aberrant behaviors in opioid medication use among patients with heterogeneous pain syndromes. Issues of validity were explored by comparing PMQ scores to a range of relevant variables measured at the outset of treatment, including measures of substance abuse, psychosocial distress, physical/functional ability, and
employment status. Issues of predictive validity will be examined in future studies, with further collection of data.

**METHOD**

**Subjects**

The core subject group was a sample of 184 consecutive patients who were newly evaluated for treatment at The Eugene McDermott Center for Pain Management (Pain Center) in Dallas, Texas, between the time period from October, 2001 through May, 2002. The sample included 98 patients who participated in the Pain Center’s interdisciplinary treatment program, which includes medical, psychological, psychiatric and physical therapy components. The sample included 86 additional patients who received only medical treatment at the Pain Center, as was deemed appropriate by their physicians.

**Procedure**

*General data-collection procedures.* Patients sought treatment at the Pain Center solely upon referral from another treating physician. Prior to their first appointment, patients completed a packet of paperwork that included consent forms, a personal data and medical history questionnaire, the PMQ, and measures of pain and functional capacity. Clerical staff members responsible for collecting these packets were coached on the importance of monitoring completeness of every patient’s paperwork; however, a very busy clinic schedule sometimes precluded this ideal, resulting in some missing data. A physician completed the initial evaluation, making a medical diagnosis and establishing a regimen for pain medication and pain management procedures. If the physician determined that the patient’s pain condition could
benefit from physical therapy and psychological interventions, a referral was made for the interdisciplinary program. Prior to psychological assessment, the patient received a packet of related paperwork, including an explanation of the behavioral medicine program, a consent form for psychological assessment and treatment, and several psychological testing instruments. Two psychologists conducted the assessments and formulated individualized treatment plans, which included individual behavioral medicine sessions, group sessions, and psychiatric medication consultation, if warranted. Results from these initial assessments comprised the data that were used in this study’s analyses.

**PMQ Development.** Items were initially developed based on literature addressing opiate misuse and its measurement [1, 9, 12, 13, 14, 15] and on input from relevant clinical personnel (nurses, physicians, psychologists) at the Pain Center. A set of 26 items was selected to reflect a range of potentially dysfunctional attitudes and aberrant behaviors surrounding the use of pain medication (see Appendix). No mention was specifically made of opioid medication in these items, so that patients taking any form of pain medication could be measured on these behaviors and attitudes. A specific effort was made to use language in item-construction that was as neutral and non-threatening as possible, in order to encourage candid responding. A readability analysis found the instrument to fall at the 7.5 grade level.

Patients responded to each item on a standard numerical-type scale, arranged in a 5-point Likert format. Numbers were not explicitly used on this Likert scale, to avoid the suggestion of any negative or positive valence for any particular response. Instead, each point on the scale was given a verbal anchor designed to reflect different degrees of a patient’s conformity with a particular behavior. Numerical values were assigned to these anchors after the test.
administration for the derivation of an item-score (e.g., “Disagree” = 0, “Somewhat Disagree” = 1, “Neutral” = 2, “Somewhat Agree” = 3, “Agree” = 4). An overall score was derived by summing the item-scores for the 26 items, with a minimum possible score of 0 (26 items × 0 points) and a maximum possible score of 104 (26 items × 4 points). Higher overall scores were assumed to reflect a greater presence of behaviors associated with potential risk for opiate misuse. To minimize the probability of a negative or positive response bias, some items (#’s 1, 2, 5, and 8) were developed to capture behaviors believed to be inversely related to risk for opioid misuse. Reverse numeration was applied to these items (e.g., “Disagree” = 4, “Somewhat Disagree” = 3, “Neutral” = 2, “Somewhat Agree” = 1, “Agree” = 0).

**Instruments and Outcome Measures**

*Beck Depression Inventory (BDI).* This 21-item multiple-choice instrument [16] has good psychometric properties in measuring behavioral signs of depression, with the following ranges of intensity: normal (0-9); mild depression (10-15); mild to moderate depression (16-19); moderate to severe depression (20-29); and severe depression (30+).

*CAGE.* Comprised of four clinical interview questions, this instrument targets behaviors and experiences related to substance abuse [17] and discriminates between known abusers and non-abusers of alcohol [18]. In the present study, the CAGE was incorporated into the pre-treatment paperwork, administered in a self-report fashion.

*The Dallas Pain Questionnaire (DPQ).* This analogue scale is a 15-item, self-report questionnaire containing items related to pain and disability [19]. Patients respond to items by
picking a point on a 10-cm line representing a range of possible answers from 0 to 10, anchored with descriptors. Higher total scores represent greater levels of disability: “mildly disabling pain” (0-39); “moderately disabling pain” (40-84); and “severely disabling pain” (85+).

**Medical Outcomes Short Form-36 Health-Status Survey (SF-36).** This 36-item self-report questionnaire assesses health-related quality of life [20]. It has eight scales and two standardized summary scales, the Mental Component Scale (MCS) and the Physical Component Scale (PCS), which were used in this study.

**Millon Behavioral Health Inventory (MBHI).** This 150-item self-report inventory provides information about a patient's coping style, psychogenic attitudes, psychosomatic correlates, and prognostic indices [21, 22, 23].

**Minnesota Multiphasic Personality Inventory (MMPI-2).** This 567-item, self-report questionnaire assesses psychiatric symptoms and personality organization [24, 25, 26, 27]. In addition to the basic scales, several specific scales were utilized in this study, including the MacAndrew Alcoholism Scale – Revised (MAC-R), the Addiction Potential Scale (APS), and the Addiction Acknowledgement Scale (AAS) [28, 29].

**Oswestry Pain Disability Questionnaire (OSW).** This self-rating scale evaluates the degree of functional impairment caused by pain in activities of daily living [30, 31, 32].
**Patient Information Form.** This clinic-specific form elicits written information, including patient demographics, employment status, education completed, date and details of injury, workers’ compensation or personal injury litigation involvement, health care utilization, medication, and chronic health problems.

**Physician Risk Assessment.** An adjunct instrument was developed for this study as a way of quantifying the physicians’ independent assessments of patient risk for opiate misuse, as determined through direct interaction with the patient. Used as one means for validating the PMQ, the Physician Risk Assessment (PRA) asked the attending physician to rate his patients on a set of six dimensions of potential risk for opiate misuse (e.g., Does this patient’s history suggest misuse of medication or another substance?). These six items were developed with input from Pain Center physicians, according to the main historical and behavioral factors they typically consider in making judgments about risk for medication misuse. Ratings for these items were captured on a five-point Likert scale reflecting increasing degrees of risk.

Inter-rater reliability was examined by asking a group of four participating pain management physicians to use this instrument in judging the risk for opiate misuse among six fictional patient case-studies, presented in written vignettes. These vignettes were developed to reflect a range of pain conditions and risk behaviors related to medication misuse. An intraclass correlation method [32] used to examine the degree of agreement in the physicians’ ratings yielded a coefficient of .74, which was deemed acceptable.

**Visual Pain Analogue (VPA).** This scale asks patients to rate their pain on a scale from 0 to 10 by marking an “X” on a 10-centimeter horizontal line, hashed at two-point intervals, with
higher numbers reflecting greater pain [23, 33]. A pain rating is derived by rounding up to the next whole number subsequent to the hash mark.

*The West-Haven-Yale Multidimensional Pain Inventory (MPI).* This self-report pain inventory measures several dimensions of pain perception and functional status among chronic pain patients [34], yielding eight subscales and an overall coping style.

**Statistical Analyses**

Generally speaking, statistical analyses in this study included descriptive analyses of demographics and instrument properties, derivation of reliability coefficients, plus two lines of statistical analyses that were used in validation studies. One type of analysis involved deriving Pearson’s correlation coefficients between the PMQ and continuous-scale validation measures. The other type of analysis sought to examine possible cut-off scores for this instrument by identifying patients falling into “high” and “low” scoring groups on the PMQ, and comparing their scores on validity measures. To define these groups, subjects falling in the lowest third of scores comprised the “Low” PMQ scoring group (L-PMQ), while all subjects falling in the highest third comprised the “High” PMQ scoring group (H-PMQ). Scores on the validity measures for these two groups were then compared through a series of independent t-tests and chi-square analyses.

**RESULTS**

**Demographic Variables**
The total sample of 184 patients was analyzed for proportional breakdowns within the categorical variables of gender, race, marital status, treatment group, disability payment status (receiving disability payments or not), litigation status (pending litigation or not), opioid status (taking opioids or not), and pain diagnoses (Table 1). Statistical analyses revealed no significant differences between the demographic profile of the study sample and that of the typical population treated at the Pain Center.

Of the 184 patients in the study sample, 66.3% were female and 33.7% were male. The mean age was 48.83 years (SD = 14.11), ranging from a minimum of 17 to a maximum of 84 years. The largest racial group was Caucasian (84.2%), while African-Americans represented the next largest group (10.3%). Hispanic, Asian and other races comprised only 5.4%, altogether. Most of the subjects were married (61.4%), although significant numbers were single (18.5%), separated/divorced (14.1%) or widowed (6.0%).

A breakdown of treatment groups indicated that 53.3% of the sample participated in the interdisciplinary treatment program, while 46.7% received only medical treatment. Statistical analyses on a subset of demographic and test variables (e.g., age, disability measures, psychological functioning) showed no significant differences between these treatment groups. The mean pain duration for the sample was 79 months (just over 6.5 years), with wide variability (SD = 109.91). Approximately 24% of the sample was receiving disability income, and nearly 15% had pending litigation related to their pain at the time of initial assessment. A heterogeneous mix of pain diagnoses was captured in the sample, with many patients holding more than one diagnosis. The most frequently represented diagnoses were lumbar (31.7%) and cervical (17.2%) back pain, and myofascial-fibromyalgia (19.4%).
Comparison of PMQ scoring groups. Statistical analyses examined differences on demographic variables between the H-PMQ and L-PMQ scoring groups. Independent t-tests compared the scoring groups on the variables of age and pain duration, while Pearson Chi-Square analyses compared the groups on distributions of gender, race, marital status, treatment group, disability payment status, litigation status, and opioid status. Results showed that the H-PMQ group contained significantly more patients who were married (67.7%) or separated/divorced (17.7%), as compared to the L-PMQ group (54.5% and 10.6%, respectively), while the L-PMQ group contained significantly more single (22.7%) or widowed (12.1%) patients, relative to the H-PMQ group (12.9% and 1.6%, respectively), \( \chi^2 (3)=8.81, p=.03 \). A significantly larger percentage of patients collecting disability payments fell in the H-PMQ group (39.0%), as compared to the L-PMQ group (15.9%), \( \chi^2 (1)=9.55, p<.01 \). No significant differences were found between groups on the variables of gender, age, race, treatment group, litigation status, and pain duration.

Just over 60% of the overall sample was prescribed opioid medications. Statistical analyses were conducted to examine how opioid medication status was related to PMQ scores. A patient’s taking of opioid pain medication was not assumed, in itself, to suggest risk for substance abuse. However, this study did seek to consider if people taking opioids reported greater frequencies of attitudes and behaviors that may indicate risk for misuse, relative to people taking non-opioid pain medications. Examining the relationship of opioid status to the H-PMQ and L-PMQ scoring groups, a chi-square analysis revealed that a significantly greater proportion
of patients on opioids were found in the H-PMQ group (70.5%) than in the L-PMQ group (44.6%), $\chi^2(1) = 8.60, p < .01$. The full range of PMQ scores also were examined, comparing the mean scores between the two groups: (a) “Yes” – on opioid medication; and (b) “No” – not on opioid medication. A significant difference was found, $t(178) = -3.11, p < .01$, revealing a higher mean PMQ score for the “Yes” group ($M = 26.47, SD = 10.07$), relative to the lower mean of the “No” group ($M = 21.75, SD = 9.73$).

**Basic Descriptive Analysis of the PMQ**

The total sample ($N = 184$) yielded a mean PMQ score of 24.60 ($SD = 10.16$), and a very similar median score of 24.25. The range was 65 points, with a low score of 2 and a high score of 67 (out of a possible maximum score of 104). Skewness was found to be .62, and kurtosis was 1.07, suggesting a reasonably close approximation to the normal curve. Measures of skewness and kurtosis falling between −1 and +1 are considered to be appropriate indicators of a normal distribution [35].

**PRA Descriptive Analysis**

Physician Risk Assessments (PRA) were completed for a subgroup ($n = 146$) of the total sample (Table 2). Missing data reflected the physicians’ busy clinic schedules, which precluded a perfect collection rate for these data. The mean PRA score was 3.97 ($SD = 5.35$), out of a possible maximum of 20 points, and the median PRA score was .75. The modal total score was 0, comprising 49.3 % of all the scores. Individual items were scored on a range of 0 to 4 points. The generally low means for PRA items reflect the high prevalence of "0" scores, indicating that,
per the physicians’ ratings, a large proportion of patients did not demonstrate any measure of these potential risks and behaviors during the initial medical evaluation.

---

**INSERT TABLE 2 ABOUT HERE**

---

**Reliability**

*Test-retest reliability.* Test-retest reliability of the **PMQ** was examined by administering the instrument to a group of 19 patients at two points, approximately one-half hour apart, during their psychological evaluation appointments. This test-retest interval was selected to minimize interference from other appointments or pain-management procedures that might have influenced responding. The coefficient of stability was derived through a Pearson’s Product Moment Correlation (Pearson’s \( r \)), which yielded a coefficient of .85. This was judged to be adequate for further psychometric analyses.

*Internal consistency.* Proposed as the measure of choice for estimating reliability [36], an internal consistency analysis was conducted after **PMQ** data were collected for the total sample. Cronbach’s alpha [37], the most widely used estimation of internal-consistency, yielded a coefficient of .73. While alpha coefficients of .80-.90 are typically desired, coefficients of greater than .70 are considered acceptable in the early stages of research on a new instrument [38].
**Item reliability.** Reliability analyses were also conducted for each of the PMQ’s 26 items. Each item was examined with a corrected item-total correlation, which shows the strength of the relationship between the individual item and the total score, subtracting the given item’s contribution. These correlation coefficients were fairly evenly dispersed between a low of -.01 (Item # 10) and a high of .50 (Item #’s 21 & 24). The low-to-moderate strength of these coefficients was judged to reflect the diverse nature of the attitudes and behaviors encompassed by this instrument. The five items with the strongest item-total correlation coefficients were Item #’s 4 (r = .42), 18 (r = .46), 21 (r = .46), 24 (r = .50), and 25 (r = .49), suggesting that these items varied most consistently with the total score. Conversely, the five items that varied in the least predictable way with the total score were Item #’s 2 (r = .13), 5 (r = <.01), 10 (r = <-.01), 11 (r = .17), and 13 (r = .09). A re-calculation of Cronbach’s alpha, subtracting these four items, yielded a coefficient of .75, suggesting that the internal consistency of the PMQ would be enhanced if these items were eliminated.

**Content Validity**

The content validity of the PMQ was examined by collecting independent ratings provided from ten pain management professionals. The raters were comprised of physicians and psychologists working at several different pain rehabilitation centers in Dallas, Texas, including a county hospital pain treatment center, a university-affiliated pain management center, a private tertiary care clinic, and a private primary and secondary care clinic. A ten-point scale was used to capture the raters’ assessments of how comprehensively the PMQ captures behaviors suggesting opiate misuse. The mean rating was found to be 9.05 (SD = .72), suggesting adequate content validity.
Construct Validity

Friedenberg [39] notes that selection of appropriate construct validation measures requires “construct explication,” or an analysis of the construct’s multiple features. This study proposed that potential risk for opioid misuse among chronic pain patients is likely to be related to a variety of behavioral and psychosocial variables, such as: 1) diminished physical and functional performance, particularly to a greater degree than is warranted by the pain condition; 2) psychological vulnerability, including depression, anxiety and poor coping; and 3) a history of problems with substances, such as alcohol, street drugs, and prescription medication.

As a subtype of construct validity, convergent validity refers to the degree to which a given measure performs in a similar, predictable way to measures of related constructs. By contrast, discriminant validity is supported when there is a negative or inconsistent relationship between the given measure and indices of behaviors unrelated or inversely related to the target construct. Convergent validity was examined with measures examining substance abuse potential, physical/functional disability and psychosocial distress. Conversely, discriminant validity was investigated with measures of physical well-being and psychosocial coping.

Convergent and discriminant validity were examined through two phases of statistical analyses. One phase of analyses involved deriving Pearson’s correlation coefficients between the PMQ and continuous-scale validation measures. The other phase sought to examine scores on convergent and discriminant validity measures relative to H-PMQ and L-PMQ scoring groups.
Physical/functional measures. L-PMQ and H-PMQ scoring groups were compared through independent t-tests on several physical/functional measures (Table 3). The DPQ, VPA and OSW served as convergent validity measures, wherein higher scores on these instruments (suggesting greater levels of pain and disability) were expected to conform to higher scores on the PMQ. By contrast, the SF-36 functioned as a discriminant measure, as higher scores on this instrument (suggesting better physical and emotional functioning) were expected to conform to lower scores on the PMQ. All results of these analyses were significant, with the H-PMQ group reporting more pain-related disability, relative to the L-PMQ group. Pearson’s r’s were calculated to examine these physical/functional measures relative to the whole range of PMQ scores. Correlation coefficients between the full range of PMQ scores and all convergent validity measures were significant at p<.01, ranging from .23 to .36. Also significant were the coefficients between the PMQ and the discriminant measures, with r = -.22, p = .04 (PCS) and r = -.35, p < .01 (MCS).

---

INSERT TABLE 3 ABOUT HERE

---

Employment status. To explore this variable relative to PMQ scores, subjects were categorized into three groups: (a) Working (WK); (b) Not Working due to pain or injury (NW-PI); and (b) Not Working due to other reasons (NW-O), such as retirement, layoff or unpaid employment in the home (i.e., homemaker). A one-way ANOVA revealed a significant difference between group means, F(93)=3.90, p<.02. The WK group had a mean PMQ of 23.61 (SD = 9.45), the NW-PI group had a mean of 28.96 (SD = 11.07), and the NW-O group had a
mean of 22.28 (SD = 9.51). Planned orthogonal comparisons were conducted between the following pairs: WK vs. NW-PI; WK vs. NW-PI & NW-O; WK & NW-O vs. NW-PI. Results showed a significant difference for the NW vs. NW-PI comparison (p = .04), and the WK & WNO vs. NW-PI comparison (p < .01). This suggests that patients who were not working due to their pain or injury had significantly higher mean PMQ scores, as compared to those who were working and those who were not working due to reasons unrelated to pain.

**MMPI-2 basic scales.** The MMPI-2 basic scales were employed to examine the relationship between psychological distress and PMQ scores (Table 4). As a first step, a MANOVA was conducted to check for a significant difference in the overall profiles of mean T-scores between L-PMQ and H-PMQ groups. A two-tailed analysis of the basic clinical scales yielded significance, Hotelling’s trace=.35, F(10, 62)=2.18, p=.03, suggesting that independent t-tests might safely be conducted for each individual scale, without employing Bonferroni corrections.

- - - - - - - - - -

**INSERT TABLE 4 ABOUT HERE**

- - - - - - - - - -

Results of t-tests showed significantly higher mean T-scores for the H-PMQ group on all scales except for scales L, K, 5, 6, 9, and 0. The strongest differences were found on Scales 1 (Hypochondriasis) and 2 (Depression). While both groups fell in the clinically-significant range on Scale 1, the H-PMQ mean was more than one-half of a standard deviation above that of the L-PMQ group. The discrepancy was equally as great on Scale 2. This trend of depression ratings
was corroborated by findings on the BDI, wherein the L-PMQ group had a mean score of 10.08 (SD = 7.46), while the H-PMQ group had a mean of 17.64 (SD = 9.67), t(70) = -3.71, p < .01. These results placed L-PMQ group in the milder range of depressive symptoms and the H-PMQ group in the moderate range of depressive symptoms. As Figure 1 illustrates, the H-PMQ group’s MMPI-2 profile, while very similar in shape to that of the L-PMQ group, shows greater intensity of symptoms and features on key clinical scales. All t-test findings were corroborated by correlations between the MMPI-2 basic scales and the full range of PMQ scores. The significant Pearson’s r coefficients ranged from .23 (Scale 4) to .35 (Scale 2).

MBHI scales. A MANOVA was conducted to compare the mean MBHI base rate scores between the L-PMQ and H-PMQ groups. The result was significant, Hotelling’s trace = .70, F(20, 51) = 1.79, p = .05, suggesting that independent t-tests could be conducted on the individual scales, without undue inflation of Type I error rates. Base-rate scores for each scale were compared between the L-PMQ and H-PMQ groups. Significant differences were revealed on several scales (Recent Stress, Premorbid Pessimism, Somatic Anxiety, Allergic Inclination, Gastro-Susceptibility, Cardiovascular Sensitivity, Pain Treatment Response and Life Threat Reactivity), with the H-PMQ group scoring significantly higher than did the L-PMQ group. Correlational analyses between MBHI scales and PMQ scores revealed significant Pearson’s r coefficients for all the scales mentioned above, ranging from .22 to .33. In addition, significant positive correlations emerged between the PMQ and the Inhibited (.22), Sensitive (.30) and
Emotional Vulnerability (.25) scales. Conversely, a significant negative correlation was found between the PMQ and the Sociable scale (-.23).

**MPI coping style.** The MPI classifies people into one of five coping styles: 1) Adaptive (AC); 2) Interpersonally Distressed (IDC); 3) Dysfunctional (DYC); 4) Anomalous; and 5) Hybrid. For this study, the latter two styles, both of which represent unclear coping styles, were collapsed into an “Other” group (OTH), resulting in four coping-style groups. A one-way ANOVA compared these four groups on mean PMQ scores, yielding a significant result, $F(96)=6.49, p<.01$. The AC group had the lowest mean PMQ score ($M=19.26, SD=8.80$), while the IDC group had the highest mean score ($M=30.63, SD=12.45$). Planned orthogonal comparisons were conducted to compare means between two pairs: AC vs. IDC & DYC, and IDC vs. DYC. Results showed a significant difference at $p < .01$ for the comparison of AC vs. IDC & DYC, where the AC group had a mean PMQ of 19.26 ($SD=8.80$) and the IDC and DYC groups had a combined mean PMQ of 29.04 ($SD=10.61$). No significant difference was found between the IDC and DYC groups.

**Substance abuse measures.** Because no well-validated or widely used measures of potential risk for opioid misuse currently exist, construct validity analyses in this study utilized several indices of general substance abuse history and potential. Among the standardized measures of substance abuse risk available for this study were the MacAndrew Alcoholism Scale - Revised (MAC-R), the Addiction Potential Scale (APS), and the Addiction Acknowledgement Scale (AAS) of the MMPI-2. Pearson’s $r$ was calculated to examine the relationship of each scale to the whole distribution of PMQ scores. These findings, ranging from .01 to .17,
suggested little or no systematic relationship between scores on the MMPI-2 supplementary substance abuse scales and the **PMQ**. Mean T-scores on these scales were compared between the L-PMQ and H-PMQ groups, showing no significant findings.

Another index of risk substance abuse used in this study was Acknowledgement of Substance Abuse History (ASAH), captured by a single, written yes-or-no question on the Patient Information Form (“Do you have a history of alcohol or drug abuse?”). Patients’ responses to this question classified them into the “Yes” group (n = 17) or the “No” group (n = 140). Because of the large difference in size between these sub-samples, a “No” group of equal size to the “Yes” group was created (matched on age, gender, and opioid status) for an independent t-test analysis. The “Yes” group had a mean **PMQ** score of 28.47 (SD = 8.90), while the “No” group had a mean of 22.47 (SD = 7.82), which was a significant difference, t(32)= -2.09, p=.05.

In this study, CAGE items were transformed to a written format and incorporated into the Patient Information Form, so as not to seem too obtrusive and “off-putting” to the patient. Due to the low frequency of responses to CAGE items (n = 79), the data were analyzed with Pearson’s Chi-Square. CAGE scores were dichotomized into two groups, according to “0” items endorsed and “1+” items endorsed. The L-PMQ and H-PMQ groups were then compared for their relative distributions between these CAGE groups. The significant findings, \( \chi^2(1)=7.21, p<.01 \), indicated that a greater proportion of the H-PMQ group (21.6%) endorsed “1+” CAGE items, while a lesser proportion of the L-PMQ (2.4%) endorsed “1+” CAGE items.

**Concurrent Validity**
As a type of criterion-related validity, concurrent validity refers to the extent to which the new measure correlates with scores on criterion measures, or measures of behavior the test is specifically designed to predict. While the other subtype of criterion-related validity, predictive validity, examines this correlation with future measures of the target behavior, concurrent validity correlates the new measure with measures obtained at about the same time. For this initial stage of psychometric development, the only criterion measures available were the concurrent measures of: 1) physician-ratings of risk for opioid misuse (PRA), obtained at initial medical evaluation; and 2) known status of opioid misuse upon the patients’ entry into the pain program, as indicated by their referring physicians or as admitted by themselves.

**Physician Risk Assessment (PRA).** Independent t-tests compared mean PRA scores, for the total instrument and each individual item, grouped by L-PMQ and H-PMQ scoring groups. Results were significant for every analysis (p<.01), suggesting that PRA scores discriminated well between L-PMQ and H-PMQ groups. For the PRA Total Score, the L-PMQ group had a mean of 1.33 (SD = 2.87), while the H-PMQ group had a mean of 7.55 (SD = 6.14), t(98) = -6.71, p<.01. The H-PMQ mean of 7.55 fell well above the total sample mean of 3.97. A series of Pearson’s r correlation coefficients, ranging from .36 (Item #5) to .45 (Item #6 and total score), corroborated these findings, and were all significant (p<.01).

**Known opioid misuse.** Because of the small number of known opioid misusers (n = 12), a sample of patients with no known history of opioid misuse was matched for age, gender and opioid status. The matched group had a mean PMQ score of 25.54 (SD = 11.54), while the opioid misusing group had a mean PMQ score of 33.88 (SD = 11.89). This difference was
found to be significant, \( t(22)=-1.76, p=.045 \), with a one-tailed \( t \)-test test, which was used given the \textit{a priori} expectation that patients with a history of opioid misuse would have higher PMQ scores.

\textbf{Factorial Validity}

\textit{Factor analysis.} While the \textbf{PMQ} was not intentionally designed to contain specific subscales, an exploratory factor analysis was conducted to investigate possible enhancements to validity by organizing and paring-down the items into a set of more focused factors. Conforming to the statistical convention suggesting a 5:1 ratio of study participants to instrument items [40], the sample size of 184 was deemed sufficient for factor-analyzing this instrument of 26 items.

\textit{Extraction of factors.} A frequently-used index for identifying the appropriate number of components is the Kaiser-Guttman criterion, which suggests retaining all components with an eigenvalue > 1.0; however, some research has demonstrated that this criterion can overestimate the number of components to retain [40]. Using a principal components analysis, nine components were found to have eigenvalues above 1.0, which was judged to be an overestimation of the number of factors represented in this instrument. The scree plot inspection yielded similarly equivocal results. While a dramatic drop between the first and second component suggested that there might be no more than one factor underlying the structure of this instrument, the continued downward slope of the plot suggested that one or more smaller factors might exist as well.
**Factor rotation.** Three separate factor rotations were therefore conducted, using the varimax method to examine the viability of two, three, and four underlying factors. Based on the conceptual cohesiveness of the items loading on the factors, as well as the strength of the factor loadings, the 2-factor analysis was judged to have the most potential for refining the PMQ's psychometric properties. With factor loadings ranging from .42 to .73, a total of 9 items were judged to load primarily on the first factor (#'s 1, 3, 4, 12, 18, 21, 22, 24 & 25). With factor loadings ranging from .45 to .55, a total of 6 items were judged to load primarily on the second factor (#'s 11, 13, 15, 16 & 17). One item (#7) appeared to load sufficiently on both factors. Loadings of .40 and greater were considered sufficient for inclusion, with the exception of Item #7, which loaded on Factor 2 at .45 and on Factor 1 at .36. Eleven items loaded significantly on neither factor. While there was some conceptual overlap between the two factors, Factor 1 was judged to capture an underlying dimension of "preoccupation with procurement of pain medication," while Factor 2 was judged to capture a more extreme set of "maladaptive behaviors and side-effects."

**Statistical Analysis of Factors**

The next step in examining the viability of these factors was to analyze their statistical properties, in order to determine if the two factors showed better reliability and validity than did the whole 26-item instrument. Thus, a subset of analyses performed on the total instrument ("Whole PMQ") was conducted separately for each of the factors. A summary of these analyses is presented here, and a more detailed explication of these results is documented elsewhere [41].

Factor 1’s coefficient alpha was found to be .77, suggesting improved internal consistency over the Whole PMQ. By contrast, Factor 2 had an alpha of .54, indicating
considerably more variability in the performance of these items relative to one another. With respect to validity analyses, both factors failed to achieve all of the significant findings yielded by the Whole PMQ; however, the pattern of results suggests potential trends that might be important. For example, Factor 1 appeared to lose significance most frequently on psychosocial measures, while retaining significance on all substance abuse measures. Factor 2 lost significance on all types of instruments, with particular losses relative to substance abuse measures. Collectively, the two Factors yielded the same pattern of significance as the Whole PMQ, with two exceptions. Neither factor achieved significance on MMPI-2/Scale 4, as did the Whole PMQ. Moreover, Factor 1 achieved significance on the MAC-R, which was not significant for the Whole PMQ.

**DISCUSSION**

This study represented the initial step in developing and validating a self-report screening instrument for assessing risk of opioid medication misuse. Much further study will be required to examine the predictive and incremental validity of the PMQ, and to develop the ideal length and appropriate norms for population subgroups. However, within the scope of the present investigation, psychometric outcomes of this study suggest that the PMQ holds promise, with considerable future refinement, as a self-report screening measure for risk of opioid misuse.

**Demographic Variables**

In the total sample of 184 patients, the “modal” subject was a married, Caucasian female, approximately 49 years old, who had been coping with a chronic pain condition, most often back pain or myofascial/fibromyalgia, for roughly 6.5 years. While the findings of this study might
readily be generalized to this type of patient, some data suggested that these findings might also be relevant for a more heterogeneous range of chronic pain patients. For example, among an array of demographic variables, only the variables of marital status and disability payment status showed significant differences in distributions between High and Low PMQ scoring groups.

The higher prevalence in the H-PMQ group of patients collecting disability is consistent with earlier proposals that people with higher levels of disability might be at greater risk for opioid misuse [1, 42]. Results from the analysis of martial status relative to PMQ score appear more complex. The H-PMQ group had relatively higher proportions of married and divorced patients, and relatively fewer single and widowed patients, as compared to the L-PMQ group. While it makes sense that divorced patients, with diminished social support, might fall into the higher risk group, it is unclear why the same would be true of married patients. One possibility is that, given the interpersonal strain that can arise with chronic pain, the marital relationships of some people in this sample might not have provided the quality of social support that can mitigate the risk of additional problems. Indeed, analysis of the MPI data indicated that the Interpersonally Distressed Coping Style was associated with the highest mean PMQ score. Thus, future investigation of the possible interplay between marital status (and, more broadly, social support) and the risk for opioid misuse appears warranted.

The non-significant findings between the interdisciplinary and medical-only treatment groups were particularly noteworthy, given the assumption that pain patients referred for interdisciplinary treatment tend to have relatively higher levels of psychosocial distress and physical disability [1]. Further investigation will be helpful in exploring the relationship between types of pain treatments and risk for opioid misuse.
Patients taking opioid medications were shown to have significantly higher PMQ scores, relative to patients not taking opioids. While this finding does not suggest that patients taking opioids are inherently at greater risk for opioid misuse, it does indicate that the opioid group tends to self-report more attitudes and behaviors that are thought to be correlates of potential misuse. These results cannot be explained as the opioid-users’ greater responsiveness to opioid-related questions, as the PMQ’s items were neutral with respect to type of pain medication.

Reliability Analyses

The interval used for assessing the PMQ’s test-retest reliability (1/2 hour) was selected to avoid the interference of any treatment on the patients’ responses to the instrument. The resulting stability coefficient of $r = .85$ must be interpreted in light of the brevity of the test-retest interval, and a follow-up analysis with a longer interval might be warranted, with further refinements of the instrument. With a Cronbach’s alpha of .73, the PMQ’s internal consistency was deemed acceptable, particularly for an instrument with heterogeneous content, in the early stages of development. All but four of the individual items contributed positively to the strength of the overall alpha. Elimination of the items with the lowest item-total correlations was found to raise the alpha to .75, suggesting this as a possible future improvement to the instrument.

Construct Validity

This study proposed that risk for opioid misuse might also be associated with extreme reports of physical disability, functional impairment, and psychosocial distress, particularly those that exceed expectation for a given pain condition. Virtually all of the convergent/discriminant analyses offered support for this position. Results for the DPQ were particularly noteworthy,
given this instrument's utility in identifying patients who give self-reports of pain that exceed expectations, given their physical findings [19]. While both PMQ scoring groups had mean DPQ scores in the "severely disabling" range, the H-PMQ group had a mean DPQ score that was nearly one standard deviation higher than that of the L-PMQ. Thus, while both groups were likely somewhat exaggerating their pain-related disability, the H-PMQ group was doing so to a substantially greater degree. This is consistent with the findings on the MBHI’s Life Threat Reactivity Scale, wherein people with higher scores tend to show poorer mental and physical functioning, relative to those with lower scores, despite the same level of actual illness and physical impairment. This suggests that the variance in PMQ scores is not entirely due to true differences in physical pain and disability and may reflect additional vulnerabilities, such as an over-reliance on pain medication.

Perhaps even more telling were the analyses surrounding work status, as a direct behavioral index of functioning. Higher PMQ scores among people unemployed due to pain demonstrates that this group tends to report more attitudes and behaviors thought to be related to risk for opioid misuse. While some of this trend might be due to true differences in physical capabilities, one might also argue that the same diminished coping strategies that undermine a patient’s motivation to return to work also place that person at greater risk for opioid misuse. The physical and work status findings are consistent with Savage’s [11] contention that correlates of opioid misuse can go beyond specific drug-seeking behaviors, and are likely revealed in other domains of life-functioning, such as debilitated capacities to cope with stress and fulfill work roles.

Similarly consistent were the results of analyses comparing PMQ scores to various psychosocial measures. Indeed, patients falling in the H-PMQ group were significantly more
likely to have Interpersonally Distressed or Dysfunctional coping styles, relative to patients in the L-PMQ group, who were more likely to have an Adaptive Coping Style. Moreover, high PMQ scorers reported significantly higher levels of psychosocial distress, including depression, anxiety, physical preoccupation and a sense of social alienation. These findings were consistent across multiple measurements, including the BDI and several scales on the MMPI-2 and MBHI.

Additional findings illustrate personality features that may be consistent with high PMQ scores. Significant results on the MBHI showed that high scorers on the PMQ tend to have a general outlook of negativity, are more likely to respond to emotional upset with physical symptoms, and are less likely to benefit from health care interventions. The moderately elevated mean on MMPI-2/Scale 4 for the H-PMQ group (M = 57.39) suggests that, while not showing pronounced antisocial behavior, this group might have tendencies toward manipulation, selective reporting and socially maladaptive behavior [43]. Results from the MPI analyses concur that high PMQ scores are more frequently associated with strained interpersonal relations and dysfunctional coping, as compared to low PMQ scores.

Construct validity analyses of the PMQ relative to measures of substance use/abuse yielded limited but thought-provoking data about the relationship of the PMQ to measures of substance use/abuse. Results of the PMQ analyses relative to the MMPI-2 substance abuse subscales (MAC-R, APS, & AAS) were not significant, but they raised important questions about the ways in which the psychological profile of an opioid medication misuser might differ from other types of substance abusers. For example, the MAC-R, designed to predict risk for alcohol abuse, is comprised of non-face-valid items that tap qualities of high energy, sociability, impulsivity and sensation-seeking. The MAC-R’s variance specific to alcohol abuse may therefore be secondary to a general life-style of risky, indulgent behavior [43]. While such
behavior seems consistent with archetypal substance abusers, it does not appear to fit the majority of chronic pain patients, who are more likely to show depression, inactivity and social isolation. This suggests the possibility of different subtypes of opioid misusers among chronic pain patients, including those who would be at risk for “cross-addiction” among a variety of substances, as well as those whose attachment is specific to opioid medications and their capacity to ease emotional, as well as physical, suffering.

Although endorsements of CAGE items were very infrequent, patients who did endorse one or more CAGE items were significantly more likely to score at the higher end of the PMQ score distribution, as compared to those who endorsed no CAGE items. Similarly, patients who acknowledged a history of drug or alcohol abuse were significantly more likely to score at the higher end of the PMQ score distribution. While limited by the small sample size and possible response bias of the missing data, these findings might suggest a trend that warrants further investigation.

**Concurrent Validity**

Studies examining the PMQ’s concurrent validity were restricted by the limited availability of relevant criterion measures and by small sample size; however, some preliminary data appeared to point to intriguing trends in the PMQ’s relationship to various indices of opioid misuse. Patients scoring high on the PMQ were rated by their physicians (via the PRA) to show, at initial medical intake, significantly more behaviors and attitudes thought to be related to risk for opioid misuse, relative to low PMQ scorers. The impact of criterion contamination on these analyses was mitigated by the fact that physicians had no knowledge of the patients’ PMQ scores when making their initial risk assessments.
Also interesting were results showing PMQ scores of patients with a known history of opioid misuse to have significantly higher mean PMQ scores, relative to those with no known history of opioid misuse. While these results show an intriguing trend, further studies with a much larger sample of known opioid misusers will be needed to examine the stability of these findings.

Factorial Validity

Overall, these findings suggest that the Whole PMQ measures a more complex panoply of phenomena than does either Factor by itself. This is to be expected, since the Whole PMQ has more items and a broader array of content. In examining a two-factor structure, Factor 2 did not appear to offer improvements in validity and reliability over the Whole PMQ. However, the same did not hold true for Factor 1. The unique benefits of Factor 1 suggested that a subset of PMQ items might exist that performs sufficiently well in measuring the full range of opioid misuse risk factors (e.g., psychosocial vulnerability, physical distress), while more effectively reflecting the variance specific to substance abuse. Thus, one approach to refining the PMQ might be to maintain it as a single-scale instrument, comprised of a smaller number of high-performing items. Alternatively, more extensive investigation into multiple factors might yield a clearer set of subscales. Further analyses are warranted to clarify the precise factor structure of this Instrument.

Conclusions and Future Directions

The continued development of this instrument offers several avenues of investigation, many of which are underway. More detailed item-analyses will need to be conducted to identify
the subset of items that perform best in identifying opioid misusers. As items are eliminated or changed, issues of reliability and factorial validity will need to be re-examined. Discriminant validity will need to be addressed by examining the amount of variance accounted for in PMQ scores by clear episodes of opioid misuse (controlling for general maladjustment to pain.) While preliminary data were offered in this study with a small subgroup of known opioid misusers, a larger sample size is needed for more in-depth analyses. Following such a sample over time will be important in distinguishing the chronic misusers of opioids from patients (i.e., the “pseudo-addicted”) whose overuse of opioids would presumably decline with treatment adjustments. Moreover, the variables of type and dosage of opioid medication will be important to consider.

Predictive validity of the PMQ will need to be explored by examining the ability of this instrument to anticipate future instances of opioid medication misuse. Physicians’ assessments of misuse, given ongoing patient contact, will continue to be an important criterion for measuring this phenomenon, so the PRA will require more extensive psychometric investigation. Predictive validity studies might also be greatly enhanced by the use of tangible behavioral and physiological measures, including requests for early prescription refills, incidences of lost prescriptions and medication, and urine drug screening. This study relied on self-report indices of substance abuse, such as the MMPI-2 subscales and the CAGE, which yielded limited information and which are not fully appropriate validation measures for opioid misuse. Indeed, clear-cut validation measures do not yet exist for this construct, so identifying appropriate criterion measures will need to be an evolutionary process.

A related target for subsequent research would be the potential heterogeneity of “opioid misusers” among the chronic pain population. While some patients in the at-risk group might show the classic, overt signs of substance abuse, other patients are likely to show subtler signs
that are more difficult to “parcel out” from the broad range of pain-related correlates. This demands that future studies investigate the potential for different subtypes among opioid misusers and examine their patterns of responding on the **PMQ**.
REFERENCES


