



Published in final edited form as:

J Pain. 2009 September ; 10(9): 932–943. doi:10.1016/j.jpain.2009.03.007.

Development and Preliminary Testing of a Computerized Adaptive Assessment of Chronic Pain

Milena D. Anatchkova, Ph.D., Renee N. Saris-Baglana, Ph.D., Mark Kosinski, M.A., and Jakob Bjorner, M.D, PhD

QualityMetric Incorporated

Abstract

The aim of this article is to report the development and preliminary testing of a prototype computerized adaptive test of chronic pain (CHRONIC PAIN-CAT) conducted in two stages: 1) evaluation of various item selection and stopping rules through real data simulated administrations of CHRONIC PAIN-CAT; 2) a feasibility study of the actual prototype CHRONIC PAIN-CAT assessment system conducted in a pilot sample. Item calibrations developed from a US general population sample (N=782) were used to program a pain severity and impact item bank (k=45) and real data simulations were conducted to determine a CAT stopping rule. The CHRONIC-PAIN CAT was programmed on a tablet PC using QualityMetric's Dynamic Health Assessment (DYHNA®) software and administered to a clinical sample of pain sufferers (n=100). The CAT was completed in significantly less time than the static (full item bank) assessment (p<.001). On average, 5.6 items were dynamically administered by CAT to achieve a precise score. Scores estimated from the two assessments were highly correlated (r=.89) and both assessments discriminated across pain severity levels (p<.001, RV=.95). Patients' evaluations of the CHRONIC PAIN-CAT were favourable.

Perspective—This report demonstrates that the CHRONIC PAIN-CAT is feasible for administration in a clinic. The application has the potential to improve pain assessment and help clinicians manage chronic pain.

Keywords

Chronic Pain; Item Response Theory; Computer Adaptive Testing; Pain Assessment

Chronic pain, defined as persistent or intermittent pain lasting at least three months, is common, produces substantial burden, and its treatment outcomes vary. Prevalence rates for the condition range from 2 to 40% worldwide^{9,10,17,18,25,49,69} An estimated 50 million people in the US are affected by chronic pain⁵³ causing substantial impact on the individuals' health-related quality of life (HRQOL);^{5,15,16,23,34,35,50} and a tremendous socioeconomic burden with annual US cost estimates ranging from \$40 to \$220 billion^{1,52,59,62,66}.

Pain is a uniquely subjective experience and there are no laboratory tests or external measures to assess its frequency, location, intensity, and impact. Clinicians must turn to patient reports when evaluating a pain treatment. There are a number of tools for measuring the severity and

© 2009 The American Pain Society. Published by Elsevier Inc. All rights reserved.

Author of correspondence Milena Anatchkova, PhD. Outcome Research Scientist QualityMetric Incorporated (QM) 24 Albion Road, Bldg 400 Lincoln, RI 02865 USA Phone: 401 334 8800, ext. 244 Fax: 401 334 8800 Email: manatchkova@qualitymetric.com.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

HRQOL impact of chronic pain, ranging from single-item Visual Analogue Scales (VAS)³¹ to long and short forms of generic or disease-specific categorical rating scales⁶⁷. Each type of tool has clear advantages and disadvantages. For example, VAS measures are practical but may not yield scores that are precise enough to distinguish between patients at particular levels of the pain severity continuum. While practical, short forms are often limited in range due to “ceiling” and “floor” effects and lack the precision to detect changes in individual scores. Alternatively, longer, comprehensive fixed-item surveys may provide more precise scores but are more costly to administer and more burdensome to respondents.

A promising strategy that can address these problems is the use of item response theory (IRT) and computerized adaptive testing (CAT) for the development of pain assessment tools^{4,8,26,54,71,78}. In combination, IRT and CAT can lead to the development of more practical and precise assessments over a wide range of pain impact and severity levels - eliminating “ceiling” and “floor” effects and making the comparability of scores from “static” and dynamic measures possible.

In the area of pain assessment there have been a number of studies that used computerized versions of existing measures^{6,29,32,39,68} or electronic diaries^{22,33,42,51,63}. Reports suggested equivalence between computer and paper and pencil modes of administration^{29,39} along with evidence that participants accept computerized assessment with ease and often prefer it to traditional measures^{12,51}. There have also been initial reports on the development of pain related item banks^{26,40} and simulated CATs^{27,28}, including the work of the Patient-Reported Outcomes Measurement Information System (PROMIS)^{13,21}. To the best of our knowledge there has been only one report on the results of a clinical validity field test of a pain CAT for patients with back pain³⁶. In this study the CAT was administered via the Internet, focused specifically on back pain and did not include a patient report.

As the first prototype system combining a CAT-based pain assessment and a patient report, CHRONIC PAIN-CAT brings the field a step closer to the practical use of this new technology in clinical settings. We developed the CHRONIC PAIN-CAT assessment using existing IRT calibrations of widely-used pain-specific items⁷ and used CAT software to select and administer the most informative and relevant items to each patient.

The goals of this paper are to report the results of two studies:

1. the evaluation of various item selection and stopping rules from real data simulated administrations of CHRONIC PAIN-CAT²;
2. a feasibility study of the actual prototype CHRONIC PAIN-CAT assessment system conducted in a pilot sample of chronic pain sufferers. The results focus on item usage, respondent burden, range of levels measured, sensitivity of the instrument to discriminate between groups, score accuracy in comparison with a full-length survey, the evaluation of the user acceptance of CAT administrations relative to a full-length survey, and patient experience using the prototype report⁵⁷.

Materials and Methods

Item bank

The *Chronic Pain Impact Item Bank* contains 45 items from widely-used instruments: the SF-36⁷⁶, SF-8⁷⁴, Nottingham Health Profile³⁰, Sickness Impact Profile²⁴, McGill Pain Questionnaire⁴⁵, Oswestry Low Back Pain¹⁹, Brief Pain Inventory¹⁴, Aberdeen Back Pain⁵⁶, EuroQOL⁶⁵, Health Insurance Experiment⁶¹, Fibromyalgia Impact Questionnaire¹¹, and Short Musculoskeletal Function Assessment⁶⁴. Based on literature review, we identified four subdomains of chronic pain: a) intensity, b) frequency, c) experience, and d) impact of pain on

function and well-being. We selected items to achieve: 1) representation of items across the 4 subdomains; 2) inclusion of items assumed to be most appropriate for patients with severe pain as well as items assumed to be most appropriate for patients with mild pain, 3) reduction in content overlap, 4) high item quality (clarity, simple everyday language and concepts, avoidance of double negatives etc.) and 5) representativeness from various surveys to enable cross-calibration of items. A sample of approximately twenty percent of the items from each tool was included and none of the cited tools has been used in its entirety. Analysis of dimensionality showed that the domain of pain experience (e.g. stabbing pain, burning pain) did not fit a unidimensionality model, and these items were subsequently excluded from the item bank. All items were scored so that low scores indicate a low level of functioning and high level of pain impact.

Item responses are scored by IRT methods. The final scores are calibrated to a metric that has a mean of 50 and standard deviation (SD) of 10 in the US general population. With norm-based scores (NBS), all scores above or below 50 can be interpreted as above or below the population norm, and each point is one-tenth of a SD unit, which facilitates estimates of effect size. NBS can be applied across measures (e.g., dynamic, static), so that comparisons are more meaningful and results are simpler to interpret in relation to population norms. Theoretically, there is no limit on the measurement range for scales scored using NBS, although 95% of scores fall within ± 3 SD units.

Study 1. Simulation study

Sample

Evaluation of the accuracy of the IRT-based algorithms for CAT administrations was conducted using computer simulation methods and real data. The pain item bank data collection was part of a major national survey of core health domains (total N = 12,050 for internet and phone interview administration). We analyzed data previously collected from a US general population survey of respondents with chronic pain (N= 782). The data were collected in 2000 by a national polling firm contracted as part of a larger study to survey the functional health and well-being of adults in the general US population. Data used in this study was collected via the Internet (AOLs Opinion Place). For comparison purposes, additional data on the same items was collected through personal telephone interviews, using random digit dialing (N=750). To avoid confounding with mode of administration, the IRT calibration and CAT simulation reported here are based only on the internet data. Screening questions were asked to achieve sampling quotas for age and gender. Internet respondents were given AOL Reward Points for completing surveys.

Data from the 782 internet participants were used in the simulation studies reported here and was previously used to develop the preliminary item calibrations. Additional details regarding sampling and data collection methods are documented elsewhere⁷². Briefly, 55% of participants were female, 14% were non-white and 27% were older than 54 years. The majority of respondents (54%) worked full-time and 49% reported incomes over \$45,000. Participants suffered from migraine (21.2%), headaches (22%), back pain (24.6%), rheumatoid arthritis (4.9%) and osteoarthritis (7.3%). Each respondent was assigned to one generic functional health or well-being domain (n = 800–1,000/domain) and administered the SF-36® Health Survey, items from that domain and up to 80 additional items from widely-used instruments hypothesized to measure the same domain, along with variables selected for purposes of validation.⁸

Procedure

To compare the psychometric merits of alternative strategies for programming CAT assessments, responses to questions selected by the CAT software are “fed” to the computer to simulate the conditions of an actual CAT assessment. Without “knowing” the response to any other item in the bank the computer uses the IRT model to select the item with the highest information function, given the patient's current score level. Given the answer to that question, the computer re-estimates the pain score and the confidence interval (CI) around that score and decides whether or not to continue testing. The selection of subsequent items in the simulation was also based on maximum information⁴⁸. The information item functions were calculated for all items using a generalized partial credit mode⁴⁷. The IRT scores for the simulation study were estimated using the weighted maximum likelihood method (WML)⁷⁷. While the responses of actual participants were used in this procedure, several CAT algorithms were simulated and evaluated. We refer to this procedure as “real data simulation”. More specifically we tested 1/ four stopping rules without content balancing and 2/ a content balancing algorithm. These algorithms are described below.

Four IRT-based CAT Chronic pain scale scores were estimated by simulating a computerized adaptive assessment based on the same item pool with different fixed lengths (2, 5, and 7 items) and a precision based rule. In the real data simulations the computer read and scored only those responses for the questions that would have been asked during a real computerized adaptive assessment. All versions of the CAT started with a pre-selected item with medium level of difficulty: “How much pain have you had during the past 4 weeks?” This item was followed by 1, 4 or 6 additional items, respectively, for the three different fixed length CATs. The total IRT (full bank) score was based on the administration of the entire 45-item bank.

In order to determine a stopping rule that would provide optimal balance between score precision and respondent burden we determined cut-off points that would be the indicators for adjustment in the levels of precision necessary in a CAT administration. The cut-off points were intended to reflect the change in precision levels at the extreme ends of the scale. Through a series of simulations done in an iterative process, the CAT stopping logic was set to use 6 different levels of precision across the score continuum without limiting the number of items to be administered. The selected score cutoff points were 30, 45, 50, 55, 62 with precision levels for each interval of 5.5, 3.6, 4.5, 5.5, 7.5 and 12.7 respectively. This precision-based simulation was also compared to the full bank score estimates and the previous scores derived by the fixed-length CAT simulations. None of these algorithms took into consideration the specific content of the selected items.

We also evaluated procedures that take into account content balancing. In the calibrated item bank items could be classified as assessing mainly pain severity, pain frequency or pain impact. In additional simulations of a 7-item CAT we ensured that each CAT assessment included items from each of these 3 subdomains. The CAT achieved this content balancing using the following procedure. The start item was set to assess pain severity, the CAT selected the second item as the most informative item within the domains of pain frequency and pain impact, without evaluating pain severity items. For the third item, the CAT chose the most informative item from the subdomain that had not yet been selected. After the selection of three items, all three subdomains had been covered, and selection of the following fourth item was performed across three subdomains based on the highest information function, the fifth item was selected from the 2 remaining subdomains and so forth.

All simulations used a version of the DYNHA® software⁷² programmed in SAS V8.01⁸. DYNHA's computerized adaptive technology uses modern measurement methods based on item response theory (IRT) to calibrate large item banks and provides estimates of measurement precision for each level of functioning. DYNHA software selects and scores only those items

required to calculate a precise score for that individual and places them on a common metric that allows comparability of scores across assessments.

Analysis plan

We used these simulations to: a) evaluate the correspondence between scores estimated from fixed form assessment and scores from simulations of computerized adaptive administration of the pain item bank, using correlation analyses, scatterplots and examining delta scores defined as the difference between the fixed form score and the scores derived from the simulated briefer assessments; b) provide estimates of the reduction in respondent burden by examining the reduction in number of items required to calculate a score and time to complete the assessment; c) define stopping rules for the CHRONIC PAIN-CAT; d) explore results at the individual level by examining individual test scores and confidence intervals at several score levels; and e) examine the discriminant validity of simulated CATs using t-tests and relative validity. We computed fixed form scores using the entire pain item bank (Full bank score), as well as separate scores for the three subdomains of pain severity, pain frequency, and pain impact.

In order to explore the discriminant validity of the simulated measures we selected participants with chronic conditions associated with pain (migraine, back pain, osteoarthritis, rheumatoid arthritis, headaches) and compared their scores to participants who did not suffer from any of the selected conditions using t-tests and relative validity coefficients. To estimate the validity of the CAT scores in discriminating among those with and without pain conditions relative to the Full bank score, relative validity was evaluated as in other studies^{37,43}. T-tests were conducted to test for differences in scores between groups and t values were transformed into F statistics. For each comparison, relative validity (RV) estimates were obtained by dividing the F statistic of the comparison CAT measure by the F statistic for the measure with largest F statistic (full item bank). The F-statistic for a measure will be larger when the measure produces a larger average separation in scores for groups being compared or has a smaller within-group variance, or both. The RV coefficient for each measure in a given test describes, in proportional terms, the empirical validity of that scale, relative to the most valid scale in that test.

Results of simulation test

Descriptive statistics for the full bank score, simulated scores from the 2-, 5-, and 7-item CATs, and precision-based simulated score of an adaptive session are presented in Table 1. It can be seen from the table that the estimates provided using different stopping rules are very similar and have equivalent standard deviations. The precision based rule used, on average, 5 items. The estimate based on the entire item bank has a slightly higher mean. Also, the maximum score was higher for the full bank. The results from the 7-item content balanced CAT were very similar to the results of the 7-item CAT with no content balancing (results not shown).

The correlations between the full bank and the 2–7 item CAT scores were high, ranging from .90 – .99 for 2–7. The scatterplots demonstrated considerable agreement between scores that increases with the number of items fielded (Figure 1). The agreement was strongest in the low and mid-score range and weakest in the high score range. The estimated scores derived through the precision based stopping logic were also highly correlated with the full bank score and were comparable to the performance of the fixed-item CATs' scores, but allowed greater flexibility for a more detailed assessment of patients with extreme scores.

To assess agreement between scores we examined delta scores. As could be expected the largest difference of 2.05 (sd=4.53) was observed between the full bank scores and the simulated CAT with 2 items. Delta scores for all other simulated CATs were comparable and smaller (CAT-5

Mean = 1.26, sd = 2.95; CAT-7 Mean = 1.15, sd = 2.44; Precision Rule Mean = 1.34, sd = 3.06). All mean differences were significantly different from zero. Comparing scores from the 7-item CAT with subdomain specific scores, the CAT scores were more similar to scores on pain impact (Mean delta= -0.41, sd = 4.32) and pain frequency (Mean delta= 0.94, sd = 3.20) and slightly less similar to pain severity subdomain scores (Mean delta= 1.63, sd = 7.36).

The magnitude of error across different score levels was also compared between the full bank score and the three fixed-item CATs (see Figure 2). As expected, the magnitude of error was the largest at the extreme ends of the scale. The two-item CAT produced larger 95% confidence intervals ($\pm 1.96 * SEM$), but the 5- and 7-item CATs compared favorably to the full bank score, although the error rate was higher across all scores ranges.

It was expected that the full item bank would be most valid in distinguishing between those with and without pain-related conditions. We expected the CAT scores to perform relatively well in comparison to the full bank scores derived from the full item bank but expected the CATs to be somewhat less precise. Results indicated that all measures had good discriminant validity and successfully differentiated between the two groups. As expected, the full bank estimate yielded the largest F value, and the 2-item CAT the lowest. The other three assessments achieved comparable results (see Table 2).

To examine the performance of the simulated CATs at the individual level several cases selected from different score levels were examined. As could be expected from the other results, cases in the 35–55 range had better agreement between assessments and narrower confidence intervals compared to cases in the high end of the continuum (see Table 3).

Based on these findings it was determined that the precision based stopping rule is the best to use in the CHRONIC PAIN-CAT application, as it was concordant with the full-length assessment, provided reductions in respondent burden that were comparable to the fixed-length simulations, and allowed for longer assessments when needed. On average 5–6 items were fielded when applying the precision based rule, leading to an expected reduction in respondent burden of approximately 88% in the field test. For the initial feasibility studies, we did not use content balancing.

Study 2. Feasibility Study

Sample

For the feasibility study, which tested an actual CAT administration, we recruited a convenience sample of adult pain sufferers (n=100) from the Dartmouth-Hitchcock Medical Center's Pain Clinic (DHMC). The study coordinator identified patients who met eligibility criteria (English-speaking adult, 18 years or older) and they were asked to participate in the study by their clinicians. The study coordinator explained the study and presented the consent information in writing and verbally to patients, obtained a signed consent form, and administered the instrument.

Participants were randomly assigned to one of two different orders of presentation of the measures, in which the sequence of the dynamic and the full-length pain-specific assessments was counterbalanced: half the sample answered the dynamic assessment first and half the sample answered the full-length assessment first. The post-assessment evaluation was administered only after the first pain assessment (either dynamic or static), so that half of participants (n=50) evaluated the dynamic form and the other half (n=50) evaluated the static form. Upon completion of the survey, participants reviewed a personalized feedback report and completed the report evaluation survey to provide feedback on the clarity, usefulness and

general outlook of the report. All procedures and the study consent process were reviewed and approved by the New England and the DHMC Institutional Review Boards.

Instruments—The application was programmed for administration via a tablet PC computer with Microsoft XP Tablet Edition operating system with a stylus pen using QualityMetric's Dynamic Health Assessment (DYNHA®) software. The application combined into one seamless assessment all the survey elements described below.

DYNHA® Chronic Pain Impact Survey (dynamic assessment): A computerized adaptive survey based on the 45-item *Chronic Pain Impact* bank as described previously. We selected a precision based stopping rule for the DYNHA® Chronic Pain Impact Survey based on the results of the simulations conducted as part of this study. The general logic of the CAT algorithm is presented in Figure 3.

The Full Bank Chronic Pain Assessment administered all 45 items from the Chronic Pain Impact bank.

SF-12v2™ Health Survey: The SF-12v2™ Health Survey is a 12-item questionnaire that measures the same eight domains of health as the SF-36® Health Survey (e.g. Physical Functioning, Role Physical, Bodily Pain, General Health, Vitality, Social Functioning, Role Emotional, and Mental Health). The SF-12v2™ Health Survey yields physical and mental summary scores that are comparable with those from the SF-36® Health Survey in all studied populations⁷⁵. As a brief, reliable measure of overall health status, the SF-12v2™ Health Survey has been used in a variety of studies as an assessment tool^{20,41,44,46,55,60} reaching over eleven hundred references in PubMed.

Level of pain severity was assessed using a numerical graphic rating scale (0–10) with the anchors of “No pain” and “Pain as bad as it could be”. The ranges to define mild, moderate and severe pain were 1–3, 4–6 and 7–10 respectively³.

Patient User's Evaluation Survey was constructed to obtain a standardized evaluation of each patient's experience completing the assessment. It was administered immediately following the first chronic pain assessment (dynamic or static depending on the randomization group). The evaluation survey included five-point Likert scale questions on helpfulness of the tool in understanding the impact of pain, relevance of the items, difficulty of completing the assessment and willingness to complete the survey again. One question that assessed the appropriateness of survey length had a dichotomous response format. The qualitative field note observations of the administrator provided further information on the patients' experience.

User's interface and procedure—The CHRONIC PAIN-CAT application begins with a brief description of the study and a log-in procedure. Upon successful login and granting consent, users are provided with standardized instructions and asked to complete the survey.

Survey components were programmed to be administered in the following order for the first randomization group: Rating of Pain Severity Item (VAS), DYNHA® Chronic Pain Impact Survey (followed by Patient User's Evaluation Survey), full-length static item bank, SF-12v2™ Health Survey, and additional modules. For the second group, the order of the dynamic and full-length assessments of chronic pain impact was reversed. Items for the dynamic and static assessment of chronic pain were administered individually (one per screen). Based on the simulation results, we set the CAT stopping logic to use a precision based stopping rule for the dynamic assessment. When the set precision level was achieved, the dynamic assessment was completed. The SF-12v2™ Health Survey, demographic items, and user's evaluation items were presented in grids on the screen to reduce respondent burden.

A meter on the bottom of the page indicated the percentage of items completed by the patient. This estimate changed with each subsequent section of the application to reflect the patient's progress toward completing the survey.

The CHRONIC PAIN-CAT application also included a prototype patient report, presented to participants upon completion of the surveys. The report provided information on: (a) patient results for pain severity and impact in relation to a normative comparison group; (b) patient results for general physical and mental health in relation to a normative comparison group; (c) a section on individual score changes in these areas, allowing longitudinal tracking of results; and (d) a brief interpretation of the results. The information in the report was organized into six areas, separated by explanatory titles. The scores were presented graphically and accompanied by a text clarifying the meaning of the graph. Patients could view the report on the screen of the tablet before providing feedback.

Statistical Analyses—As part of the feasibility test several features of the CHRONIC PAIN-CAT were evaluated including item usage, respondent burden, range of measured levels (ceiling and floor effects), and a preliminary test of validity in discriminating across chronic pain severity levels.

Respondent burden was assessed by comparing the number of items required to calculate a score and the average amount of time (in seconds) per administration of the dynamic and static assessment using t-tests. Counterbalancing in the order of the dynamic and static assessment administrations allowed us to conduct both within group and between group comparisons.

In order to evaluate “ceiling” and “floor” effects we examined the data from the CAT for cases where all administered items received the highest or the lowest score.

To evaluate measurement accuracy we examined the descriptive characteristics and the plots of 95% confidence intervals ($\pm 1.96 * SEM$) against NBS scores for the two assessments. We also computed delta scores as the difference between scores based on the 45 item bank and the dynamic theta estimate for each respondent, computed the mean of these scores and tested it against the null hypothesis of no difference. Finally we examined some cases at different levels of theta to explore performance of the assessments at individual level.

Discriminant validity was evaluated by examining the ability of the full bank and dynamic pain assessments to distinguish between groups of patients with various degrees of pain severity determined through the numeric rating scale. We used general linear models to evaluate the impact of severity and order of presentation on the final scores.

Results of feasibility study

Sample characteristics

Data were collected during a four month period in the fall of 2006. The sample (N=100) was primarily comprised of White (94.6%), non-Hispanic (98%), middle-aged women (63%). The majority of participants were married (59%) and 25% had previously participated in studies involving computerized surveys. Most participants had received some training after school or attained a college degree (66%) and 67% reported severe pain.

Item Usage

On average, the dynamic survey was typically completed in 5–6 items (range: 4 – 19). Participants with mild to moderate levels of pain severity were administered a maximum of 6 items per assessment. A single administration in the severe group required 19 items. Abbreviated item stems and the frequency with which each item was administered by

DYNHA® are shown in Table 4. The dynamic assessment was programmed to administer the same first item (“global item”) to each participant. Of the remaining 44 items in the bank, 27 items (61%) were selected and administered by the CAT system. The 5 items that were administered with the highest frequencies across all severity groups were related to impact of pain on the ability to perform or restriction on routine daily activities.

Respondent burden

As expected, participants completed the dynamic assessment in significantly less time than the static assessment of the full item bank ($t(94) = 22.5, p < .001$). The average completion time for the dynamic assessment was less than a minute and a half ($M = 80.6s, SD = 34.9s, range = 39.9–229s$). In comparison, the average completion time for the static assessment was approximately 10 minutes ($M = 603.6s, SD = 239.6s, range = 341–1795s$). The dynamic assessment used, on average, 5.57 items, which represents a decrease in respondent burden of approximately of 88%.

Range of Levels Measured – “Ceiling” and “Floor” Effects

None of the participants in the clinical feasibility study scored at the “ceiling” (highest observable score based on administered items) or the “floor” (lowest observable score based on administered items) on static and dynamic assessments.

Score Accuracy and Discriminant Validity

At a group level, dynamic assessment achieved roughly equivalent measurement precision (or accuracy) (95% CI) in score estimation with substantially fewer items administered than the full-length assessment (see Table 5). The mean of the delta scores (difference between full bank and dynamic scores for each respondent) was .3 ($sd=2.58$) and was not significantly different than zero, indicating no difference in the two methods of estimation. Individual dynamic case scores in the available score ranges demonstrated good correspondence with full bank scores (see Table 3).

Measurement precision also was evaluated by plotting the confidence intervals against NBS scale scores. Figure 4 shows that for both measures, precision was greatest at the score range of 30 – 45. Also, both measures had less precision in the lower score range. As expected, full bank scores were more precise than scores based on the dynamic assessment but the two scores were highly correlated ($r=.89$).

In this sample, 4% reported mild, 29% reported moderate and 67% reported severe pain. We expected that the group of participants with more severe pain would have lower mean scores on the pain assessments (indicating a lower level of functioning and higher level of pain impact). Although the sample sizes are uneven and the mild group had only 4 participants, both dynamic and static assessments discriminated between each pain severity level; pain impact scores decreased monotonically with increased levels of pain severity. Also, within each of the three severity groups, mean scores on the dynamic and static assessments were roughly equivalent (Table 6). The order effect was non-significant for both the static ($F(1, 91) = .05, p > .05$) and the dynamic assessment ($F(1, 92) = .17, p > .05$).

Participants’ Evaluations

The overall evaluation of the assessment was positive – the majority of participants found the measure to be useful, relevant, of appropriate length and easy to complete. There were no significant differences between the evaluations of the dynamic and static surveys.

Participants also completed a report evaluation survey to assess how helpful and understandable the report is and whether it would be used to initiate a discussion with a

physician. These evaluations were predominantly positive as well with no significant differences in the evaluations of the dynamic and static survey reports.

The observations and field notes taken by the study coordinator provided additional valuable information regarding the acceptance of the CHRONIC PAIN-CAT administered via tablet PC by this specific patient population. Eighty-five of the participants in the study provided comments and suggestions for improvement, which were recorded in the field notes. Examination of this qualitative data revealed several areas for improvement. Almost half of the participants expressed some confusion related to the feedback report, ranging from poor formatting and feeling confused by the presentation of data to several accounts of feeling depressed by the feedback. Some technical features related to the use of the tablet PC (stylus and the need to scroll down) as a platform also elicited some negative feedback. Almost half of the sample had difficulty scrolling down the pages and about one fourth reported frustration with the use of the stylus. The pop-up screens with messages regarding problems with administration were also confusing for some. Finally, some participants were concerned with the repetitiveness of the questions, an expected reaction given the design of this study, where all participants answered both dynamic and static assessments, resulting in an overlapping content. Based on the field observations of the study coordinator, it also became apparent that the use of tablet PC as a platform has some specific limitations for people who are left handed, using a wheel chair or have visual problems. These findings will be considered in our future efforts to improve the application.

Discussion

We sought to achieve major innovations in the standardization of chronic pain assessment and in the technology used to collect, process and display results through the development of a computerized chronic pain assessment. The result of this effort was a prototype patient-based CHRONIC PAIN-CAT assessment, with demonstrated evidence of administrative feasibility and psychometric performance. The dynamic assessment of pain, completed in 80 seconds on average, provided a substantial reduction in respondent burden compared with the full-length item bank, while on average, providing very similar scores. In our feasibility test, the mean dynamic and static scores were very similar within groups of patients with mild, moderate and severe pain and demonstrated the comparability of the two assessments. Likewise, both versions discriminated between respondents with differing levels of pain severity. The DYNHA® Chronic Pain Impact Survey used 61% of all available bank items, with a higher number of items required for participants with severe pain. These findings suggest that the DYNHA® Chronic Pain Impact Survey is a viable alternate to the full static survey form for administration in settings where longer forms are not desired.

In the initial stages of the work a series of simulations were conducted to examine the performance of various stopping rules to be implemented in the final application. Although dynamic assessments produced scores with lower precision than full-length static assessment, on average, only 5–6 tailored items were needed to reach acceptable levels of precision, representing an 88% reduction in respondent burden. All assessments were most accurate in the middle score range of the scale and showed substantial increases in measurement error towards the lower and upper end points. These results support the potential of CATs to lower respondent burden for patients but also indicate possible room for improvement in the item bank through inclusion of additional items covering the more subtle and more severe levels of pain impact. The use of norm based scoring facilitates the improvement of item banks, since it allows inclusion of additional items without leading to changes in the interpretation of final results. Comprehensive, high quality item banks covering a wide range of the construct's continuum will allow clinicians to take advantage of the ability of CATs to improve precision of assessment through narrow confidence intervals in the cases where clinical decisions need

to be made. The potential of precise assessment in combination with capability for immediate reporting of results to patients and clinicians can turn computerized tests into valuable tools of clinical practice.

The results of the user evaluation indicated that the majority of the study participants favorably evaluated the dynamic assessment in terms of its length, difficulty, relevance and appropriateness. However, there were no significant differences between groups evaluating the dynamic and static assessments. In both groups, the predominant evaluation of the report was also positive, but some respondents reported difficulty with the interpretation of the report and doubts regarding its usefulness. In addition, the experience of participants was evaluated through the study coordinator's field notes, which revealed some specific problems with the administration of the tool. Some patients experienced difficulties using the stylus, scrolling down the screen when necessary, or inputting numbers. Also, many respondents felt the survey was too long, but this impression is most likely due to the study design which involved completion of both the dynamic and static assessment for all participants.

Limitations

The current study has some limitations. As the study was planned as a feasibility test for the CHRONIC PAIN-CAT, the sample size of this study is rather small and comprehensive testing with a larger and more diverse sample is required. The feedback we received from patients and the clinical staff points to the need for some improvements in the technical solutions used for the delivery of the dynamic assessment. The study also indicated that the patient report used in the study needs substantial improvement to make it more informative and clear for patients. Lastly, there was some degree of mismatch between the intended target population for this tool (patients with chronic pain) and the field sample composition (67% reporting severe pain), which may help explain some of the problems and negative comments in the field. The main concern for patients in the severe stage of a chronic pain condition is to alleviate the immediate discomfort of acute pain. In this state the impact of chronic pain on quality of life is of lesser importance and the use of even brief self-assessment of this impact may be viewed as burdensome or irrelevant by pain sufferers. Regardless of these limitations, the results of the study demonstrate that a dynamic pain assessment system is feasible to administer, and reduces respondent burden with a degree of accuracy that is comparable to a lengthier static pain assessment. Some preliminary evidence of the external validity of the tool was found, but additional work in this area is still needed. Although the tool needs some further improvements, it shows promise for use with chronic pain patient populations in medical practices, clinics and health management organizations.

Future Research

Future research efforts should include a full evaluation of the current pain item bank for measurement gaps and item content (redundancies, relevance) to ensure that the full continuum of pain severity and impact is covered. In the current study not all items included in the bank were selected for administration, suggesting that items initially developed and tested in the tradition of the classical test theory may not necessarily be optimal for dynamic assessments under the IRT model. Future efforts need to focus on additional evaluation of items, extending and fine-tuning the bank by creating new items with the potential to cover the extremes of the pain continuum and eliminating items with potential redundancy. In addition, the technical platform and interface of the tool will need to be evaluated in usability studies with direct input from the potential users of the final product to make the assessment user-friendly. This work should address the use of tablet PC as a platform and the specific presentation of the content on the screen (e.g. no need to scroll, clear instructions). In addition, the target population for this assessment should be clearly defined to avoid burdening patients in acute pain with assessments that are more appropriate for chronic pain sufferers.

In broader context, the CHRONIC PAIN-CAT follows a recent trend of development of computerized adaptive assessments for use in clinical practice across various conditions (e.g. osteoarthritis³⁸, headache⁷³, diabetes⁵⁸). Computerized adaptive testing has some remarkable advantages – it provides a solution to the tradeoff between measurement precision and burden for respondents; allows immediate feedback to patients and clinicians; and reduces resources devoted to coding and scoring assessments³⁸. As CATs have the potential to be administered through alternative platforms (internet, PDA, telephone), they can potentially save resources for traveling and clinic visits for patients with chronic conditions. The increasing availability of computers in households and clinical offices can facilitate the use of CATs in routine clinical practice, which in turn would streamline, simplify and enhance the self-assessment process, ultimately leading to improved health care for millions of patients with chronic conditions. The results of the feasibility study of CHRONIC PAIN-CAT presented in this report bring patients one step closer to these improvements.

Acknowledgments

This research was supported by Small Business Innovation Research Grant #1R43AR052251-01A1 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The authors would like to thank Kevin Spratt, Ph.D., John C. Baird, Ph.D. and John E. Ware, Jr., Ph.D. for their comments and suggestions on an earlier version of this manuscript.

Reference List

1. American Academy of Orthopaedic Surgeons. Musculoskeletal conditions in the United States. American Academy of Orthopaedic Surgeons Bulletin 1999;27:34–36.
2. Anatchkova M, Kosinski M, Saris-Baglana R. How many questions does it take to obtain valid and precise estimates of the impact of pain on a patient's life? The Journal of Pain 2007;8(4 Supplement 1):S72.
3. Anderson, K.; Syrjala, K.; Cleeland, C. How to assess cancer pain.. In: Turk, DC.; Melzack, R., editors. Handbook of Pain Assessment. The Guilford Press; New York, NY: 2001.
4. Bayliss MS, Dewey JE, Dunlap I, Batenhorst AS, Cady R, Diamond ML, Sheftell F. A study of the feasibility of Internet administration of a computerized health survey: the headache impact test (HIT). Qual Life Res 2003;12:953–961. [PubMed: 14651414]
5. Becker J, Schwartz C, Saris-Baglana RN, Kosinski M, Bjorner JB. Using Item Response Theory (IRT) for Developing and Evaluating the Pain Impact Questionnaire (PIQ-6tm). Pain Medicine 2007;8:129–144.
6. Berry DL, Trigg LJ, Lober WB, Karras BT, Galligan ML, Ustin-Seymour M, Martin S. Computerized symptom and quality-of-life assessment for patients with cancer part I: development and pilot testing. Oncol Nurs Forum 2004;31:E75–E83. [PubMed: 15378104]
7. Bjorner JB. Final bodily pain item bank calibrations. QM Internal Report. 2006
8. Bjorner JB, Kosinski M, Ware JE Jr. Calibration of an item pool for assessing the burden of headaches: an application of item response theory to the headache impact test (HIT). Qual Life Res 2003;12:913–933. [PubMed: 14651412]
9. Blyth FM, March LM, Brnabic AJ, Jorm LR, Williamson M, Cousins MJ. Chronic pain in Australia: a prevalence study. Pain 2001;89:127–134. [PubMed: 11166468]
10. Breivik H, Collett B, Ventafridda V, Cohen R, Gallacher D. Survey of chronic pain in Europe: prevalence, impact on daily life, and treatment. Eur J Pain 2006;10:287–333. [PubMed: 16095934]
11. Burckhardt CS, Clark SR, Bennett RM. The fibromyalgia impact questionnaire: development and validation. J Rheumatol 1991;18:728–733. [PubMed: 1865419]
12. Carlson LE, Specia M, Hagen N, Taenzer P. Computerized quality-of-life screening in a cancer pain clinic. J Palliat Care 2001;17:46–52. [PubMed: 11324185]
13. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH

- Roadmap cooperative group during its first two years. *Med Care* 2007;45:S3–S11. [PubMed: 17443116]
14. Cleland CS. The Brief Pain Inventory, a Measure of Cancer Pain and its Impact. *Quality of Life Newsletter* 1994;9:5–6.
 15. Colombo GL, Caruggi M, Vinci M. [Quality of life and treatment costs in patients with non-cancer chronic pain]. *Recenti Prog Med* 2004;95:512–520. [PubMed: 15598088]
 16. Dogra S, Hahn B, King-Zeller S. Impact of chronic pain on quality of life. *Pain Medicine* 2000;1:196.
 17. Elliott AM, Smith BH, Hannaford PC, Smith WC, Chambers WA. The course of chronic pain in the community: results of a 4-year follow-up study. *Pain* 2002;99:299–307. [PubMed: 12237208]
 18. Eriksen J, Jensen MK, Sjogren P, Ekholm O, Rasmussen NK. Epidemiology of chronic non-malignant pain in Denmark. *Pain* 2003;106:221–228. [PubMed: 14659505]
 19. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy* 1980;66:271–273. [PubMed: 6450426]
 20. Fleishman JA, Cohen JW, Manning WG, Kosinski M. Using the SF-12 health status measure to improve predictions of medical expenditures. *Med Care* 2006;44:154–163. [PubMed: 16625065]
 21. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol* 2005;23:S53–S57. [PubMed: 16273785]
 22. Gaertner J, Elsner F, Pollmann-Dahmen K, Radbruch L, Sabatowski R. Electronic pain diary: a randomized crossover study. *J Pain Symptom Manage* 2004;28:259–267. [PubMed: 15336338]
 23. Gerbershagen HU, Lindena G, Korb J, Kramer S. Health-related quality of life in patients with chronic pain. *Schmerz* 2002;16:271–+. [PubMed: 12192436]
 24. Gilson BS, Gilson JS, Bergner M, Bobbit RA, Kressel S, Pollard WE, Vesselago M. The sickness impact profile. Development of an outcome measure of health care. *Am J Public Health* 1975;65:1304–1310. [PubMed: 1200192]
 25. Gureje O, Von Korff M, Simon GE, Gater R. Persistent pain and well-being: A World Health Organization Study in Primary Care. *Journal of the American Medical Association* 1998;280:147–51. [PubMed: 9669787]
 26. Hahn EA, Cella D, Bode RK, Gershon R, Lai JS. Item banks and their potential applications to health status assessment in diverse populations. *Med Care* 2006;44:S189–S197. [PubMed: 17060827]
 27. Hart DL, Cook KF, Mioduski JE, Teal CR, Crane PK. Simulated computerized adaptive test for patients with shoulder impairments was efficient and produced valid measures of function. *J Clin Epidemiol* 2006;59:290–298. [PubMed: 16488360]
 28. Hart DL, Mioduski JE, Werneke MW, Stratford PW. Simulated computerized adaptive test for patients with lumbar spine impairments was efficient and produced valid measures of function. *J Clin Epidemiol* 2006;59:947–956. [PubMed: 16895818]
 29. Heuser J, Geissner E. [Computerized version of the pain experience scale: a study of equivalence]. *Schmerz* 1998;12:205–208. [PubMed: 12799973]
 30. Hunt, S.; McEwen, J.; McKenna, S. *Measuring Health Status* Croom Helm. 1986.
 31. Huskisson, EC. Visual analog scales.. In: Melzack, R., editor. *Pain Management and Assessment*. Raven Press; New York, New York: 1983. p. 33-37.
 32. Jamison RN, Fanciullo GJ, Baird JC. Computerized dynamic assessment of pain: comparison of chronic pain patients and healthy controls. *Pain Med* 2004;5:168–177. [PubMed: 15209971]
 33. Jamison RN, Raymond SA, Slawsby EA, McHugo GJ, Baird JC. Pain assessment in patients with low back pain: comparison of weekly recall and momentary electronic data. *J Pain* 2006;7:192–199. [PubMed: 16516825]
 34. Katz N. The impact of pain management on quality of life. *J Pain Symptom Manage* 2002;24:S38–S47. [PubMed: 12204486]
 35. Kerr S, Fairbrother G, Crawford M, Hogg M, Fairbrother D, Khor KE. Patient characteristics and quality of life among a sample of Australian chronic pain clinic attendees. *Intern Med J* 2004;34:403–409. [PubMed: 15271174]
 36. Kopec JA, Badii M, McKenna M, Lima VD, Sayre EC, Dvorak M. Computerized adaptive testing in back pain: validation of the CAT-5D-QOL. *Spine* 2008;33:1384–1390. [PubMed: 18496353]

37. Kosinski M, Bjorner JB, Ware JE Jr, Batenhorst A, Cady RK. The responsiveness of headache impact scales scored using 'classical' and 'modern' psychometric methods: a re-analysis of three clinical trials. *Qual Life Res* 2003;12:903–912. [PubMed: 14651411]
38. Kosinski M, Bjorner JB, Ware JE Jr, Sullivan E, Straus WL. An evaluation of a patient-reported outcomes found computerized adaptive testing was efficient in assessing osteoarthritis impact. *J Clin Epidemiol* 2006;59:715–723. [PubMed: 16765275]
39. Kvien TK, Mowinckel P, Heiberg T, Dammann KL, Dale O, Aanerud GJ, Alme TN, Uhlig T. Performance of health status measures with a pen based personal digital assistant. *Ann Rheum Dis* 2005;64:1480–1484. [PubMed: 15843456]
40. Lai JS, Dineen K, Reeve BB, Von Roenn J, Shervin D, McGuire M, Bode RK, Paice J, Cella D. An item response theory-based pain item bank can enhance measurement precision. *J Pain Symptom Manage* 2005;30:278–288. [PubMed: 16183012]
41. Lam CL, Tse EYY, Gandek B. Is the standard SF-12 Health Survey valid and equivalent for a Chinese population? *Qual Life Res* 2005;14:539–547. [PubMed: 15892443]
42. Marceau LD, Link C, Jamison RN, Carolan S. Electronic Diaries as a Tool to Improve Pain Management: Is There Any Evidence? *Pain Medicine* 2007;8:101–109.
43. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247–263. [PubMed: 8450681]
44. McKenzie DP, Ikin JF, McFarlane AC, Creamer M, Forbes AB, Kelsall HL, Glass DC, Ittak P, Sim MR. Psychological health of Australian veterans of the 1991 Gulf War: An assessment using the SF-12, GHQ-12 and PCL-S6. *Psychological Medicine* 2004;34:1419–1430. [PubMed: 15724873]
45. Melzack R. The McGill Pain Questionnaire: Major properties and scoring methods. *Pain* 1975;1:277–99. [PubMed: 1235985]
46. Muller-Nordhorn J, Nolte CH, Rossnagel K, Jungehulsing GJ, Reich A, Roll S, Villringer A, Willich SN. The use of the 12-item short-form health status instrument in a longitudinal study of patients with stroke and transient ischaemic attack. *Neuroepidemiology* 2005;24:196–202. [PubMed: 15802924]
47. Muraki E. A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement* 1992;16:159–176.
48. Muraki E. Information functions of the generalized partial credit model. *Appl Psychol Measur* 1993;17:351–363.
49. Nickel R, Raspe HH. [Chronic pain: epidemiology and health care utilization]. *Nervenarzt* 2001;72:897–906. [PubMed: 11789433]
50. Niv D, Kreitler S. Pain and quality of life. *Pain Practice* 2001;1:150–161. [PubMed: 17129291]
51. Palermo TM, Valenzuela D, Stork PP. A randomized trial of electronic versus paper pain diaries in children: impact on compliance, accuracy, and acceptability. *Pain* 2004;107:213–219. [PubMed: 14736583]
52. Panel on Musculoskeletal Disorders and the Workplace CoBaSSaENRC. *Musculoskeletal Disorders and the Workplace: Low Back and Upper Extremities*. The National Academies Press; 2001.
53. Phillips DM. JCAHO pain management standards are unveiled. *Joint Commission on Accreditation of Healthcare Organizations*. *Journal of the American Medical Association* 2000;284:428–9. [PubMed: 10904487]
54. Revicki DA, Cella DF. Health status assessment for the twenty-first century: item response theory, item banking and computer adaptive testing. *Qual Life Res* 1997;6:595–600. [PubMed: 9330558]
55. Rozario PA, Morrow-Howell NL, Proctor EK. Changes in the SF-12 among depressed elders six months after discharge from an inpatient geropsychiatric unit. *Qual Life Res* 2006;15:755–759. [PubMed: 16688507]
56. Ruta DA, Garratt AM, Wardlaw D, Russell IT. Developing a valid and reliable measure of health outcome for patients with low back pain. *Spine* 1994;19:1887–1896. [PubMed: 7997920]
57. Saris-Baglama R, Anatchkova M, Kosinski M. A clinical feasibility study of a computerized adaptive test for chronic pain (CHRONIC PAIN-CAT). *The Journal of Pain* 2007;8(4 Supplement 1):S73.

58. Schwartz C, Welch G, Santiago-Kelley P, Bode R, Sun X. Computerized adaptive testing of diabetes impact: a feasibility study of Hispanics and non-Hispanics in an active clinic population. *Qual Life Res* 2006;15:1503–1518. [PubMed: 17031500]
59. Sheehan J, McKay J, Ryan M, Walsh N, O'Keeffe D. What cost chronic pain? *Ir Med J* 1996;89:218–219. [PubMed: 8996949]
60. Singh A, Gnanalingham K, Casey A, Crockard A. Quality of Life assessment using the Short Form-12 (SF-12) questionnaire in patients with cervical spondylotic myelopathy - Comparison with SF-36. *Spine* 2006;31:639–643. [PubMed: 16540866]
61. Stewart, AL.; Ware, JE., Jr.; Brook, RH. A Study of the Reliability, Validity, and Precision of Scales to Measure Chronic Functional Limitations Due to Poor Health. RAND Corporation; 1977.
62. Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. *Journal of the American Medical Association* 2003;290:2443–54. [PubMed: 14612481]
63. Stinson JN, Petroz GC, Tait G, Feldman BM, Streiner D, McGrath PJ, Stevens BJ. e-Ouch: usability testing of an electronic chronic pain diary for adolescents with arthritis. *Clin J Pain* 2006;22:295–305. [PubMed: 16514331]
64. Swiontkowski MF, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. *J Bone Joint Surg Am* 1999;81:1245–1260. [PubMed: 10505521]
65. The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Dec* 16;1990
66. Thomsen AB, Sorensen J, Sjogren P, Eriksen J. Economic evaluation of multidisciplinary pain management in chronic pain patients: a qualitative systematic review. *J Pain Symptom Manage* 2001;22:688–698. [PubMed: 11495715]
67. Turk, DC.; Melzack, R. Handbook of Pain Assessment. The Guilford Press; New York, NY: 2001.
68. VanDenKerkhof EG, Goldstein DH, Lane J, Rimmer MJ, Van Dijk JP. Using a personal digital assistant enhances gathering of patient data on an acute pain management service: a pilot study. *Can J Anaesth* 2003;50:368–375. [PubMed: 12670814]
69. Verhaak PF, Kerssens JJ, Dekker J, Sorbi MJ, Bensing JM. Prevalence of chronic benign pain disorder among adults: A review of the literature. *Pain* 1998;77:231–9. [PubMed: 9808348]
70. Wainer, H. Computerized adaptive testing: A primer. Vol. Second Edition. Lawrence Erlbaum Associates; Mahwah NJ: 2000.
71. Ware JE Jr. Conceptualization and measurement of health-related quality of life: comments on an evolving field. *Arch Phys Med Rehabil* 2003;84:S43–S51. [PubMed: 12692771]
72. Ware JE Jr, Bjorner JB, Kosinski M. Practical implications of item response theory and computerized adaptive testing: A brief summary of ongoing studies of widely used headache impact scales. *Med Care* 2000;38:73–82.
73. Ware JE Jr, Kosinski M, Bjorner JB, Bayliss MS, Batenhorst A, Dahlof CG, Tepper S, Dowson A. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. *Qual Life Res* 2003;12:935–952. [PubMed: 14651413]
74. Ware, JE., Jr.; Kosinski, M.; Dewey, JE.; Gandek, B. How to Score and Interpret Single-Item Health Status Measures: A Manual for Users of the SF-8 Health Survey (With a Supplement on the SF-6 Health Survey). QualityMetric Incorporated; Lincoln, RI: 2001.
75. Ware, JE., Jr.; Kosinski, M.; Turner-Bowker, DM.; Gandek, B. How to Score Version 2 of the SF-12 (r) Health Survey (With a Supplement Documenting Version 1). QualityMetric Incorporated; Lincoln, Rhode Island: 2002.
76. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83. [PubMed: 1593914]
77. Warm TA. Weighted likelihood estimation of ability in item response theory. *Psychometrika* 1989;54:427–450.
78. Wilkie DJ, Judge MK, Berry DL, Dell J, Zong S, Gillespie R. Usability of a computerized PAINReportIt in the general public with pain and people with cancer pain. *J Pain Symptom Manage* 2003;25:213–224. [PubMed: 12614956]

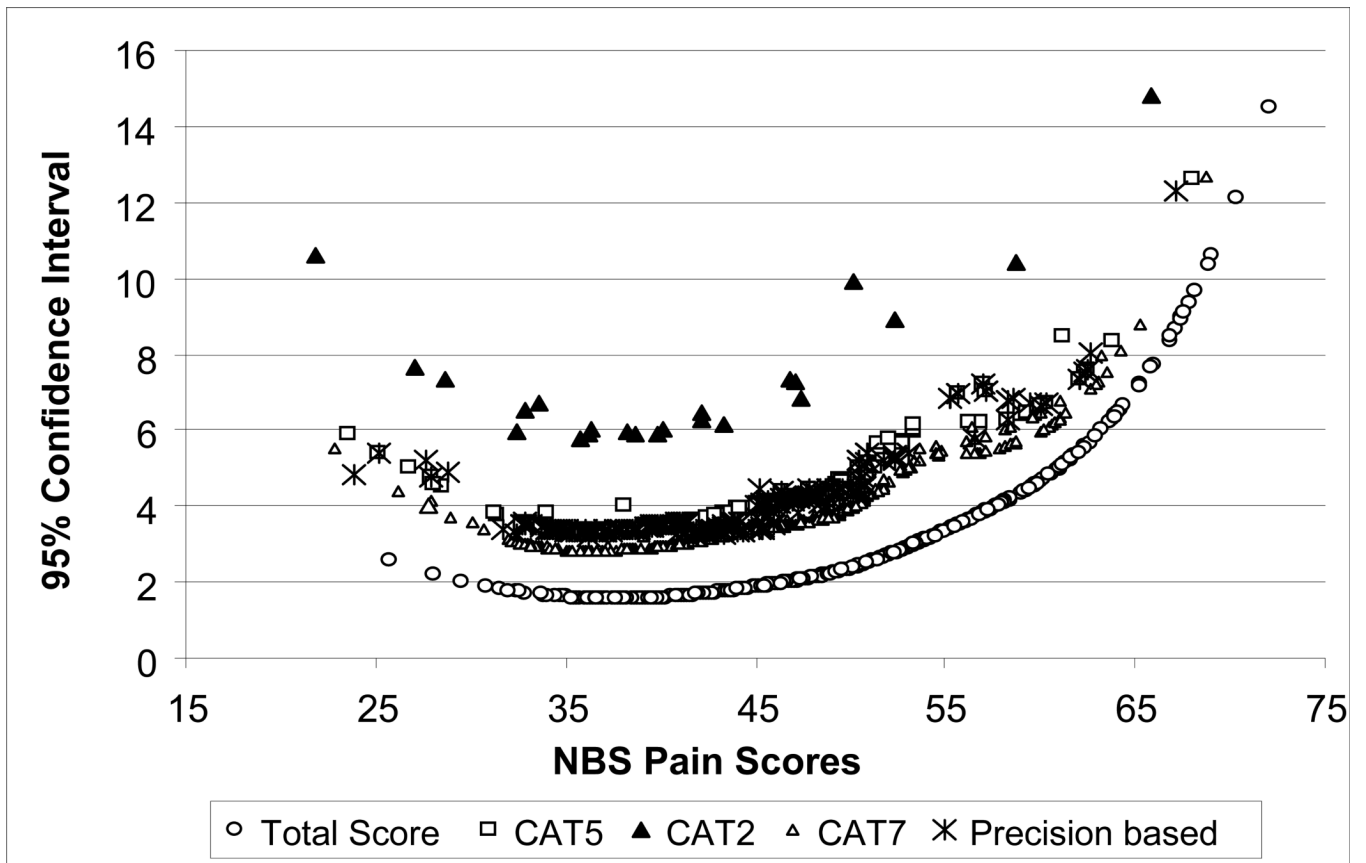


Figure 1.
Scatter plots of Simulated CAT Scores vs. Full Bank Scores

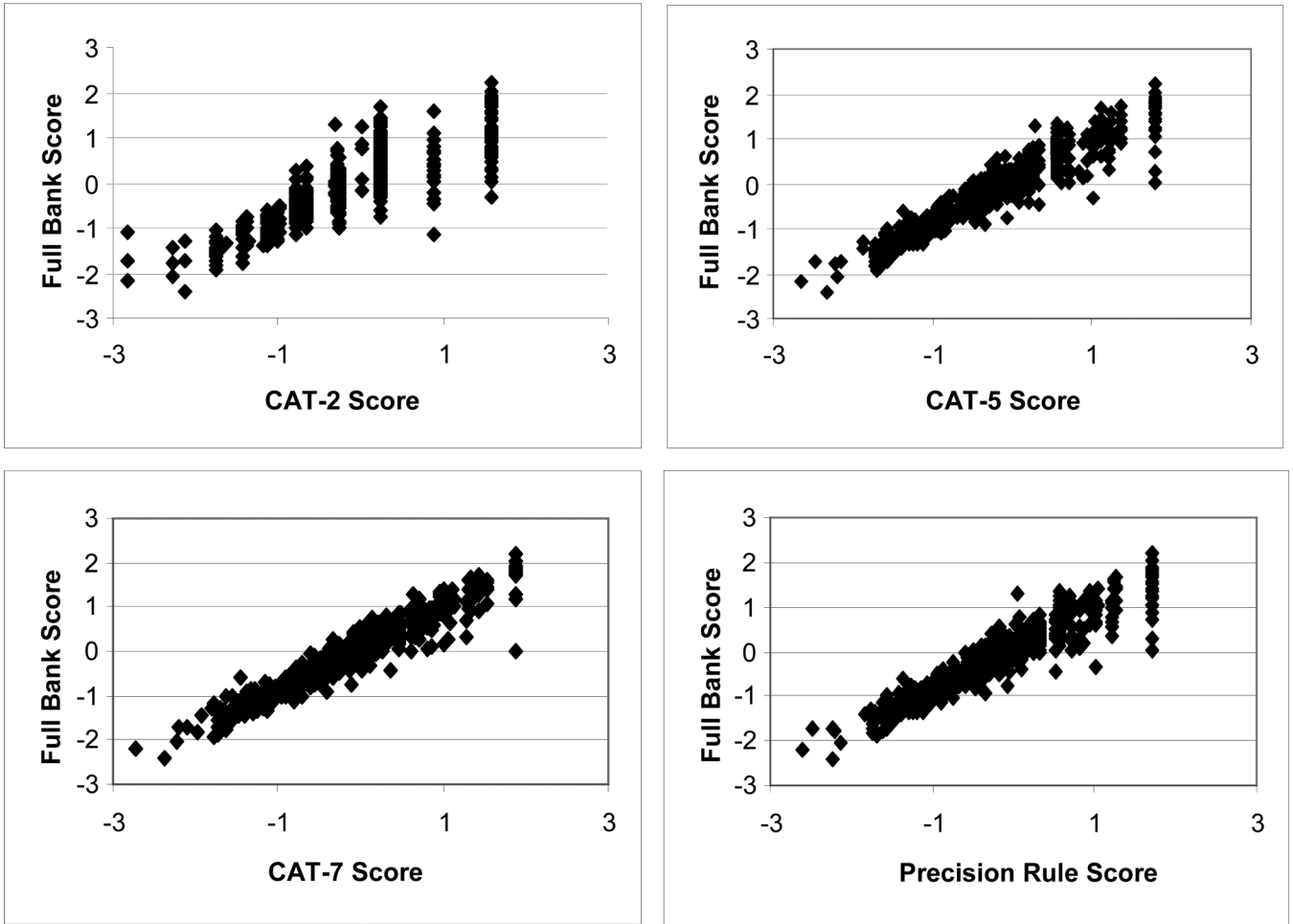
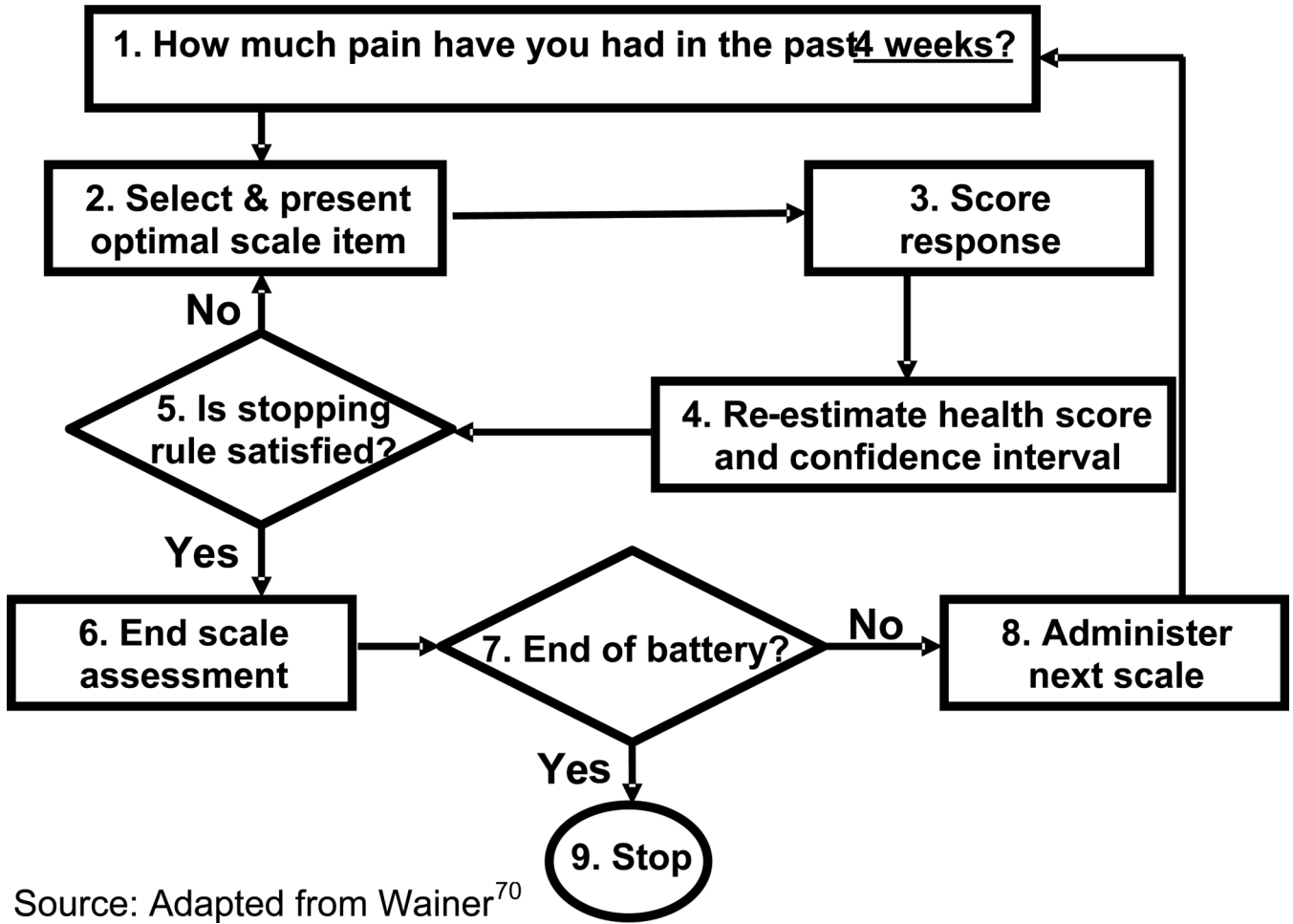


Figure 2. Comparison of 95% CIs for Simulated CATs (with 2, 5, and 7 items), Precision based estimate and Full Bank Scores.



Source: Adapted from Wainer⁷⁰

Figure 3. Logic of Computerized Adaptive Testing (adapted from Wainer et al (2000))

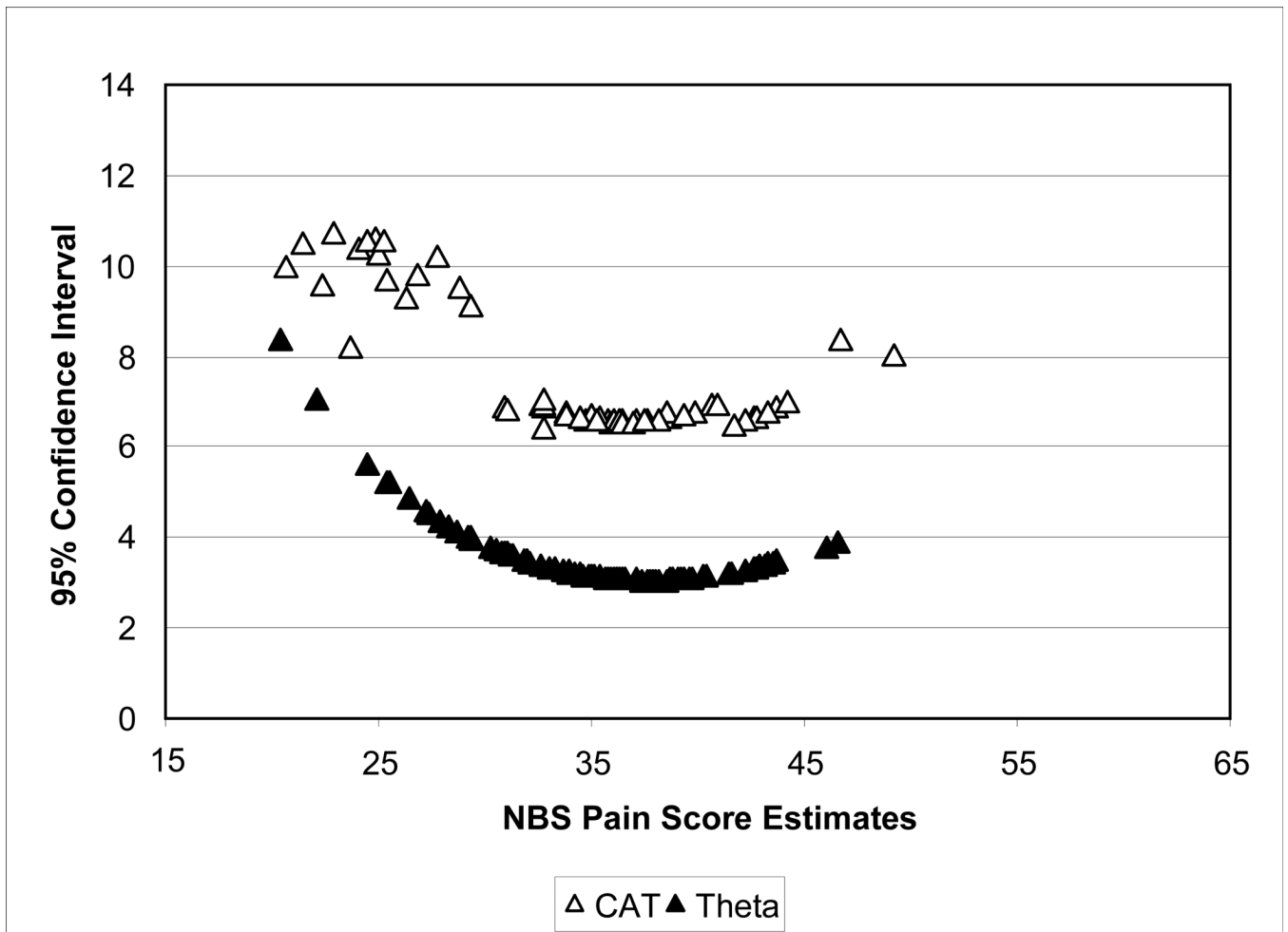


Figure 4.
Comparison of 95% CIs for Dynamic and Full Bank Scores

Table 1
Descriptive Statistics for Simulated Chronic Pain Survey Stopping Rules

	Full bank score	CAT-2	CAT-5	CAT-7	Precision Rule
Mean	52.5	50.4	51.2	51.3	51.1
SD	10.7	9.9	10.2	10.3	10.0
Minimum	25.8	21.8	23.6	22.8	23.9
Maximum	72.1	65.9	68	68.8	67.1

Table 2

Discriminant Validity for Simulated CATs

	Pain (N=367)	No Pain (N=415)	<i>t</i>	<i>p</i>	<i>F</i>	$R^2 = (T^2 / Total)^2$
<i>Full bank score</i>						1
<i>CAT - 2</i>	46.9(8.0)	57.4(10.4)	-15.60	<.001	243.38	0.68
<i>CAT - 5</i>	46.0(8.1)	54.3(9.7)	-12.86	<.001	165.31	0.82
<i>CAT - 7</i>	46.3(8.1)	55.5(9.9)	-14.17	<.001	200.74	0.85
<i>Precision based</i>	46.3(8.0)	55.7(10.0)	-14.39	<.001	207.05	0.85
	46.3(8.0)	55.4(9.6)	-14.25	<.001	203.03	0.83

Table 3

Individual norm based scores for selected cases

Full bank	Real data simulation cases						Precision rule CAT			Feasibility test cases	
	CAT-2		CAT-5		CAT-7		Precision rule CAT			DYHNA	
Score	95% CI	Score	95% CI	Score	95% CI	Score	95% CI	Score	95% CI	Score	95% CI
32.0	1.8	27.1	7.6	27.9	4.8	30.1	3.6	27.9	4.8	30.9	3.5
37.2	1.6	38.7	5.9	38.3	3.4	38.2	2.9	38.3	3.4	38.2	3.4
41.4	1.6	35.8	5.8	40.8	3.5	40.0	2.9	40.8	3.5	41.7	3.3
46.6	2.0	43.4	6.1	45.1	4.0	46.0	3.5	45.1	4.0	46.7	4.3
52.4	2.8	52.3	8.9	52.7	5.7	51.2	4.3	50.7	4.5	no data available	no data available
57.9	4.0	52.3	8.9	55.7	6.9	57.1	5.8	55.7	6.9	no data available	no data available
58.0	4.1	58.8	10.4	56.2	6.2	54.7	5.3	55.3	6.9	no data available	no data available
65.2	7.2	65.9	14.8	63.8	8.4	64.3	8.1	67.1	12.3	no data available	no data available

Table 4

Item Usage by DYNHA® Chronic Pain Impact

Abbreviated Item Stem	Item Administrations by VAS Pain Severity Group						Total Sample	
	Mild		Moderate		Severe		n	%
	n	%	n	%	n	%		
How much pain ...during the past 4 weeks?	4	100%	28	100%	64	100%	96	100%
... pain make simple tasks hard to complete?	4	100%	28	100%	52	81.3%	84	87.5%
pain limit you in performing usual daily activities..	4	100%	28	100%	51	79.7%	83	86.5%
... pain interfere with general activities....	3	75.0%	21	75.0%	34	53.1%	58	60.4%
pain interfere with ... normal work....?	1	25.0%	12	42.9%	36	56.3%	49	51.0%
restricted in performing routine daily activities because .pain activities	1	25.0%	11	39.3%	25	39.1%	36	37.5%
...pain limit... ability to concentrate on work	1	25.0%	7	25.0%	16	25.0%	24	25.0%
... pain keep you from enjoying social gatherings	3	75.0%	5	17.9%	5	7.8%	13	13.5%
How much bodily pain have you had	1	25.0%	1	3.6%	16	25.0%	18	18.8%
...leisure activities... affected by pain ...					14	21.9%	14	14.6%
...pain limits travel...	1	3.6%	1	3.6%	13	20.3%	14	14.6%
...pain left you too tired to do work or daily activities	1	3.6%	1	3.6%	9	14.1%	10	10.4%
...pain restrict you from going to social activities...	2	7.1%	2	7.1%	5	7.8%	7	7.3%
.... Throbbing pain....					5	7.8%	5	5.2%
Mean and range of items administered		5.5 (5-6)		5.17 (5-6)		5.75 (4-19)		5.57 (4-19)

Note:13 additional items were used in the severe sample;; All items use a 4 week recall period

Table 5
Feasibility Study - Descriptive Statistics for Dynamic and Static Assessments

	DYNAMIC (CAT) n=100	Static (k=45) n=100
Mean score (SD)	34.74 (5.75)	35.16 (5.12)
95% Confidence Interval	33.57 – 35.90	34.11 – 36.20
% at “ceiling”	0	0
% at “floor”	0	0

Feasibility Study - Static and Dynamic Scores by Level of Pain Severity

Table 6

	Mild (n=4)	Moderate (n=29)	Severe (n=66)	F(2, 91)	p	RV
DYNAMIC (CAT)	44.7	37.1	33.1	13.1	0.0001	1
Static (K=45)	44.2	36.9	33.8	12.5	0.0001	.95