FOCUS ON TOMORROW
RESEARCH FUNDED BY WORKSAFEBC

Measuring Health Outcomes: An Empirical Comparison of Adaptive and Standard Questionnaires

June 2005

Principal Investigator/Applicant
Dr. Jacek Kopec

RS2002/03-DG16

WORKING TO MAKE A DIFFERENCE

Measuring Health Outcomes: An Empirical Comparison of Adaptive and Standard Questionnaires

Final Report
June, 2005

Jacek A. Kopec
Maziar Badii
Mario McKenna
Eric C. Sayre
Viviane Dias Lima
Pam Rogers
Marcel Dvorak

Main Results

- This is the first study of an IRT-based computerized adaptive questionnaire to measure health-related quality of life in patients with back pain.
- We found that the adaptive questionnaire is feasible, reliable, valid, and efficient.
- Measurement properties of a relatively short adaptive questionnaire are comparable to those of standard disease-specific and generic measures.
- These results suggest that our adaptive questionnaire could replace current outcome measures in back pain

Executive Summary

Background

Modern psychometric methods based on item response theory (IRT) can be used to improve health outcome instruments. One of the most important applications of IRT is adaptive (or dynamic) assessment, also known as computerized adaptive testing (CAT). In CAT, different respondents answer different questions, and the questions are selected by a computer from a large "item bank" of previously calibrated questions. Scores are obtained using maximum likelihood estimation. CAT has been used in educational testing and more recently in psychology and health assessment, but has not been applied to back pain. In a previous study, we developed item banks for 5 domains of health-related quality of life (HRQL), namely Walking, Handling Objects, Usual Daily Activities, Pain / Discomfort, and Feelings. We also developed a method to calculate overall health utility score from domain-specific scores. Building on our previous research, the purpose of the current study was to develop a functional CAT system (adaptive questionnaire) and test it in a sample of patients with back pain.

Objectives

The specific objectives of the study were a) to develop a computerized adaptive questionnaire to measure five domains of health-related quality of life and overall health utility in patients with back pain; b) to assess feasibility, reliability, validity and efficiency of the computerized adaptive questionnaire; and c) to compare the measurement properties of the adaptive questionnaire with standard back pain questionnaires and a generic health questionnaire.

Methods

The computer-administered web-based (online) questionnaire included the new adaptive questionnaire, the modified Oswestry Disability Questionnaire (MODQ), Roland-Morris Questionnaire (RMQ), SF-36 Health Survey, Von Korff Pain Scale (VKPS), Neurogenic

Symptoms Scale (NSS), standard questions about pain and its treatment and demographic information. Patients from the Vancouver General Hospital Combined Spine Program diagnosed with mechanical back pain by the treating physician, who were 18-74 years of age and able to communicate in English were eligible to participate. Subjects who agreed to participate were asked to complete the questionnaire twice. They were given the options to complete the questionnaire at home, by a telephone interview, or in person at the Arthritis Research Centre of Canada (ARC). Scores for the adaptive questionnaire were computed using IRT methods. Test-retest reliability was measured by the intraclass correlation coefficient (ICC). Validity assessment included ceiling and floor effects, correlations between instruments, and discrimination between groups that were expected to differ in their average scores. Relative efficiency (RE) for each instrument was calculated by comparing its F-statistic to that of the most discriminating instrument.

Results

There were 215 patients who completed the baseline questionnaire and of those, 84 patients completed the follow-up questionnaire. To minimize patient burden, moderate precision was required for the adaptive questionnaire and the number of questions was set to between 3 and 7. The actual average number of questions answered per domain was 5.2. Mean domain scores for the adaptive questionnaire were all in the 41-45 range, i.e., 0.5-0.9 standard deviations below the population mean. The mean overall health utility score was 0.87 (SD=0.10) and ranged from 0.29 to 0.99. Test-retest ICCs ranged from 0.83 (Feelings) to 0.92 (Pain / Discomfort) and 0.93 for the overall health utility score. For comparison, ICCs were 0.92 for the MODQ, RMQ and Physical Component Summary (PCS-36), 0.90 for NSS and 0.86 for VKPS. Among the domains of the adaptive questionnaire, the highest ceiling effect was seen for Daily Activities (6%). For Pain / Discomfort the ceiling was 0.5%, compared to 2% for MODQ and 5% for RMQ. Scores on the adaptive questionnaire correlated as anticipated with other measures of HRQOL. The adaptive questionnaire discriminated well according to the level of satisfaction with current symptoms and duration of current episode as well as between patients with and without sciatica and those receiving vs. not receiving disability compensation. The average RE for between-group discrimination was 0.87 for the Pain / Discomfort domain of the adaptive questionnaire, 0.67 for Daily Activities, 0.62 for Walking and 0.53 for the overall health utility score. For comparison, average REs were 0.89 for MODQ, 0.80 for RMQ, 0.59 for VKPS, and 0.59 for PCS-36.

Conclusions

The adaptive questionnaire is feasible, reliable, valid, and efficient. Despite a relatively small number of questions, the Pain / Discomfort domain was as reliable and discriminating as the MODQ and RDQ and better discriminating than the PCS-36. Other domains were only slightly less discriminating than the disease-specific measures. Based on our results, the adaptive questionnaire should be considered a standard measure in patients with back pain. Future studies should be carried out to confirm our results and assess longitudinal validity (responsiveness) of the adaptive instrument.

Research Problem

Questionnaires to measure health outcomes are important tools in clinical and population health research[1,2]. However, a number of conceptual and methodological problems have been identified in the current measures[3-9]. Modern psychometric methods based on item response theory (IRT) can be used to improve existing measures or develop new ones[5,10,11]. Examples of IRT models being used for scale validation can be found in the quality of life literature[12-17]. One of the most important applications of IRT is adaptive (or dynamic) assessment, also known as computerized adaptive testing (CAT)[18]. In CAT, different respondents answer different questions, and the questions are selected by a computer. CAT has been used for a number of years in educational testing and more recently in psychology[18] and health assessment[11,19], but has not been applied to back pain.

Back pain (BP) remains a very significant public health problem and is a leading cause of disability in persons less than 40 years of age[20]. There is a strong need for improvement and standardization of outcomes in this very large group of patients[7,21-24]. A large number of disability questionnaires have been used in back pain research[7,9,26-36]. The Oswestry Disability Questionnaire (ODQ)[29] consists of 10 items assessing the level of pain and interference with several physical and role activities. Roland and Morris[30] selected 24 items from the Sickness Impact Profile and added the phrase "because of my back" to develop a back-specific scale (RMQ). The Quebec Back Pain Disability Scale (QBPDS) is a 20-item scale of physical disability associated with back pain[21]. The Lumbar Spine Questionnaire promoted by the North American Spine Society (NASS) incorporates a modified version of the ODQ and questions about neurogenic symptoms, satisfaction, and expectations met[36]. Most back-specific instruments measure pain intensity, limitations in simple physical activities, and the impact of pain on complex role activities. Usually, a single overall score is provided. A variety of response formats are implemented, such as yes/no, ordered categories, visual analogue scales, or numerical scales.

Most of the instruments have been tested for validity, reliability and responsiveness. However, head-to-head comparisons between different measures are relatively rare. Kopec et al compared the QBPDS with the RMQ, ODQ, and the SF-36 Physical Function Scale[21]. Several authors have compared the RMQ and ODQ[21,23,26,29,30]. A group of experts led by Deyo[21] recommended the use of the latter two measures plus a generic measure, such as the SF-36[28], in clinical studies of low-back pain. They listed several other questions that could be used; some of them have been incorporated into the NASS instrument[36].

All these measures suffer from problems stemming from the methodological framework within which they were developed, namely classical measurement theory. These problems include the selection of domains to be measured, validity of the scales for each domain, precision of score estimation, ceiling and floor effects, and measurement efficiency. Furthermore, proliferation of back-specific instruments and lack of standardization make it difficult to compare the results across studies[7,8] and comparative data needed to select an appropriate outcome instrument are often unavailable[8,9] . Comparison of scores from instruments that measure the same concept but have different items is fraught with problems because the items vary in their meaning and measurement properties.

In the current study we developed an IRT-based computerized adaptive questionnaire to measure five domains of health-related quality of life (HRQOL) and overall health utility in patients with BP. We also assessed the reliability and validity of this new instrument in a sample of patients with BP and compared it with standard disease-specific and generic measures of HRQOL.

Methodology

Item Response Theory

An IRT model specifies how the concept being measured and item properties are related to a person's item responses[10,38-40]. This allows the investigator to estimate the probability that a respondent will choose a particular option for a given item as a function of the measured attribute. A large number of items measuring the same underlying concept can be calibrated, and their statistical properties can be described. Because IRT models include item properties, IRT-based scores are comparable for any set of calibrated items[10,18,39,40]. IRT-based methods have been recommended for item selection, scale development, and score estimation[4,5,39-41]. Numerous examples of IRT models being used for scale validation can be found in the quality of life literature[41-46]. Kopec et al used a non-parametric IRT approach to evaluate and select items for a scale of disability in patients with back pain[16].

IRT methods form the foundation for computerized adaptive testing (CAT)[18]. In CAT, the computer selects each question from a pool of calibrated items, according to the respondent's previous responses. As a result, the items are more relevant to the respondent, and the measurement process is more efficient. Even though different respondents answer different questions, the estimated scores are comparable since the items in the pool have been calibrated on the same scale. CAT has been used for a number of years in educational and psychological testing[18,45,46] and its validity in those applications has been documented [10,18,39,40,45,46]. Ware and colleagues at QualityMetric have developed a CAT version of the SF-36 and a measure of headache impact[47-48].

Item banks

In adaptive assessment, items are selected from an item bank, i.e., a pool of calibrated items. Desirable properties of an item bank are unidimensionality, lack of differential item functioning, and optimal distribution of discriminating powers throughout the full spectrum of trait levels[18]. As part of a study funded by the Canadian Arthritis Network (CAN), we developed item banks for 5 domains of HRQL relevant to musculoskeletal diseases, including BP. The domains are referred to as Walking (difficulty running, walking, climbing stairs), Handling Objects (difficulty carrying, lifting, grasping, holding objects), Usual Daily Activities (problems in work, domestic, recreational activities), Pain / Discomfort (interference with activities), and Feelings (depression, anxiety). The items were derived primarily from validated instruments, although additional items were developed. The main criteria in item selection and development were conceptual unidimensionality and a broad range of levels for each domain being measured. The items were pilot tested and some items were re-worded or deleted. This process resulted in a 221-item calibration questionnaire. The calibration questionnaire was administered by mail to a sample from the general population (n=643) and a second sample whereby we asked the oldest person in the household to respond (n=335). The questionnaire was also administered to patients with RA recruited from a cyclosporine clinic (n=303) and patients with OA waiting for hip or knee replacement (n=337).

Several IRT models were investigated, ranging from the restrictive Rasch model[38] to Muraki's generalized partial credit model[39]. A non-parametric approach was also explored[49,50]. With the generalized partial credit model chosen, item banks for each domain were selected from the calibration questionnaire, based on information function, dimensionality, differential item functioning, and distribution of item locations. The final 5 item banks contain 182 items: Walking - 35, Handling Objects - 36, Daily Activities - 36, Pain / Discomfort - 33 and Feelings – 42.

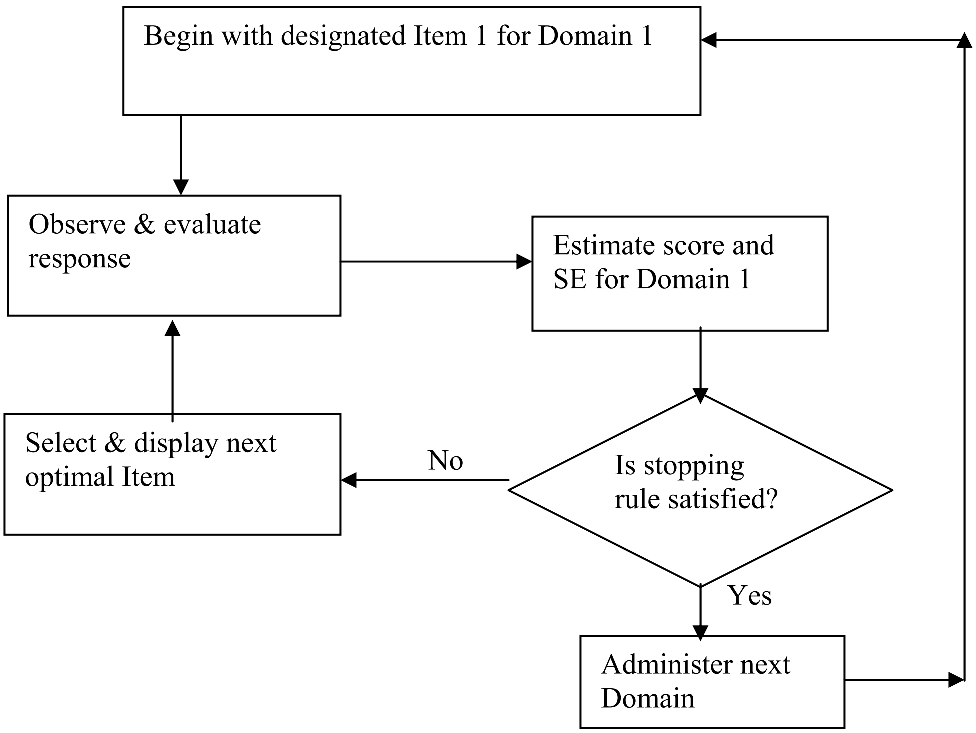
In the CAN-funded study we also measured rating scale values and standard gamble preferences (health utilities) associated with different levels of HRQOL, as measured by questions from the 5 item banks. Two preference assessment studies were conducted, with a total sample of 278 individuals randomly selected from the general population.

In the present application, we used the same item banks to develop a computer algorithm for CAT and administer the questionnaire to patients with BP. It should be noted that the domains measured by the questionnaire are highly relevant to BP and the questions are domain-specific, not disease-specific. In our original database used for item calibration, 543 individuals (33%) reported back pain.

Technical description of CAT

A key concept in CAT is item information function (IIF)[10,18]. The derivation of IIFs is critical for adaptive assessment because the selection of items in CAT is based on this property. IIF measures the precision for each item in estimating the score over the entire spectrum of each domain. For a given level of the attribute measured, denoted as θ, information is defined as the expected value of the inverse of the error variance, i.e., $I(\theta) = E(1/\sigma_e^2)$, where σ_e denotes standard error of measurement[18,p.162]. The algorithm for adaptive assessment is described in a flow chart below (adapted from Ref [18]). A key component of the algorithm is the IIF database for all items[18]. Once θ is estimated from the response to the first item (using maximum likelihood estimation), the IIF database is consulted and the computer selects the most informative item for a given θ. This procedure is repeated until a desired level of precision is attained (or another stopping rule is satisfied).

A flowchart describing computerized adaptive assessment



Patient recruitment

The questionnaire was initially tested on employees and volunteers at the Arthritis Research Centre (ARC). Next, feasibility of computer administration was pilot tested in a sample of 6 patients with back pain treated by one of the investigators (Dr. Badii). Only those patients who had internet access were asked to complete the online questionnaire. Patients were later contacted by telephone to receive feedback on the ease of usability and navigation of the questionnaire.

Following minor modifications to the questionnaire based on patient and research team feedback, full-scale testing and recruitment began. Subjects were recruited from patients who saw Dr. Badii between 1999 and 2004 at the Vancouver General Hospital Combined Spine Program. This list contained the names of 3,767 people. Of those, 1,031 were treated for mechanical back pain. The list was then screened for patient contact information (complete address and phone number), resulting in 848 names. Additional eligibility criteria included age between 18 and 74 years, ability to communicate in English, and mental competence.

An introductory letter from the Spine Program and ARC was mailed out in blocks of approximately 100 letters based on the above criteria. The letter included the purpose of the study, time commitment on the patient's part, patient rights and relevant contact information. The letter also listed the online survey URL (Internet address). Approximately three to four days later, a follow-up phone call was made by a trained interviewer to assess the interest of the subject in participating in the study. A maximum of seven calls were placed before the individual was removed from the contact list. This process was repeated until the block was completed upon which a new block of letters was mailed out. Those who agreed to participate in the study and had Internet access were asked to complete the online questionnaire from home. For those who did not have Internet access or lacked the necessary skills, a time was arranged to complete the online questionnaire either in person at ARC or over the telephone with a trained interviewer. Each participant was offered an honorarium of $20, while those who traveled to ARC would receive $25.

All patients who completed the questionnaire were invited to complete the questionnaire a second time approximately 3-5 days later via an e-mail reminder if the subject had Internet access, by telephone if the person had completed the questionnaire over the phone, or at ARC. E-mail reminders contained an embedded hyperlink to the online questionnaire.

Data analysis

Scores for the MODQ, RMQ and SF-36 (8 domains plus physical and mental component scores, PCS-36 and MCS-36) were calculated using available guidelines[29,30,47]. For the adaptive questionnaire, we obtained domain-specific IRT-based scores from Parscale[51] (maximum likelihood estimates) and overall utility scores based on our model. Domain scores, as well as scores for the SF-36, MODQ and NSS were standardized to population norms (mean=50 and SD=10), higher scores indicating better health. The RMQ used a 0-24 scale and the VKPS used a 0-100 scale, with higher scores indicating poorer health. Utility scores ranged from -1 to +1. Measures of location and dispersion (means, ranges and standard deviations) for all instruments were calculated. Ceiling and floor effects were measured by the percentage of subjects with the maximum and minimum possible score. We used data from the baseline and follow-up assessments to calculate the test-retest reliability coefficient for each domain. Test-retest reliability was measured by the intraclass correlation coefficient (ICC)[52p.26]. Cross-sectional construct validity was evaluated in terms of correlations between the adaptive questionnaire and established measures of HRQOL and discrimination between groups that were expected to differ in their average scores[1]. Specifically, we considered such variables as age, sex, satisfaction with current symptoms, duration of the current episode of back pain, receiving disability or workers' compensation, and presence of sciatica. Discrimination was assessed by the F-statistic. Relative efficiency (RE) for each instrument was calculated by comparing its F-statistic to that of the most discriminating instrument[53].

Research Findings

Patient recruitment

We mailed 848 letters, with 98 letters being returned. Of the 848 patients who had letters mailed to, 129 were ineligible. Of these, 343 telephone numbers were not in service or missing. Of the eligible subjects with valid telephone numbers, 215 completed the questionnaire, 6 started the questionnaire but failed to complete it, while 142 refused or were not interested in participating. This resulted in a participation rate of 57.2% (Table 1). Table 2 shows the questionnaire completions according to method. At baseline, 151 subjects completed the online questionnaire at home, 38 completed it via a telephone interview, and 26 subjects came to ARC to complete it in person. The questionnaire was completed twice by 84 patients (39%). At follow-up, 54 subjects completed the questionnaire at home, 11 completed it via a telephone interview, and 19 subjects came to ARC to complete it in person.

Descriptive statistics

As shown in Table 3, 55.3% of the participants were men, 36.3% were <45 years old and 21.9% were 65+. About 42% had college or university education, 43.7% were currently working and 14.9% were receiving disability or workers' compensation benefits. Only 15.4% were very or somewhat satisfied with their symptoms, while 44.2% were very dissatisfied. About ¾ of the subjects experienced >5 episodes of back pain, 47.0% reported the current episode to last >1 year and 23.3% had undergone surgery for back pain.

Performance of adaptive algorithm

The minimum number of questions per domain was set to 3 and the maximum was set to 7. The IRT-based standard error (SE) of the individual score was set to 0.15 for Walking, 0.15 for Handling Objects, 0.20 for Daily Activities, 0.15 for Pain and 0.25 for Feelings. These settings were determined by previous analyses of our database which suggested that most people should be able to achieve such levels of precision by answering 3-7 questions. The stopping rule was defined as either responding to the maximum number of questions allowed or attaining the required SE for a given domain, whichever was reached first. The average numbers of questions actually answered were 5.7 for Walking, 6.6 for Handling Objects, 4.6 for Daily Activities, 4.4 for Pain / Discomfort and 5.0 for Feelings. Most people answered 7 questions in the Handling Objects domain, 5 in the Walking domain and 4 in the remaining domains.

Score distribution

The distributions of scores for the 5 domains of the adaptive questionnaire and the overall health utility score are presented in Figures 1-6. Descriptive statistics for the adaptive questionnaire as well as other measures of HRQOL are presented in Table 4. The mean overall health utility score was 0.87 (SD=0.10) and ranged from 0.29 to 0.99. Mean domain scores for the adaptive questionnaire were all in the 41-45 range, i.e., 0.5-0.9 standard deviations below the population mean, the lowest scores being observed for Pain / Discomfort and Daily Activities. For comparison, the mean score for the MODQ was 36.8. Among the SF-36 domains, the lowest score was 39.4 (Physical Function). The score for Bodily Pain (SF-36) was 40.5, similar to the score for Pain / Discomfort in the adaptive questionnaire (41.1), while the score for Mental Health (SF-36) was 48.3, compared to 44.9 for Feelings in the adaptive questionnaire.

The results of internal consistency and test-retest reliability testing for all measures are presented in Table 5. Cronbach's alpha coefficients for the domains were high and ranged from 0.91 for Feelings to 0.96 for Pain / Discomfort. For comparison, the alpha coefficients were 0.96 for both MODQ and RMQ. In the SF-36, alpha ranged from 0.82 for Role Emotional to 0.98 for Physical Function.

Test-retest reliability coefficients for the adaptive questionnaire ranged from 0.83 (Feelings) to 0.92 (Pain / Discomfort) and 0.93 for the overall health utility score. For comparison, both the MODQ and RMQ had a test-retest reliability of 0.92. Test-retest reliabilities for the SF-36 domains ranged from 0.70 (Role Emotional) and 0.74 (Social Function) to 0.95 (Physical Function). The coefficients for PCS-36 and MCS-36 were 0.92 and 0.74, respectively.

Validity

Ceiling and floor effects are presented in Table 6. There was no ceiling (good health) or floor (poor health) effect for the health utility score on the adaptive questionnaire, and no floor effect for any of the domains. Ceiling effects for the domains ranged from 0% for Feelings and 0.5% for Pain / Discomfort to 6.1% for Daily Activities. For comparison, the MODQ had a ceiling effect of 1.9% and no floor effect, whereas the RMQ had a ceiling effect of 5.1% and a floor effect of 0.5%. All SF-36 domains had some ceiling effect and all except General Health had some floor effect. The ceiling effect was very strong for some SF-36 domains, for example, 59.5% for Role Emotional and 34.4% for Social Functioning, although no ceiling or floor effects were seen for PCS-36 and MCS-36.

Correlations between the adaptive questionnaire and other measures of HRQOL are presented in Table 7. The pattern of correlations was consistent with expectations. For example, strong correlations were seen between Pain and MODQ (0.81), Pain and PCS-36 (0.82), and Daily Activities and PCS-36 (0.83). The Feelings domain correlated 0.72 with MCS-36 and 0.33 to 0.48 with measures of pain and physical function. The Handling Objects domain showed slightly lower correlations with other measures than the Pain / Discomfort, Daily Activities and Walking domains. Correlations with the NSS were generally lower compared to those with pain or physical function measures. Correlation plots of Pain / Discomfort with MODQ and RMQ are shown in Figures 7 and 8, respectively. These plots display a characteristic non-linearity for the extreme scores suggesting a better discrimination by the adaptive questionnaire at extreme levels of the attribute measured.

Males scored better than females on all measures of physical function and pain but only the Walking and Handling Objects domains of the adaptive questionnaire and the PCS-36 showed a statistically significant difference (Table 8). These scales also discriminated between different age groups (Table 9).

Satisfaction with current symptoms was strongly related to all adaptive domains (Table 10). The relationship between satisfaction and overall health utility is shown in Figure 9. Those receiving disability or workers' compensation reported lower scores on all measures compared with the reminder of the subjects. Figure 10 shows the difference between these groups for the overall utility score. Patients with acute pain (<1 week) were less disabled and had less pain than those with longer duration (data not shown). Mean health utility was 0.92 for those with <1 week duration compared to 0.83 for patients whose pain lasted >1 year. Patients with sciatica (as determined from the NSS) reported significantly lower scores on all domains, including overall health utility scores, than those without sciatica. Finally, those who had undergone surgery for back pain reported significantly more neurogenic symptoms than those who did not have surgery but their levels of pain in general were similar (data not shown).

In Table 11 we show the F-statistics and RE for between-group discrimination according to 4 variables, namely satisfaction with current symptoms, presence of sciatica, duration of current episode and disability compensation. The average RE was 0.87 for the Pain / Discomfort domain of the adaptive questionnaire, 0.67 for Daily Activities, 0.62 for Walking and 0.53 for the overall health utility score. For comparison, average REs were 0.89 for MODQ, 0.80 for RMQ, 0.59 for VKPS, and 0.59 for PCS-36. The Feelings domain displayed much lower RE, similar to MCS-36, as one would expect.

Implications for Future Research on Occupational Health

The adaptive approach to measurement has the potential to improve outcome assessment in clinical research and practice[11]. However, there have been very few comparisons of adaptive and standard (fixed) questionnaires. To our knowledge, this is the first study of an IRT-based adaptive questionnaire in back pain, the first study to compare adaptive and standard questionnaires in terms of test-retest reliability and relative efficiency, and the first study to use telephone administration for an adaptive HRQL questionnaire.

We have demonstrated that our adaptive questionnaire is feasible to use in patients with back pain. Second, we have demonstrated that the adaptive algorithm performs well, as different patients answered different questions according to their level of HRQOL. Third, we provided descriptive information on the distribution of scores for each domain and overall health utilities in a sample of patients with back pain from a referral care centre. Fourth, we showed that the adaptive questionnaire is reliable, valid and efficient in this group of patients.

The advantages of a multi-dimensional generic measure compared to a one-dimensional disease-specific measure are well known and include the ability to assess different domains (including emotional function) and greater comparability of scores across diseases and populations. Our adaptive questionnaire also provides an overall utility score that can be used in economic analyses. Additional advantages of a computerized adaptive questionnaire are that it can be implemented on the Internet and connected to a database. Use of web-based data collection methods allows data to be collected through remote access and allows for easier follow-up and real-time data integrity checks.

A potential disadvantage of generic instruments is a lower ability to detect small differences between individuals or small changes in disease impact over time. When assessing the properties of our instrument against other measures it must be stressed that we restricted the maximum number of questions and did not require a very high precision of measurement. These parameters were set to relatively low values to limit patient burden, as the questionnaire was quite long. As a result, the average patient answered 26 adaptive questions (5.2 questions per domain) compared with 36 questions in the SF-36 (8 domains), 10 questions in the MODQ (single score) and 24 in the RMQ (single score).

Despite those restrictions, our data indicate that the adaptive questionnaire is as reliable as the best disease-specific measures. The ceiling effect was zero for the overall score and was minimal for most domains except Daily Activities. Correlation plots suggest that the adaptive instrument should be able to distinguish between persons with very high (or very low) scores, who may display ceiling (or floor) effect on standard measures. While we did not assess longitudinal validity (responsiveness) in the current study, the Pain / Discomfort domain discriminated as well as the MODQ or RMQ in cross-sectional data, but with fewer questions. As expected, other domains were slightly less effective, most likely because pain (and its impact on activity) is the most important concern in this group of patients. The psychometric properties of the adaptive domains could be improved further by changing the stopping rule to allow more questions and require higher precision of the estimates (this can be done very easily by the system administrator). Also, the number of questions per domain can be adjusted according to the objectives of a given study. For example, if pain is the primary outcome, one can increase the level of precision for that domain and limit the number of questions for other domains.

Currently, most authors advocate the use of both a generic and a disease specific instrument in studies of back pain. A preferable approach may be to apply a multi-domain adaptive instrument. It should be noted that a dynamic (adaptive) version of the SF-36 has been developed; however, it is not widely available. Our adaptive instrument has advantages over the adaptive version of SF-36. Specifically, it measures 5 domains of activity, and separates lower and upper body functions. Most importantly, it provides an overall health utility score. (A utility score may be obtained from the conventional SF-36 through SF-6D, but not from the adaptive version. Moreover, the SF-6D score is based on only 6 questions whereas our utility score uses all information in the questionnaire and therefore is more precise). Finally, our questionnaire has been validated in patients with back pain.

Limitations

Web-based questionnaire administration is a relatively new technique and some limitations should be noted. A large proportion of subjects indicated that the URL for logging in to the system was long and cumbersome, which made it difficult for subjects to type it into their web browser. This may have resulted in a failure to access the webpage. This was partially resolved for those individuals who possessed an e-mail address to which an embedded hyperlink could be sent. However, even in this instance, it was not a completely satisfactory solution, as not all e-mail software programs are capable of accepting html and some rely on plain text. Furthermore, some subjects indicated that they did not receive follow-up e-mails from the study coordinator despite a confirmation receipt being received. This was most likely due to 'junk mail' filters on individual e-mail programs being set high. This would result in an unknown user's e-mail address not being recognized and moved to the 'junk mail" folder.

The initial registration page was considered quite difficult by some subjects. For those who completed the survey at home, it required subjects to enter his/her name, e-mail address, and contact information (optional). Once entered, the subject was then sent a username and password to his/her e-mail account. This two-step registration process required the subject to retrieve this information and enter it into the webpage before being able to access the questionnaire. Furthermore, the information e-mail was sometimes mistaken as 'junk mail' and the subject was unable to proceed any further. As a consequence, we simplified the registration process into a single registration page where subjects entered their e-mail address (required) and name (optional), with the email validation step being skipped. This greatly increased usability and navigation ease, but made tracking completions by subject difficult as some people chose to leave the name field blank.

Some participants commented that the questionnaire was quite long and required them to sit for an extended period of time, causing discomfort. Furthermore, for the non-adaptive sections of the online questionnaire, there was no ‘no response’ option allowing a subject to skip a question that he/she did not wish to answer. Therefore, if a subject did not indicate a response to the question, he or she was unable to continue. This could have resulted in some lost subjects who may have wished to continue but deemed a particular question as inappropriate or private in nature. These limitations can easily be addressed in future applications.

The response rate at baseline was close to 60% and almost 40% of the participants filled out the follow-up questionnaire. Although every attempt was made to accommodate subjects who were not computer literate or proficient with Internet browser software by allowing them to complete the questionnaire over the telephone or to come to ARC, lack of computer skills may have created a potential barrier to participation. Subjects were told to complete the follow-up questionnaire within 3-5 days following the initial completion, but it was difficult to achieve compliance. For these reasons, our sample may not be representative of all patients in the Spine Program eligible for the study and those who completed the questionnaire twice may not be representative of all participants. Furthermore, patients in a referral treatment center are not representative of all patients with back pain. Future studies in other patient groups are needed to confirm our results. At the same time it should be clear that the participation rate in this study was not greatly different from what one would expect using standard methods of questionnaire administration. Interestingly (and unexpectedly), most patients were able to access the questionnaire at home and opted for this method of administration.

Policy and Prevention

Implications arising from research

The most important practical implication of this research is that our generic multi-domain computerized adaptive questionnaire can potentially replace both disease-specific (MODQ, RMQ) and generic (SF-36) instruments. It provides a reliable, valid and efficient assessment of outcomes and is flexible in terms of questionnaire length and measurement precision. Its wide use would simplify outcome assessment in back pain, reduce the cost of data collection, and improve data quality. It would also facilitate score interpretation and comparisons across studies and conditions, and allow economic analyses of back pain therapies in terms of quality-adjusted life years.

Relevant user groups

The most relevant user groups for this research are researchers and clinicians involved in outcome measurement, including clinical trial researchers and epidemiologists. We hope these users will find out questionnaire useful and that it will become a preferred tool for outcome assessment in back pain.

Policy related interactions

Our ultimate goal is to improve the quality of life of persons suffering from back pain by providing researchers and clinicians with a better tool to assess health outcomes. While the effect of such tools on health policy and treatment outcomes is indirect and may take a substantial amount of time to materialize, we believe it is important for health policy makers to be aware of the progress in outcome assessment and to appreciate the role of modern technology in developing, scoring, and administering health-related quality of life questionnaires. As health researchers, we interact on a regular basis with policy makers and welcome any opportunity to discuss these issues.

Dissemination / Knowledge Transfer

Our plan for KTE (knowledge translation and exchange) involves the following components:

- publication of the results of our study in an international peer-reviewed journal (for example Spine or Quality of Life Research) in 2006
- presentation of the results to an international audience at the next meeting of the ISOQOL (in 2006)
- presentation of the results at local research rounds and other forums (following journal publication)
- posting of the results on the Arthritis Research Centre website and other websites (following journal publication)
- development of a website to support the use of the adaptive questionnaire by researchers around the world (following journal publication)
- supporting future efforts to translate the instrument to other languages
- addressing intellectual property, copyright, and access issues associated with the dissemination and use of the adaptive questionnaire

The most important component of our KTE strategy is to publish the results in a peer-reviewed journal and present them at the ISOQOL meeting. We will not post the results on the Internet (even as an abstract or summary) prior to journal publication and ISOQOL presentation.

References

1. Streiner DL, Norman GR: Health Measurement Scales. A Practical Guide to Their Development and Use. 2nd Edition, Oxford, Oxford University Press, 1995

2. Wolfson MC. Social Proprioception: Measurement, Data, and Information from a population Health Perspective. In: Evans RG, Barer ML, Marmor TR (Editors). Why Are Some People Healthy and Others Not? The Determinants or Health of Populations. . Aldine de Gruyter, New York, 1994, pp 287-316

3. Fisher WP Jr. Measurement-related problems in functional assessment. Am J Occup Ther. 1993; 47(4):331-8.

4. McHorney CA. Generic health measurement: past accomplishments and a measurement paradigm for the 21st century. Ann Intern Med 1997, 15;127(8 Pt 2):743-50

5. 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(6):595-600

6. Wood-Dauphinee S. Assessing quality of life in clinical research: from where have we come and where are we going? J Clin Epidemiol. 1999 Apr;52(4):355-63.

7. Zanoli G, Stromqvist B, Padua R, Romanini E.Lessons learned searching for a HRQoL instrument to assess the results of treatment in persons with lumbar disorders. Spine 2000 Dec 15;25(24):3178-85

8. Patrick DL, Deyo RA. Generic and disease specific measures in assessing health status and quality of life. Med Care 1989, 27 (Supplement): S217-S232a

9. Kopec JA. Measuring Functional Outcomes in Persons with Back Pain: A Review of back-specific questionnaires. Spine. 2000, 25: 3110-3114.

10. Lord FM: *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale, NJ, Lawrence Erlbaum Associates, Publishers, 1980

11. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. Med Care 2000 Sep;38(9 Suppl):II28-42

12. Engberg A, Garde B, Kreiner S. Rasch analysis in the development of a rating scale for assessment of mobility after stroke. Acta Neurologica Scandinavica 1995; 91(2):118-27

13. Fisher WP Jr. Foundations for health status metrology: the stability of MOS SF-36 PF-10 calibrations across samples. J La State Med Soc 1999; 1(11):566-78

14. Haley SM, McHorney CA, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. J Clin Epidemiol 1994; 47(6):671-84

15. McArthur Dl, Cohen MJ, Schandler SL. Rasch analysis of functional assessment scales: An example using pain behaviors. Arch Phys Med Rehabil 1991; 72: 296-304

16. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther 1999; 53(1):83-90

17. Kopec JA, Esdaile JM, Abrahamowicz M et al. The Quebec Back Pain Disability Scale: conceptualization and development. J Clin Epidemiol. 1996; 49(2):151-61.

18. Wainer H. *Computerized Adaptive Testing: A primer*. Lawrence Erlbaum Associates, Publishers, Hillsdale, NJ; 1990

19. Zhao J, Bjorner JB, Kosinski M, Laliberte KA, Ware Jr.JE, (Abstract). Precision and brevity: vitality assessment using computerized adaptive testing methods. Quality of Life Research 2001;10(3) 216

20. Deyo RA, Cherkin D, Conrad D, Volinn E. Cost, controversy, crisis: low back pain and the health of the public. Annu Rev Public Health 1991; 12:141-56.

21. Deyo RA, Battie M, Beurskens AJ, et al. Outcome measures for low back pain research. A proposal for standardized use. Spine 1998; 23(18):2003-13

22. Kopec JA, Esdaile JM, Abrahamowicz M, Abenhaim L, Wood-Dauphinee S, Lamping D, Williams JI. The Quebec Back Pain Disability Scale: Measurement properties. Spine 1995, 20: 341-352.

23. Kopec JA, Esdaile JM. Spine Update. Functional disability scales for back pain. Spine 1995, 20: 1943-1949.

24. Kopec JA, Esdaile JM. Occupational role performance in persons with back pain. Disability and Rehabilitation. 1998, 20: 373-379.

25. Torrance GW. Measurement of health state utilities for economic appraisal: A review. J Health Econ 1986; 5:1-30.

26. Deyo RA. Measuring the functional status of patients with low back pain. Arch Phys Med Rehabil 1988; 69:1044-1053

27. Deyo RA, Andersson G, Bombardier C et al. Outcome measures for studying patients with low back pain. Spine 1994; 19(18 Suppl):2032S-2036S.

28. Ware JE, Sherbourne CD: The MOS 36-item Short Form Health Survey (SF-36). Medical Care 1992, 30:473-483

29. Fairbank JCT, Couper J, Davies JB, O'Brian JP: The Oswestry low back pain disability questionnaire. Physiotherapy 1980; 66:271-273

30. Roland M, Morris R: A study of the natural history of back pain. Part I: Development of a

reliable and sensitive measure of disability in low back pain. Spine 1983; 8: 141-144

31. Million R, Hall W, Nilsen KH, Baker RD, Jayson MIV: Assessment of the progress of the back pain patient. Spine 1982; 7:204-212

32. Waddell G, Main CJ. Assessment of severity in low back disorders. Spine 1984, 9:204-208,

33. Greenough CG, Fraser RD. Assessment of outcome in patients with low-back pain. Spine 1992; 17:36-41

34. 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(17):1887-96

35. Manniche C, Asmussen K, Lauritsen B et al. Low Back Pain Rating scale: validation of a tool for assessment of low back pain. Pain 1994; 57(3): 317-26

36. Daltroy LH, Cats-Baril WL, Katz JN et al. The North American Spine Society lumbar spine outcome assessment instrument: reliability and validity tests. Spine. 1996; 21(6):741-9.

37. Stewart AL, Ware JE Jr. Measuring Function and Well-Being. The Medical Outcome Study Approach. Duke University Press, Durham, 1992

38. Rash G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: Nielson and Lydiche, Copenhagen; 1960

39. Van Der Linden WJ & Hambleton RK (Eds.). *Handbook of Modern Item Response Theory,* Springer Verlag, NY; 1997

40. Embretson S, Reise SP. *Item Response Theory for Psychologists.* Lawrence Erlbaum Associates, Mahwah, NJ; 2000

41. Fisher WP Jr, Harvey RF, Taylor P et al. Rehabits: a common language of functional assessment. Arch Phys Med Rehabil 1995; 76(2):113-22

42. Granger CV, Hamilton BB, Linacre JM et al. Performance profiles of the functional independence measure. Am J Phys Med Rehabil 1993; 72(2):84-9

43. McArthur Dl, Cohen MJ, Schandler SL. Rasch analysis of functional assessment scales: An example using pain behaviors. Arch Phys Med Rehabil 1991; 72: 296-304

44. Velozo CA, Kielhofner G, Lai JS. The use of Rasch analysis to produce scale-free measurement of functional ability. Am J Occup Ther 1999; 53(1):83-90

45. Urry VW. Tailored testing: A successful application of latent trait theory. J Educational Measurement 1977, 14 181-186

46. Howard EP. Applying the Rasch Model to test administration. J Nurs Educ 1985; 24(8): 340-3

47. Ware JE Jr, at al. Measuring health for consumers and patients (Internet: www.amIhealthy.com)

48. Ware JEJr, Kosinski M, Bjorner JB, Bayliss MS, Batenhorst A, Dahlöf CGH, Tepper S, Dowson A. Applications of computerized adaptive testing (CAT) to the assessment of headache impact. Qual Life Res 12 (8): 935-952, 2003

49. Ramsay JO. Testgraf. A Program for the Graphical Analysis of Multiple Choice Test and Questionnaire Data. McGill University, Montreal, Canada, 1995

50. Ramsay JO. Kernel smoothing approaches to nonparametric item characteristic curve estimation. Psychometrika 1991, 56: 611-630

51. Muraki E, Bock RD. Parscale: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks. Chicago: Scientific Software Int.; 1993

52. Fleiss JL: The Design and Analysis of Clinical Experiments. New York: John Wiley & Sons, 1986.

53. Liang M, Larson M, Cullen K et al. Comparative measurement efficiency and sensitivity of five health status instruments for arthritis. Arthritis Rheum 28:542-547, 1985.

Tables and Figures

Table 1. Recruitment of subjects into the study

Type of subject	N	%
Completed all questions	215	25.4
Failed to complete all questions	6	0.7
Failed to meet one or more eligibility criteria	129	15.2
Telephone number not in service or missing	343	40.4
Unable to reach subject after seven attempts	13	1.5
Not willing to participate	142	16.7
Total	848	100.0

Table 2. Frequency of different methods of questionnaire administration

Method of administration	Baseline		Follow-up	
	N	%	N	%
On Internet at Home	151	70.2	54	64.3
Telephone Interview	38	17.7	11	13.1
On Internet at ARC	26	12.1	19	22.6
Total	215	100	84	100

Table 3. Characteristics of patients in the study

Variable	N	%
Sex		
Male	119	55.3
Female	96	44.7
Age		
<45	78	36.3
45-54	50	23.3
55-64	40	18.6
65+	47	21.9
Education		
High school or less	42	19.5
Trade / technical school	79	36.7
College / university	90	41.9
	4	1.9
Employment		
Currently working	100	43.7
Retired	40	17.5
Unemployed	38	16.6
Other	12	5.2
Not stated	39	17.0
Satisfaction with symptoms		
Very satisfied	12	5.6
Somewhat satisfied	21	9.8
Neutral	32	14.9
Somewhat dissatisfied	55	25.6
Very dissatisfied	95	44.2
Receiving disability / workers' comps		
Yes	32	14.9
No	182	84.7

Number episodes of back pain		
None	15	7.0
1-5	37	17.2
>5	163	75.8
Duration of current episode		
<1 week	50	23.3
1-6 weeks	36	16.7
6-12 weeks	11	5.1
12 weeks - 1 year	17	7.9
>1 year	101	47.0
Surgery for back pain		
Yes	50	23.3
No	165	76.7

Numbers may not add up to 100% due to missing values

Table 4. Descriptive statistics for the scores on the HRQOL instruments (N=215)

Instrument	Mean	Median	SD	Minimum	Maximum
Adaptive Questionnaire					
Walking	44.57	45.24	10.30	21.25	67.06
Handling Objects	44.80	43.31	8.41	27.07	66.31
Daily Activities	42.94	42.57	10.12	23.92	64.78
Pain / Discomfort	41.05	39.94	7.07	21.82	63.57
Feelings	44.90	45.28	9.16	8.16	68.18
Health Utility	0.869	0.885	0.099	0.288	0.987
SF-36					
Physical Function	39.39	40.20	11.10	14.94	57.03
Role Physical	40.53	42.16	11.91	17.67	56.85
Bodily Pain	40.51	41.41	9.61	19.86	62.12
General Health	43.60	43.40	10.56	18.61	63.90
Vitality	46.85	48.97	11.16	20.87	70.82
Social Function	44.01	45.94	12.49	13.22	56.85
Role Emotional	49.41	55.88	9.93	9.23	55.88
Mental Health	48.25	50.01	11.18	7.77	64.09
PCS-36	38.02	38.55	11.25	12.68	58.87
MCS-36	51.17	54.75	11.17	15.69	74.24
MODQ	36.82	37.60	12.71	2.66	57.72
RMQ	10.47	11.00	6.64	0.00	24.00
VKPS	34.61	31.25	25.57	0.00	100.00
NSS	43.99	47.46	12.07	11.74	56.38

Footnote to Table 4

SF-36 = 36-Item Short-Form Health Survey, Version 2.

PCS = Physical Component Score

MCS = Mental Component Score

MODQ = Modified Oswestry Disability Questionnaire

RMQ = Roland-Morris Questionnaire

VKPS = Von Korff Pain Scale

NSS = Neurogenic Symptoms Scale

For the adaptive questionnaire, SF-36, MODQ and NSS higher scores indicate better health.

For RMQ and VKPS higher scores indicate worse health.

Adaptive questionnaire, SF-36, MODQ and NSS scores were standardized to population norms (mean = 50, SD=10).

VKPS scores were standardized to 0-100.

RMQ sores can range from 0 to 24.

Health utility scores can range from -1 to +1.

Table 5. Internal consistency and test-retest reliability of the instruments

Instrument	Cronbach's alpha	Intraclass Correlation Coefficient	95% CI
Adaptive Questionnaire			
Walking	0.947	0.900	0.833,0.941
Handling Objects	0.949	0.905	0.842,0.944
Daily Activities	0.937	0.881	0.803,0.930
Pain / Discomfort	0.960	0.924	0.872,0.955
Feelings	0.911	0.830	0.723,0.898
Health Utility	0.963	0.930	0.881,0.959
SF-36			
Physical Function	0.975	0.950	0.916,0.971
Role Physical	0.919	0.852	0.758,0.912
Bodily Pain	0.957	0.919	0.864,0.952
General Health	0.943	0.890	0.817,0.935
Vitality	0.891	0.802	0.681,0.881
Social Function	0.846	0.736	0.583,0.838
Role Emotional	0.818	0.697	0.528,0.812
Mental Health	0.918	0.850	0.755,0.911
PCS-36	0.958	0.920	0.866,0.953
MCS-36	0.847	0.740	0.589,0.841
MODQ	0.955	0.916	0.859,0.951
RMQ	0.957	0.919	0.864,0.952
VKPS	0.929	0.857	0.766,0.915
NSS	0.948	0.895	0.826,0.938

For explanations see footnote to Table 4.

Table 6. Ceiling and floor effects

Instrument	Ceiling Effect (%)	Floor Effect (%)
Adaptive Questionnaire		
Walking	1.86	0
Handling Objects	2.33	0
Daily Activities	6.05	0
Pain / Discomfort	0.47	0
Feelings	0	0
Health Utility	0	0
SF-36		
Physical Function	2.79	0.47
Role Physical	13.49	5.58
Bodily Pain	1.86	1.40
General Health	0.47	0
Vitality	1.40	1.40
Social Function	34.42	1.86
Role Emotional	59.53	0.47
Mental Health	3.72	0.47
PCS-36	0	0
MCS-36	0	0
MODQ	1.86	0
RMQ[a]	5.12	0.47
VKPS[a]	10.23	0.93
NSS	19.07	0.47

For explanations see footnote to Table 4.

[a]Scores for RMQ and VKPS were reversed for the calculation of ceiling and floor effects so that ceiling always refers to good health and floor to poor health

Table 7. Correlations between the adaptive questionnaire and other measures of HRQOL

Instrument	PCS-36	MCS-36	MODQ	RMQ	VKPS	NSS
Adaptive Questionnaire						
Walking	0.79	0.31	0.76	-0.76	-0.53	0.53
Handling Objects	0.69	0.20	0.65	-0.66	-0.41	0.38
Daily Activities	0.83	0.28	0.77	-0.77	-0.62	0.46
Pain / Discomfort	0.82	0.32	0.81	-0.74	-0.73	0.51
Feelings	0.33	0.72	0.48	-0.45	-0.47	0.30
Health Utility	0.67	0.49	0.75	-0.70	-0.64	0.55

For explanations see footnote to Table 4.

Table 8. Mean scores (SD) on all instruments according to sex

Instrument	Males	Females	P-value
Adaptive Questionnaire			
Walking	46.49 (9.82)	43.02 (10.45)	0.014
Handling Objects	48.26 (8.82)	42.01 (6.93)	0.000
Daily Activities	44.27 (10.36)	41.87 (9.84)	0.085
Pain / Discomfort	41.97 (7.57)	40.32 (6.59)	0.089
Feelings	44.24 (8.95)	45.42 (9.32)	0.347
Health Utility	0.879 (0.091)	0.861 (0.104)	0.184
SF-36			
PCS-36	40.65 (10.64)	35.90 (11.33)	0.002
MCS-36	50.23 (11.11)	51.93 (11.21)	0.269
MODQ	38.57 (13.37)	35.41 (12.03)	0.070
RMQ	9.57 (7.06)	11.18 (6.23)	0.077
VKPS	32.24 (25.00)	36.55 (25.97)	0.221
NSS	44.34 (12.26)	43.7 (11.96)	0.702

For explanations see footnote to Table 4.

Table 9. Mean scores (SD) on all instruments according to age

Instrument	<45	45-54	55-64	65+	p-value
Adaptive Quest.					
Walking	47.63 (10.17)	45.45 (9.47)	43.22 (11.87)	39.69 (7.92)	0.000
Handling Objects	47.19 (7.31)	45.15 (8.86)	43.84 (9.66)	41.27 (7.30)	0.001
Daily Activities	44.83 (9.89)	43.01 (10.41)	41.34 (12.51)	41.10 (7.28)	0.154
Pain / Discomfort	42.14 (7.32)	40.86 (6.21)	39.80 (8.03)	40.53 (6.61)	0.338
Feelings	44.10 (7.50)	44.52 (8.88)	44.64 (12.4)	46.83 (8.74)	0.426
Health Utility	0.885 (0.075)	0.870 (0.107)	0.836 (0.147)	0.868 (0.062)	0.079
SF-36					
PCS-36	40.71 (10.87)	39.91 (9.71)	34.70 (12.68)	34.37 (10.74)	0.002
MCS-36	49.43 (10.11)	50.52 (11.56)	51.39 (12.25)	54.58 (11.07)	0.090
MODQ	38.72 (13.04)	37.43 (11.63)	33.98 (15.09)	35.46 (10.70)	0.223
RMQ	9.35 (6.89)	10.06 (6.23)	11.73 (7.07)	11.66 (6.09)	0.150
VKPS	33.81 (26.95)	34.80 (25.26)	36.44 (26.39)	34.20 (23.48)	0.962
NSS	45.6 (10.95)	43.85 (11.87)	42.62 (14.36)	42.61 (12.00)	0.472

For explanations see footnote to Table 4.

Table 10. Mean scores (SD) on all instruments according to satisfaction with current symptoms

Instrument	Very Satisfied	Somewhat Satisfied	Neutral	Somewhat Dissatisfied	Very Dissatisfied	P-value
Adaptive Quest.						
Walking	54.00 (8.78)	48.33 (9.87)	47.65 (10.06)	45.35 (8.81)	41.06 (10.16)	< 0.001
Handling Objects	53.48 (7.58)	46.14 (7.69)	47.65 (8.60)	45.24 (8.97)	42.19 (7.16)	< 0.001
Daily Activities	57.82 (10.45)	49.08 (10.05)	46.20 (8.43)	43.14 (8.93)	38.49 (8.33)	< 0.001
Pain / Discomfort	52.90 (7.49)	45.69 (5.15)	44.22 (5.91)	41.23 (5.64)	37.36 (5.66)	< 0.001
Feelings	51.64 (6.05)	51.93 (6.71)	47.86 (6.70)	46.46 (6.55)	40.58 (9.87)	< 0.001
Health Utility	0.9541 (0.0481)	0.9256 (0.0402)	0.9115 (0.0442)	0.8865 (0.0581)	0.8202 (0.1186)	< 0.001
SF-36						
PCS-36	50.33 (10.82)	43.71 (9.96)	41.40 (10.10)	38.55 (10.75)	33.76 (10.22)	< 0.001
MCS-36	56.74 (4.81)	56.55 (9.86)	55.65 (7.23)	54.68 (8.36)	45.75 (12.22)	<0.001
MODQ	49.87 (10.55)	44.31 (10.98)	44.95 (8.49)	38.57 (9.70)	29.78 (11.95)	<0.001
RMQ	3.75 (5.07)	6.38 (6.48)	6.50 (4.81)	9.87 (5.55)	13.94 (6.01)	<0.001
VKPS	11.25 (19.55)	17.98 (18.06)	20.47 (19.48)	31.45 (24.31)	47.98 (22.89)	<0.001
NSS	52.79 (7.44)	46.68 (7.43)	48.76 (7.70)	45.18 (10.72)	39.98 (13.84)	<0.001

For explanations see footnote to Table 4.

Table 11. Discrimination (F-statistic) and relative efficiency (RE) according to satisfaction with current symptoms, duration of current episode, compensation, and sciatica

Instrument	Satisfaction F (RE)	Duration F (RE)	Compensation F (RE)	Sciatica F (RE)	Mean RE
Adaptive Quest.					
Walking	7.61 (0.27)	7.54 (0.35)	23.06 (0.87)	36.82 (1.00)	0.62
Handling Objects	7.36 (0.26)	4.97 (0.23)	17.70 (0.67)	15.95 (0.43)	0.40
Daily Activities	18.33 (0.65)	11.83 (0.55)	24.19 (0.91)	20.46 (0.56)	0.67
Pain / Discomfort	28.33 (1.00)	16.19 (0.75)	21.72 (0.82)	33.69 (0.91)	0.87
Feelings	13.97 (0.49)	3.10 (0.14)	5.26 (0.20)	4.36 (0.12)	0.24
Health Utility	14.71 (0.52)	9.90 (0.46)	18.86 (0.71)	16.20 (0.44)	0.53
SF-36					
PCS-36	10.74 (0.38)	13.46 (0.63)	21.56 (0.81)	19.78 (0.54)	0.59
MCS-36	12.38 (0.44)	0.68 (0.03)	6.88 (0.26)	4.73 (0.13)	0.22
MODQ	22.00 (0.78)	21.49 (1.00)	26.26 (0.99)	29.04 (0.79)	0.89
RMQ	19.41 (0.69)	16.45 (0.77)	26.53 (1.00)	26.69 (0.72)	0.80
VKPS	18.29 (0.65)	18.82 (0.88)	11.51 (0.43)	14.57 (0.40)	0.59

For explanations see footnote to Table 4.

Figure 1. Distribution of the overall health utility score in the study population

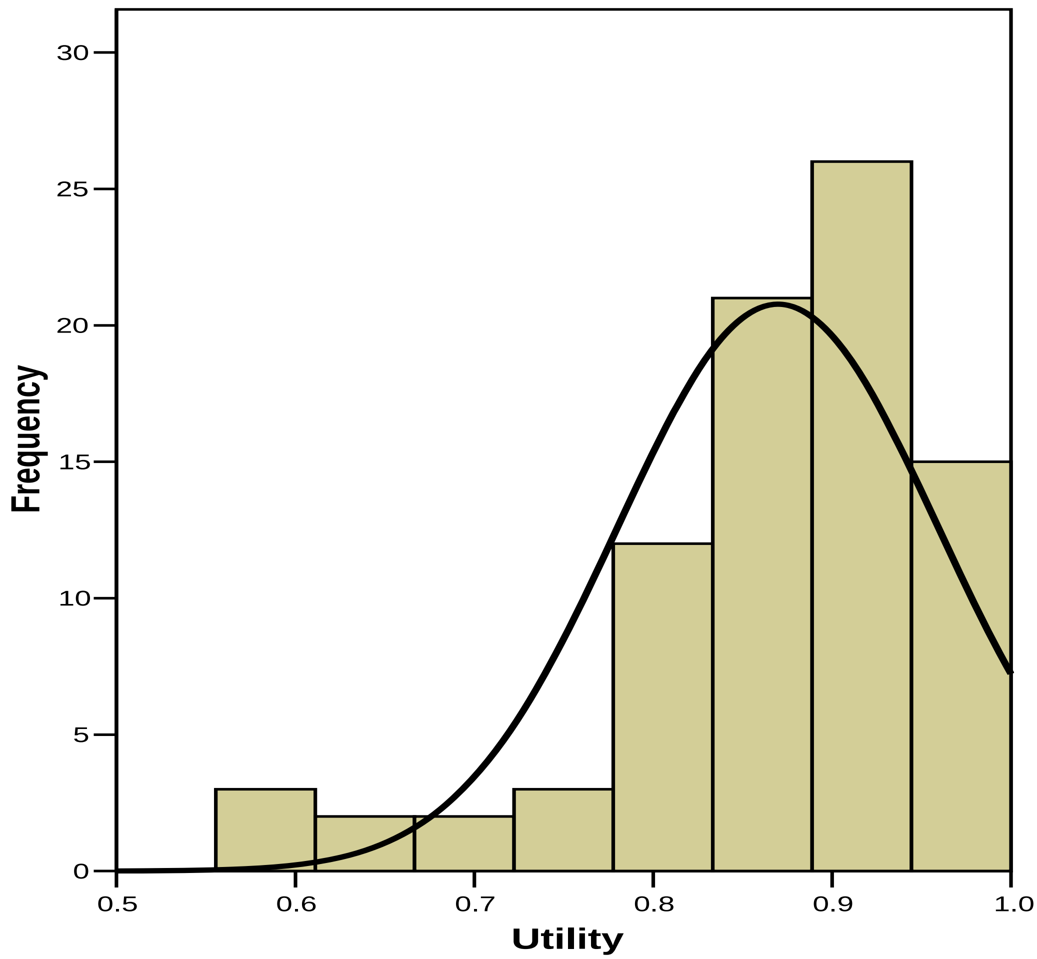


Figure 2. Distribution of scores for the Walking domain of the adaptive questionnaire in the study population

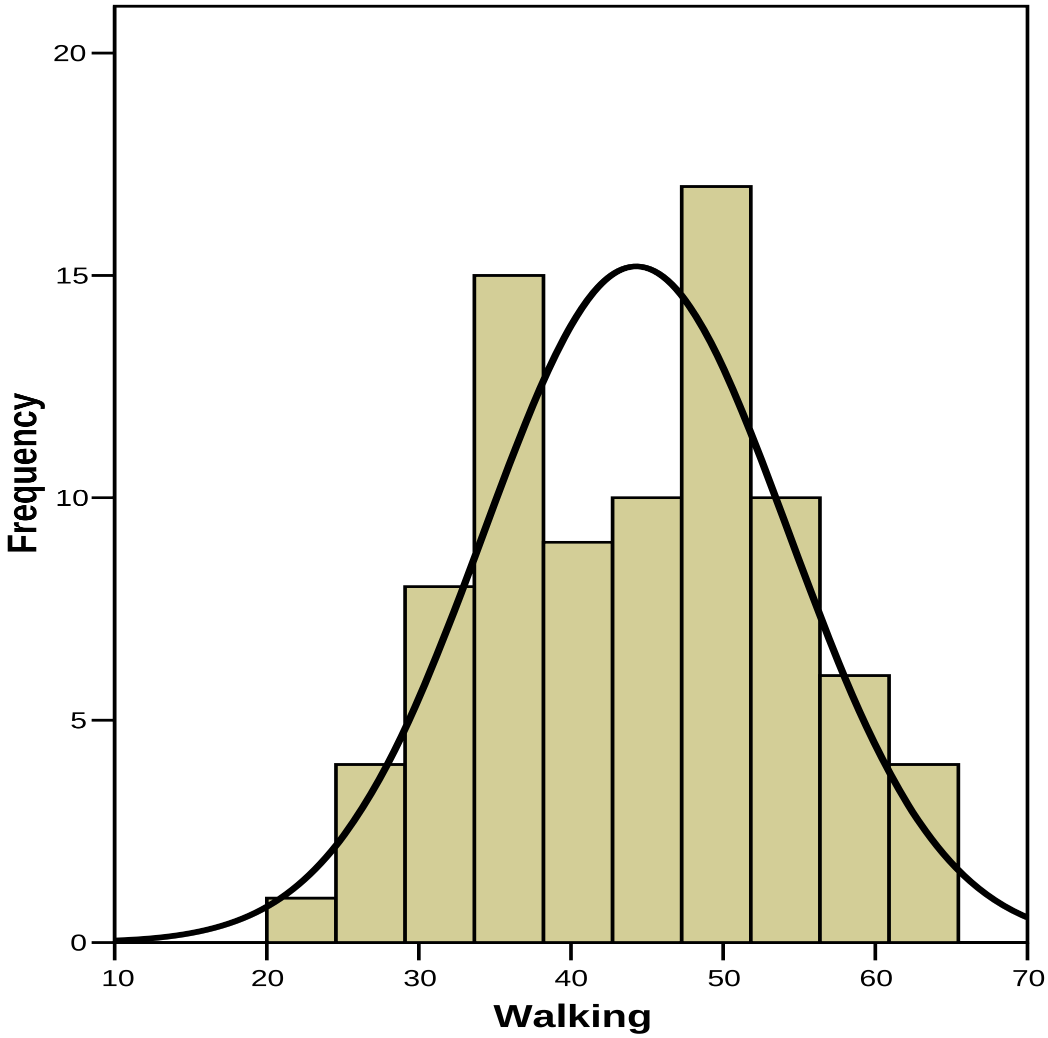


Figure 3. Distribution of scores for the Handling Objects domain of the adaptive questionnaire in the study population

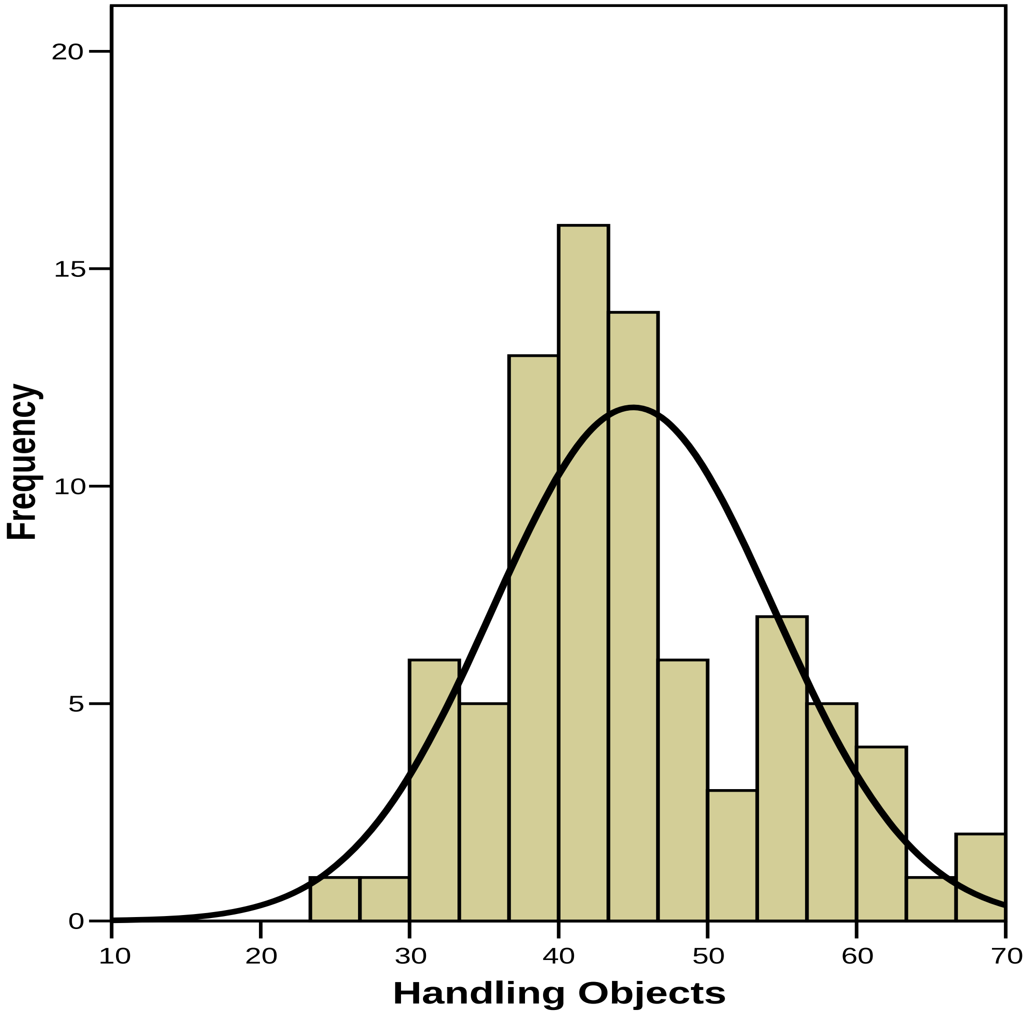


Figure 4. Distribution of scores for the Daily Activities domain of the adaptive questionnaire in the study population

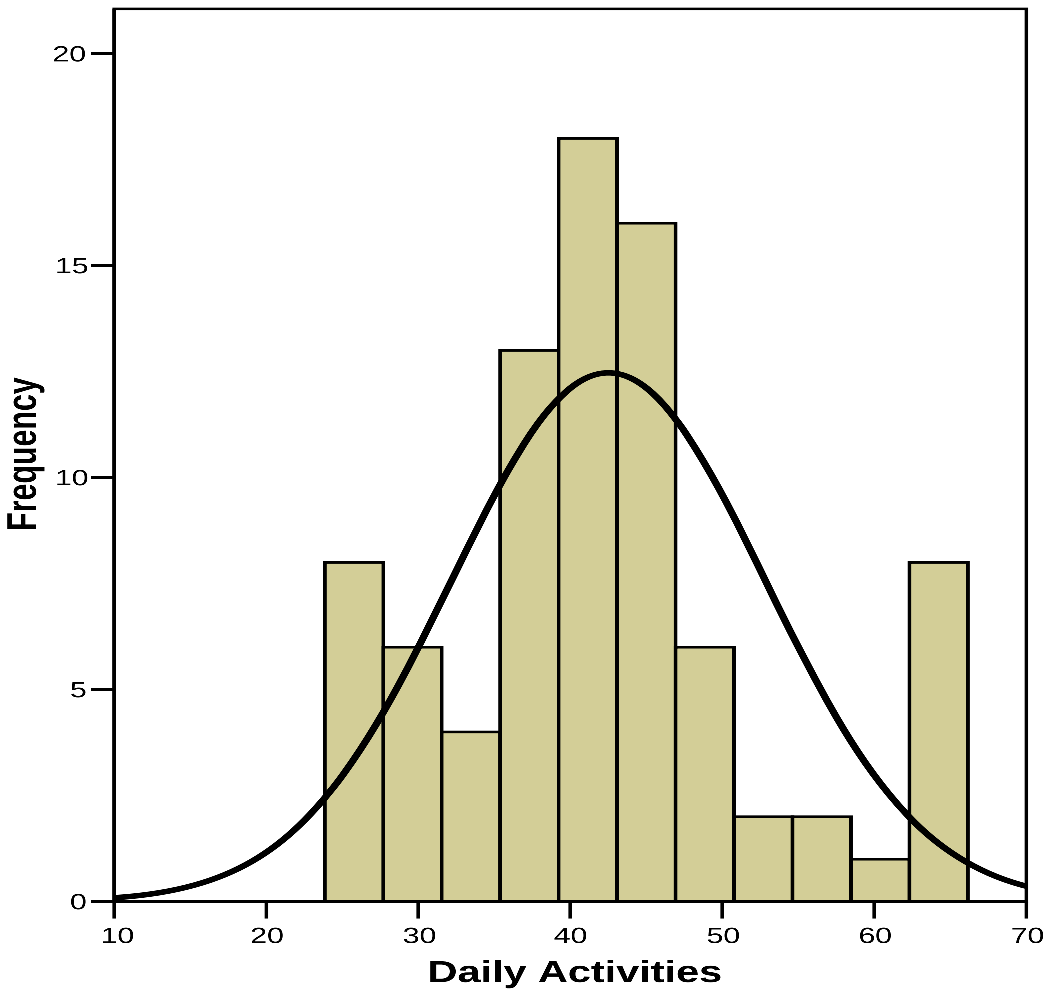


Figure 5. Distribution of scores for the Pain / Discomfort domain of the adaptive questionnaire in the study population

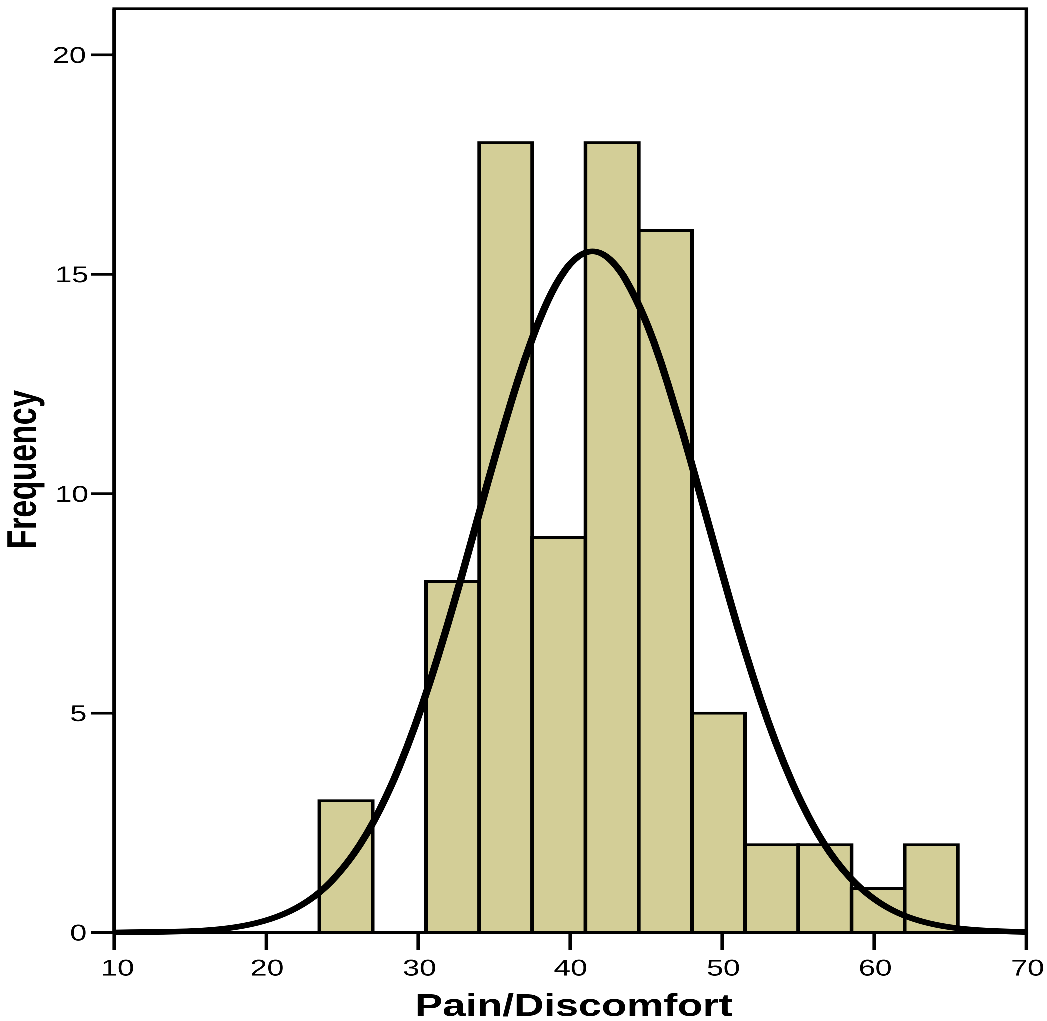


Figure 6. Distribution of scores for the Feelings domain of the adaptive questionnaire in the study population

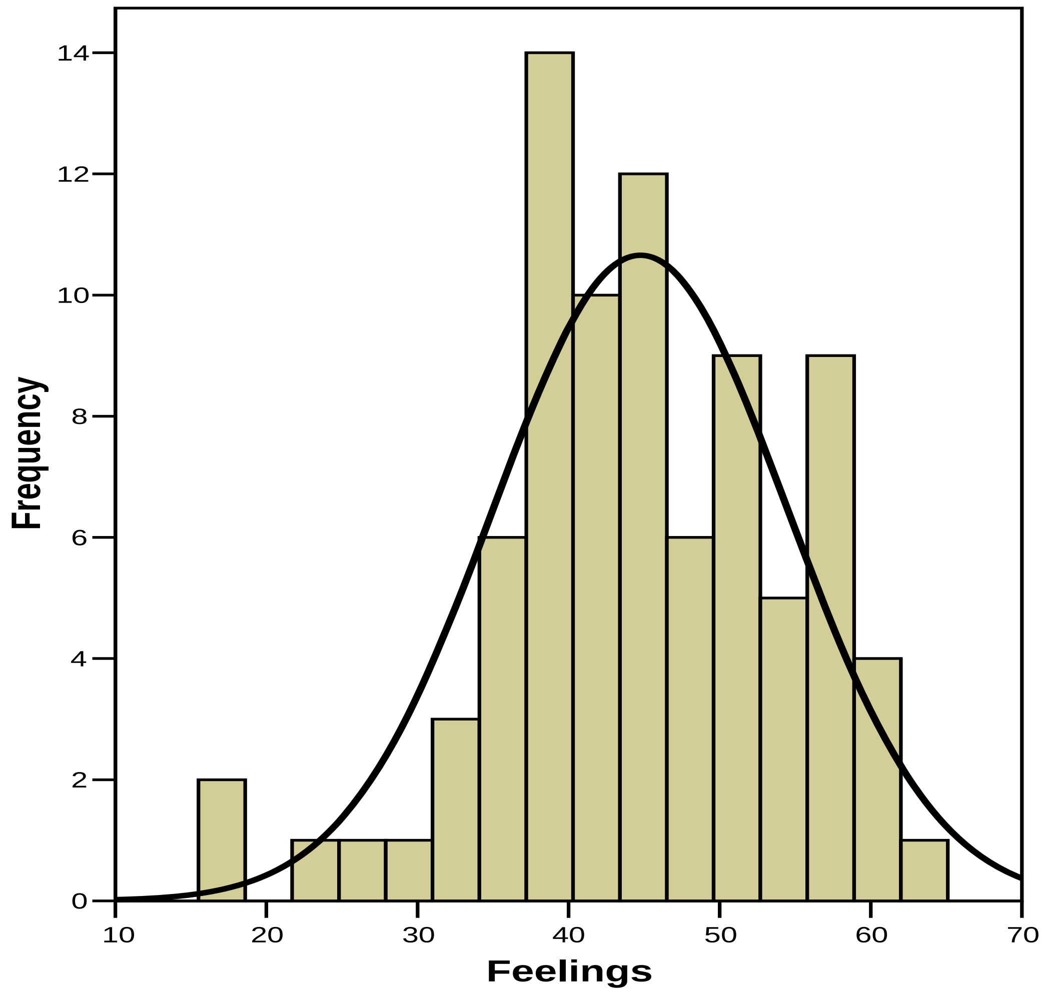


Figure 7. Correlation between Pain / Discomfort (adaptive) and Modified Oswestry Disability Questionnaire (Pain/Disability Scale)

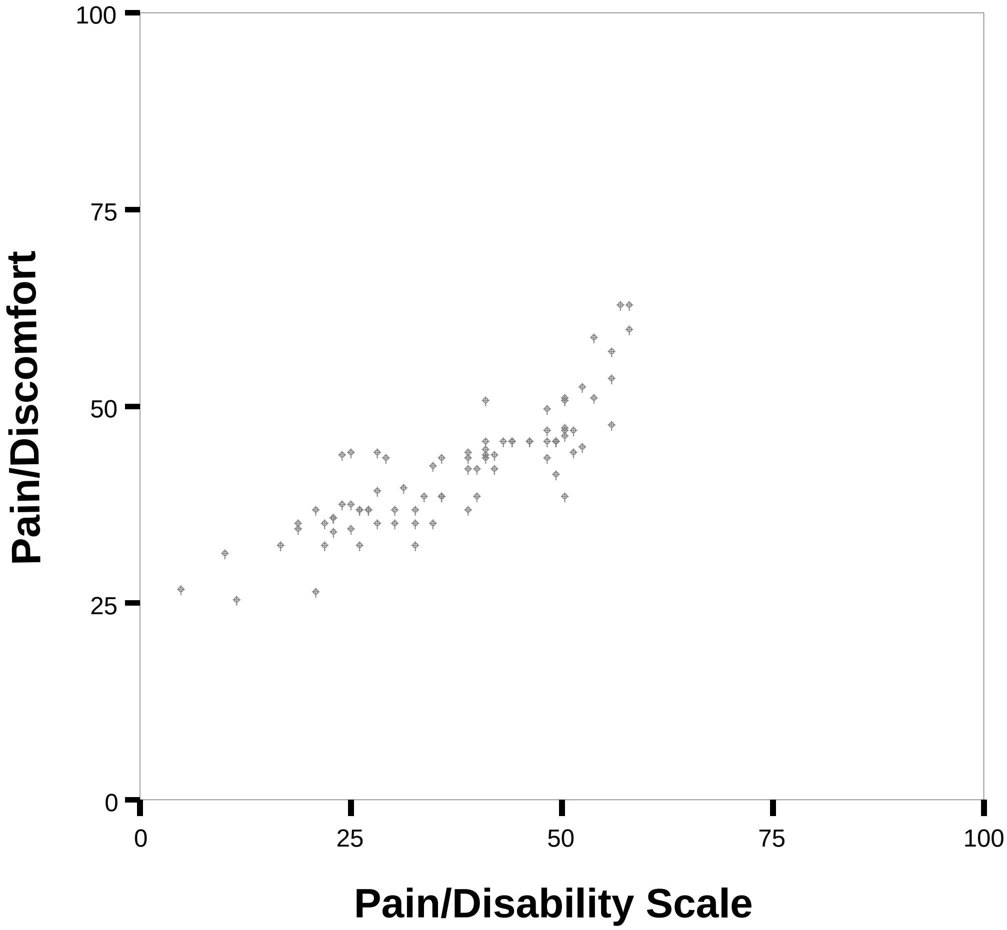


Figure 8. Correlation between Pain / Discomfort (adaptive) and Roland-Morris Questionnaire

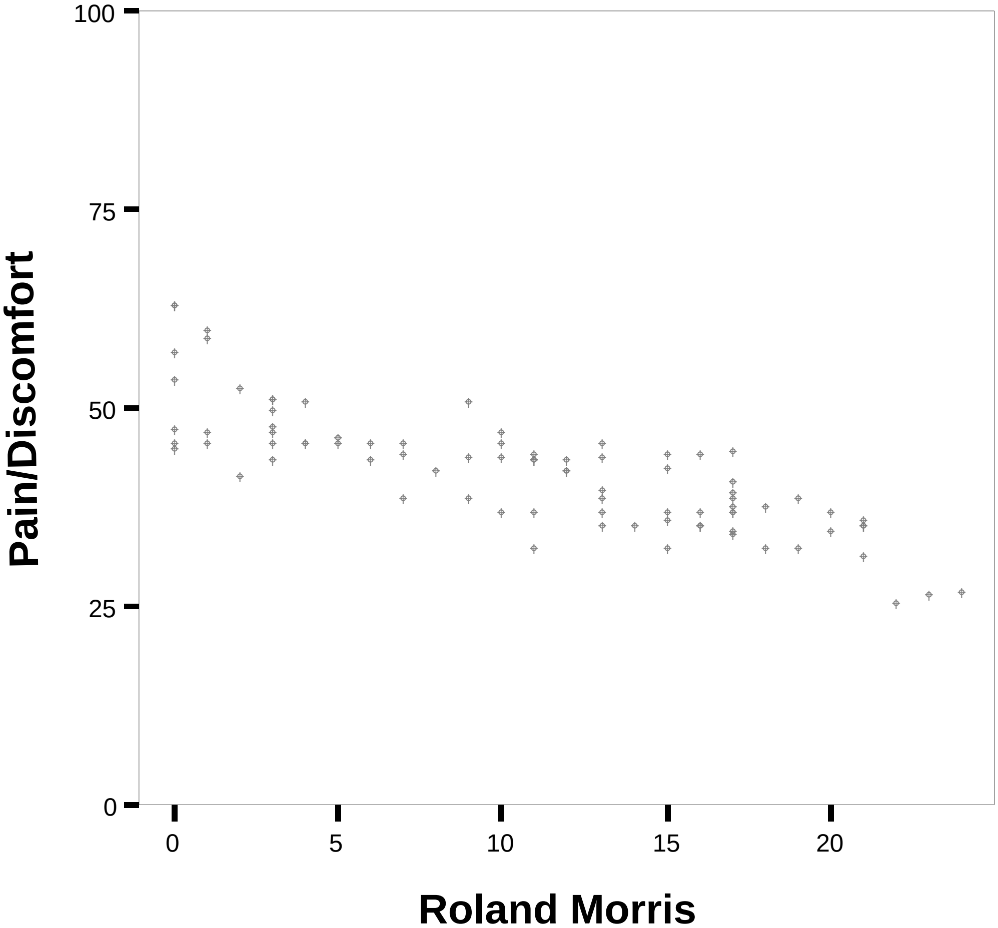


Figure 9. Relationship between overall health utility (adaptive) and satisfaction with current symptoms

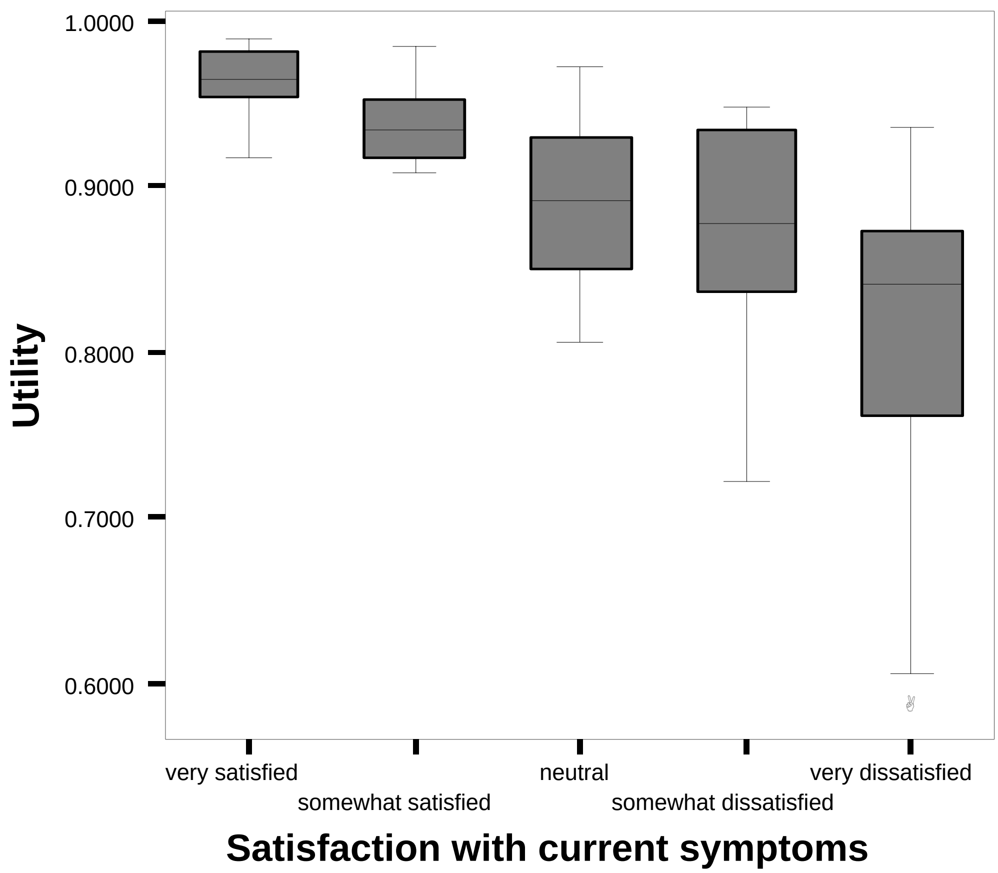
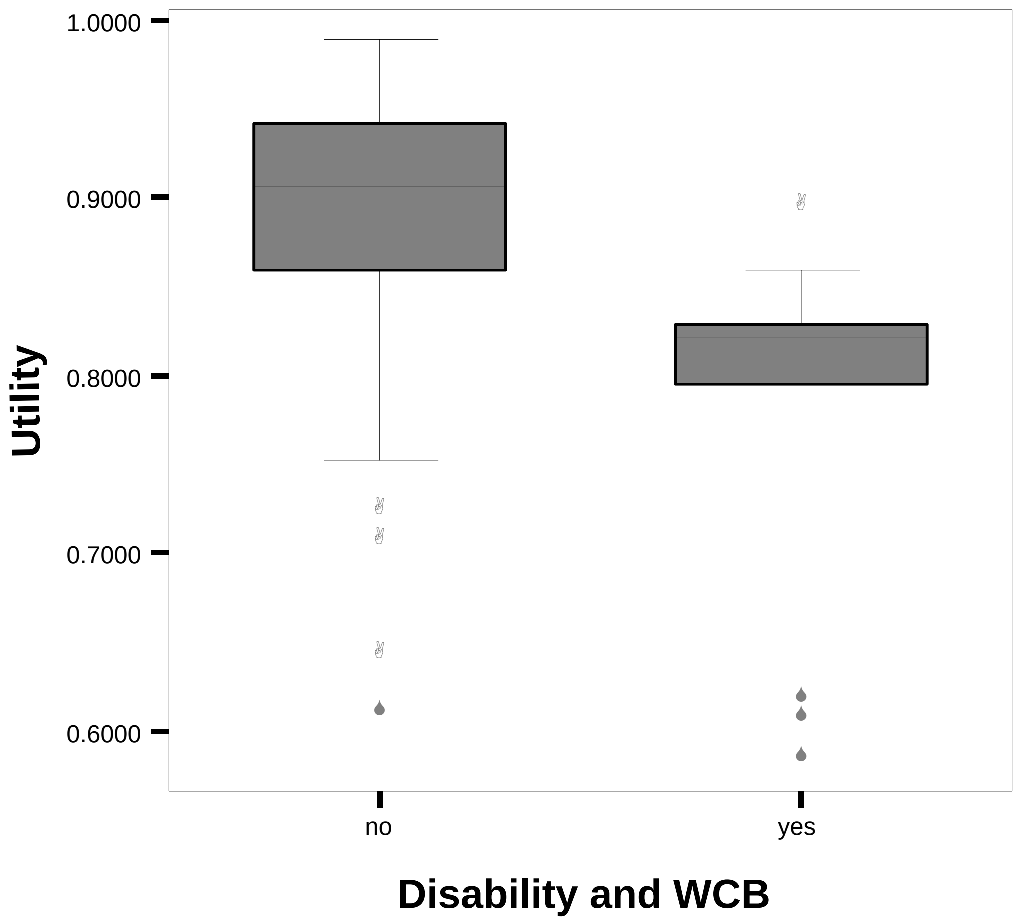


Figure 10. Relationship between overall health utility (adaptive) and receiving disability or workers' compensation for back pain



All rights reserved. The Workers' Compensation Board of B.C. encourages the copying, reproduction, and distribution of this document to promote health and safety in the workplace, provided that the Workers' Compensation Board of B.C. is acknowledged. However, no part of this publication may be copied, reproduced, or distributed for profit or other commercial enterprise or may be incorporated into any other publication without written permission of the Workers' Compensation Board of B.C.

Additional copies of this publication may be obtained by contacting:

Research Secretariat
6951 Westminster Highway
Richmond, B.C. V7C 1C6
Phone (604) 244-6300 / Fax (604) 244-6295
Email: resquery@worksafebc.com