Alex Papou\textsuperscript{1}, Salma Hussain\textsuperscript{2}, Daniel McWilliams\textsuperscript{3}, Weiya Zhang\textsuperscript{3} and Michael Doherty\textsuperscript{3}

Responsiveness of SF-36 Health Survey and Patient Generated Index in people with chronic knee pain commenced on oral analgesia: analysis of data from a randomised controlled clinical trial

\textsuperscript{1}Department of Rheumatology, Queen's Medical Centre, Derby Road, Nottingham, NG7 2UH, United Kingdom
\textsuperscript{2}Frimley Park Hospital, Portsmouth Road, Frimley, GU16 7UJ, United Kingdom
\textsuperscript{3}Division of Rheumatology, Orthopaedics and Dermatology, University of Nottingham, Academic Rheumatology, Clinical Sciences Building, Nottingham City Hospital, Hucknall Road, Nottingham, NG5 1PB, United Kingdom

Corresponding author

Weiya Zhang

Academic Rheumatology, Clinical Sciences Building, City Hospital, Nottingham, NG5 1PB, United Kingdom

Email Weiya.Zhang@nottingham.ac.uk

Phone 0115 823 1756

Fax 0115 823 1757
Introduction

Health-related quality of life (HRQoL) instruments are important in the assessment of efficacy of various treatments and healthcare delivery decisions, and are of particular significance in people with chronic painful conditions such as osteoarthritis (OA) which is known to progress with age and significantly influence people's lives.

Over the last decades a variety of HRQoL instruments have been developed, the majority being self-completed questionnaires. These instruments can be divided into three groups: generic, which are used for a broad range of conditions; disease-specific, which contain items particularly influenced by the specific problems associated with the disease of interest; and individualised patient-centred instruments. The SF-36 (Short Form Health Survey consisting of 36 items) is a widely used generic instrument, validated in many populations, which allows comparisons between different diseases and to the population 'norm' for scales [1]. The Patient Generated Index (PGI) is an individualised tool where people are able to choose and rate the most important areas where their HRQoL is affected [2]. Its validity, reliability and responsiveness have been examined in various conditions, including rheumatic diseases [3].

Significant difficulties arise when various HRQoL instruments are being compared with each other and various instruments have been developed to ensure adequate methodology and reporting, such as COSMIN criteria and the CONSORT Patient-Reported Outcome (PRO) extension [6]. Responsiveness has been proposed as criterion for selecting HRQoL instruments for clinical trials [9]. There is a limited literature assessing responsiveness of different HRQoL instruments in musculoskeletal conditions. Only a few studies have compared responsiveness between instruments in patients with osteoarthritis [10] but none have included the PGI.

The objectives of this study were: (1) to assess the responsiveness of the SF-36 and PGI in people with knee pain predominantly due to OA who were given oral paracetamol and/or ibuprofen, in the context of a community-based randomised controlled trial (RCT); and (2) to perform content analysis of the SF-36 and PGI aiming to identify differences between the instruments and causes of different responsiveness.
Methods

Participants and study design

This study was undertaken in people who participated in an RCT evaluating the comparative efficacy and safety of widely available over-the-counter analgesics: paracetamol, ibuprofen or the combination of the two which took place in the UK from June 2007 to July 2008 (ISRCTN registry, www.isrctn.com, trial registration number ISRCTN77199439) [13]. Participants in this trial were randomly allocated to one of 4 oral treatments which were taken three times daily for 13 weeks: paracetamol 1000 mg tds, ibuprofen 400 mg tds, paracetamol 500 mg plus ibuprofen 200 mg tds, or paracetamol 1000 mg plus ibuprofen 400 mg tds. The majority were recruited by postal questionnaire in Nottinghamshire and the remainder from general practice lists or from local advertisement. The subjects’ written consent was obtained according to the Declaration of Helsinki and the study has been approved by the Southampton & South West Hampshire Research Ethics Committee A in March 2007 (ref: 07/Q1702/19). Inclusion criteria were: age 40 years and older; knee pain for most of the past 3 months and for 4 of the preceding 7 days; Steinbrocker functional classification of I-III [14]; and pain affecting the index knee (after a washout period if currently taking analgesics) of 30-80mm on a 100mm visual analogue scale over the previous 48 hours for at least one of number of specified daily activities (walking on a flat surface, going up/down stairs, at night, sitting, lying, standing upright). For further details of the original study including randomisation and blinding methods, sample size calculation, statistical analyses and trial outcomes, etc., see Doherty et al 2011 [13].

HRQoL assessments

In addition to pain and functional assessments which have been reported in the study [13], each participant was asked to complete the SF-36 version 2 and the PGI questionnaires at baseline, day 10, week 7 and week 13. The participants had access to the research nurse for clarification of how to complete the questionnaires if required. The data were collected and transferred to SPSS for Windows version 20 where statistical calculations were performed. Both overall results and the results according to treatments were analysed.
The SF-36 Each of the 8 scale scores were derived using standard methodology [15] and were then norm-transformed using the weightings and transformations recommended by Jenkinson 1999 [16]. Missing data were imputed with the average value for the remaining items on the subscale, unless more than 50% of items were missing (in which case the subscale was not calculated) [15]. This produces values between 0 and 100 with the average population in the UK having an average score of 50 and a standard deviation of 10 in each domain. The lower the score the worse is the condition reported. The two summary scores - Mental Component Score (MCS) and Physical Component Score (PCS) were calculated from the corresponding scale scores.

The PGI This is an individualised self-reported 18-item questionnaire that comprises 3 parts (Online Resource Table S1). In part 1 responders are asked to list any 5 areas of their life affected by the knee pain and its treatment, together with a sixth predefined area headed 'All other areas of your life affected by your knee pain and its treatment'. In part 2 participants are asked to score the impact that each of the six components has on their life, ranging from 0 ('as bad as it could possibly be') to 6 ('as good as it could possibly be'). In the final part they are asked to weight the importance of each area using 10 points to distribute as they choose. When the participants completed the PGI at day 10, week 7 and week 13 they were reminded of the areas they entered at baseline but not the weightings they afforded to each. The PGI score was calculated using a standard approach which gives a value ranging from 0 being the worst quality of life possible to 6 being the best [2].

Responsiveness

The responsiveness and 95% confidence intervals (CI) were measured as a standardised response mean (SRM), which is calculated as:

\[ SRM = \frac{\bar{X}_0 - \bar{X}_1}{SD_{01}} \]

where \( \bar{X}_0 \) is the mean at baseline and \( \bar{X}_1 \) is the mean at follow-up and \( SD_{01} \) is the standard deviation of the change [17]. This will allow for comparison between different scoring systems. The data were interpreted using the usual consensus
that absolute responsiveness values between 0 and 0.2 represent no response, values of 0.2 to 0.5 indicate a weak response, values of 0.5 to 0.8 indicate a moderate response and values above 0.8 represent a strong response [18].

Content analysis

An attempt was made to identify prevailing domains in the PGI, their frequency, and areas of life which are of most concern for the patients. Initially we attempted to stratify PGI responses using the 8 domain scores from the SF-36 questionnaire but this was impossible since the different HRQoL areas in the SF-36 were intersecting or too broad. Therefore all individual PGI responses were aggregated into other, more discrete areas, and their frequencies were calculated. Items asked about in the SF-36 survey were compared to those in the PGI domains and individual responses in order to detect items present in the PGI but not in SF-36.

Results

Characteristics of the study population

Baseline demographic of the participants appear in Table 1. Of 884 people completed both SF-36 and PGI questionnaires at baseline, 783 (11.4% missing) and 776 (12.2% missing) at day 10, 647 (26.8% missing) and 642 (27.4% missing) at week 7, and 807 (8.7% missing) and 597 (32.5% missing) at week 13 completed the SF-36 and PGI respectively. Having analysed four treatment groups separately we found no statistically significant difference in responsiveness between them, therefore a combined analysis was undertaken for this report. Of the 884 participants, 85% satisfied American College of Rheumatology criteria for knee OA (1986) and 63% had radiographic evidence of knee OA (definite osteophyte and definite narrowing in at least one compartment).

Responsiveness
In general, the SF-36 has shown no or weak responsiveness in the majority of the domains (Table 2). Some moderate responses were observed in the Bodily Pain Score for which the SRM values ranged from 0.49 to 0.63 upon different time points. In contrast, PGI had weak responsiveness at day 10 but moderate responsiveness throughout the rest of the study period (SRM 0.41 to 0.61) (Table 2). Apart from the Bodily Pain Score, none of the other SF-36 scores were comparable to PGI in terms of the responsiveness to the analgesics used in this trial.

**Content analysis of the SF-36 and the PGI**

A total of 3356 items were recorded within the PGI at baseline. It was found that all responses could be mapped into 13 main domains and their frequencies were then calculated (Figure 1). The most frequent areas for the participants with knee pain appeared to be: mobility (31.0%); sports and activities (14.9%); and house work and gardening (12.9%). It was apparent that many responses in the PGI were highly individualised and would not be captured within the specific topics listed within the SF-36. There were multiple individual responses in each of the 13 PGI domains which are not covered by SF-36 questions, such as many individual sports activities, gardening, family contact, driving, hobbies, etc. (Online Resource Table S2). Several domains appeared not to be represented in the SF-36 survey at all (family and relationships, driving, shopping, and hobbies/leisure/holidays).

**Discussion**

This study focused on middle-aged and older community-derived people with chronic knee pain, the majority of whom had clinical and radiographic features of knee OA. The results show that following commencement of paracetamol and/or ibuprofen in an RCT setting the PGI is more responsive than the SF-36. Only the SF-36 Bodily Pain scale had comparable responsiveness values. Five out of eight SF-36 domains and one of two summary scores showed no response to oral analgesics at any time point and only 14 out of the 36 questions asked in the SF-36 appeared relevant in calculations of scale scores which responded to oral analgesics.
The SF-36 is probably one of the most widely used and best studied generic HRQoL tools. Many of the SF-36 questions are relevant for lower limb OA. For example, the SF-36 Physical Function score is calculated based on ten items, three of which assess walking, two assess climbing, with other items assessing kneeling, bending, running, etc. The SF-36 Bodily Pain scale asks about pain severity and functional limitations due to pain. Although it is generally considered a reliable tool for the assessment of HRQoL in arthritis [20], debate continues as to whether the SF-36 is sufficiently responsive for OA and whether it captures all relevant information. For example, Rannou et al (2007) have shown that the two summary SF-36 scores (the MCS and PCS) are not optimal for use in OA due to multiple intrinsic issues, especially with respect to mixing of data [21]. Salaffi et al (2005) concluded that a disease-specific instrument WOMAC (Western Ontario and McMaster Universities Index of Osteoarthritis) is preferred to SF-36 for evaluating patients with lower limb OA [12]. Similarly in a study by Angst et al (2001) the WOMAC was significantly more responsive than the SF-36 in assessing function in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention, and for both instruments, the pain scales were more responsive than the function scales [10].

Although the PGI was developed in the 1990s it has been relatively little studied or used. It was shown to have good validity and responsiveness in graded structured reviews [4] but there is only a very limited literature comparing the PGI and other HRQoL instruments. We consider that the main merit of this instrument is that it can exclude the 'noise' of questions which are not of direct concern to the individual person, and capture those areas of life which are important for people but often not represented in generic HRQoL tools. More than half of the items that patients were entering into the PGI did not have corresponding questions in the SF-36 survey, although we understand some of these items could have possibly been captured indirectly by the SF-36 general questions. Interestingly, pain per se was not the predominant complaint mentioned in the PGI in our group, and issues of mobility and various activities were of far more concern for the patients (Figure 1). We believe this reflects the flexibility and ability of the PGI to prioritise items, and this not only can improve responsiveness but also comes closer to what we conceptualise as 'HRQoL' which is always specific to an individual [22]. Further study of the PGI and other individualised instruments in OA, and other conditions, would seem warranted.
The study has several caveats. Firstly, a disease (OA) specific questionnaire was not included so assessment of performance between the three types of HRQoL instruments (generic, disease-specific, individualised) could not be undertaken. Such an assessment in OA would be of great interest. Secondly, we used only one widely recognised quantitative method (SRM) to assess the responsiveness to one treatment modality, specifically over-the-counter doses of established oral analgesics. Other methods for assessing responsiveness and other treatments with different efficacy potential and different associated contextual responses may result in different outcomes in HRQoL instruments. The same applies to content analysis as evaluating content validity still remains a challenge and various methods exist [23]. Thirdly, although OA is a chronic condition the length of our study was only 13 weeks so the relative performance of the two HRQoL instruments over a longer time period remains unknown. Finally, the PGI is a relatively complex instrument which participants can have difficulties with if there is no access to appropriate guidance and support.

**Conclusions**

The PGI has been shown to be responsive to analgesic treatment in community-derived people with knee pain. In contrast only one domain in the SF-36 (Bodily Pain) is responsive to this treatment. The PGI is able to elicit HRQoL areas which are not captured by generic instruments and appears to be a sensitive tool in the assessment of HRQoL in knee pain and OA. It appears to merit further study by inclusion in future OA clinical trials.

**Compliance with Ethical Standards**

Funding source: There is no specific grant associated with this study. Reckitt Benckiser Healthcare International Ltd funded the original randomised controlled trial which provided the data for analysis but the conception and design of the current nested study was undertaken independently to the funder with their approval.

Disclosure of potential conflicts of interest: The authors declare no conflicts of interest.
Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent: Informed consent was obtained from all individual participants included in the study.

Acknowledgments

The authors are grateful to Reckitt Benckiser Healthcare International Ltd for financial support for the initial RCT and to the other investigators and research nurses involved in the trial which generated the data analysed in the present study.
References


**Figure legend**

Figure 1

Areas of life affected by knee pain as self-reported by the patients in baseline Patient Generated Index (PGI).