Health-Related Quality of Life in Adults with Osteogenesis Imperfecta

Chaya N. Murali, MD[1,2], **Brady Slater, MS**[1], **Salma Musaad, MD, PhD**[3,4], **David Cuthbertson, MS**[5], **Dianne Nguyen, BS**[1], **Alicia Turner, MSN, RN**[1], **Mahshid Azamian, MD**[1], **Laura Tosi, MD**[6], **Frank Rauch, MD**[7], **V. Reid Sutton, MD**[1,2], **Brendan Lee, MD, PhD**[1,2], **Members of the BBD Consortium**[*], **Sandesh C.S. Nagamani, MBBS, MD**[1,2]

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

[2]Texas Children's Hospital, Houston, Texas, USA

[3]Department of Pediatrics-Nutrition, Baylor College of Medicine, Houston, TX, USA

[4]United States Department of Agriculture/Agricultural Research Service Children's Nutrition Research Center, Houston, TX, USA

[5]College of Medicine, University of South Florida, Tampa, FL, USA

[6]Bone Health Program, Children's National Health System, Washington, D.C., USA

[7]Shriner's Hospital of Children and McGill University, Montreal, QC, Canada

[8]Department of Orthodontics and Dentofacial Orthopedics, University of Missouri-Kansas City, Kansas City, MO, USA

[9]Craniofacial Anomalies & Regeneration Section, National Institute of Dental and Craniofacial Research, Bethesda, MD, USA

[10]Pediatric Orthopedic Surgery, University of Nebraska Medical Center, Omaha, NE, USA

[11]Division of Medical Genetics, Alfred I DuPont Hospital for Children, Wilmington, DE, USA

[12]Department of Orthopedic and Sports Medicine, University of Washington, Seattle, WA, USA

Corresponding Author: Sandesh C.S. Nagamani, MD, Telephone number: 713-798-8335, nagamani@bcm.edu.
[*]Members of the Brittle Bone Disease Consortium (not listed in the author list): Francis Glorieux[7], Jean-Marc Retrouvey[8], Janice Lee[9], Paul Esposito[10], Maegen Wallace[10], Michael Bober[11], David Eyre[12], Danielle Gomez-Reynolds[13], Gerald Harris[14], Tracy Hart[15], Mahim Jain[16], Deborah Krakow[17], Eric Orwoll[18], Lindsey Nicol[18], Cathleen Raggio[19], Peter Smith[20].

Ethics Declaration:

The study procedures were approved by the Institutional Review Boards of all participating clinical sites (Baylor College of Medicine (Houston, TX), Kennedy Krieger Institute (Baltimore, MD), Nemours/Alfred I. DuPont Hospital for Children (Wilmington, DE), Oregon Health & Science University and Shriners Hospital for Children (Portland, OR), Shriners Hospital for Children (Chicago, IL), Shriners Hospital for Children (Montreal, QC), University of California Los Angeles (Los Angeles, CA), University of Nebraska Medical Center (Omaha, NE), Hospital For Special Surgery (New York, NY), Shriners Hospital for Children (Tampa, FL), and Children's National Medical Center (Washington, D.C.)) and informed consent was obtained from all participants or their legal guardians as required by the IRB of all participating clinical sites.

Conflict of Interest

The authors do not have any conflict of interest to declare.

Data Availability Statement

Participant-level data will not be available as these are governed by patient-privacy laws and NIH Rare Diseases Clinical Research Network's policies.

[13]Shriners Hospital for Children, Tampa, FL, USA

[14]Marquette University and Medical College of Wisconsin, USA

[15]Osteogenesis Imperfecta Foundation, Gaithersburg, MD, USA

[16]Departments of Bone and Osteogenesis Imperfecta, Kennedy Krieger Institute, Baltimore, MD, USA

[17]Departments of Orthopedic Surgery and Obstetrics and Gynecology, David Geffen School of Medicine, University of California, Los Angeles, CA, USA

[18]Department of Medicine, Division of Endocrinology, Oregon Health & Science University, Portland, OR, USA

[19]Hospital for Special Surgery, New York, NY, USA

[20]Shriner's Hospitals for Children, Chicago, IL, USA

Abstract

Patient-reported outcome measures (PROMs) are increasingly utilized as endpoints in clinical trials. The Short Form Health Survey-12 (SF-12v2) is a generic PROM for adults. We sought to evaluate the validity of SF-12v2 in adults with osteogenesis imperfecta (OI). Physical and mental health-related quality of life (HRQoL) were assessed in a large cohort of adults in a multicenter, observational, natural history study. Physical HRQoL scores were correlated with the Gillette Functional Assessment Questionnaire (GFAQ). We calculated sample sizes required in clinical trials with crossover and parallel-group designs to detect clinically meaningful changes in physical HRQoL. Three hundred and two adults with OI types I, III, and IV were enrolled. Physical HRQoL scores in the study population were lower than population norms. Physical HRQoL scores moderately correlated with GFAQ for OI types I and IV. We found no correlations between mental and physical HRQoL. From a clinical trial readiness perspective, we show that SF-12v2 reliably measures physical function in adults with OI and can be utilized in crossover trials to detect meaningful physical HRQoL changes with small sample sizes. This study shows that SF-12v2 can be used to measure changes in physical HRQoL in response to interventions in OI.

Keywords

osteogenesis imperfecta; quality of life; natural history; sample size

Introduction

In the clinical care and investigational settings, it is important to monitor not only "traditional" clinical and laboratory biomarkers of disease activity, but also outcomes related to lived experiences of patients. Thus, over the recent years, the use of patient-reported outcome measures (PROMs), which inquire about multiple facets of patients' well-being, has been increasing. PROMs typically comprise validated questionnaires that can assess physical symptoms, health-related quality of life (HRQoL), and functional and psychological status. As health status is assessed through direct input from patients or their proxies, most measures can be administered without the direct involvement of a healthcare

provider. This flexibility of administration allows for remote and repeated utilization of PROMs which not only empowers patients by providing autonomy to report their experiences, but also generates reliable data for both clinical care and clinical investigation[1,2]. PROMs can be broadly divided into: 1) generic measures, which can be used in healthy individuals as well as individuals affected by various disorders, and 2) disease-specific measures, which are applicable only to individuals with specific health conditions[3]. Generic measures offer the flexibility of administration to multiple groups and may yield comparisons among multiple disease groups and with healthy individuals. Disease-specific measures offer the precision of inquiring about health outcomes applicable to particular diagnoses, thus providing a more targeted readout of patient experience.

While changes in clinical endpoints or laboratory biomarkers can reflect the impact of a clinical intervention, these measures alone may not fully reflect the effect of an intervention on the day-to-day lives of patients and caregivers. PROMs can be powerful tools to evaluate patient experiences during drug development, and the use of PROMs in product labeling can assist prescribers and patients in understanding the potential impact of a drug. The Food and Drug Administration (FDA) guidance on the use of PROMs in both common and rare diseases places emphasis on the reliability, validity, and feasibility of the measure as well as the ability to detect a meaningful change[4]. However, there are few PROMs that have been validated in rare disease populations and thus their use in rare disease clinical trials can be challenging[4]. Not surprisingly, between 2002 and 2017, only 17 percent of orphan drug labels contained a patient-reported outcome and only one-fifth of these utilized a disease-specific, validated scale. Over the same 15-year period, less than half of rare disease pivotal clinical trials utilized PROMs as primary or secondary endpoints[5]. As developing a disease-specific PROM requires significant effort, one approach to overcome this limitation is to adapt and use existing PROMs for rare diseases. This approach is consistent with recommendations from large collaborative networks such as The International Rare Disease Research Consortium[6].

Encompassing over 400 conditions, rare bone diseases account for 5 percent of all birth defects[7,8]. Osteogenesis imperfecta (OI) is a prototypic rare bone disorder characterized by bone fragility, joint laxity, cardiopulmonary manifestations, muscle weakness, hearing loss, dental abnormalities, and scoliosis[9–13]. To address the paucity of validated PROMs for rare skeletal disorders, our group recently investigated the utility of the Pediatric Outcomes Data Collection Instrument (PODCI), a generic functional status PROM used in children with musculoskeletal disorders, in children with osteogenesis imperfecta (OI)[14–16]. Using data from a large, multicenter, observational, natural history study of OI conducted by the National Institutes of Health (NIH) Rare Disease Clinical Research Network's (RDCRN) Brittle Bone Disorders Consortium (BBDC), we analyzed PODCI responses from 418 children with OI. We showed that this generic PROM was a useful measure of physical function in children with OI and that the physical function scores correlated with corresponding scales of an observer-rated outcome measure, the Brief Assessment of Motor Function[14,17]. Using these data, we were also able to estimate the sample sizes that would be required in clinical trials to detect clinically meaningful changes in physical function in children with OI.

There is a significant need to evaluate similar PROMs in adults with OI, especially since many interventional clinical trials in OI are focused solely on adult participants. The study endpoints in such trials typically include changes in areal bone mineral density (aBMD) and bone remodeling markers[18,19]. However, these measures do not assess the effect of therapies on physical and psychological well-being. In this study, we aimed to assess the validity and reliability of an existing generic PROM for adults, the 12-item Short Form Health Survey, version 2 (SF-12v2), in the OI population. The SF-12 was derived from the 36-item Short Form Health Survey (SF-36), which was developed from the Medical Outcomes Study[20]. The SF-36 has been utilized in clinical studies that have enrolled adults with OI; however, most of these studies have been limited by small sample sizes[21–25]. To our knowledge, the SF-12v2 has not been utilized in the OI population, but compared to the SF-36, it is shorter, has less respondent burden, and thus may be better suited for serial administration in clinical trial settings. Utilizing data from the natural history study of the BBDC, we compared scores on the Physical Component Summary (PCS) of the SF-12v2 with integer scores on item 1 of the Gillette Functional Assessment Questionnaire (GFAQ), which was utilized as an observer-reported measure of mobility in the BBDC[26]. We also conducted analyses to assess for correlations between physical and mental HRQoL, as well as sample sizes needed to detect clinically significant differences in PCS scores.

Materials and Methods

Study Population

The data were collected from participants enrolled in the Longitudinal Study of OI, a multicenter, observational, natural history study (NCT02432625) being conducted by the NIH RDCRN's BBDC. The clinical sites of the BBDC include: Baylor College of Medicine (Houston, TX), Kennedy Krieger Institute (Baltimore, MD), Nemours/Alfred I. DuPont Hospital for Children (Wilmington, DE), Oregon Health & Science University and Shriners Hospital for Children (Portland, OR), Shriners Hospital for Children (Chicago, IL), Shriners Hospital for Children (Montreal, QC), University of California Los Angeles (Los Angeles, CA), University of Nebraska Medical Center (Omaha, NE), Hospital For Special Surgery (New York, NY), Shriners Hospital for Children (Tampa, FL), and Children's National Medical Center (Washington, D.C.). Data were collected in a systematic manner across all sites in accordance with the manual of operations (MOO). All data were captured using online case report forms and were housed and managed by the RDCRN's Data Management and Coordinating Center. Individuals with a diagnosis of OI that was made based on clinical and/or radiographic features or by molecular analysis were enrolled in the study. The OI severity was classified using clinical criteria, which were generally based on Sillence clinical criteria: mild, non-deforming (type I); perinatally lethal (type II); progressively deforming (type III); and moderate OI with normal sclerae (type IV)[27]. The study procedures were approved by the Institutional Review Boards of all participating clinical sites and informed consent was obtained from all participants or their legal guardians.

Data Collection

The following data were collected for analyses presented in this manuscript: age, gender, OI type (I, III, IV, V, or Other), SF-12v2 responses, and response to item 1 of the GFAQ.

Analyses were conducted only on data collected at the enrollment visit. Bisphosphonate treatment status was not extracted for the purposes of this study due to the potential for unequal access to IV bisphosphonate therapy, in addition to the potential for confounding, given that those with increased clinical severity may be more likely to receive IV bisphosphonates.

SF-12v2 evaluates health-related quality of life (HRQoL) in adults through a 12-item questionnaire. The items comprise eight scales: Physical Functioning (PF), Role-Physical (RP, the extent to which physical health limits the respondent's role or activities), Bodily Pain (BP), General Health (GH), Vitality (VT, the respondent's energy level), Social Functioning (SF), Mental Health (MH), and Role-Emotional (RE, the extent to which mental health and psychological well-being limit the respondent's role or activities). BP, GH, VT, and SF consist of one item each. The other scales consist of two items each. Item response values are recoded and recalibrated, and a raw score, 0–100 score, and Z-score are computed for each scale. The Z-score from each scale is multiplied by its associated factor score coefficient to compute the Physical Component Summary (PCS), an overall measure of physical HRQoL, and the Mental Component Summary (MCS), an overall measure of mental HRQoL. Finally, the PCS and MCS scores are transformed into norm-based scores with a mean of 50 and standard deviation of 10 in the general US population[28–30].

The GFAQ (Supplemental File 1) is an observer-rated measure of functional mobility consisting of three items. Item 1 assesses mobility on a scale of 1 to 10. Item 2 assesses which factors limit the patient's mobility. Item 3 assesses which actions the subject is able to do beyond basic mobility[26]. The GFAQ was completed by trained research personnel according to the study MOO.

Statistical Analyses

Data extraction was performed by one author, D.C. SF-12v2 scoring and conversion of GFAQ item 1 responses to numeric responses were performed by one author, C.N.M. SF-12v2 responses were scored using the Optum PRO CoRE scoring software according to the user's guide[28–30]. Optum PRO CoRE provides norm-based scores for PCS and MCS, as well as raw scores for the eight scales. PCS, MCS, and scale score responses range from 0–100. PCS and MCS were converted to norm-based scores with a mean of 50 and standard deviation of 10 in the general US population. Higher scores indicate better quality of life, more energy, less pain, fewer physical or emotional limitations, and better general health. Healthy U.S. population norms for PCS and MCS are 50.8 and 50.0, respectively[31].

PCS and MCS mean scores were compared among the OI subtypes using the Kruskal-Wallis test. To confirm the differences in PCS and MCS scores among the OI subtypes, we used generalized linear modelling. In these models, PCS or MCS score was the dependent outcome measure and OI subtype was the independent variable. These models were adjusted by gender. Significance level of pairwise comparisons was adjusted using Tukey's method. To compare the mean PCS and MCS scores in OI with the general population means of 50.8 and 50.0, respectively, we used a one sample t-test. Comparisons were conducted in the overall sample and by OI subtype. Correlations between PCS and GFAQ item 1, and PCS and MCS were performed using the Polychoric and Spearman Correlation Coefficients,

respectively. For correlation coefficients, 0.30 was interpreted as a weak positive relationship; 0.50 as a moderate positive relationship; and 0.70 as a strong positive relationship. For clinical trial readiness analysis, sample sizes needed to detect clinically meaningful differences were calculated for the 2-sample parallel group design using the Wilcoxon-Mann-Whitney test and for the crossover design using the Wilcoxon Signed-Rank test at an alpha error of 0.05 and a power of 0.8. Analyses were conducted using SAS version 9.4 (SAS Institute, INC., Cary, NC).

Results

Study Population

Data extraction was completed on 8/28/2019, by which time 320 adults (age 18 years and older) were enrolled in the BBDC longitudinal study. Overall, 173 individuals with OI type I, 49 with OI type III, 80 with OI type IV, 5 with OI type V, and 13 with other OI types (VI, VII, and unclassified) were enrolled. The cohort had a higher proportion of female participants (67% female vs 33% male). Participant characteristics are detailed in Table 1. Due to low participant numbers in OI type V and "other" OI types, primary analyses were limited to individuals with OI types I, III, and IV. Lumbar spine (LS) areal bone mineral density (aBMD) Z-scores at enrollment were available in 141 adults with OI type I, 22 adults with OI type III, and 48 adults with OI type IV (Table 1). Of the 302 individuals with OI types I, III, and IV, DNA testing results were available for 74. Of these, all but two had pathogenic variants in *COL1A1* or *COL1A2*. *SF*-12v2 data were available on all but two individuals in the entire cohort.

Individuals with OI Report Lower Physical Health-Related Quality of Life but not Mental Health-Related Quality of Life

Mean SF-12v2 0–100 scores for the eight scales and norm-based scores for PCS and MCS for OI types I, III, and IV are shown in Table 2. Among the eight scales, only PF (Physical Function) differed significantly among all three OI types (66.6 OI type I vs 64.3 OI type III vs 64.5 OI type IV, p<0.05). The overall mean PCS score in adults with OI types I, III, and IV was significantly lower than mean PCS score in healthy adults in the U.S. (44.8 vs 50.8; p<0.001)[31]. However, the overall mean MCS score in adults with OI types I, III, and IV was numerically higher than the mean MCS score in healthy adults (51.4 vs 50; p=0.014)[31]. As the severity of OI is expected to influence the physical and mental HRQoL, we compared the mean PCS and MCS scores among OI subtypes and with the general population norms. The mean PCS scores in OI types I, III, and IV were significantly lower than the mean in healthy adults in the U.S. (46.4, 39.9, and 44.3, respectively, vs 50.8; p<0.001). Individuals with OI type III (severe form) had PCS scores that were significantly lower than those with OI type I (mild form) and OI type IV (moderate form) (Figure 1a). The mean MCS scores in OI types I and IV did not differ significantly from the mean in healthy adults in the U.S. (50.1 and 51.9, respectively, vs 50.0), but the mean scores in OI type III, the most severe form of OI, was higher than the mean in healthy individuals (55.2 vs 50; p=0.002) (Figure 1b).

Correlation between PCS Scores and GFAQ

In order to assess the validity of the SF-12v2 PCS score as a measure of physical functioning in adults with OI, we correlated the scores from this measure with the integer response to item 1 of the GFAQ, which ranges from 1 to 10. The mean GFAQ scores for OI types I and IV were 9.3 and 7.8, respectively, compared to 2.3 in OI type III, where we observed a "floor effect." Three-fourths (76%) of participants with OI type III had GFAQ scores of 1 or 2. Correlation coefficients between these two measures were 0.68 in OI type I and 0.50 in OI type IV ($p<0.0001$). However, no statistically significant correlation was detected between PCS score and GFAQ in OI type III.

Correlation between PCS Scores and LS aBMD

We performed spearman correlation analyses to assess whether PCS scores, as a measure of physical functioning, correlated with LS aBMD. We found no meaningful correlation between these parameters in any OI type. The correlation coefficients in OI types I, III, and IV were −0.2325 (p=0.0055), −0.1417 (p=0.5402), and −0.00432 (p=0.9768), respectively.

Correlation between physical HRQoL and mental HRQoL

To determine whether there is a correlation between mental health-related quality of life and physical health-related quality of life in adults with OI, we correlated PCS and MCS scores. Correlation coefficients ranged from −0.11 to 0.17, and none were statistically significant. Overall, there were no clear correlations between PCS and MCS scores in any OI type.

Clinical Trial Readiness

In order to adapt generic PROMs in rare disease clinical trials, the lack of information on mean scores and inter-individual variation must be addressed. Thus, using the means and standard deviations in our study population, we determined the sample sizes required to detect clinically meaningful changes in PCS for OI types I, III, and IV. Table 3 displays necessary sample sizes for a range of mean paired differences in PCS scores for parallel group and crossover designs.

Discussion

Currently, there are no disease-specific PROMs to assess functional health status or health-related quality of life in individuals with osteogenesis imperfecta, and thus generic PROMs including the Short Form Health Survey 36 (SF-36), The World Health Organization Quality of Life measure (WHOQOL-BREF), and the International Physical Activity Questionnaire (IPAQ) have been used in clinical studies[21–25]. However, the validity of these generic PROMs in this rare disease population has not been systematically analyzed. Such limitations make it challenging to utilize generic PROMs in clinical trials involving OI. The SF-12v2 is derived from the SF-36, but has the advantage of decreasing the respondent burden[20]. Therefore, to understand the utility of SF-12v2 in OI, we utilized data from a large cohort of adults to answer the following questions: 1) Is the physical component summary (PCS) score consistent with the known clinical severity of OI types?; 2) Does the PCS correlate well with a validated observer-reported assessment of physical function?; 3) Are there differences in mental HRQoL among individuals with different types of OI?; 4) Is

there any correlation between physical HRQoL and mental HRQoL in OI?; and 5) What are the sample sizes required to detect a range of differences in PCS in individuals with OI?

Our group has previously validated a generic PROM, the PODCI, for use in children with OI[14]. Similar to our findings on the PODCI, this study revealed that PCS scores were representative of the known clinical severity of OI, with the mean score in OI type III (severe) being the lowest (worst) and mean scores in OI type I (mild) being the highest. Mean PCS scores in all adults with OI types I, III, and IV were statistically lower than mean PCS scores in healthy adults; however, these were within one standard deviation of the expected mean of 50. This may reflect adaptation to the physical limitations of OI in adults. Our results are consistent with SF-36 PCS scores reported by Hald and colleagues, who also found that the overall PCS scores were higher in OI type I than in OI type III[22]. The fact that SF-12v2 and SF-36 PCS scores correlate with OI severity in two separate independent populations implies that PCS as measured by these two PROMS could be a surrogate marker of at least some domains of physical function. While some generic musculoskeletal PROMs assess various domains of physical functioning such as upper limb and lower limb functions and fine motor abilities, the SF-12v2 comprises only 12 items that inquire about generic mobility activities such as ability to climb stairs and complete household tasks like vacuuming. Thus, whereas SF-12v2 can assess the mobility domain of physical function, it cannot differentiate between upper extremity and lower extremity function.

When determining whether to utilize a PROM in a clinical trial, it is important to determine whether patient responses are consistent with observer-reported judgements. Thus, we examined correlations between item one of the GFAQ, which assesses mobility, with PCS scores. Modest correlations were observed in OI types I and IV, whereas no correlation was observed in OI type III. This may be due to the mobility focus of these scales. Item one of GFAQ only assesses mobility, and as mobility is much more severely affected in individuals with OI type III; the "floor effect" on the GFAQ scores impacts any such analysis.

The mean MCS scores in individuals with type I and IV OI were similar to the mean scores from the healthy population in the U.S.; however, the mean score in individuals with OI type III, the most severe form, was higher than the general population norms. Hald and colleagues also observed that individuals with OI type III had the highest MCS scores while using SF-36 as a PROM[22]. The MCS scores for both SF-12v2 and SF-36 are calculated by adding a small negative weight to physical scales and a small positive weight to mental scales. Therefore, low physical SF-12v2 scores, as seen on OI type III, result in higher MCS scores than expected[28,29,33]. In the general healthy population, correlations between PCS and MCS scores are not widely known and in fact, PCS scores tend to decrease with increasing age while MCS scores tend to increase[31]. Similar to the general population, in our OI subjects, there were no significant correlations between PCS and MCS scores. Given these limitations, we suggest that in clinical trials, it would be more appropriate to use PCS and not MCS as a secondary endpoint for improvement in patients' lived experience.

Given the rarity of osteogenesis imperfecta, clinical trials in this population will always have a relatively small number of participants. Thus, we sought to determine the feasibility of using the SF-12v2 as a secondary endpoint. The minimum clinically important differences

(MCIDs) for PCS in the OI population are not known, thus MCIDs were extrapolated from literature pertinent to other musculoskeletal conditions including knee arthroplasty, anterior cruciate ligament repair, knee osteoarthritis, limb reconstruction, and early rheumatoid arthritis, in which MCIDs range from 2 to 7[34–36]. For OI, we calculated sample sizes required to detect mean score differences of 2, 5, and 10 on PCS. With a crossover design, even sample sizes of ~60 can detect differences in PCS of 2 or greater; this would translate to a small effect size and thus it would be reasonable to expect such increases with an effective therapeutic agent. However, with a parallel group design, one would require at least modest effect sizes (score of 5 or more) to detect statistically significant differences; such effect sizes may be hard to achieve.

The results from this study should be interpreted in the context of its strengths and limitations. The strengths include the large number of participants, adequate representation of individuals with mild, moderate, and severe forms of OI, standard process of data collection across sites, and availability of an observer-rated scale with which the patient reported data could be compared. One limitation is that the data are cross-sectional and thus we cannot comment about variability with time. The BBDC is collecting longitudinal data and we hope to address this limitation in future studies. A second limitation is that we cannot discern whether treatment with bisphosphonates had an influence on the overall HRQoL. Previous studies have shown that bisphosphonates can have a positive effect on pain and physical function[21,37–39]. The challenges with performing a similar analysis in our cohort are the following: 1) as this is a non-interventional study, individuals would have received bisphosphonates at differing time points and for variable durations, 2) the use or non-use of bisphosphonates may be driven not only by the phenotype but also by unequal access to treatment, and 3) individuals with severe OI are more likely to have received bisphosphonates than individuals with mild OI; this interaction between OI type and treatment makes it difficult to incorporate these two as independent variables in analysis.

In summary, we show that the SF-12v2 can be used to reliably assess the mobility domain of physical function in adults with OI. We demonstrate that PCS can be used as a secondary outcome measure in clinical trials of adults with OI and that in crossover designs, it would be feasible to detect clinically meaningful differences with small sample sizes. However, we also demonstrate that with the limited number of items on the SF-12v2, it is still an incomplete assessment of physical functioning, especially in individuals with OI. We thus recommend that future efforts to develop and validate OI-specific PROMs, both in the pediatric and adult settings, should have an emphasis on questions pertinent to this rare disease population.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. US Food and Drug Administration. Guidance for Industry: Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims, 2009. https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/UCM193282.pdf. Accessed January 30, 2019.

2. Slade A, Isa F, Kyte D, et al. Patient reported outcome measures in rare diseases: a narrative review. Orphanet J Rare Dis. 2018;13(1):61. [PubMed: 29688860]

3. Cohen JS, Biesecker BB. Quality of life in rare genetic conditions: a systematic review of the literature. Am J Med Genet A. 2010;152A(5):1136–1156. [PubMed: 20425818]

4. US Food and Drug Administration. Rare Diseases: Common Issues in Drug Development Guidance for Industry DRAFT GUIDANCE, 2019. https://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/guidances/default.htm. Accessed August 3, 2020.

5. Lanar S, Acquadro C, Seaton J, Savre I, Arnould B. To what degree are orphan drugs patient-centered? A review of the current state of clinical research in rare diseases. Orphanet J Rare Dis. 2020;15(1):134. [PubMed: 32493385]

6. Morel T, Cano SJ. Measuring what matters to rare disease patients – Reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis. 2017;12(1).

7. Bonafe L, Cormier-Daire V, Hall C, et al. Nosology and classification of genetic skeletal disorders: 2015 revision. Am J Med Genet A. 2015;167A(12):2869–2892. [PubMed: 26394607]

8. Tosi LL, Warman ML. Mechanistic and therapeutic insights gained from studying rare skeletal diseases. Bone. 2015;76:67–75. [PubMed: 25819040]

9. Rossi V, Lee B, Marom R. Osteogenesis imperfecta: advancements in genetics and treatment. Curr Opin Pediatr. 2019;31(6):708–715. [PubMed: 31693577]

10. Tam A, Chen S, Schauer E, et al., A multicenter study to evaluate pulmonary function in osteogenesis imperfecta. Clin Genet. 2018;94(6)502–511. [PubMed: 30152014]

11. Basel D, Steiner RD. Osteogenesis imperfecta: recent findings shed new light on this once well-understood condition. Genet Med. 2009;11(6):375–385. [PubMed: 19533842]

12. McKiernan FE. Musculoskeletal manifestations of mild osteogenesis imperfecta in the adult. Osteoporos Int. 2005;16(12):1698–1702. [PubMed: 15902417]

13. Pillion JP, Shapiro J. Audiological findings in osteogenesis imperfecta. J Am Acad Audiol. 2008;19(8):595–601. [PubMed: 19323351]

14. Murali CN, Cuthbertson D, Slater B, et al. Pediatric Outcomes Data Collection Instrument is a Useful Patient-Reported Outcome Measure for Physical Function in Children with Osteogenesis Imperfecta. Genet Med. 2020;22(3).

15. American Academy of Orthopaedic Surgeons. The American Academy of Orthopaedic Surgeons Outcomes Instruments: Normative Values from the General Population, Norms Base Scoring and Original Standard Raw Scores.; 2000.

16. Daltroy LH, Liang MH, Fossel AH, Goldberg MJ. The POSNA pediatric musculoskeletal functional health questionnaire: report on reliability, validity, and sensitivity to change. Pediatric Outcomes Instrument Development Group. Pediatric Orthopaedic Society of North America. J Pediatr Orthop. 1998;18(5):561–571. [PubMed: 9746401]

17. Cintas HL, Siegel KL, Furst GP, Gerber LH. Brief assessment of motor function: reliability and concurrent validity of the Gross Motor Scale. Am J Phys Med Rehabil. 2003;82(1):33–41. [PubMed: 12510183]

18. Glorieux FH, Devogelaer JP, Durigova M, et al. BPS804 Anti-Sclerostin antibody in adults with moderate osteogenesis imperfecta: results of a randomized phase 2a trial. J Bone Miner Res. 2017;32(7):1496–1504. [PubMed: 28370407]

19. Orwoll ES, Shapiro J, Veith S, et al. Evaluation of teriparatide treatment in adults with osteogenesis imperfecta. J Clin Invest. 2014;124(2):491–498 [PubMed: 24463451]

20. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220:233. [PubMed: 8628042]

21. Feehan AG, Zacharin MR, Lim AS, Simm PJ. A comparative study of quality of life, functional and bone outcomes in osteogenesis imperfecta with bisphosphonate therapy initiated in childhood or adulthood. Bone. 2018;113:137–143. [PubMed: 29787832]

22. Hald JD, Folkestad L, Harslof T, Brixen K, Langdahl B. Health-related quality of life in adults with osteogenesis imperfecta. Calcif Tissue Int. 2017;101:473–478. [PubMed: 28676897]

23. Widmann RF, Bitan FD, Laplaza FJ, Burke SW, DiMaio MF, Schneider R. Spinal deformity, pulmonary compromise, and quality of life in osteogenesis imperfecta. Spine (Phila Pa 1976). 1999;24(16):1673–1678. [PubMed: 10472101]

24. Widmann RF, Laplaza FJ, Bitan FD, Brooks CE, Root L. Quality of life in osteogenesis imperfecta. Int Orthop. 2002;26(1):3–6. [PubMed: 11954845]

25. Balkefors V, Mattsson E, Pernow Y, Saaf M. Functioning and quality of life in adults with mild-to-moderate osteogenesis imperfecta. Physiother Res Int. 2013;18(4):203–211. [PubMed: 23281201]

26. Novacheck TF, Stout JL, Tervo R. Reliability and validity of the Gillette Functional Assessment Questionnaire as an outcome measure in children with walking disabilities. J Pediatr Orthop. 2000; 20: 75–81 [PubMed: 10641694]

27. Patel RM, Nagamani SCS, Cuthbertson D, et al. A cross-sectional multicenter study of osteogenesis imperfecta in North America – results from the linked clinical research centers. Clin Genet. 2015;87(2):133–140. [PubMed: 24754836]

28. Optum©. (2019). PRO CoRE version 1.4 Smart Measurement© System: Users' Guide. Johnston, RI: Author.

29. Ware J, Kosinski M, Keller SD. SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. 1995 (12).

30. Fleishman JA, Selim AJ, Kazis LE. Deriving SF-12v2 physical and mental health summary scores: a comparison of different scoring algorithms. Qual Life Res. 2010;19:231–241. [PubMed: 20094805]

31. Gandek B, Ware JE, Aaronson NK, et al. Cross-validation of item selection and scoring for the SF-12 health survey in nine countries: results from the IQOLA Project. International Quality of Life Assessment. J Clin Epidemiol. 1998;51(11):1171–1178 [PubMed: 9817135]

32. Ablon J Personality and stereotype in osteogenesis imperfecta: Behavioral phenotype or response to life's hard challenges? Am J Med Genet. 2003;122A:201–214. [PubMed: 12966520]

33. Ware J, Kosinski MA, Keller SD. SF-36 Physical and Mental Health Summary Scales: A User's Manual. 1994 (12).

34. Clement ND, Weir D, Holland J, Gerrand C, Deehan DJ. Meaningful changes in the Short Form 12 physical and mental summary scores after total knee arthroplasty. Knee. 2019;26(4):861–868. [PubMed: 31109846]

35. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. J Clin Epidemiol. 2017;89:188–198. [PubMed: 28676426]

36. Nwachukwu BU, Chang B, Voleti PB, et al. Preoperative Short Form Health Survey score is predictive of return to play and minimal clinically important difference at a minimum 2-year follow-up after anterior cruciate ligament reconstruction. Am J Sports Med. 2017;45(12):2784–2790. [PubMed: 28727937]

37. Seikaly MG, Kopanati S, Salhab N, et al. Impact of alendronate on quality of life in children with osteogenesis imperfecta. J Pediatr Orthop. 2005;25(6):786–791. [PubMed: 16294137]
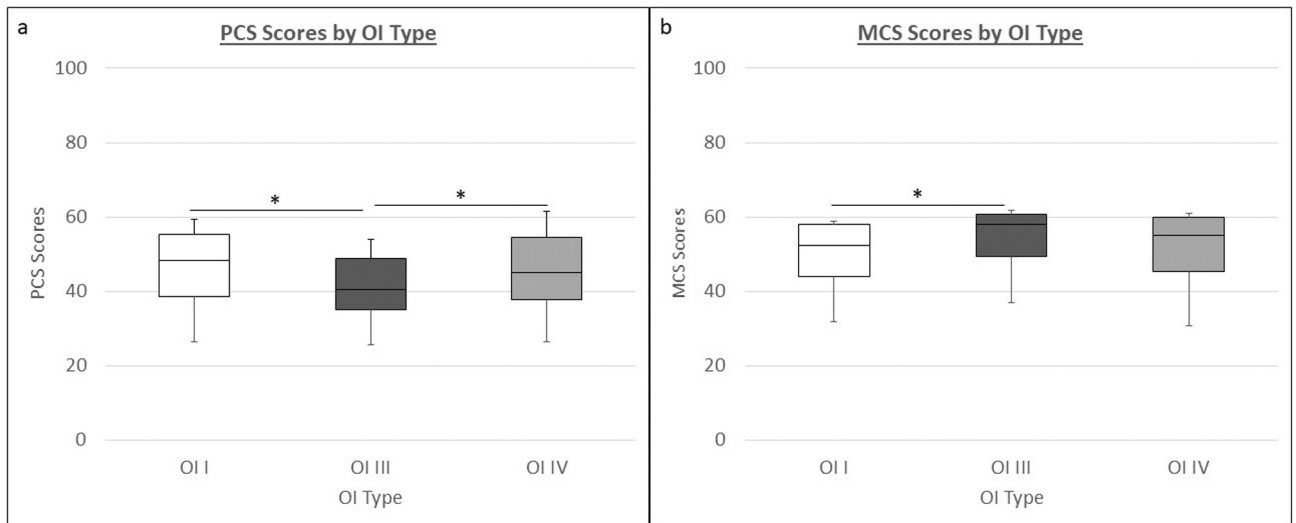
38. Tsimicalis A, Boitor M, Ferland CE, et al. Pain and quality of life of children and adolescents with osteogenesis imperfecta over a bisphosphonate treatment cycle. Eur J Pediatr. 2018;177(6):891–902. [PubMed: 29637375]

39. Rodriguez Celin M, Simon JC, Krzack JJ, et al. Do bisphosphonates alleviate pain in children? A systematic review. Current Osteoporosis Reports. 2020(3).

Figure 1: PCS (a) and MCS (b) in adults with OI.
The box plots depict the medians and interquartile ranges while the whiskers depict the 5th and 95th centiles. (*$p<0.05$, ANOVA)

Table 1:

Characteristics of adults enrolled in the BBDC Longitudinal Study.

	OI I	OI III	OI IV	OI V	Other/ Unclassified	Total
N	173	49	80	5	13	320
Male n (%)	45 (26)	17 (35)	33 (41)	3 (60)	7 (54)	105 (33)
Female (%)	128 (74)	32 (65)	47 (59)	2 40)	6 (46)	215 (67)
Mean age in yrs at enrollment	39.8	30.7	36	42.2	32.5	37.2
Median age (IQR) in yrs at enrollment	37.8 (28.8–46.9)	27.3 (21.3–32.5)	33.4 (23.1–46.6)	46.8 (31.3–50.4)	29.1 (23.8–33.6)	33.7 (25.3–46.6)
Median L-spine aBMD Z-score (IQR) at enrollment [*]	−1.8 (−2.6 – −0.8)	−3.2 (−4.3 – −2.4)	−2.6 (−3.0 – −1.9)			

[*] L-spine aBMD Z-scores available in 141, 22, and 48 individuals with OI types I, III, and IV, respectively.

Table 2:

Mean scores and standard deviations of all SF-12v2 scales, PCS, and MCS by OI type. Mean norm PCS and MCS scores shown for comparison are from Gandek et al[31]

	OI I	OI III	OI IV	Norm
N	173	49	80	
Physical Domains Mean (SD)				
PF	66.6 (34.9)	20.8 (29.3)	54.6 (34.5)	
RP	65.0 (30.0)	64.3 (30.0)	64.5 (27.8)	
BP	67.5 (28.7)	66.8 (30.4)	65.9 (29.8)	
GH	65.8 (24.1)	68.3 (22.6)	66.3 (24.1)	
PCS	46.4 (10.6)	39.9 (7.06)	44.3 (9.49)	50.8 (8.9)
Mental Domains Mean (SD)				
VT	53.9 (22.8)	56.3 (23.3)	55.1 (28.9)	
SF	77.8 (25.6)	76.6 (28.4)	76.6 (25.4)	
RE	78.3 (24.8)	83.9 (23.8)	80.0 (24.6)	
MH	67.7 (20.2)	69.5 (19.3)	70.1 (20.7)	
MCS	50.1 (9.66)	55.2 (8.97)	51.9 (10.2)	50.0 (9.5)

PF Physical Functioning, RP Role-Physical, BP Bodily Pain, GH General Health, PCS Physical Component Summary, VT Vitality, SF Social Functioning, RE Role-Emotional, MH Mental Health, MCS Mental Component Summary.

Table 3.

Sample sizes needed to detect the listed mean paired differences using two-sided Wilcoxon-Mann-Whitney test (two sample parallel group design) and paired Wilcoxon Signed-Rank test (crossover design) at a 2-sided significance level (α) of 0.05 and at 80% power.

	SF12v2 assessment		Total sample size required for a two-sample parallel group design to detect a mean difference of			Total sample size required for a one-sample crossover design to detect a mean difference of		
	Mean score	SD	2	5	10	2	5	10
Type I collagen-related OI (Types I, III, and IV)								
PCS	**44.8**	**10.1**	824	136	36	55	11	6
OI type I								
PCS	**46.4**	**10.6**	926	152	40	55	11	6
OI Type III								
PCS	**39.9**	**7.1**	418	70	20	36	9	5
OI type IV								
PCS	**44.3**	**9.5**	744	122	34	44	10	5