

**Table 1** Estimated second-order factor loadings and correlations with SIS

MQOL-E domains	Factor loading	Correlation with SIS
Physical	0.70	0.68
Psychological	0.68	0.45
Existential	0.83	0.52
Relationships	0.58	0.30
Health care	0.45	0.32
Cognition	0.66	0.31
Feeling like a burden	0.40	0.31
Physical surroundings	0.50	0.31

Item Scale (MQOL-SIS) measuring global QOL. A revised version presented at the ISOQOL conference in 2014, MQOL-R, covers the same domains with some improvements. We recently created MQOL-Expanded (MQOL-E) to add coverage of contributors to QOL found to be important in studies since MQOL's creation. Six items covering new domains (cognition; physical surroundings; feeling like a burden; quality of health care) were added to MQOL-R for a total of 8 hypothesized domains and 21 items. The aim for this report was to examine the measurement structure of the MQOL-E and specifically determine the: (1) presence of a second-order latent factor representing overall QOL, (2) relative importance of the MQOL-E domains in relation to the second-order factor, and (3) correlations of the MQOL-E domains and second-order factor with MQOL-SIS. **METHODS:** Data came from four Canadian samples of people with life-threatening illnesses from various care settings ( $N = 869$ ). Multiple imputation was used for missing data. Confirmatory factor analysis was applied to evaluate the first- and second-order factor models of the MQOL-E; standardized loadings for the second-order factor structure were estimated to evaluate relative importance. Correlations were calculated between the SIS and (a) the 8 MQOL-E latent factor domains, and (b) the second-order factor (overall QOL). **RESULTS:** There was good fit for the first-order (CFI = .963; RMSEA = .033) and second-order (CFI = .932; RMSEA = .042) factor models (Table 1). Standardized second-order factor loadings ranged from .83 (existential) to .40 (feeling like a burden). Correlations between the MQOL-SIS and the domains ranged from .68 (physical) to .30 (relationships) (Table 1), and .66 with the second order factor. **CONCLUSIONS:** The results support the use of an MQOL-E Total score (the mean of the 8 submeasure scores). Existential wellbeing, along with the physical and psychological domains, is an important contributor to QOL in people with life-threatening illness. MQOL-E is conceptually related to but not identical to global QOL.

### (2036) Comparison of computerized adaptive testing and fixed-length short forms of the Prosthetic Limb Users Survey of Mobility (PLUS-M™)

Dagmar Amtmann, PhD, University of Washington, Seattle, WA, United States; Jiseon Kim, University of Washington, Seattle, WA, United States; Hyewon Chung, Chungnam National University, Daejeon, Korea, South; Ryoungsun Park, University of Texas, Austin, TX, United States; Rana Salem, University of Washington, Seattle, WA, United States; Brian J. Hafner, University of Washington, Seattle, WA, United States

**AIMS:** The Prosthetic Limb Users Survey of Mobility (PLUS-M™) is an item bank developed to measure mobility of people

with lower limb amputation who use prosthetic limb(s) to ambulate. The purpose of this study was to compare the performance and efficiency of computerized adaptive testing (CAT) to two fixed-length short forms (SFs) (7- and 12-item). **METHODS:** PLUS-M item bank includes 44 items calibrated to item response theory (IRT) and the scores are on a T-metric ( $M = 50$ ,  $SD = 10$ ). The CAT and SFs were administered to adults with lower limb amputation. The minimum number of CAT items administered to each participant was 4 and maximum was 12. The CAT stopped when the standard error (SE) dropped below 3 or 12 items had been administered. We examined the number of items administered, standard errors (SEs) of the final scores, and the correlations between the CAT- and the SF-based scores. **RESULTS:** 199 lower limb prosthesis users responded to the CAT and SFs. On average, the CAT administered 5 items (range 4–12) with 59.8 % responding to 4 items, 20.6 % to 5 items, and 8.5 % to 6 items. 7 participants who chose the highest or lowest response to most items responded to the maximum number of items. The CAT was more efficient for scores between 40 and 60, with an average of 4 items ( $n = 155$ ,  $SD = 0.7$ ; range = 4–7) administered. Average PLUS-M scores were: CAT 50.8 ( $SD = 8.6$ ); 7-item SF 50.5 ( $SD = 7.7$ ); and 12-item SF 50.3 ( $SD = 7.9$ ). Average SEs associated with these scores were 2.7 ( $SD = 0.3$ ) for CAT; 2.9 ( $SD = 0.5$ ) for the 7-item SF; and 2.2 ( $SD = 0.5$ ) for the 12-item SF. For participants with  $SE > 3.0$ , the averages scores were 64.8 ( $n = 5$ ; an average of 12 items administered) for CAT; 59.0 ( $n = 60$ ) for the 7-item SF; and 62.6 ( $n = 16$ ) for the 12-item SF. Pearson correlations between the CAT score and the 7- and 12-item SFs scores were both 0.9. **CONCLUSIONS:** PLUS-M CAT required fewer items and estimated scores with greater precision than either fixed-length SF. The CAT could be made more efficient by implementing different stopping rules for those who select the highest or the lowest response category.

### (2038) Refining items in the Long-Term Conditions Questionnaire (LTCQ): cognitive interviews, stakeholder feedback and translatability assessment

Caroline Potter, University of Oxford, Oxford, United Kingdom; Cheryl Hunter, University of Leeds, Leeds, United Kingdom; Laura Kelly, University of Oxford, Oxford, United Kingdom; Elizabeth Gibbons, University of Oxford, Oxford, United Kingdom; Crispin Jenkinson, University of Oxford, Oxford, United Kingdom; Ray Fitzpatrick, University of Oxford, Oxford, United Kingdom; Michele Peters, PhD, University of Oxford, Oxford, United Kingdom

**AIMS:** This research sought to clarify and reduce items within the 23-item draft Long-Term Conditions Questionnaire (LTCQ), a patient-reported outcome measure (PRO) designed to be applicable in mental and physical health and multi-morbidity. **METHODS:** The 23 items were generated through literature reviews, professional stakeholder interviews, and interviews with 42 patients who had at least one long-term condition (LTC). Thirteen of these patients (age range 36–88 years) participated in a cognitive interview. The interviews were audio-recorded and analysed for comprehensiveness, acceptability, interpretability, and feasibility of individual items and the LTCQ as a whole. The draft questionnaire was also sent to the professional stakeholders who had informed initial item generation; 13 stakeholders from across health and social care provided comments. A professional translatability assessment was undertaken to assess the LTCQ's potential for use in diverse languages: Arabic, French, Polish, Punjabi, simplified Chinese, and Urdu in addition to the original English. **RESULTS:** Participants took 6 min on average to complete the questionnaire and expressed general support for the paper-based, self-administered format. All participants found the item content

relevant, but suggestions were made for deleting repetitive or ambiguous items and for simplifying the item content. Professional stakeholders also found the item content highly relevant and made suggestions for additional questions or item revisions. The translatability assessment raised no fundamental concerns but suggested minor revision to five items for greater clarity. Based on these three strands of work, six of the initial 23 items were deleted, six items were substantially revised to clarify meaning (including one original item being split into two new items), five items were refined through minor wording changes, and six items were retained unchanged. Items were also re-ordered as patients found it difficult to switch repeatedly between positively and negatively phrased items. This work resulted in a revised LTCQ of 18 items. **CONCLUSIONS:** The LTCQ is acceptable to people living with LTCs, and its potential value to professional stakeholders across health and social care has been reconfirmed. Next steps will include further cognitive interviews with patients ahead of testing via a large-scale survey.

**(2040) Age-specific factor structures can improve the sensitivity of patient reported outcome measures: evidence from the Grow up Gothenburg study**

Sarah Hitz, MSc, Hamburg University of Applied Sciences, Hamburg, Germany; Ebba Brann, PhD student, Inst. of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; Kerstin Albertsson-Wikland, MD, PhD, Professor, Inst. of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; Zita Schillmöller, PhD, Hamburg University of Applied Sciences, Hamburg, Germany, John Chaplin, PhD, AFBPsS, CPsychol., Inst. Clinical Sciences, Sahlgrenska Academy at the University Gothenburg, Sweden, Gothenburg, Sweden

**AIMS:** To confirm and test the factor structure of the Gothenburg well-being scale for children (GWBC) in a school based pre-adult population (18–19 year olds). The GWBC was developed in 1994 as a measure of quality of life to be used in clinical studies of short stature. The 34 items were derived from a pool of 49 items following an exploratory factor analysis. It is a patient-reported questionnaire consisting of VAS (scales with bipolar adjectives as endpoints derived from a general school population (9–13 years). The GWBA has six domains: mood, self-esteem, physical condition, psychosocial functioning, stability and coping. It has been used consistently in studies of children's quality of life in short stature. **METHODS:** 5002 students (51 % boys) in the final grade (school year 12) of high school in Gothenburg completed the questionnaire as part of a larger survey. A confirmatory factor analysis (CFA) (Amos 20.0) was carried out. **RESULTS:** Assumptions of normality were confirmed on all items. The CFA showed fit discrepancies (Chi square test  $p < 0.001$ ) with the original. The GFI and AGFI did not pass the threshold for acceptability (GFI = 0.80; AGFI = 0.77). The standardised root mean square residual (SRMR) was too high (SRMR = .075). RMSEA had an almost acceptable fit value (RMSEA = 0.08). An alternative model for adolescents (GWBA) was developed excluding 7 items and including 8 items previously excluded from the original pool of 49 adjectives. Internal consistency was acceptable (Cronbach's alpha  $> 0.75$ ). The domains constituted unitary concepts. Scale inter-correlations within the GWBC scale ranged from 0.17 to 0.53 indicating that the new model (GWBA) had generally higher correlated domains (0.277–0.861). **CONCLUSIONS:** The factor structure developed on a pre-teen population was not relevant to an older pre-adult population. This indicates the necessity of age specific instruments. The Gothenburg Well-being scale for Adolescence (GWBA) can be used in pre-adult populations to measure quality of life and well-being. It has applications in several areas of research including gender analysis and disability assessment.

**(2042) Development of thresholds for clinical importance for four key domains of the EORTC QLQ-C30**

Johannes M. Giesinger, Department of Psychiatry and Psychotherapy, Medical University of Innsbruck, Innsbruck, Austria; Wilma Kuijpers, The Netherlands Cancer Institute, Amsterdam, Netherlands; Teresa Young, Lynda Jackson Macmillan Centre, Mount Vernon Cancer Centre, Northwood, Middlesex, United Kingdom; Krzysztof Tomaszewski, Department of Anatomy, Jagiellonian University Medical College, Krakow, Poland; Elizabeth Friend, Basingstoke & North Hampshire Hospital, Basingstoke, United Kingdom; August Zabernigg, Kufstein County Hospital, Kufstein, Austria; Bernhard Holzner, Medical University of Innsbruck, Innsbruck, Austria; Neil K. Aaronson, The Netherlands Cancer Institute, Amsterdam, Netherlands

**AIMS:** The EORTC QLQ-C30 questionnaire is increasingly being used to monitor cancer patients' quality of life in daily clinical practice. In such a setting the interpretation of questionnaire scores can be facilitated by using thresholds for clinical importance. Our study aimed at identifying thresholds for clinical importance for four EORTC QLQ-C30 scales that are often affected by cancer and its treatment: Physical Functioning (PF), Emotional Functioning (EF), Pain (PA) and Fatigue (FA). **METHODS:** For our study we recruited cancer patients at the Netherlands Cancer Institute (the Netherlands), Kufstein County Hospital (Austria), Mount Vernon Cancer Centre (UK), Basingstoke & North Hampshire Hospital (UK) and at Jagiellonian University Medical College (Poland). No restrictions were placed on diagnosis or type of treatment. Patients completed the QLQ-C30 and, for each of the four QLQ-C30 domains (PF, EF, PA and FA), three anchor items assessing burden, limitations in daily activities and need for help. We merged the three anchor items into a dichotomous external criterion to estimate thresholds for clinical importance. Statistical analysis relied on Receiver Operator Characteristic (ROC) curves. **RESULTS:** We recruited 548 patients with a mean age of 60.6 years. 54 % were female and the most frequent diagnoses were breast (26 %), colorectal (13 %) and lung cancer (12 %). The QLQ-C30 scales showed high diagnostic accuracy with regard to predicting burden, limitations and need for help related to PF, EF, PA and FA (all areas under the curve were above 0.86). We were able to estimate thresholds for clinical importance for these four QLQ-C30 scales. **CONCLUSIONS:** Thresholds for clinical importance will facilitate interpretation of scores at both an individual and group level. When used in daily clinical practice, these thresholds can help to identify patients with clinically relevant problems requiring further exploration and possibly intervention by health care professionals. The thresholds also allow converting QLQ-C30 scores into prevalence rates.

**(2044) Improving patient centered care: implementation of PROMIS in a surgical spine practice**

Richard L. Skolasky, ScD, Johns Hopkins University, Baltimore, MD, United States; David Li, Johns Hopkins University, Baltimore, MD, United States; Brian J. Neuman, MD, Johns Hopkins University, Baltimore, MD, United States

**AIMS:** Patients provide important information as experts in disease experience. Improved understanding of symptoms will result from combined patient-provider perspectives. Existing patient reported outcomes (PROs) were developed with variable levels of patient input. Significant changes in diagnosis and treatment of spine disease and patient expectations have occurred. It is critical to accurately assess patient experiences to integrate their assessment as measures of treatment outcome. The goals are: prioritize important aspects of disease experiences and identify corresponding PROMIS measures;