the pretest OTC Medicine Safety Assessment. A principal component factor analysis revealed that the Rasch dimension explained 25.9 % of the variance which supports the multidimensionality of the measure. Items were well targeted to the population and able to separate students along the performance continuum (reliability = 0.81, separation = 2.04) with higher scores indicative of higher knowledge at pretest. There were no items above the infit criteria of 1.5 indicating that students responded consistently to items of similar difficulty. CONCLUSIONS: The OTC Medicine Safety Assessment demonstrates strong psychometric characteristics while successfully discriminating between levels of student knowledge and understanding. These characteristics make it an appropriate tool for assessing knowledge and behavior with respect to OTC medicines.

(203.5) Burnout among surgeons—the problem and evolving solutions

Juliane Bingener, Mayo Clinic, Rochester, MN, United States; Tait Shanafelt, Mayo Clinic; Susan Hallbeck, Mayo Clinic; Jeff Sloan, PhD, Mayo Clinic, Rochester, MN, United States

AIMS: Burnout is common among American surgeons with national incidence rates as high as 40 % and even higher rates reported by single institutions or specialties. Burnout has a number of personal consequences including depression, anxiety, substance abuse, and an adverse impact on personal relationships. Burnout also has important professional consequences including cynicism in dealing with patients and colleagues, loss of meaning in work, and increased risk of major medical errors. This presentation reports on evolving assessment and intervention efforts. METHODS: Burnout assessments were included in routine staff satisfaction surveys. A brief qualitative survey was developed to assess burnout causes and to facilitate intervention creation. Research projects were initiated to deal with burnout related to surgical practices. One project involved using the Surgical Task Load Index (Surg-TLX) was used to evaluate 6 sub-scales evaluating mental, physical and temporal demands, situational awareness, task complexity and operating room distractions. The Surg-TLX was used to evaluate the impact of short pauses and stretches during long surgical procedures. Another project assessed alleviation of physical pain of surgeons. RESULTS: The overall burnout rate among surgeons was 42 %, with variation across units. Factors contributing to burnout included difficulties with work/life balance, excess administrative work, and overscheduling. Over 75 % of 33 surgeons participating in 90-s stretching breaks during 118 cases wanted to continue incorporating this approach into their operations. 50 % reported stretching breaks improved their physical performance and 30 % reported improved mental focus during surgery. 57 % of surgeons reported physical pain at the end of the operating day (back (87 %), neck (60 %), and feet (63 %)). 55 % reported pain affected their work and 57 % stated it affected their life outside of work. Over 90 % of surgeons invited to try a fatigue reducing mat tried and endorsed use of the mat. CONCLUSIONS: Evidence-based interventions are needed to reduce burnout, improve professional satisfaction, and reduce occupation-related pain among surgeons. The preliminary results reported here identify several practical strategies to assist surgeons. Additional studies are needed to identify individual and institutional approaches to reduce burnout and enhance surgeon well-being and function.

204: PRO Development: From Conception through Revision

(204.1) Development and validation of the Myeloma Patient Outcome Scale (MyPOS): A quality of life questionnaire for use in the clinical care of people with multiple myeloma

Thomas Osborne, MA MBBS MRCP, King's College London, London, United Kingdom; Christina Ramsenthaler, Cicely Saunders Institute, King's College London, London, United Kingdom; Susanne De Wolf-Linder, Cicely Saunders Institute, King's College London, London, United Kingdom; Caty Pannell, Cicely Saunders Institute, King's College London, London, United Kingdom; Richard J. Siegert, Auckland University of Technology, Auckland, New Zealand; Polly Edmonds, Department of Palliative Care, King's College Hospital; Steve Schey, Department of Haematological Medicine, King's College Hospital; Irene J. Higginson, Cicely Saunders Institute, King's College London, London, United Kingdom

AIMS: A systematic literature review showed that existing quality of life (QOL) questionnaires developed or validated for use in multiple myeloma do not capture all the issues important to patients, and may not be well suited to clinical use. The aim was therefore to develop and validate a disease-specific questionnaire for use in the clinical setting to aid the assessment of QOL in people with multiple myeloma. METHODS: Phase 1: Qualitative interviews with myeloma patients ($n = 40$), two patient focus groups ($n = 7$, $n = 4$) and one healthcare professional focus group ($n = 6$) to identify the issues important to QOL, develop a theoretical model and identify problems with existing questionnaires. Phase 2: Development of the MyPOS questionnaire and pre-testing using cognitive interviews ($n = 12$). Phase 3: National, multi-centre, cross-sectional survey in a clinically representative sample of myeloma patients recruited from 14 hospital trusts across England ($n = 380$). Psychometric evaluation including assessment of acceptability (time to complete), internal consistency (Cronbach's alpha), structural validity (exploratory and confirmatory factor analysis and item response theory modeling), construct validity (known-group comparisons) and criterion validity (comparison to the EORTC QLQ-C30 and MY20 and the EuroQOL EQ-5D-3L). RESULTS: A conceptual model of QOL in myeloma is presented to form the basis for item development. Patients highlighted the importance of questions about sexual function, health service factors and the impact of symptoms on function. The MyPOS was developed with a combination of structured (27) and open (6) questions. Items were further refined and acceptability confirmed using cognitive interviews. Psychometric evaluation showed good internal consistency ($a = 0.89$). Factor analysis confirmed three subscales for Symptoms & Function; Emotional Wellbeing and Support factors. The MyPOS and its subscales showed a strong ability to distinguish between clinically different groups (performance status ($F = 26.33$, $p < 0.001$), disease stages ($F = 11.89$, $p < 0.001$), on/off treatment ($t = 3.415$, $p < 0.001$), and good convergent and discriminant validity to hypothesized subscales of EORTC and EQ-5D. CONCLUSIONS: The MyPOS is a reliable and valid instrument for use in myeloma patients at all stages of disease. Further longitudinal work is needed to assess responsiveness to change and test–retest reliability.

(204.2) The Plus-M: Item Bank Of Mobility For Prosthetic Limb Users

Dagmar Amtmann, PhD, University of Washington, Seattle, WA, United States; Daniel C. Abrahamson, University of Washington, Seattle, WA, United States; Sara Morgan, University of Washington, Seattle, WA, United States; Rana Salem, University of Washington, Seattle, WA, United States; Robert L. Askew, University of Washington, Seattle, WA, United States; Robert Gailey, University of Miami, Miami, FL, United States; Ignacio Gaunaurd, VA Miami, Miami, FL, United States; Andre Kajlich, University of Washington, Seattle, WA, United States; Brian Hafner, University of Washington, Seattle, WA, United States

AIMS: Measurement of mobility in people who use lower limb prostheses for ambulation is challenging because general measures of mobility often are not appropriate for prosthetic limb users. The aim of this study was to develop the Prosthetic Limb Users Survey of Mobility (PLUS-M)- a standardized outcome measure suitable for measuring mobility in prosthetic limb users. PLUS-M is intended to be used for research and clinical purposes, including comparative effectiveness studies, documentation of health outcomes, and facilitation of evidence-based treatment decisions. METHODS: Modern psychometric methodology was used to develop PLUS-M. A candidate item bank was developed with input from literature review, clinicians, researchers, end-users, and other stakeholders. Five possible responses range from "without any difficulty" to "unable to do." Items were tested in cognitive interviews with prosthetic limb users and administered to a large sample of people with unilateral amputation. IRTPRO was used to fit a 2 parameter Item Response Theory (IRT) model. RESULTS: One hundred five candidate items were administered to 1091 prosthetic limb users with lower limb amputation due to trauma or dysvascular reasons. Confirmatory factor analysis results supported unidimensionality. Items with poor discrimination or less-than-optimal fit were removed from the bank, resulting in a final item bank of 44 calibrated items. PLUS-M score is a T-score ($M = 50$, $SD = 10$), centered on the development sample mean. The PLUS-M score is reliable at 0.9 or higher from 3SD below and 2SD above the mean. Both 7- and 12-item short forms (SFs) are available and their scores are highly correlated ($r > 0.96$) with the full bank score. The PLUS-M computerized adaptive test (CAT) administered on average 5 items to reach standard error of 3.0. The correlations of PLUS-M scores with scores from similar and dissimilar domains were in the expected directions and magnitudes. CONCLUSIONS: PLUS-M is a clinically meaningful and psychometrically sound measure of mobility for prosthetic limb users. IRT calibrations allow for flexible modes of administration by paper or CAT on phones, tablets, or computers. Full item bank calibrations and short forms are freely available and ready for use in clinical care and research (www.plus-m.org).

(204.3) Development of the EORTC emotional functioning item bank for computer adaptive testing (CAT)

Morten Aa Petersen, MSCi, Bispedjerg Hospital, Copenhagen, Denmark; Eva-Maria Gamper, Innsbruck Medical University, Innsbruck, Austria; Anna Costantini, Sant'Andrea Hospital. Sapienza University of Rome. Italy; Johannes M. Giesinger, PhD, Innsbruck Medical University, Innsbruck, Tyrol, Austria; Bernhard Holzner, PhD BE, Univ. Hospital, Innsbruck, Innsbruck, Tyrol, Austria; Colin Johnson, Mchir, Dept of Surgery, Southampton, Hants, United Kingdom; Monika J. Sztankay, Innsbruck Medical University, Innsbruck, Austria; Teresa E. Young, BSc, Mount Vernon Hospital, Northwood, United Kingdom; Mogens Groenvold, MD, PhD, University of Copenhagen, Copenhagen, Denmark

AIMS: Emotional functioning (EF) is a key aspect of health-related quality of life (HRQOL). Hence, precise measurement of EF is essential for a comprehensive description of a patient's HRQOL. Using computerized adaptive testing (CAT) based on item response theory (IRT) an instrument can be adapted to the individual while maintaining comparability of scores across patients. This allows for more precise and appropriate measurement. To utilize these advantages, the EORTC Quality of Life Group is developing an EF item bank for CAT measurement which will be within the HRQOL framework of the EORTC QLQ-C30 questionnaire. METHODS: Based on literature search and evaluations by international samples of experts and cancer patients 38 candidate items were developed. The psychometric properties of the items were evaluated in a large international sample of cancer patients. This included evaluations of dimensionality using factor analysis methods for categorical data, calibration of IRT model and evaluation of item fit, investigation of differential item functioning (DIF) using logistic regression methods and evaluation of the practical impact of DIF findings, and evaluation of the measurement precision/statistical power of the CAT measure particularly in comparison with the QLQ-C30 EF scale. RESULTS: Responses were obtained from 1,023 cancer patients from four countries (Austria, Denmark, Italy, and the UK). Factor analysis showed that 24 items could be included in a unidimensional model with acceptable fit (RMSEA = 0.089, CFI = 0.91 and TLI = 0.99). Tests of item fit indicated good fit to a generalized partial credit model for all 24 items. Twelve of 216 tests for DIF were significant. However, none of the significant findings seemed to have more than negligible impact on the estimation of EF. Evaluations indicated that the CAT measure may reduce sample size requirements by up to 25 % compared to the QLQ-C30 EF scale. CONCLUSIONS: Based on thorough psychometric evaluations we have established an EF item bank of 24 items. This will allow for more precise and flexible measurement of EF, while maintaining backward compatibility with the QLQ-C30 EF scale.

(204.4) The thyroid-related quality of life measure ThyPRO has good responsiveness and ability to detect relevant treatment effects

Torquil Watt, MD PhD, Copenhagen University Hospital, Copenhagen, Denmark; Per Cramon, MD, Copenhagen University Hospital, Copenhagen, Denmark; Laszlo Hegedüs, Odense University Hospital, Odense, Denmark; Jakob B. Bjorner, MD PhD, QualityMetric; Steen J. Bonnema, Odense University Hospital, Odense, Denmark; Åse K. Rasmussen, Copenhagen University Hospital Rigshospitalet, Copenhagen, Denmark; Ulla Feldt-Rasmussen, Copenhagen University Hospital, Copenhagen, Denmark; Mogens Groenvold, MD, PhD, University of Copenhagen, Copenhagen, Denmark

AIMS: The purpose of this study was to evaluate responsiveness of the thyroid-related quality of life (QOL) instrument ThyPRO in patients undergoing relevant clinical treatments for benign thyroid diseases, and to compare it with responsiveness of the generic SF-36 Health Survey V2 METHODS: A sample of 435 patients undergoing treatment completed ThyPRO and SF-36 Health Survey at baseline and 6 months after treatment initiation. Responsiveness was evaluated in three thyroid patient-groups: Patients with hyper- ($n = 66$) and hypothyroidism ($n = 84$) rendered euthyroid, and patients with a clinically detectable non-toxic goiter treated with surgery or radioactive iodine and remaining euthyroid ($n = 62$). Changes in QOL were evaluated in terms of effect size and compared to the changes predicted by clinical experts. The responsiveness of equivalent scales from ThyPRO and SF-36 Health Survey V2 were compared with the relative validity index. RESULTS: The ThyPRO demonstrated good responsiveness across the whole range of QOL aspects in patients with hyper- and hypothyroidism. Responsiveness to treatment of non-toxic goiter was also demonstrated for physical and mental symptoms and overall QOL, but not for impact on social life or cosmetic complaints, in contrast to clinicians' predictions. For all comparable scales except one, the ThyPRO was more responsive to treatment than the SF-36 Health Survey. CONCLUSIONS: The ThyPRO was responsive to treatment across the range of benign thyroid diseases. We suggest implementing this measurement instrument as a patient-reported outcome in clinical studies and in clinical management.

(204.5) Psychometric properties of the Functional Assessment of Chronic Illness Therapy—Spiritual Well-Being, Expanded

Hayley S. Whitford, PhD, Cancer Council Australia, Sydney, Australia; Simeon B. W. Jones, BA (Hons), University of South Australia,