---

**Proceedings Paper:**
Young, T. orcid.org/0000-0001-8467-0471, Carlton, J. orcid.org/0000-0002-9373-7663, Powell, P. et al. (1 more author) (2025) Proceedings of the patient reported outcome measures (PROMs) research conference, Sheffield 2023 22nd June 2023, University of Sheffield, Sheffield, UK. In: Quality of Life Research. Proceedings of the Patient Reported Outcome Measures (PROMs) Research Conference, 22 Jun 2023, Sheffield, United Kingdom. Springer Nature, pp. 1-12. ISSN: 0962-9343. EISSN: 1573-2649.

https://doi.org/10.1007/s11136-024-03887-7

---

Quality of Life Research (2025) 34 (Suppl 1):S1–S12
https://doi.org/10.1007/s11136-024-03887-7

ABSTRACTS

# Proceedings of the Patient Reported Outcome Measures (PROMs) Research Conference, Sheffield 2023

## 22nd June 2023, University of Sheffield, Sheffield, UK

Published online: 7 February 2025


## Introduction

**Proceedings: PROMs research Conference, University of Sheffield, 22nd June 2023**

**Tracey Young[1*], Jill Carlton[1], Phil Powell[1], Amanda Lane[1].**

**[1]Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, United Kingdom.**

[*]Corresponding author t.a.young@sheffield.ac.uk.

The 7th UK Patient Reported Outcome Measures (PROMs) Research Conference (#UKPROMS) held on the 22nd June 2023 brought together 145 academics, clinicians and PhD students, representing 51 organisations. This was the first time since the Covid-19 pandemic and subsequent lockdowns that the conference had been held in-person. The conferences held in 2021 at The University of Sheffield, England, United Kingdom, and 2022 at the Welsh Value in Health Centre, Wales, United Kingdom had been held virtually.

The aim of the conference was to look at *PROMS Across the Lifespan* and bring together leading experts and early career researchers to engage in the latest advances in the field of PROMs research and implementation. The conference focused on researcher-led activities on methodologies around the development, testing and use of PROMs in different contexts and settings.

## Conference summary

The conference was held in person over one day at the University of Sheffield. Sponsorship: Publication of this supplement was sponsored by the School of Medicine and Population Health, University of Sheffield. All content was reviewed and selected by the PROMs Research Scientific Advisory Committee, which held full responsibility for the abstract selections. The day was structured around a mixture of facilitated plenary sessions, parallel sessions and poster displays by PROMs researchers predominantly from across the UK. There were 26 parallel sessions, three poster sessions of which 46 posters were on display and two plenary sessions. The oral presentations covered topics spanning *'PROMs Across the LifeSpan'*.

**Professor Melanie Calvert**, Professor of Outcomes Methodology, Institute of Applied Health Research, University of Birmingham, gave a presentation on *The challenges in the field of PROs in the next 5 years*, discussing the value of PROs, the challenges of data collection, and the work required to maximise the use of PRO data to inform care and effective communication. **Professor Joanne Greenhalgh**, Professor of Applied Social Research Methodology in the School of Sociology and Social Policy at the University of Leeds and the President of the International Society for Quality of Life Research (ISOQOL), chaired a session looking at research on *The challenges relating to measuring outcomes at the three extremes—in childhood, life limiting conditions and older people*. This included **Lucy Coombes**, from the Florence Nightingale Faculty of Nursing, Midwifery & Palliative Care, King's College London, who presented her research on *Measuring health-related outcomes in childhood*. **Dr James van Oppen**, Clinical Research Fellow in Emergency Medicine, Department of Health Sciences, University of Leicester, who discussed his research on *PROMs with older people*. **Professor Fliss Murtagh**, Professor of Palliative Care and Director of the Wolfson Palliative Care Research Centre, University of Hull, who concluded by presenting her research on *The Challenges in measuring outcomes in advanced and life-limiting conditions*.

The conference offered three prizes which were judged and awarded as follow:

- Best PhD poster was awarded to **Hannah Worboys,** University of Leicester, Leicester, United Kingdom.
- Best Oral presentation that best communicates their research to the public as judged by patient and public involvement collaborators was awarded to **Raveen Jayasuriya**, Sheffield Children's Hospital, Sheffield, United Kingdom and University of Sheffield, Sheffield, United Kingdom.
- Best oral presentation by an Early Career Researcher, sponsored by Oxford University Innovation Ltd was awarded to **Raveen Jayasuriya**, Sheffield Children's Hospital, Sheffield, United Kingdom and University of Sheffield, Sheffield, United Kingdom

## Ethics declarations

Conflict of interest: there is no conflict of interest regarding the research, authorship, or publication of this article.

Consent for publication: informed consent was obtained.

## Abstracts for oral presentations

### A1 Patient-centred development and validation of the Patient-Reported Impact of Dermatological Diseases (PRIDD) measure.

**Rachael Pattinson[1], Nirohshah Trialonis-Suthakharan[2], Jennifer Austin[3], Allison FitzGerald[3], Tim Pickles[1], Matthias Augustin[2], Chris Bundy[1]**

**[1]Cardiff University, Cardiff, United Kingdom. [2]University Medical Center Hamburg-Eppendorf, Hamburg, Germany. [3]GlobalSkin, Ottawa, Canada**

Background: Dermatological conditions are highly prevalent worldwide and significantly impact upon quality of life. Existing patient-reported outcome measures (PROMs) do not comprehensively capture the full impact of dermatological diseases on patients' lives and therefore underestimate patient burden. Aim: To develop and validate a new measure of the impact of dermatological conditions on patients' lives called PRIDD (Patient-Reported Impact of Dermatological Diseases) in close collaboration with patients. Methods: GRIDD is a mixed methods study consisting of five phases: (1) COSMIN systematic review. (2) Qualitative interview study developing the conceptual framework of impact and generating items. (3) Delphi study eliciting consensus from patients on which impact items to prioritise for inclusion in PRIDD. (4) Cognitive interview study evaluating content validity, acceptability, and feasibility. (5) Psychometric testing. Adults ($\geq$ 18 years) living with a dermatological condition worldwide were recruited through GlobalSkin's unique global membership network of over 200 patient organisations from 67 countries and representing 57 dermatological disease areas. Results: 2221 people representing 90 conditions from 61 countries participated. (1) None of the 36 PROMs evaluated in the systematic review were recommended for use as the 'gold standard.' (2) The conceptual framework depicted impact as a multifaceted construct involving physical, life responsibilities, psychological and social impacts. (3) The Delphi study reduced the item pool of 263 to a 27-item draft of PRIDD through patient consensus. (4) Cognitive interviews produced a 26-item PRIDD with evidence of content validity, feasibility, and acceptability from patients. (5) A confirmatory factor analysis refined the conceptual framework. PRIDD fits the Rasch model and met the COSMIN criteria for structural validity, internal consistency, construct validity, and test–retest reliability. Measurement error and responsiveness will be tested in a future study. Conclusion: PRIDD is the only theory-led impact measure developed in partnership with patients and patient organisations worldwide meeting the gold-standard COSMIN criteria. PRIDD will greatly enhance patient perspectives by providing quantifiable patient impact data for better decision-making at the individual, national and global levels, with higher prioritisation of dermatological conditions.

### A2 Collaborative research methods to develop a patient-reported outcome measure (PROM) for people with communication disability who use augmentative and alternative communication (AAC).

**Katherine Broomfield[1,2], Georgina Jones[3], Simon Judge[4,5], Deborah James[1]**

**[1]Manchester Metropolitan University, Manchester, United Kingdom. [2]North Bristol NHS Trust, Bristol, United Kingdom. [3]Leeds Beckett University, Leeds, United Kingdom. [4]University of Sheffield, Sheffield, United Kingdom. [5]Barnsley Hospitals NHS Foundation Trust, Barnsley, United Kingdom**

Background: People use augmentative and alternative communication (AAC) because they have a communication disability. Communication disabilities result from conditions that people are born with or acquire as an adult; conditions that are usually lifelong and can be life-limiting. AAC strategies include paper and powered external devices such as picture books, spell charts, and computer-based systems that produced synthetic speech from messages entered or stored within them. People who use AAC are not consistently involved in decisions about the AAC devices that are recommended by healthcare professionals. PROMs can be useful to support clinician-person communication and facilitate shared decision-making. There are currently no patient-reported outcome measures (PROMs) for this group and no clearly defined quality of life parameters. The Unspoken Voices project was concerned with understanding more about the experiences of using AAC to develop the content for a PROM. Method: We adopted a collaborative research approach consisting of a multi-disciplinary research team, an active public involvement group of people who use AAC and their family and carers, and a qualitative methodology that oriented the project towards principles of person-centredness. We carried out 2 systematic reviews and 2 phases of qualitative data collection to develop the content for a PROM. Results: Our collaborative research approach resulted in the iterative evolution of the research project. Applying principles of person-centred research supported the co-production of the content for a PROM and the co-creation of meaning-making in relation to the concepts and items developed. We also proposed a theoretical framework that can be used to support dialogue during clinical interactions and facilitate the negotiation of the individual outcomes in the PROM. Conclusion: Collaborative research practices resulted in an authentic and person-centred PROM development research project. We generated multi-dimensional concepts in a framework that reflects the influence of context and time on outcome items. The results reflect the complexity of communication using AAC and the changeable nature of experience. PROMs can support mutuality in clinical interactions and can enable conversations that attribute individual meaning and relevance to PROM concepts and items. Person-centred, collaborative research is essential to developing clinical tools to record person-centred outcomes.

### A3 Co-designing and testing an electronic patient reported outcome measure package with patients for use in clinical trials in pulmonary hypertension.

**Joseph Newman[1,2], Elisa Ferrer-Mallol[3], Shaun Clayton[4], Mary Ferguson[4], Iain Armstrong[4], Elin Haf Davies[3], Mark Toshner[1,2]**

**[1]University of Cambridge, Cambridge, United Kingdom. [2]Royal Papworth Hospital, Cambridge, United Kingdom. [3]Aparito Ltd, Wrexham, United Kingdom. [4]Pulmonary Hypertension Association UK, Sheffield, United Kingdom**

Background: There are multiple patient reported outcome measures (PROMs) in use with pulmonary hypertension patients. There is both overlap in conceptual coverage between these PROMs and unique questions, with no tool able to capture the full spectrum of disease impact. PROMs are increasingly used as endpoints in clinical trials with regulators encouraging the use of measures of feel and function. EmPHasis-10, as the shortest, has gained particular interest as the disease-specific PROM most amenable to digitisation for use in decentralised trials. Aims: 1. To quantify the perceived breadth of emPHasis-10; 2. To domain map any limitations to other established

PROMs to complement the coverage of emPHasis-10. 3. To configure and beta-test the package of ePROMs for patient usability. Methods: A national online patient survey was iteratively designed with and promoted by Pulmonary Hypertension Association UK. It included specific questions on PROMs in PH. The subsequent domain mapped ePROMs were configured on the Atom5$^{TM}$ mobile phone application and beta-tested with patients. Online feedback was captured quantitatively by a questionnaire and qualitatively by semi-structured questioning in a focus group. Results: Survey results were received from 112 patients. 94% of patients reported that emPHasis-10 covers most (50%) or all (43%) of the ways that PH affects their life. Free-text comments on aspects of QoL not covered by emPHasis-10 included mental health, activities, exercise, drug side effects, sleep and some specific physical symptoms. Domain mapping showed that the inclusion of EQ-5D-5L, GP5 of FACT-G and the JSEQ efficiently covered most of these aspects with minimal additional questions. App usability testing of these ePROMs with 9 patients was generally favourable. Conclusions: Involvement of patients with the co-design of a novel ePROM package has ensured broad domain coverage and a highly functional interface on the app. This tool needs evaluation in a longitudinal study to evaluate construct validity, acceptability and adherence.

## Session 1B

### A4 Generating a UK Value Set for the Weight- Specific Adolescent Instrument for Economic Evaluation (WAItE)

**Tomos Robinson[1], Sarah Hill[2], Yemi Oluboyede[2]**

**[1]Newcastle University, Newcastle, United Kingdom. [2]PHMR, Newcastle, United Kingdom**

Background: Adolescent obesity is a public health problem in the UK. The Weight-Specific Adolescent Instrument for Economic Evaluation (WAItE) has been developed as the first weight-specific health-related quality of life measure appropriate for economic evaluation. However, the WAItE currently cannot be used to generate quality adjusted life years (QALYs), which are the basis of cost-utility analysis. Aims: This main aim of this study aimed to generate a preference-based scoring algorithm for the WAItE. A secondary aim was to compare the value sets generated using the discrete choice experiment-visual analogue scale (DCE-VAS) and Time Trade-Off (TTO) anchoring methods. Methods: A discrete choice experiment (DCE) was administered to a representative sample of the UK adult general population (n = 1005) via an online survey to value health states described by the WAItE classification system. Multinomial Logit and Mixed Logit models were used to analyse the data. The DCE-VAS and TTO anchoring methods were used to anchor the value set on to the 0–1 QALY scale. The VAS data for the DCE-VAS was collected as part of the DCE study, and the TTO data was collected from a standalone online study (n = 42). Results: Interim results show that although the DCE produced some counterintuitive results with regards to the monotonicity of the WAItE, a value set was able to be generated using the TTO anchoring technique. The value of the 'PITS' state (the worst health state possible from the WAItE classification system) generated from the external TTO study and used in the interim value set was 0.229. The DCE-VAS anchoring technique produced similar results. Final results (including a full comparison of the two anchoring techniques) will be presented at the conference. Conclusions: The algorithm developed from this study can be used to generate health state utility values for any study using the WAItE, enabling the estimation of weight-specific QALYs in the context of adolescent obesity.

### A5 Discrete Choice Experiments in Health State Valuation: Progress and New Trends

**Haode Wang[1], Donna Rowen[1], Clara Mukuria[1]**

**[1]University of Sheffield, Sheffield, United Kingdom**

Background: Discrete choice experiments (DCEs) are increasingly used in health state valuation studies. This systematic review updates the progress and new findings of DCE studies in health state valuation, covering the period from June 2018 to December 2022. The review reports what methods are currently being used in DCE studies to value health, and study design characteristics and reviews DCE health state valuation studies published in the Chinese language for the first time. Methods: English language databases PubMed and Cochrane, and Chinese language databases Wanfang and CNKI were searched using the self-developed search terms. Health state valuation or methodology study papers were included if the study used DCE or Case 3 Best–Worst Scaling (BWS) to generate a value set for a preference-based measure. Key information extracted included DCE study design strategies applied, methods for anchoring the latent coefficient onto a 0 to 1 QALY scale and data analysis methods. Results: 65 studies were included, one Chinese language publication and 64 English language publications. The number of health state valuation studies using DCE rapidly increased in recent years, and were conducted in a broader range of countries that prior to 2018. Wide usage of DCE with duration attributes, D-efficient design and models accounting for heterogeneity continued in recent years. Although more methodological consensus was found than in studies conducted prior to 2018, this consensus may be driven by valuation studies for common measures with an international protocol. Valuing longer measures with wellbeing attributes attracted attention and more realistic design strategies (e.g., inconstant time preference, efficient design and implausible states design) were identified. Conclusion: The use of DCEs in health state valuation continues to grow dramatically with methodological progress. However, study design is driven by international protocols and method selection is not always justified. There is no gold standard for DCE design, presentation format or anchoring method. More qualitative and quantitative methodology studies are recommended to evaluate the effect of new methods before researchers make methodology decisions.

### A6 Understanding Public Views on the Normative Decisions Made when Valuing Health-Related Quality of Life Patient Reported Outcome Measures in Children and Young People

**Philip Powell[1], Anju Keetharuth[1], Clara Mukuria[1], Koonal Shah[2], Donna Rowen[1]**

**[1]University of Sheffield, Sheffield, United Kingdom. [2]National Institute for Health and Care Excellence, London, United Kingdom**

Background: Developing methodology for measuring and valuing health related quality of life (HRQoL) in children and young people is a priority area for the National Institute for Health and Care Excellence (NICE). Items from HRQoL patient reported outcome measures (PROMs) can be 'valued' by eliciting relative preferences from people for living in different states of HRQoL. One of NICE's current priorities is to better understand public opinion on the normative questions of whose preferences should be elicited (adults, children, or both) and from which perspective (who should be imagined is living in different states of HRQoL). Methods: Opinions of the adult (18+) UK public (N = 32) were elicited using online semi-structured focus groups, featuring a breadth of age, sex, ethnicities, and responsibility

for children under 18. Participants were provided with bespoke informational material on valuation and their views were elicited on who should be asked and with what perspective when valuing HRQoL in children. Participant understanding was probed and measured. Selected arguments for and against different positions (e.g. the 'taxpayers' argument) were presented and participants views recorded. Results were analysed using Framework Analysis. Results: The UK adult public demonstrated near-to-universal agreement that young people should be involved in valuation in some form, yet this differed depending on age or maturity. There was a lot of support for combining involvement from children and adults (e.g., parents). There was little support for the 'taxpayers' argument for asking adults. Mixed opinions were given on perspective, with most support for an 'own' perspective. Most participants thought that people should know the task is about valuing children's HRQoL for ethical reasons. Conclusion: Informed views from the UK public on who should be asked and with what perspective when valuing child HRQoL PROMs have been elicited and may differ from normative positions adopted in current valuation protocols.

## Session 1C

### A7 Using cognitive interview and think aloud techniques to assess content validity for a new Patient Reported Outcome Measure for Rheumatoid Arthritis Disease Activity

**Tim Pickles[1], Rhiannon Phillips[2], Mike Horton[3], Karl Bang Christensen[4], Ernest Choy[1]**

**[1]Cardiff University, Cardiff, United Kingdom. [2]Cardiff Metropolitan University, Cardiff, United Kingdom. [3]University of Leeds, Leeds, United Kingdom. [4]University of Copenhagen, Copenhagen, Denmark**

Background: Disease Activity (DA) monitoring is a standard of care in Rheumatoid Arthritis (RA). A systematic review of Patient Reported Outcome Measures for RA DA following COSMIN guidelines demonstrated a lack of sufficient evidence for content validity and thus concluded that none can be recommended for use. Rasch measurement theory was used on data collected from 691 RA patients in South Wales to develop a valid item pool of 12 items, covering tenderness and swelling, disease activity, pain, physical functioning and stiffness. General health and fatigue items were seen to measure a separate construct. Aim: The aim of this study is using cognitive interview and think aloud techniques to look at how participants thought about answering questions relating to their RA. Method: A sampling framework was devised using a maximal variation approach to sample participants from varying age categories, gender and education level. This framework included 20 participants, with a proviso that a minimum of 10 were to be held with continuation to 20 unless data saturation was reached. Individual interviews with participants took place from November 2022 to February 2023. Cognitive interview and think aloud techniques were implemented as participants went through a copy of the final 12 items along with 6 similar items and 6 general health and fatigue items. Participants were then asked about the item instructions and response options, and specifically whether they could distinguish RA symptoms from those of other conditions, whether it was suitable to consider tenderness and swelling together within an item and whether they felt that fatigue and general health were separate issues to their disease activity. Interviews were audio-recorded, transcribed verbatim, and analysed thematically using the approach described by Braun and Clarke. Results: Analysis will focus on comprehensiveness, comprehensibility and relevance and is currently underway. Conclusions: Cognitive interview and think aloud techniques are important in developing valid measurement. This research will determine the content validity of this new set of items. Once confirmed, this will allow for the development of the item pool into a computer adaptive test, which would provide streamlined assessment of a patient's RA DA.

### A8 Is my instrument inclusive and accessible? Checking language variation and age appropriateness of an eczema patient reported outcome measure.

**Michaela Gabes[1], Aviël Ragamin[2], Arabella Baker[3], Gesina Kann[4], Theresa Donhauser[5,4], Daniela Gabes[4], Laura Howells[3], Kim S. Thomas[3], Jart A. F. Oosterhaven[6], Suzanne G.M.A. Pasmans[2], Marie-Louise A. Schuttellaar[7], Christian Apfelbacher[5]**

**[1]Otto-von-Guericke_University Magdaburg, Magdeburg, Germany. [2]Erasmus MC University Medical Center, Rotterdam, Netherlands. [3]University of Nottingham, Nottingham, United Kingdom. [4]University of Regensburg, Regensburg, Germany. [5]Otto-von-Guericke University Magdeburg, Magdeburg, Germany. [6]University of Groningenen, Groningen, Netherlands. [7]University of Groningen, Groningen, Netherlands**

Background: Eczema is a common, inflammatory, itchy skin condition characterised by periods of increased disease activity and relative remission, affecting both children and adults. Patient-reported eczema control is an important outcome when evaluating treatments. The Recap of atopic eczema (RECAP) is a patient-reported outcome measure (PROM) assessing eczema control and it is part of the core outcome set for eczema. This instrument was developed and validated in the UK. There is a self-reported and a proxy-reported version in English, Dutch and German. However, it is unclear whether the self-reported version shows adequate content validity when completed by young people in these languages. Aims: To assess the content validity (comprehensibility, relevance and comprehensiveness) of the English, German and Dutch versions of the self-reported RECAP in young people with eczema and to identify the most appropriate age cut-off for self-completion. Methods: We conducted 23 semi-structured cognitive interviews with young people from 8 to 16 years, using the "think-aloud" method. In Germany and the Netherlands, participants were recruited in dermatology clinics and in the UK through social media. Interviews were audio recorded, transcribed verbatim and analysed in the three languages, using a problem-focused coding manual. Transcripts were coded by two independent reviewers in each country. Themes were translated into English and compared across the three countries. Results: Significant age-related comprehensibility issues with the last three items of the questionnaire occurred with young people aged 8 to 11 years, causing difficulties in completing RECAP without assistance. However, older children had only minor problems and were able to complete the questionnaire by themselves. The self-reported version of RECAP has sufficient content validity for self-completion in young people aged 12 years and above. Conclusions: The self-reported version of RECAP is appropriate for use from the age of 12 years and above. The proxy-version is suitable for children younger than 12 years.

### A9 Throwing a fit: exploring tensions between model fit and content validity in PROM development.

**Jill Carlton[1], Philip Powell[1]**

**[1]University of Sheffield, Sheffield, United Kingdom**

Background: Existing guidance on the development of patient reported outcomes measures (PROMs) highlights the importance (and evidence) of good content validity. This is normally achieved in initial item design, through qualitative methods. However, in subsequent item selection stages (i.e., when assessing psychometric performance) item content coverage can reduce, potentially compromising content validity. PROM developers face difficult decisions during final item selection to ensure face validity whilst demonstrating good psychometric performance (model fit) and considering acceptability of the final instrument to the target population. Aims: To demonstrate the implications of two item selection approaches ('data driven' vs. 'face validity driven') in a worked example of a new PROM for hypoglycaemia. Methods: A conceptual framework was used to inform development of candidate items for a new PROM designed to measure the impact of hypoglycaemia upon health-related quality of life (HRQoL). 40 candidate items were tested in a psychometric survey. Psychometric analysis and item response theory analysis were undertaken to identify which of the candidate items were suitable for inclusion in the final PROM. Two approaches were applied. In Approach A, items were selected purely on 'performance', i.e., data driven. In Approach B, items were selected based upon 'performance', stakeholder engagement and consideration of results from item generation, i.e., face validity driven. Results: Approaches A and B will be discussed, and the resultant PROMs shown. Relative advantages and disadvantages of both approaches will be highlighted, alongside key challenges of where evidence may be conflicting, and decisions may be difficult. Conclusions: This presentation is a precursor to a planned study designed to determine the relative importance and weighting(s) of 'evidence' to be considered during PROM development. In other words, to identify consensus on what to prioritise in situations of conflicting evidence. Potential collaborators will be sought from national and international stakeholders in PROM development.

## Session 2A

### A10 Including young children in the development and testing of patient reported outcome (PRO) instruments: a systematic review of children's involvement in qualitative methods.

**Victoria Gale[1], Jill Carlton[1]**

**[1]The University of Sheffield, Sheffield, United Kingdom**

Background: Qualitative research during the development/testing of patient reported outcome measures (PROMs) is recommended to support PROM content validity. However, it is unclear if and how young children ($\leq$ 7 years) can be involved in this research because of their unique cognitive and communicative needs. Aims. To investigate the involvement of children $\leq$ 7 years in qualitative research for PROM development/testing. To identify which elements of PROM development/testing and which subjective health concepts have been explored with this age group, and by which qualitative methods. Methods: Three electronic databases were searched (updated 29th June 2022) with no date restrictions. Included studies had samples of at least 75% aged $\leq$ 7 years or reported distinct qualitative methods for children $\leq$ 7 years in primary qualitative research for concept elicitation, content generation, cognitive interviewing, or content validity testing. Articles not in English and PROMs that did not enable children $\leq$ 7 years to self-report were excluded. Data on study type, subjective health concepts, and qualitative methods were extracted and synthesised narratively. Methods were compared to recommendations from guidance. Results: Of 19 included articles, 15 reported concept elicitation research and 4 reported cognitive interviewing. Compared to concept elicitation studies, cognitive interviewing studies typically reported less methodological detail and fewer recommendations from guidance, including methodological adaptations to meet young children's needs. Most cognitive interviewing studies were limited in scope with respect to content validity; clarity was generally the focus with comprehensiveness and relevance being less explored. Conclusions: Establishing content validity of PROMs is essential and cognitive interviews are recommended to help achieve this. However, this review found cognitive interviews with young children are limited in frequency, scope, and reported methodological detail, potentially impacting the content validity of PROMs for this age group. It is unclear whether this is because they are too challenging, or because they have not been suitably adapted for this age group.

### A11 Identifying determinants to predict patient adherence to brace treatment in scoliosis: development of a questionnaire based on Protection Motivation Theory

**Raveen Jayasuriya[1,2], Paul Norman[2], Daniel Hind[2], Ashley Cole[1]**

**[1]Sheffield Children's Hospital, Sheffield, United Kingdom. [2]University of Sheffield, Sheffield, United Kingdom**

Background: Successful full-time bracing in adolescent idiopathic scoliosis is adherence dependent. Predicting adherence is crucial to develop complex interventions which increase treatment adherence. Factors affecting adherence behaviour in bracing have not been evaluated using a theory-based framework. Methods: Stage 1: Semi-structured interviews explored adolescent and parent perspectives of the psychological effects of bracing. The topic guide and directed content analysis of transcripts operationalised protection motivation theory (PMT) as the underpinning psychological framework. Stage 2: Brace Adherence Prediction Questionnaire (BAPQ) development through item generation, reduction, PPI and expert panel review. Results: Fifteen patients and nineteen parents participated in interviews. Appraising the threat of the diagnosis was difficult as scoliosis was an unknown for half. With information from surgeons and independent research, consistent themes of severity of their scoliosis and vulnerability for future prognosis emerged. The response cost of treatment adherence was dominant and consistent between participants, despite good belief in response efficacy of bracing. The self-efficacy strategies that patients employed to increase adherence, and the maladaptive response rewards of non-adherence were variable. Eight key themes spanning the seven domains of PMT were identified from content analysis. A long list of 90 items were generated from the content analysis, which was reduced to 38 items, which coved the 8 themes and 7 domains. BAPQ-38 was reviewed by a scoliosis PPI group, and comments actioned to produce BAPQ-32 which was subjected to expert panel (surgeons, orthotists, specialist nurses, physiotherapists) critique. The final version of BAPQ-32 was checked by two expert PPI members, and then approved by a REC for validation through the BASIS study, which is prospectively collecting objective adherence data using thermal sensors. Conclusion: The qualitative study demonstrates that PMT provides a coherent psychological framework to explore the factors affecting brace adherence in AIS. This questionnaire intervention targets PMT constructs as intermediate outcomes in long and complex causal chains, which will

contribute to preventing scoliosis curve progression via improving brace adherence outcomes. Dyadic interview data, PPI and expert review has informed a de novo prediction questionnaire for stratifying patient adherence, for future validation through a prospective study.

### A12 Developing a questionnaire for children with hearing loss – the York Binaural Hearing Related Quality of Life – Youth (YBHRQL-Y)

**Sarah Somerset[1,2], Adam Pedley[1], Padraig Kitterick[3]**

**[1]University of Nottingham, Nottingham, United Kingdom. [2]Nottingham Biomedical Research Centre, Nottingham, United Kingdom. [3]National Acoustics Laboratory, Sydney, Australia**

Background: Within hearing research patient reported outcome measures exist for use with children. None have been designed with these children at the heart of development. Aim: We aim to re-operationalise the existing domains in the YBHRQL (Summerfield, Kitterick et al. 2022) into the YBHRQL-Y for use in young people with hearing loss. Method: We used the qualitative pre-test interview approach (Buscle, Reiter et al. 2022) to re-operationalise the domains for young people with hearing loss. The construct validity of the YBHRQL-Y was then assessed by administering alongside existing measures (YBHRQL, HUI3, CHU-9D) at two time points 2 weeks apart to parent/child dyads via JISC. The time trade-off technique was used to obtain preference weights for the 27 health states in the YBHRQL-Y via interviews with young adults aged 18 to 24 years. The EuroQol group's EQ-VT protocol was followed. Interviews were conducted online. Results: The YBHRQL-Y consists of 5 response options, uses simple scene setting vignettes and has short sentence answers as per direction of children with hearing loss. The measure within the research setting has shown itself to have construct validity and to be reliable. It remains to be seen within the wider Both Ears (BEARS) trial how it will perform within the clinical setting. This is a unique measure that has been developed with children at its heart which is critical to future development of measures within and beyond the hearing field. Conclusion: This measure is a key development in the field as it is hearing specific, designed with children with hearing loss and a preference-based measure. The measure is currently being applied within the BEARS clinical trial to evaluate its performance within the clinic.

## Session 2B

### A13 Recommendations for HRQoL data collection in dementia populations

**Hannah Hussain[1], Anju Keetharuth[1], Allan Wailoo[1], Donna Rowen[1]**

**[1]University of Sheffield, Sheffield, United Kingdom**

Background: Measuring health-related quality of life (HRQoL) in dementia populations presents with its own set of specific challenges. There is a need to consider pragmatic and controllable aspects to optimise data collection, and there maximise data quality and validity. Cognitive impairment in people living with dementia (PlwD) had led to an increased reliance on proxies to provide HRQoL outcome reports, which has the potential to introduce variation in assessments, and might not be the most accurate or appropriate way of assessing HRQoL. This paper uses existing trial data to evaluate pragmatic aspects of data collection in this population, aiming to provide recommendations for use in future dementia studies. Methods: Data were acquired from three previously conducted dementia clinical trials, "REMCARE", "ACTIFCARE" and "EPIC", and analysed by rater-type (patient, proxy). We examined convergent validity and inter-rater agreement by dimension and dementia severity stage. Data related to the administration method of measures were appraised where available. Results: The three studies included PlwD-caregiver dyads spanning from very mild–severe dementia. Convergent validity of EQ-5D dimensions with external measures of dementia symptomatology were consistently stronger with proxy-assessments. However, the findings indicate adequate association in mild dementia for PlwD-self-assessments. There were differences between proxy-types, with staff-proxy assessments producing a higher degree of inter-rater agreement with PlwD-self-assessments. For informal proxies, agreement was strongest with spouses (compared with offspring). Of the administration methods, face-to-face had fewer missing data at baseline. Conclusions: The results suggest that self-assessment of HRQoL may be suitable in mild dementia, but at more severe stages a proxy may be required. The type of proxy is pivotal and should be someone with close everyday contact with the PlwD; this may be a staff member for institutionalised PlwD, and typically a spouse for PlwD that are community-dwelling. Face-to-face administration is recommended.

### A14 Participant-reported questionnaire modules to measure the use of informal care, social care and personal expenses: development and preliminary validation.

**Kirsty Garfield[1], Eileen Sutton[1], Gail Thornton[2], Samantha Husbands[1], Ailsa Cameron[1], William Hollingworth[1], Sian Noble[1], Joanna Thorn[1]**

**[1]University of Bristol, Bristol, United Kingdom. [2]Public contributor, Bristol, United Kingdom**

Background: Measuring resource use is the first step in assessing treatment or service costs in cost-effectiveness analyses. ModRUM, a generic, modular resource-use measure, designed for collecting self-report healthcare resource-utilisation data, has recently been developed (bristol.ac.uk/modrum) and tested rigorously. However, when considering interventions across the life course, a broader perspective may be required; for example, social care resource use may be relevant. Aim: To develop additional modules to bolt-on to ModRUM for measuring social care, informal care (i.e. unpaid care provided by relatives, friends or the community) and personal expenses. Methods: A rapid review, supplemented by an online survey of social-care professionals, was conducted to create a long-list of potential items. Focus groups were held with people who access social care and academic health economists. The health economists were asked to prioritise items from the long-list; the results guided both focus group discussions. Items were selected following qualitative analysis, drawing on methods of constant comparison. Draft modules were developed and tested via in-depth interviews with health economists. Think-aloud testing was undertaken with people who access social care. Results: The review identified a long-list of approximately 200 items after grouping similar items. 24 social care professionals responded to the online survey; one item was renamed to reflect real usage. Five health economists and four people who accessed social care participated in focus groups. Focus-group feedback shaped the social care and informal care modules, but suggested that the scope of personal expenses was too broad. The personal expenses module was therefore changed to focus on aids and adaptations, which can be costly. Input from ten service user participants via think-aloud interviews and 12 health economist interviews guided improvements to ensure the draft modules are as relevant and user-friendly as

possible. For example, an explanation of what "personal care" means was added. Conclusions: Patient-reported informal care, social care and aids/adaptations modules have been developed for use alongside ModRUM, a rigorously developed, standardised UK healthcare-utilisation questionnaire. These modules extend the use of ModRUM to studies undertaking economic analyses from broader perspectives than healthcare. The new modules will shortly undergo pilot testing.

### A15 The impact of different researchers to capture patient reported Quality of Life measures in a dementia randomised controlled trial.

**Rachel Evans[1], Zoe Hoare[1], Paul Brocklehurst[1], Jean Ryan[1]**

**[1]Bangor University, Bangor, United Kingdom**

Background: Capturing changes in Quality of Life (QoL) within randomised controlled trials (RCTs) can be complex. The precision and accuracy of scales to measure change is crucial and consideration needs to be given to potential measurement errors when collecting outcomes. Many RCTs use multiple researchers to collect data, which has the potential to introduce variation in measurements. Aim: The aim of the research was to assess two research questions 1. Does the use of a different researcher at time points impact upon the outcome measure? 2. Does the (biological) gender of the attending researcher impact upon the outcome measure? Methods: A previously conducted study assessing the impact of reminiscence therapy, 'REMCARE' (Reminiscence groups for people with dementia and their family caregivers), provided the platform for this exploratory analysis. Two variables were created: 'researcher continuity' (the same researcher attended, or different researchers undertook the assessments) and 'researcher gender continuity' (same or different researcher and if different whether the researcher gender was consistent or not). ANCOVA (Analysis of Covariance) models were run on two QoL measures, the QoL-AD (Quality of Life in Alzheimer's Disease) and QCPR (Quality of the Caregiving Relationship). Results: 330 PwD-carer dyads were included in the analysis. A statistically significant effect was found on the researcher continuity variable for at follow-up 1 but not at follow-up 2. Similarly, a statistically significant effect was seen at follow-up 1 for both measures on the researcher gender continuity variable. At follow-up 2, there was a statistically significant finding on the QCPR measure. Conclusions: These exploratory results indicate the possible impact of researcher continuity and/or researcher gender continuity on QoL outcomes in dementia studies. Further research is required to explore this further and establish causality. If demonstrated, this would have implications for the planning of future empirical studies collecting QoL measures.

## Session 2C

### A16 Comparison of statistical methods for the analysis of patient-reported outcomes (PROs) in randomised controlled trials (RCTs): a simulation study

**Yirui Qian[1], Richard Jacques[1], Laura Flight[1], Stephen Walters[1]**

**[1]ScHARR, University of Sheffield, Sheffield, United Kingdom**

Background: Patient-reported outcome (PRO) scores are typically bounded, ordinal, and skewed, which complicates the decision on what statistical methods to use to analyse them. An inappropriate analysis can result in unreliable estimates and fail to provide accurate and robust results for decision-making on the use of health interventions. Aim: This study aims to use simulation analysis to compare the accuracy and robustness of different statistical methods for analysing PROs in randomised controlled trial (RCT) settings. Methods: Monte Carlo simulation was conducted to compare the performance of multiple linear regression (MLR), median regression (Median), ordered logit model (OL), beta-binomial regression (BB), and fractional logistic regression (Frac) for different possible categorical values of PROs. Predefined treatment difference was set at zero under the null hypothesis. The standardised treatment difference was used as the target estimand to solve the issue that the estimated treatment differences from these methods are based on different scales. Results: Under the null hypothesis, Median produced higher bias of estimated treatment difference than other methods, with larger empirical standard error and smaller coverage of 95% CIs. Tobit performed similarly to MLR. The coverage of transformed-scale based methods (OL, BB, and Frac) became poorer for PRO scores with higher number of possible categorical values. BB tended to have less bias than OL and Frac under the null hypothesis. Conclusions: MLR is robust despite the bounded and discrete nature of PROs, and is thus recommended as a simple method to use for PRO analysis. Median is not advised for the analysis of ordinal categorical PROs, even when the data is treated as continuous. Transformed-scale based methods can be considered when treating PRO as ordinal data with smaller number of possible categorical values. Future work will compare the performances of these methods under a series of alternative hypotheses.

### A17 Statistical analyses and interpretation of patient-reported outcomes in early phase dose-finding oncology trials: A methodological review

**Emily Alger[1], Anna Minchom[1], Olalekan Lee Aiyegbusi[2], Christina Yap[1]**

**[1]The Institute of Cancer Research, London, United Kingdom. [2]Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom**

Background: Within early phase dose-finding trials, the recommended phase 2 dose (RP2D) is determined by dose limiting toxicities (DLT) graded by clinicians during the DLT assessment period (usually up to cycle 1 or 2). As patients receive cycles of treatment beyond the DLT period, lower grade toxicities often under-reported by clinicians may become intolerable for patients. It is increasingly important to assess patients' health-related quality of life (HRQoL) alongside clinicians' assessment of toxicity using patient-reported outcomes (PROs) to ensure that the assessment of tolerability can continue beyond initial treatment cycles. Aims: to evaluate current statistical methods to analyse PRO data and the use of PROs to inform dose recommendations in published dose-finding oncology trials. Methods: Dose-finding oncology trials published from 2016 to 2021, which reported a PRO within their dose-escalation phase were included. We extracted 27 eligible papers indexed in PubMed. Study characteristics extracted included: PRO measures used, frequency of assessment, statistical techniques used and whether the PRO was involved in the dose-selection decisions. Results: Of the trials which reported a PRO as an endpoint (n = 21), most were identified as secondary endpoints (19, 90%). Each trial utilised a median of 2 (Range: 1–5) PRO measures and median of 5 (2–24) PRO assessments. Nearly all (25, 93%) trials were conducted in adult populations, two were conducted in pediatric populations. Only 3 (12%) trials used PROs to confirm the tolerability of the suggested RP2D, and no trials used PROs to inform dose (de-)escalation decisions. Half of the papers only used descriptive statistical techniques, however three papers employed more advanced inferential techniques such as time to HRQoL deterioration and

mixed-effect models. Conclusion: Most published trials which reported a PRO did not utilise this data to inform dose decisions. For those that did, this was to confirm RP2D tolerability rather than actively considered in dose-escalation or RP2D decisions. Furthermore, with no template for analysis methods for PROs within (de-)escalation decisions, comparison between study findings and recommended treatment dosages is unduly complex. Therefore, there is a crucial need for guidance to introduce more advanced PRO analysis methods to the dose-finding setting.

### A18 Development of core outcome set for use in adult primary glioma phase III interventional trials and identification of those that can be patient-reportable - a mixed methods study.

**Ameeta Retzer[1,2,3], Elin Baddeley[4], Stephanie Sivell[4], Annmarie Nelson[4], Helen Bulbeck[5], Kathy Seddon[6], Robin Grant[7], Richard Adams[8], Colin Watts[9], Olalekan Lee Aiyegbusi[1,2,3,10], Pamela Kearns[9,3,11], Samantha Cruz Rivera[1,2,3], Linda Dirven[12,13], Melanie Calvert[1,2,3,14,10], Anthony Byrne[4]**

**[1]Centre for Patient Reported Outcomes Research, Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom. [2]National Institute for Health Research (NIHR) Applied Research Collaboration West Midlands, Birmingham, United Kingdom. [3]National Institute for Health Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, United Kingdom. [4]Marie Curie Palliative Care Research Centre, Division of Population Medicine, Cardiff University School of Medicine, Cardiff, United Kingdom. [5]Brainstrust - the brain cancer people, London, United Kingdom. [6]Cardiff University, Cardiff, United Kingdom. [7]Department of Clinical Neurosciences, Royal Infirmary of Edinburgh, Edinburgh, United Kingdom. [8]Centre for Trials Research, Cardiff University, Cardiff, United Kingdom. [9]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom. [10]NIHR Birmingham-Oxford Blood and Transplant Research Unit (BTRU) in Precision Transplant and Cellular Therapeutics, UoB, Birmingham, United Kingdom. [11]Cancer Research UK Clinical Trials Unit, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, United Kingdom. [12]Department of Neurology, Leiden University Medical Center, Leiden, Netherlands. [13]Department of Neurology, Haaglanden Medical Center, The Hague, Netherlands. [14]Midlands Health Data Research UK, Birmingham, United Kingdom**

Background: Glioma patients' poor prognosis and high symptom burden has led to greater emphasis on their quality of survival, through maintaining neurocognitive and physical function, and overall health-related quality of life (HRQOL) throughout the disease trajectory. Patient-reported outcomes (PROs) allow insight into how treatment affects patients' perceived functioning, complementing other outcome data such as survival and radiological response, aiding physicians and patients in clinical decision-making. This requires effective data synthesis and meta-analyses, underpinned by consistent trial outcome measurement, analysis and reporting. Development of a core outcome set (COS) may contribute to a solution. Methods: A COS for glioma trials was developed through: (1) trial registry review and systematic review of qualitative studies; (2) interviews with glioma patients and caregivers; (3) outcome list de-duplication; (4) two-round Delphi process facilitating outcome list rating; (5) a consensus meeting to finalise the COS. Patient-reportable COS outcomes were identified by cross-referencing the outcomes in the final COS with outcomes and outcome measures reported in the trial registry review. Results: A COS consisting of 19 outcomes grouped into seven outcome domains (Survival, Adverse Events, Activities of Daily Living, HRQOL, Seizure Activity, Cognitive Function and Physical Function) was finalised by 13 participants at the consensus meeting. Use of six PRO measures (PROMs) were reported in the registry review. Cross-referencing PROM items with COS sub-domains identified overlap with five of the seven core outcomes (adverse events, activities of daily living, HRQOL, seizure activity, neurocognitive function, physical function). Conclusion: A COS for glioma trials was developed, comprising seven outcome domains. Sub-domains of 6/7 outcomes included in the final COS can be patient-reported. Given stakeholder interest in PROs and prioritising the patient perspective, opportunities to collect data in this way should be promoted and taken. Additional research will identify appropriate measurement tools, promote accessible PRO delivery, and further validate this COS.

## Session 3A

### A19 Development and validation of the kidney symptom burden questionnaire (KSB-Q): Rasch Analysis

**Derek Kyte[1,2], Benjamin Fletcher[2], Mike Horton[3], Sarah Damery[4], Olalekan Aiyegbusi[2,4,5,6,7], Nicola Anderson[2,4,5], Melanie Calvert[2,5,6,7,8,4], Paul Cockwell[2,9,10], James Ferguson[2,5], Muirne Paap[11], Chris Sidey-Gibbons[12], Neil Turner[13], Anita Slade[2,4,5]**

**[1]School of Allied Health and Community, University of Worcester, Worcester, United Kingdom. [2]Centre for Patient Reported Outcomes Research (CPROR), Institute for Applied Health Research (IAHR), University of Birmingham, Birmingham, United Kingdom. [3]Leeds Psychometric Laboratory for Health Sciences, University of Leeds, Leeds, United Kingdom. [4]National Institute for Health Research Applied Research Collaboration West Midlands (ARCWM), Institute for Applied Health Research (IAHR), University of Birmingham, Birmingham, United Kingdom. [5]National Institute for Health Research Birmingham Biomedical Research Centre, University of Birmingham, Birmingham, United Kingdom. [6]NIHR Oxford-Birmingham Blood and Transplant Research Unit (BTRU) in Precision Therapeutics, Birmingham, United Kingdom. [7]Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, United Kingdom. [8]National Institute for Health Research Surgical Reconstruction and Microbiology Research Centre, University of Birmingham, Birmingham, United Kingdom. [9]Department of Renal Medicine, Queen Elizabeth Hospital Birmingham, University Hospitals Foundation Trust Birmingham, Birmingham, United Kingdom. [10]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom. [11]Department of Child and Family Welfare, Faculty of Behavioural and Social Sciences, University of Groningen, Groningen, Netherlands. [12]Section of Patient-Centered Analytics, University of Texas MD Anderson Cancer Center, Houston, USA. [13]Centre for Inflammation Research, University of Edinburgh, Edinburgh, United Kingdom**

Background: Measurement of the symptoms that matter most to patients with Chronic Kidney Disease (CKD), currently requires concurrent completion of multiple patient-reported outcome measures (PROMs). This may lead to questionnaire burden: a widely recognised threat to adherence. Moreover, many PROMs used in CKD lack evidence of important measurement properties and were not developed using robust contemporary psychometric methods. Aim: Development and validation of a single kidney symptom burden questionnaire (KSB-Q) using Rasch Analysis. Methods: This mixed

methods prospective study was undertaken between October 2019 and September 2022 at 4 sites across England (Birmingham, London, Sheffield and Nottingham). First, a conceptual framework underpinning the KSB-Q was developed, following a global systematic review and meta-analysis. Second, the draft item pool was field-tested in a population including adults (≥ 18 years) with CKD stages 3–5. Third, a Rasch measurement theory approach was used to psychometrically assess the measurement characteristics of the item set using RUMM2030 software. The minimum sample size for the field test was 250 respondents, which provides 99% confidence that Rasch item calibrations and person measures are stable within ± 0.50 logits. Finally, in order to evaluate content validity, cognitive debriefing interviews were conducted by an experienced interviewer with expertise in measure development. Results: A total of 1464 KSB-Q surveys were posted out to patients and 350 returned questionnaires (24% response rate) were included in the analysis. The sample included 61% male respondents; 70% were white, with 26% reporting other ethnic backgrounds. The Rasch analysis indicated that, following a change to the scoring structure, items representing 9 key symptom areas (fatigue, pain, memory/concentration, poor sleep, skin problems, gastrointestinal problems, dizziness, restless legs and shortness of breath) formed a valid (chi-square p = 0.06), well-targeted (PSI = 0.73, Cronbach's alpha = 0.86), unidimensional (1.6% significant p = 0.05 t-tests) measure of kidney symptom burden. Cognitive debriefing and item bank survey responses provided evidence of content validity encompassing relevance, comprehensiveness and comprehensibility. Conclusions: The KSB-Q represents a short, accessible, PROM with evidence of strong psychometric properties.

## A20 Assessment and validation of the Modified COVID-19 Yorkshire Rehabilitation Scale (C19-YRSm) patient-reported outcome measure for Long COVID or Post-COVID syndrome

**Mike Horton[1], Román Rocha Lawrence[1,2], Ruairidh Milne[3,4], Manoj Sivan[1,5], LOCOMOTION Consortium[6]**

**[1]University of Leeds, Leeds, United Kingdom. [2]ELAROS 24/7 Ltd., Sheffield, United Kingdom. [3]Person with Long Covid, Southampton, United Kingdom. [4]University of Southampton, Southampton, United Kingdom. [5]Leeds Teaching Hospitals NHS Trust, Leeds, United Kingdom. [6]NIHR Study, Leeds, United Kingdom**

Background: The C19-YRS was the first condition-specific, validated scale published for patient assessment and monitoring in Long COVID or Post-COVID syndrome. At the UK PROMs Conference 2022, we presented the evolution of the original C19-YRS to the modified version, which was based on a combination of psychometric evidence, clinical content relevance and feedback from patients and healthcare professionals. Aims: This study aims to psychometrically assess and validate the modified C19-YRS using newly collected data from a large-scale, multi-centre NIHR study looking to optimise Long-COVID treatments and services across the NHS (LOCOMOTION). Methods: 720 patients (68% Female; mean age = 47, SD 12.6) completed the C19-YRSm scale digitally on the LOCOMOTION study ELAROS PROMs platform. The psychometric properties of the Symptom Severity (SS) and Functional Disability (FD) subscales were assessed using a Rasch Measurement Theory framework, where all individual scale items were assessed for model fit, targeting, reliability, unidimensionality, local dependency, response category functioning and differential item functioning (DIF) by age group and sex. Results: Rasch analysis revealed promising psychometric properties of the modified SS and FD subscales, with both displaying good targeting and reliability (SS: PSI = 0.80, Cronbach's alpha = 0.82; FD: PSI = 0.78, Cronbach's alpha = 0.81), evidence that the modified 4-point response category structure was successful, and no indication of DIF by age group or sex. However, some individual measurement anomalies were noted, indicating underdiscriminations for the Cough, Smell/Taste (both SS) and Communication (FD) items (Fit Residuals > 4.0). Minor pairwise local dependency was apparent between the Fatigue and Post-Exertional Malaise items of the SS scale (Q3 value = 0.225 above average). Conclusion: Although some minor anomalies are apparent, the modifications to the original C19-YRS appear to have strengthened the measurement characteristics, and the clinical and conceptual relevance. Further analytic and multidisciplinary (psychometric, clinical and patient perspective) collaborative work is needed to determine further developments.

## A21 A psychometric evaluation of the Musculoskeletal Health Questionnaire (MSK-HQ), assessing validity and sensitivity to change in inflammatory arthritis.

**Nikita Arumalla[1], James Galloway[1], Joanna Ledingham[2], Toby Garrood[3], Sam Norton[1]**

**[1]Academic Rheumatology, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom. [2]Rheumatology, Queen Alexandra Hospital, Portsmouth, United Kingdom. [3]Rheumatology, Guy's and St Thomas' NHS Trust, London, United Kingdom**

Background: The Musculoskeletal Health Questionnaire (MSK-HQ) is a patient reported outcome measure developed for use across the spectrum of musculoskeletal diseases. This study assesses the validity of the MSK-HQ and sensitivity to change in inflammatory arthritis in a national prospective cohort. Methods: The study sample included patients recruited to the National Early Inflammatory Arthritis Audit (NEIAA) between May 2018 and March 2020, with a diagnosis of inflammatory arthritis, returning a baseline PROM with at least half of the MSK-HQ items completed. Patients also completed PROMs at 3 and 12 months, contemporaneously to their clinic visits. Convergent validity was assessed in relation to the Health Assessment Questionnaire (HAQ)-II, Patient Health Questionnaire (PHQ)-4 and Disease Activity Score (DAS)-28, construct validity assessed using confirmatory factor analysis and MCID was estimated based on standard error of the measurement (SEM) and one-third of a standard deviation (SD). Results: Of 13,129 patients recruited to NEIAA in this period, 5106 met the inclusion criteria. Of these 73% had rheumatoid arthritis, 13% psoriatic arthritis, 2% axial spondyloarthritis, and 12% undifferentiated arthritis. The MSK-HQ total score was approximately normally distributed, without floor or ceiling effects. The MSK-HQ correlated well with the HAQ-II (r = − 0.79), PHQ-4 (r = − 0.66) and moderately with the DAS-28 (r = − 0.42). A unidimensional structure for the MSK-HQ was confirmed only when items 12 and 13, corresponding to disease understanding and self-efficacy, were excluded. The MSK-HQ total score demonstrated good sensitivity to change with baseline to 12-month change having a large effect size (standardised response mean 1.01; 95%CI 0.95 to 1.07). MCID for the overall sample was 3.9 based on SEM and 3.6 assessed with one-third of an SD, both rounding to an integer of 4-points across the inflammatory arthritis subtypes. Conclusion: This study provides evidence for the validity and sensitivity to change of the MSK-HQ in patients with inflammatory arthritis. In these patients, a change of more than 4 units is likely to be clinically meaningful. The MSK-HQ has high convergent and construct validity and is sensitive to change, providing a valuable tool for clinical care and research studies.

## Session 3B

### A22 Improving the uptake of electronic patient reported outcome measures in a specialised axial spondyloarthropathy clinic, Sheffield Teaching Hospitals

**Judith Jade[1], Zoe Cox[1], Emily Fox[1], Rachel Tattersall[1], Lisa Dunkley[1]**

**[1]Sheffield Teaching Hospitals, Sheffield, United Kingdom**

Background: Sheffield Teaching Hospitals (STH), UK, has a specialised axial spondyloarthropathy (axSpa) clinic run by a rheumatology consultant and physiotherapist. BASDAI and BASFI patient reported outcome measures are used to assess disease activity and response to treatment. This quality improvement project aimed to increase the use of electronic BASDAI and BASFI (ePROMs) in the axSpa clinic. A multi-pronged approach has been taken since March 2020 to increase ePROMs completion and improve their use. Clinicians invited each patient to join the electronic patient messaging system MyPathway (MP) and sent them a link after their appointment. QR codes were added to all rheumatology patient letters, encouraging patients to register. A pathway was set up that automatically sent a prompt to patients registered on MP to complete ePROMs prior to clinic appointments. Clinicians began logging into MP to view scores during appointments to provide patients with real-time feedback. Methods: A mixed methods approach was used to assess the uptake of ePROMs over time. We tracked MP registration and BASDAI completion rates as the key outcome measures. Results: The total number of axSpa patients seen in the specialised clinic who have registered with MP has increased from 56 (35.9%) in January 2019 to 200 (58.9%) in September 2022. 80% of patients completed more than one BASDAI in 2022, compared to 24% in 2019, as illustrated on a run chart. Patients can complete BASDAI forms sent to them in a previous month, therefore the completion rate some months exceeds 100%. Conclusion: In a dedicated focus group, patients reported that ePROMs were convenient, and provided a useful record to refer back to. There is concern that the importance of BASDAI and BASFI scores at an individual patient level is not clear. This may be rectified by more discussion with clinicians in appointments, to add nuance to these scores. There was also concern that sleep and other generic health measures are not covered in the BASDAI or BASFI. The EQ-5D generic ePROM is sent to all patients, but there seems to be little awareness of this in the axSpa clinic.

### A23 Design and delivery of remote electronic patient reported outcomes (ePROs) for personalisation of therapy in pulmonary arterial hypertension (PAH)

**Frances Varian[1,2], Charlotte Pearson[1,2], PHoenix Trial Investigators[1], Elisa Ferrer[3], Rothman Alex[1,2]**

**[1]University of Sheffield, Sheffield, United Kingdom. [2]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom. [3]Aparito, Wrexham, United Kingdom**

Background: Trends toward PROs as a secondary clinical endpoint in pulmonary arterial hypertension (PAH) are gaining momentum, particularly with validation of EmpPHasis-10 as an independent prognostic measure (1). Electronic patient-reported outcome (ePRO) systems for symptom monitoring in cancer patients have shown quality of life and survival benefits in controlled trials (2). Utility of remote health-related quality of life (HRQoL) measures for PAH trials have yet to be evaluated. Delivery of remote ePROs within a randomised clinical trial (RCT) has the capacity to empower patients to report factors impacting their life, without external influence, at their convenience. Aims: 1. To evaluate engagement with weekly remote ePROs through a digital mobile application. 2. Validation of the remote PHoenix-PRO medication compliance questionnaire. 3. Consideration of remote ePROs in evaluation of treatment efficacy within trial design. Method: PHoenix is a randomised $2 \times 2$ cross-over trial for approved drug and device therapies for PAH. Physiological data can be analysed remotely following implantation of two monitors; one in the pulmonary artery, and another under the skin over the heart which monitors heart rate and rhythm. A mobile platform was chosen to integrate specific questionnaires to evaluate ePRO. These include validated generalised anxiety and depression (GAD) and patient health screening questionnaires (PHQ), EMPHASIS-10, and questionnaires designed specifically to record medication side effects and medication compliance (Phoenix-PRO). ePROs are asked weekly with each medication change. Internal validity of Phoenix-PRO medication compliance is evaluated. A nested side effects study explores free text versus symptom selection. Responses can be evaluated in conjunction with daily physiological data from remote pulmonary artery pressure (PAP) and heart rate (HR) established through implanted monitors. Results: Results to be presented of early trial data with sample of weekly ePROs to evaluate (1) engagement with regular remote monitoring of ePROs (2) internal validity of Phoenix-PRO as a remote measure of medication compliance (3) addition of remote ePROs for personalisation of PAH treatments. Conclusions: This trial design is a new opportunity to evaluate ePROs and physiological measurements. This initial sample evaluates feasibility and utility of remote ePROs for personalisation of therapy for patients with PAH.

### A24 Realist and Social Return on Investment Evaluation of the use of Patient Reported Outcomes in Four Value Based Healthcare Programmes

**Jane Noyes[1], Leah McLaughlin[1], Carys Jones[1], May Bianchim[1], Ellie Crane[1], Barbara Neukirchinger[1], Adele Cahill[2], Dale Evans[2], Gareth Roberts[2]**

**[1]Bangor University, Bangor, United Kingdom. [2]Aneurin Bevan UHB, Cardiff, United Kingdom**

Background: There is growing recognition that in order to remain sustainable, the National Health Service must deliver the best patient outcomes within available resources. This focus on outcomes relative to cost is the basis of Value Based Healthcare (VBHC) and has led to interest in the use of patient reported outcomes (PROMs) to measure patient perspectives on the impact of a health condition on their lives. Aim: To understand what works about PROMs collection, for whom, in what contexts, and why in a VBHC context. We are also assessing the social value of integrating PROMs collection into routine care. Setting: Every Health Board in Wales is now required to collect PROMS as part of routine care. We are evaluating the VBHC programmes implemented in a lead Health Board. Methods: We are conducting a three stage mixed-methods study comprising of a realist evaluation integrated with Social Return on Investment analysis (SROI) across four VBHC programmes: Parkinson's disease, epilepsy, heart failure and cataract surgery. The study is organised in four worstreams. Workstream 1: Development of logic models, informed by a rapid review, documentary analysis, patient and public involvement (PPI), staff and key stakeholder engagement. Workstream 2: Realist evaluation building on multiple data sources from stages 1 and 3 to test and refine the programme theories that arise from the logic model development. Workstream 3: SROI analysis using interview data with patients, staff and carers, stakeholder and

PPI engagement, anonymised routinely collected data, and questionnaires to populate a model that will explore the social value generated by the implementation of PROMs. Findings across stages will be validated with key stakeholders. Results: We have developed the initial programme theories, evaluation framework and impact maps and are currently completing final data collection to test these theories. Initial findings show which aspects of the programmes are working, for whom and in what contexts. We have identified evidence of theory and implementation failure in some contexts as well as PROMS working as intended in other contexts. The SROI analysis is ongoing. Conclusions: Findings will lead onto actionable recommendations to further improve implementation at scale across Wales.

## Session 3C

### A25 Development of a patient-reported outcome measure for gastrointestinal recovery: The PRO-diGI study.

**Georgina Jones[1], Daniel Baker[2], Gabrielle Thorpe[3], Debby Hawkins[2], Sue Blackwell[4], Robert Arnott[5], Stephen Chapman[6], Deena Harji[7], Matt Lee[2]**

**[1]Leeds Beckett University, Leeds, United Kingdom. [2]Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, United Kingdom. [3]University of East Anglia, Norwich, United Kingdom. [4]Patient Expert, Sheffield, United Kingdom. [5]Patient Expert, Oxford, United Kingdom. [6]Leeds Teaching Hospitals, Leeds, United Kingdom. [7]Manchester University NHS Foundation Trust, Manchester, United Kingdom**

Background: Patients undergoing abdominal surgery often experience gastrointestinal dysfunction, which can be associated with substantial morbidity, mortality and healthcare costs. Recovery of gastrointestinal function following surgery is poorly measured. Since no patient-focused outcome measure exists, comparisons of interventions across studies are limited to a variable selection of clinician-reported outcomes. Consequently, interventions that may be highly effective to improve gastrointestinal function cannot be reliably detected, which impedes their implementation into clinical practice. Aim: To develop the first patient-reported outcome measure to assess gastrointestinal recovery following abdominal surgery. Methods: The PRO-diGI study consisted of three phases: Phase 1: Qualitative interviews of patients admitted to hospital for major gastrointestinal, gynaecological, or urological surgery, or with intestinal obstruction. Phase 2: Development of questionnaire and face validity testing with patients as in Phase 1, and expert clinicians internationally. Phase 3: The questionnaire was distributed in a survey to eligible patients from hospital sites across England. Participants completed the questionnaire as soon after abdominal surgery as they were able. Factor analysis and internal consistency reliability was performed to confirm the domain structure of the questionnaire and reduce the number of items. Results: Phase 1: 29 patients were interviewed and reported 26 key symptoms or experiences across 6 themes including impact on life, mental health, and general physical wellbeing, diet and appetite, abdominal symptoms, nausea and vomiting, and bowel function. Phase 2: 18 patients and 15 clinicians assessed a 44-item questionnaire, confirming good face validity. Phase 3: 297 participants were recruited. Factor analysis identified 21 questions across five domains relating to enjoyment of food, ability to eat and drink, bowel function, nausea and vomiting, and general wellbeing, for inclusion in the final tool. Cronbach's alpha exceeded 0.70 for all five domains. The final tool shows correlation with self-rating of gut function, GI-2, and operative approach. Conclusion: This study has developed a multi-dimensional patient reported outcome measure which can be used in studies of gastrointestinal recovery. Further psychometric testing to assess test–retest reliability, construct validity (known groups) and responsiveness is in progress.

### A26 A scoping review and semi-structured interviews to generate a list of outcome measures used to evaluate the clinical effectiveness of interventions in children with nocturnal enuresis.

**Nicole O'Connor[1,2], Eugenie Johnson[1,2], Sheila A Wallace[1,2], Fiona Pearson[2,3], Luke Vale[1,4]**

**[1]Cochrane Incontinence, Newcastle upon Tyne, United Kingdom. [2]Evidence Synthesis Group, Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom. [3]NIHR Innovation Observatory, Newcastle upon Tyne, United Kingdom. [4]Health Economics Group, Population Health Sciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom**

Background: To develop a list of enuresis outcome measures that have been reported in randomised trials and systematic reviews and identified as important by people with lived experience of enuresis. Listing enuresis outcome measures is the first methodological step in developing a core outcome set to ensure relevance, quality, and comparability across trials. Aims: To generate a list of candidate outcome measures. Methods: Following recommended COMET methodology a convergent parallel mixed methods research design was undertaken. This comprised a scoping review and semi-structured interviews. The scoping review searched nine electronic databases for outcome measures that have been reported in randomised controlled trials and systematic reviews of effectiveness published from 2017 to 2022 (search date: 11 May 2022).

Semi-structured interviews took place on-line (May–June 2022) with a convenience sample of six participants recruited by snowball technique utilising existing communication channels with patient organisations and professional bodies. Interviews were audio recorded, transcribed verbatim, and analysed using open abductive coding and the framework method. Findings were integrated into the five core areas of the COMET taxonomy along with the identified published outcome measures. Results: During the scoping review, 2210 records were screened at title and abstract stage, 207 records at full text with 140 records included. Due to high volume of identified literature a date restriction was placed and of the included studies, 87 were labelled 'awaiting classification' 15 as 'ongoing studies' and 38 articles underwent data extraction and synthesis. Two hundred and sixty-two codes were generated from the interview transcripts, these were arranged into themes and integrated with the scoping review findings into four of five core areas of the COMET taxonomy. Conclusions: This study relied on single screening, data extraction, qualitative data coding and analysis this increases chances of error during these processes comparative to their undertaking in duplicate. Frequency of wet nights is a priority outcome measure. However, emotional and social functioning are identified as being important to people with lived experience but are infrequently reported in randomised controlled trials and systematic reviews. Core outcome sets should consider their inclusion to increase their uptake in future clinical research ensuring its relevance to all stakeholders.

**Acknowledgements** Key funding and support were provided by the following organisation, the National Institute of Health Research (NIHR) Applied Research Collaborations (ARC) Yorkshire and

Humber. The conference was endorsed by the International Society of Quality of Life (ISOQOL) and supported by the NIHR Yorkshire and Humber Research Design Service, and sponsored by Oxford University Innovation Ltd and InSpired Health Outcomes.



**Funding** Young and Lane are funded by the National Institute for Health Research (NIHR) Applied Research Collaboration at the University of Sheffield. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department for Health and Social Care.