Health-related quality of life in pulmonary arterial hypertension trials: a review and conceptual map

Frances Varian¹,², Rebecca Burney¹,², Charlotte Pearson¹,², Ze Ming Goh¹,², Joseph Newman³,⁴, Gregg Rawlings⁵, Hamza Zafar¹,², David G Kiely², AA Roger Thompson¹,², Robin Condliffe², Mark Toshner³, Ciara McCormack⁶, Iain Armstrong²,⁸, Tessa Peasgood⁶, Jill Carlton⁶, Alex Rothman¹,²

1. Division of Clinical Medicine, School of Medicine and Population Health, (formerly Faculty of Medicine, Dentistry and Health), University of Sheffield, Beech hill Rd, Broomhall, Sheffield, S10 2RX, UK
2. Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, S10 2JF, UK
3. Royal Papworth Hospital, Papworth Rd, Trumpington, Cambridge, CB2 0AY, UK
4. Victor Phillip Dahdaleh Heart and Lung Research Institute, University of Cambridge, UK
5. Clinical and Applied Psychology Unit, Cathedral Court, 1 Vicar Lane, Sheffield, S1 2LT, University of Sheffield, UK
6. Department of Sport Science and Nutrition, Maynooth University, Maynooth, Co. Kildare, Ireland
7. Sheffield Centre for Health and Related Research (ScHARR), School of Medicine and Population Health, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA, UK
8. Pulmonary Hypertension Association United Kingdom, Sheffield, UK

Corresponding author: Alex Rothman
Address: Division of Clinical Medicine, School of Population and Health, Beech Hill Road, Sheffield, S10 2RX, United Kingdom
Email: a.rothman@sheffield.ac.uk

Target Journal: European Respiratory Journal
Total number of tables/figures: 5
Word count: 4655 (max 5000)
Type of contribution: systematic review
Key words: health-related quality of life, pulmonary arterial hypertension, outcome measures, psychometrics, quality of life measurement, systematic review

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Conflicting interests

Professor DG Kiely and Dr I Armstrong were involved in the derivation of EmPHasis-10 but remained independent in the risk of bias analysis for the COSMIN review. Professor Jill Carlton is a co-investigator for the UK EQ-5D-5L study team. Dr Tessa Peasgood is a member of EuroQol and involved in research development of the EuroQol Health and Wellbeing instrument. Other authors are not affiliated with PROMs evaluated in this review. The authors (FV/RB/CP/ZMG/JN) performing the data collection, selection and evaluation of review articles have no conflicts of interest.

Funding

Wellcome Trust Clinical Research Career Development Fellowship (AMKR: 206632/Z/17/Z), BHF Intermediate Fellowship (AART: FS/18/13/33281), MRC Confidence in Concepts (AMKR), Medtronic External Research Program Award (AMKR), MRC Experimental Medicine grant (AMKR/MT/ DK: MR/W026279/1), BHF Clinical Research Training Fellowship (HZ/AMKR: FS/CRTF/23/24465, MT/JN: FS/CRTF/22/24390). The research was carried out at the National Institute for Health and Care Research (NIHR) Sheffield and Cambridge cardiorespiratory Biomedical Research Centres. AR is grateful to Richard Hughes, whose generous philanthropic support has helped to make this work possible.

Disclosures

AART research funding: Heart Research UK, Janssen-Cilang Ltd, British Heart Foundation, and honoraria from Janssen-Cilad Ltd for lectures and education. AMKR: research funding: Wellcome Trust Clinical Research Career Development Fellowship (206632/Z/17/Z), Medical Research Council (UK) Experimental Medicine Award (MR/W026279/1), NIHR Biomedical Research Centre Sheffield, Contribution in kind: Medtronic, Abbott, Endotronix, Novartis, Janssen. Research support and consulting: NXT Biomedical, Endotronix, SoniVie, Neptune, Gradient. DGK has received personal funding from the NIHR Biomedical Research Centre Sheffield, research funding from Ferrer, GSK and Janssen and consulting and educational funding from Acceleron, Attivant, Ferrer, Gossamer, Janssen, MSD and United Therapeutics. FV has received educational funding from Janssen and is a Medical Research Council (UK) clinical fellow. JN Research funding: British Heart Foundation, education and travel funding Aparito Ltd, United Therapeutics, Pulmonary Vascular Research Institute. MT: Research funding: NIHR Biomedical Research Centre Cambridge, NIHR HTA. Personal support: GCK and Jansen. RC has received honoraria for speakers’ fees and conference travel from Janssen. All others: none.
Abstract

Introduction

Health-related quality of life (HRQoL) in pulmonary arterial hypertension (PAH) is valued as an outcome measure by patients, clinicians and regulators. Despite the incorporation of HRQoL in trials of PAH therapies, there is limited data on their suitability, accuracy and reliability.

Method

We report a systematic review following PRISMA guidelines (PROSPERO ID: CRD42024484021). Selection of PROMs included those powered to detect a minimal clinically important difference (MCID). Measurement properties were evaluated according to COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and graded by recommendation for use. An a priori framework was then used to develop a ratified conceptual model from patient interviews and surveys to map the content of PROMs to HRQoL.

Results

Screening of 896 records was performed after removal of duplicates. Of 43 trials with a HRQoL endpoint, 20 selected an instrument with a MCID. Of these, only 8 trials were adequately powered. Three different PROMs (EuroQoL-5D-5L, Short Form-36, Living with Pulmonary Hypertension Questionnaire (LPHQ)) were used. For COSMIN measurement property evaluation, 389 records were screened and 21 were included; EmPHasis-10 was also evaluated due to its inclusion in forthcoming trials. Using COSMIN criteria EmPHasis-10 and LPHQ can be recommended (Grade A) for use in clinical trials in PAH. However, SF-36 and EQ-5D-5L (Grade B) require further study. Conceptual mapping from 8045 patients showed disease-specific instruments uniquely capture self-identity and autonomy.

Conclusion

To improve evaluation of HRQoL outcomes, future PAH therapy trials require appropriate PROM selection, with adequate power, and consideration of conceptual mapping.
Graphical abstract: health-related quality of life outcomes in pulmonary arterial hypertension trials

COSMIN CONsensus-based standards for the Selection of health-Measurement INstruments, EQ-5D-5L EuroQol-5D-5L; HRQoL health-related quality of life; LPHQ Living with Pulmonary Hypertension Questionnaire, MCID minimal clinically important difference; PAH pulmonary arterial hypertension; PROM patient reported outcome measure, QALY quality adjusted life year, RCT randomised controlled trial, SF-36 36-item Short Form survey.

Plain language summary (242 of max 300)

Individuals living with pulmonary hypertension want to know which treatments improve their quality of life related to their health. We use questionnaires to capture the experiences of people living with pulmonary hypertension. Examples of this used in clinical practice are EmPHasis-10. We reviewed all the clinical trials in pulmonary hypertension to see which questionnaires were used to measure health-related quality of life. Some questionnaires may be better at capturing the experience of living with pulmonary hypertension than others. We found 20 clinical trials used a questionnaire that could detect a change in health-related quality of life in pulmonary hypertension. However, only 8 trials were designed to detect a significant treatment impact. We then evaluated these questionnaires against current best practice guidelines to ensure they are fit for purpose. EmPHasis-10 and the Living with Pulmonary Hypertension Questionnaire are preferred from the four evaluated in this study. The final part of this study was to look at what quality of life means for those living with pulmonary hypertension. Data from 8045 patients across the world was used to draft a health-related quality of life
framework. We then finalised this design with professionals and patients. This framework can be used in the future to help understand how the well a questionnaire captures things important to those with lived experience of pulmonary hypertension. This will help us to better understand treatments that improve quality of life for people living with pulmonary hypertension.

**Background**

Endpoints in randomised controlled trials (RCT) have traditionally focussed on physiological measures including functional markers such as 6-minute walk distance (6MWD).(1,2) However, approaches prioritising clinician-derived endpoints (3–5) can undervalue the patient voice. Integral to assessment of health-related quality of life (HRQoL), patient-reported outcome measures (PROMs) are an instrument developed to capture and quantify the experience of living with a health condition. Improvement in HRQoL is an important treatment goal for clinicians, regulators and patients, yet it is often not examined in clinical trials.(6–10) Furthermore, significant advancements in the diagnosis and treatment of PAH mean people are living longer, with a focus not only on length of life, but also quality.

Comparison of the cost-effectiveness of interventions is usually based on Quality Adjusted Life Years (QALYs). To allow such a calculation, PROMs used to describe and assess HRQoL also need a value set. In combination these are termed a preference-weighted measure (PWM). Value sets are based on the views or preferences of the public and/or patients and vary by country to reflect sociocultural differences.(11,12). A PWM scores each health state described by the PROM as a single value or ‘utility index’ on a scale, such that 1 represents full health, and zero represents death. A score below zero indicates a health state considered worse than being dead. The index score of a health state can be combined with time spent in that state to estimate QALYs. QALYs are an important outcome for regulatory and clinical decision-making and therefore dependent upon robust PROMs.(13)

There are many challenges in validating PROMs for accurate measurement of HRQoL and for use as PWM. Condition-specific measures may offer greater sensitivity to changes in HRQoL than generic PROMs however evidence is limited.(14) The condition-specific PROMs for PAH include CAMPHOR, EmPHasis-10, Living with Pulmonary Hypertension Questionnaire (LPHQ) and PAH-SYMptoms and imPACT (PAH-SYMPEACT).(15) Sensitivity to change must be interpreted with respect to being clinically meaningful. Multiple standards, including those of the US Food
and Drug Administration (FDA), specify that PROMs should have an established minimal clinically important difference (MCID). (16–22) A generic PROM (e.g. SF-36) may be used, providing the instrument has been validated in the population of interest to include a MCID. In addition, the choice of PROM should follow guidance developed using international Delphi approaches (18) and be evidence-based. (23) PROMs used as HRQoL outcome measures in PAH clinical trials have yet to undergo psychometric evaluation using the COConsensus-based standards for the Selection of health Measurement INstruments (COSMIN) guidance. (17,23) COSMIN guidance supports the identification of PROMs that can detect meaningful change within the health condition of interest, and aids decision-making for recommendations for use. HRQoL endpoints can be further enhanced by identifying HRQoL concepts captured by the PROM. (24) Developing a conceptual framework aids visualisation of important aspects of HRQoL for people living with PAH. (25,26)

Aims and Objectives

This is the first systematic review of PROMs for adults with PAH (15,27–29) to 1) evaluate MCIDs, and 2) compare measurement properties in accordance with COSMIN guidance; including evaluation of psychometric performance and grading recommendation for use. (19–21,30). To advance HRQoL outcomes in PAH, we undertake a literature review to develop a conceptual framework to inform relevant HRQoL constructs from the patient perspective.

Methods

Systematic searches

The protocol for the systematic review of PAH RCTs was registered on PROSPERO (CRD42024484021). The additional COSMIN evaluation is not independently registered. Methodology adhered to the Cochrane Handbook of Systematic Reviews of Interventions, and COSMIN guidance. (31) Reporting structure followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (see online supplementary figure E1) and PRISMA-COSMIN outcome measurement instruments (online supplementary figure E4 and E9). MEDLINE (1980 to December 2023) and Cochrane Library (2002 to December 2023) were searched for RCTs evaluating the effectiveness of any intervention for PAH designed to improve a clinical outcome measure as determined by the FDA endpoints “feel, function or survive”. Inclusion and exclusion criteria are registered on PROSPERO. After removal of
duplicates, one author (FV) screened the titles and abstracts of articles for relevance before reviewing the full text for eligibility. Where there was uncertainty about the relevance of an article, a second author (JN) reviewed the title and abstract/main text. A third author was available to adjudicate discrepancies. This process was repeated for the second PRISMA-COSMIN search and reporting structure followed (supplementary figure E3 and E4). PRISMA-COSMIN studies included pulmonary hypertension (PH) comprising group 1 and group 4 patients to maximise psychometric property evaluation. Forward and backward searches were performed on eligible articles for both searches, and citation searching performed on systematic reviews identified.

Data Extraction

Five authors (FV, RB, CP, ZMG, JN) extracted information independently from all RCTs using a pre-determined template. This included sample and trial characteristics, primary and secondary outcome measures and results, and details of HRQoL PROMs used. Primary and secondary endpoints were categorised into measures of how a patient ‘feels’, ‘functions’ or ‘survives’ as per FDA recommendations for clinical trial endpoints.(32) HRQoL endpoints were classified as ‘feel’ (e.g. EQ-5D-5L), ‘function’ (e.g. any form of exercise parameter and World Health Organisation Functional Class (WHO FC)) and ‘survive’ (inclusive of clinical worsening events and mortality and not restricted to survival analyses).

Risk of Bias and Strength of Evidence

Two authors (ZMG, RB) assessed the systematic review risk of bias (RoB) using the Cochrane RoB2 Toolkit, and strength of evidence according to the Grading of Recommendations Assessment Development Evaluation (GRADE) criteria. The COSMIN RoB checklist was completed by two other authors (FV & CP). (30,31) Any disagreements were discussed until consensus was reached. Further bias assessment (e.g. Egger’s) was not appropriate due to the large heterogeneity of study interventions and low number of PROM instruments limiting interpretation. A summary of the overall strength of PROM recommendation is made by grading into one of three categories: (A) PROM can be trusted for use with sufficient evidence of psychometric properties; (B) PROM has potential to be recommended for use but insufficient to meet A or C categories; (C) PROMs with high-quality evidence that a measurement property is insufficient and therefore should not be recommended for use.(17,33,34). A description of terms used in the COSMIN evaluation is available in supplementary table E5.
It is recommended that the MCID is considered for sample size calculations for comparing HRQoL outcomes. To determine whether trials were sufficiently powered for the chosen PROM, the MCID for each instrument was obtained (supplementary table E1). If data were unavailable specifically for PAH, a MCID was searched for respiratory conditions (1) and heart failure (2) to maximise PROM inclusion. A second author (JN) confirmed absence of MCID using search criteria in PubMed and the PROMs Data Archive: [instrument_name] AND [MCID OR MID OR Minimal].

Computation of MCID mean and standard deviation in GPower v3.1 was used to estimate significance, one-tailed test. Trials insufficiently powered to detect a meaningful change in HRQoL as defined by the MCID calculations were excluded from analysis. The MCID for six-minute walk distance (6MWD) was defined at a threshold of 33m, as correlated with the physical functioning item of the Short-Form 36 (SF-36). Meta-analysis was undertaken per therapy and by PROM and was calculated in SPSSv28.1.

Scoping review for conceptual framework and patient and public involvement and engagement (PPIE)

A scoping literature review was conducted independently by two authors (FV & RB) to map HRQoL concepts on to PROMs in PAH. An a priori framework from generic health and wellbeing instruments and HRQoL model was used to inform the PAH HRQoL conceptual framework. Published studies using primary and secondary analytical methods and grey literature, such as surveys asked by Pulmonary Hypertension Associations, were included (supplementary table E6). To corroborate previous synthesis from a qualitative systematic review, a random selection of four publications was appraised in detail for blinded thematic content agreement prior to evaluation of the synthesised systematic review. Themes were not duplicated in final synthesis. Subthemes were extracted and seven a priori themes considered; final subthemes were weighted from most to least commonly reported. Key professional stakeholders from centres in the UK and Ireland then ratified the framework followed by PPIE obtained from representatives from Pulmonary Hypertension Association UK (PHA UK) and patient volunteers registered within Sheffield’s local PPIE PAH network. The form asked to “consider what quality of life means” before reviewing the conceptual framework for anything missing. Participation was entirely voluntary without
reimbursement. PROMs from the COSMIN review were then mapped to the conceptual
framework to visualise instrument scope.

Results

Systematic review of valid HRQoL endpoints and their psychometric properties

The systematic search identified 896 unique records. After screening, 178 remained for full-text
review. Overall, 90 potentially eligible RCTs were identified with a clinical endpoint. 73% (n=66)
included pharmacological interventions which were categorised into ‘feel’, ‘function’ or ‘survive’ and these were mapped (supplementary figure E2). This demonstrated a
predominance of functional endpoints, with an increasing trend toward emphasis on survival.
The scope of HRQoL or ‘feel’ was limited to secondary endpoints.

In total, 43 RCTs with a HRQoL endpoint were considered for final inclusion (supplementary
table E2). All studies showed some risk of bias (supplementary table E3). The strength of all
studies with a HRQoL endpoint was ‘moderate’ (supplementary table E4). There was no
evidence of patient involvement in selection of PROM in any PAH RCT (supplementary table
E2).(18)

Regarding condition-specific PROMs, a valid MCID for EmPHasis-10, LPHQ and CAMPHOR was
found, but not for PAH-SYMPACT (55). All available MCID values and methods of derivation are
included in supplementary table E1. Figure 1 shows that 20 of the 43 RCTs with a HRQoL
endpoint selected an instrument with an MCID for PAH. Of these, only 8 trials met the full
inclusion criteria with adequate power to detect a meaningful change in HRQoL (Table 1).(56–
99) PROMs meeting final inclusion (Table 1) were SF-36, EQ-5D-5L, LPHQ and Minnesota Living
with Heart Failure (MLWHF). A utility index is available for EQ-5D-5L for the PAH population, but
not a specific MCID. A sample size was therefore conservatively estimated from a comparable
6MWD of 35m from an interstitial lung disease population to maximise inclusivity.(100) All trials
powered for HRQoL selected 6MWD as their primary endpoint (Table 1). Bosentan (EARLY)(86)),
IV epoprostenol (PACES)(88), and inhaled treprostinil (TRIUMPH-I)(85) did not meet their
primary endpoint and showed no improvement in the SF-36 physical functioning domain (Table
1). Significant improvements in 6MWD for ambrisentan (ARIES2)(101) and exercise (EU-TRAIN-
01)(102) were reported, however only the MCID was met for the role-physical domain of SF-36
in EU-TRAIN-01(102).
PATENT-1(103) and PATENT-2(104) (riociguat vs placebo) were the only RCTs available for combined meta-analysis (Figure 2). Two PROMs (EQ-5D-5L and LPHQ) were completed by the same patients. EQ-5D-5L overall appeared less sensitive to changes in HRQoL (Cohen’s d ES=0.24, SE=0.08, p<0.001) compared to LPHQ (ES=-0.48, SE=0.11, p<0.001) (Figure 2).

Although an exploratory endpoint, it is unclear which country-specific value sets was used for EQ-5D-5L as recommended by reporting practices.(12,105–108) All trials reported statistical significance (p<0.05) between arms rather than MCID, requiring additional interpretation as a valued endpoint. The open-label extension study supported a sustained improvement in HRQoL with riociguat compared to placebo as measured by LPHQ.(104) However this was not true for all dose regimes. There was no significant change in HRQoL for the group receiving 2.5mg dose of riociguat as measured by EQ-5D-5L (Figure 2)(109). EQ-5D-5L was sensitive to change in the 1.5mg subgroup, which had a statistically higher proportion of patients in WHO FC III compared to II (Fisher’s exact p<0.05) and this was also clinically meaningful (MCID is +0.017) at 12 months (0.13±0.24) compared to the 2.5mg group (+0.06±0.24).(104)

For COSMIN evaluation, 369 eligible articles were screened for psychometric properties with additional citation searching (n=20) from 3 systematic reviews. EmPHasis-10 was considered relevant for inclusion as recruitment for two RCTs is underway and has an estimated MCID.(110,111) 21 studies demonstrated measurement properties (supplementary figure E4).(15,35,42,112–129) MLWHF for Pulmonary Hypertension (MLWHF-PH)(113) was later renamed LPHQ and therefore these instruments are pooled for evaluation.(15,35,112,113) SF-36 is available as either a PWM or PROM. The MCID for SF-36 is not specifically reported for mental health, pain, general health and role-emotional domains but is available for physical functioning, role-physical, energy-fatigue and social functioning.(115) EQ-5D-5L is also PWM used in PAH,(130) and the instrument compared in derivation of LPHQ.(35) However we found no studies validating psychometric properties of EQ-5D-5L in adults with PH, limiting further review. All RCTs identified from the initial systematic review were included within the COSMIN review as a measure of PROM ‘responsiveness’.(17,30,33,34)

PROM suitability for the PAH population in accordance with COSMIN guidance
PROM design includes how comprehensively the instrument covers HRQoL for the population of interest, also known as content validity. A description of terms structuring COSMIN analysis is available in supplementary table E5. The characteristics and measurement properties of the PROMs selected are outlined in Table 2 with full details in supplementary tables E7 and E8. For adequate content validity, all instruments require post-hoc cognitive interviewing with patients and experts. LPHQ was the only instrument to perform post-hoc saturation analysis to confirm the relevance of the final instrument from the patient’s perspective. No cognitive interviewing for content validation in the PAH population has been performed for SF-36 or EQ-5D-5L.

Statistical analysis must be performed in an appropriate sample size to evaluate the internal structure of the PROM. Appropriate statistical analysis for the ‘model fit’ must also be reported. (131) Internal structure comprises structural validity, determined by appropriate statistical analysis, internal consistency, defined by response agreement between items (e.g. Cronbach’s alpha ≥0.7), and measurement invariance, which determines how well the PROM performs across different groups (i.e. potential response variations e.g. age, gender, BMI).

Statistical analysis validating the psychometric structure of SF-36 in PH was not reported and was incomplete for LPHQ and EmPHasis-10. Internal consistency is well evidenced for LPHQ, EmPHasis-10 and SF-36. This is not relevant for EQ-5D-5L as items are not inter-related, with only 1 item per domain (Table 2). (132) No studies adequately considered measurement invariance. One study suggested EmPHasis-10 may vary by demographic and clinical characteristics, (119) but there is an absence of further testing. While multiple translations are available for EQ-5D-5L and SF-36, there is insufficient psychometric validation (cross-cultural validity) (133) to support use of these languages in the PAH population (Table 2, online supplementary table E8). EmPHasis-10 demonstrates strong linguistic validation. (121,123,127,134) Developed in the UK and Ireland, EmPHasis-10 is the only PROM validated cross-culturally in US, China, Japan, Italy and Turkey. (118,119,121,123,126,127) LPHQ is available in English only, though was derived in the US, France and Germany.

PROM structure considers how the questionnaire should be scored and interpreted. Each score should be structurally validated using psychometric statistical analyses to check accuracy. LPHQ has a multifactor structure with physical, emotional and total scores (35), as is SF-36 with eight domains and physical and mental component scores, whereas EmPHasis-10 was derived as a unidimensional structure (i.e. a single, total score) (118). Reporting of the structural validity of these PROMs does not meet current requirements (supplementary table E8). (23) However, a
recent analysis of EmPHasis-10 structured into three scoring components (breathlessness
(three items), fatigue (three items) and independence (four items)) does meet COSMIN
requirements.(125) Further evaluation is required to consider the clinical relevance of
interpreting EmPHasis-10 in this way.

Test-retest reliability is essential for defining the natural score variation of the PROM during a
period of stability. If the mean variation exceeds the MCID, then the PROM becomes invalid.

Test-retest reliability is indeterminate for LPHQ and SF-36, with limited evaluation of the
smallest detectable change or limits of agreement (supplementary table E8). Only two SF-36
domains (physical functioning and general health) meet adequate test-retest reliability in PH,
with wide confidence intervals and standard errors of measurement in others raising
concerns.(114,115)

Hypothesis testing comprises the final COSMIN analyses to evaluate 1) how well the PROM
correlates with other instruments (convergent validity), 2) discriminates subgroups e.g. WHO
FC (construct validity) and 3) responds to intervention (responsiveness). It is the most widely
tested psychometric property (supplementary table E8). All instruments show correlation with
others. Criterion validity additionally assesses sensitivity and specificity of the instrument;
however, it is challenging to achieve without a ‘gold standard’ measure. Performance of
instruments across PAH subgroups requires improvement for SF-36 and EQ-5D-5L. The SF-36
physical component score correlates with 6MWD (r=0.62, p<0.001);(42) however, other SF-36
domains and component scores show no or inconsistent relationship with WHO FC and
6MWD.(42,112,115,129) LPHQ appears to correlate well with WHO FC (r=0.61)(128) but
response to changes in WHO FC requires further evaluation.(35,112,128) EmPHasis-10 has
been shown to accurately discriminate WHO FC, and has good correlation with 6MWD, however
treatment responsiveness lags behind, with much-anticipated RCTs
underway.(118,119,122,124) Snapshot haemodynamics have yet to show strong correlation
with any PROM.(42,112,113,117–119,123).

Summary COSMIN instrument recommendations are grade A for LPHQ and EmPHasis-10 and
grade B for SF-36 and EQ-5D-5L. No PROMs received a grade C recommendation. However the
overall quality of evidence for LPHQ and Emphasis-10 is low, and for SF-36, very low (Table 2).

Mapping the HRQoL conceptual framework
Improving HRQoL matters to people living with pulmonary hypertension. Surveys report this as the most important treatment focus (52-83%) over other outcomes such as life expectancy (33-75%), n=1196, UK, Canada,(10,46) HRQoL concepts of interest may vary between clinical and trial applications,(21,30,44,135) however, recognising their relationship to PROMs is key for appropriate instrument selection. A conceptual framework developed from the Wilson and Cleary HRQoL concepts of interest may vary between clinical and trial applications, however, recognising their relationship to PROMs is key for appropriate instrument selection. A conceptual framework developed from the Wilson and Cleary and subsequent models(25,26,45) was inductively modified to reflect concepts of HRQoL. These subthemes (e.g. ‘stigma’) were identified from combined questionnaires and surveys of 8045 patients from around the world.(14,119,136) Demographics (where available) were reflective of the disease prevalence with a female predominance (79%, n=4700). Average age of patients was 55 years (range 24-80 years) and 88% self-reported to be Caucasian (supplementary table E6).

Figure 3A summarises the conceptual framework, with six themes and 25 subthemes identified. The framework was ratified by 6 PH consultants, 2 PH clinical fellows, 1 clinical nurse specialist, 1 physiotherapist, 1 clinical psychologist and PPIE obtained from two PHA UK representatives and five patients with relevant demographic representation. One patient commented specifically that they “never really thought about quality of life before their diagnosis” and they “think mental health is a big thing and this is affected differently and sometimes unexpectedly each day”. No additional themes or subthemes were identified.

The most frequently reported concepts in Figure 3A are presented in bold, with most-to-least common left-to-right and those overlapping representing similar weighting. The most reported impacts were activity, sadness/depression, self-worth, sense of loss, treatment- and financial burden. Cultural variation was evident for this latter subtheme and more commonly discussed in surveys and interviews of those living in Canada, USA and China compared the UK and Europe.

PROMs were then mapped onto the conceptual framework (Figure 3B). No single PROM covers all subthemes directly. EmPHasis-10 and LPHQ cover all main themes. Two commonly reported themes, self-identity and autonomy, are not specifically captured by EQ-5D-5L or SF-36. In addition, EQ-5D-5L does not capture impact on relationships. LPHQ is the only PROM to directly ask about treatment burden by including items on side effects but may also include items that are less impactful in this patient group (e.g. diet).
Discussion

This is the first systematic review to evaluate meaningful changes in HRQoL in RCTs in patients with PAH. Based on rigorous methodology using COSMIN guidance both EmPHasis-10 and LPHQ receive a grade A recommendation for use, whereas SF-36 and EQ-5D-5L receive a grade B recommendation. Of these PROMs EmPHasis-10 provides the broadest scope internationally and is validated in three continents. Whilst SF-36 is the most frequently used PROM in PAH RCTs to date, only 2 of 8 SF-36 domains meet current psychometric guidance. Currently no PROMs used in PAH RCTs are adequate for PAH-QALY calculation. To aid future work, we developed a conceptual framework which allows visualisation of what PROMs measure to capture aspects of HRQoL important to people living with PAH. Six themes and 25 subthemes were identified by researchers and ratified in the conceptual framework. Whereas both EmPHasis-10 and LPHQ likely capture all major themes, two major themes (self-identity and autonomy) are unlikely to be captured by SF-36 and EQ-5D-5L. Further study mapping the PROMs to this conceptual framework from the patients’ perspective is required. Complementary psychometric approaches will aid future selection of the most appropriate PROMs to measure HRQoL outcomes in clinical trials.

A PROM with a MCID is a meaningful trial endpoint.

PROMs should be resilient to the day-to-day variability in HRQoL. There will be a natural change in score, without a significant change in HRQoL and this may vary depending on disease severity. Meaningful change in HRQoL therefore may not equate to statistical difference. (137) The MCID for PAH PROMs has traditionally been anchored using the SF-36 physical functioning domain to changes in 6MWD. CAMPHOR is the only PROM to include PAH patient opinion in derivation of a MCID. (122,138,139) As illustrated by the conceptual framework, 6MWD alone is unlikely to adequately benchmark all aspects of change in HRQoL. (139,140) There is further inaccuracy in over-simplifying differences based on average distributions. (137,141) Multiple MCIDs should ideally be anchored over many individual timepoints to improve sensitivity. (137,139,140) Other factors influencing MCID include direction of change (improvement or deterioration) and individual baseline value. Neither of which may be synonymous with predicted disease outcome. (141) While highly valuable for trial endpoints, MCIDs evaluated should be interpreted with caution, and within the context of measurement error. (16,137,139,141) Measurement error includes PROM scores undertaken during stable
conditions (test-retest reliability). SF-36 has shown the variable performance in this regard. As the most widely adopted PROM in PAH RCTs, five of the eight domains have a valid MCID however only two (physical functioning and general health) meet adequate test-retest reliability.

*Psychometric measurement properties validate HRQoL outcomes*

HRQoL is a multifactorial construct with diurnal, daily and lifelong variability. Perception varies across the patient’s lifespan. Changes in values and priorities (response shift) depends upon ‘pre-diagnosis’, ‘transitioning through diagnosis’ and ‘duration living with PH’. The latter group reportedly face challenges with recognising disease progression and monitoring the condition.(51,142) Registry data shows consistent performance of EmPHasis-10 in patients with recent diagnoses (<6 months) but other time points are lacking.(119) For consistency, a PROM must perform regardless of ‘time since diagnosis’. This is known as response shift and has not been studied in PROMs for people living with PAH to date.

Further complexity is introduced by variations in HRQoL perceptions with age, gender, and disease severity.(54) Age and gender have been shown to influence PROMs.(135,143) These factors require further assessment in the PAH population.(118,136,143,144) Perceptions and response to limitations in activity also vary with individual coping strategies and personality types.(142) Responses may therefore differ depending on the choice of PROM. No PROMs used in PAH trials have specifically addressed variations in activity perceptions in longitudinal subgroup analyses. Understanding PROM performance also requires consideration of PROMs across subgroups (e.g. WHO FC)(137), known as measurement invariance. As shown by the meta-analysis, EQ-5D-5L may be less responsive to changes in WHO FC II compared to WHO FC III, potentially underestimating the HRQoL treatment benefit in this subgroup. Similar comparison was shown with EQ-5D-5L and CAMPHOR.(145) While exemplifying the importance PROM selection, combining PROMs in a trial setting offers useful comparison of responsiveness.

Development of the conceptual framework helps recognise important HRQoL captured by PROMs. All PROMs capture limitations in activities, however two major themes identified (self-identity and autonomy) are unlikely to be captured by SF-36 and EQ-5D-5L (Figure 3B). While LPHQ has received critique for poor symptom saturation,(27,35) the conceptual framework
does not support omitted symptoms of ‘palpitations’ or ‘problems with limbs’ as impactful. (6–9,35,46–52) Moreover, while some symptoms may be less relevant (e.g. diet/appetite), ‘time in hospital’ and ‘side effects’ are uniquely captured. LPHQ is also the only instrument to consider financial impact, which may have cultural relevance. (7,9,47,49) It is unclear whether PROMs adequately capture treatment burden (a key subtheme) in PAH, or whether this cross-loads with other concepts. Future cognitive interviewing should consider utilising the conceptual framework to elicit patient interpretation of PROM questions, in addition to modelling perceptions across the disease course. Mapping the PROM questions to framework will also establish likely relationships to concepts, helping to solidify PROM content validity from the patients’ perspective.

How do we advance HRQoL endpoints in PAH?

PROMs offer a descriptor for the patient voice, and this should be their primary purpose. Delivering a valid HRQoL trial endpoint requires appropriate PROM selection with a patient-centred MCID, and prioritisation of PPIE preferences which are reported in line with recommendations. (18) Greater consistency in PROM selection will improve knowledge of therapeutic outcomes according to lived patient experience. As a minimum, PAH clinical trials should select PROMs with grade A recommendation for use. PROMs further offer health economic evaluation to support regulatory decision-making. Following COSMIN review, neither generic PWM (EQ-5D-5L and SF-36) can be recommended, and therefore a condition-specific PWM with strong psychometric properties is preferentially considered. (145,146) CAMPHOR (147) is currently the only condition-specific PWM with a value set however, this is underutilised in RCTs and yet to undergo COSMIN evaluation. Future development of PWMs in PAH should focus on either improving PROMs with a B grade recommendation and/or developing a value set for those with a grade A recommendation. This will support robust evaluation for QALY outcomes.

Strengths and Limitations

Our systematic review of recent publications was designed with rigour, using multiple reviewers, a minimum of dual coders, and triangulation to enhance quality. Nevertheless, data
informing the conceptual framework was not analysed at source and therefore may be subject to bias. However, following UK PPIE opinion, there were no additional concepts added to the framework and based on the authors experience in international studies in PAH, we consider the framework to be relevant for other countries. As with adaptation of PROMs cross-culturally, future research is required to ensure individual concepts are applicable to the chosen area. This process could offer further understanding of cultural differences in people living with PAH.

Analysis of instrument power was based on MCID; although, as discussed, this may be inadequate, potentially over- or under-estimating the RCTs included. Furthermore, these estimates are calculated for between-group rather than within-patient differences and insufficient to base regulatory decisions.\(^{(16)}\) However, this is currently the only available measurement criteria for estimating sufficient responsiveness, and useful for calculating study size.\(^{(17,33,34)}\) CAMPHOR has an MCID but did not meet inclusion due to insufficiently powered historical or forthcoming RCTs. As this is currently the only PAH-specific PWM,\(^{(145,147)}\) independent COSMIN analysis is warranted. Finally, it is recognised that all PROMs considered in this analysis were developed prior to COSMIN guideline recommendations, and therefore some of the methodological concerns may be overstated due to missing publication details rather than instrument flaws. Despite these challenges and low quality of evidence, two instruments still achieved a grade A recommendation, showing promise for future HRQoL endpoints.

**Conclusion**

Global use of suitable PROMs in PAH occurred in 20 of 43 RCTs. Interpretation of HRQoL requires a MCID, yet only 8 trials were adequately powered to detect a meaningful change. All MCIDs require further validation, taking into consideration directionality and disease severity. Language availability is not necessarily concordant with cultural validity, and this should be considered in international and multi-centre RCT settings. LPHQ outperforms EmPHasis-10 regarding responsiveness whereas EmPHasis-10 demonstrates the strongest reliability and cross-cultural validity, however both can be recommended for use. SF-36 and EQ-5D-5L should be used with caution until further examination in people living with PAH. The conceptual framework, ratified without iterations, shows LPHQ and EmPHasis-10 are more likely to capture autonomy and self-identify. This should be developed further by mapping PROM items to the framework from the patient perspective. Concurrent cognitive interviewing is required for all
PROMs to strengthen content validity. HRQoL outcomes should focus on appropriate PROM selection, powered for a valid MCID, with the aim of continued psychometric development and health economic analyses.
<table>
<thead>
<tr>
<th>Study</th>
<th>N=</th>
<th>Demographics (Age (y), Female (%))</th>
<th>Intervention</th>
<th>PROM†</th>
<th>Additional Endpoints</th>
<th>Primary outcome met?</th>
</tr>
</thead>
<tbody>
<tr>
<td>EU-TRAIN-01</td>
<td>129</td>
<td>55±12.7 77.6% 52.3±12.4 69%</td>
<td>Exercise training</td>
<td>SF-36 (all)</td>
<td>Primary 6MWD Secondary WHO FC, CPET</td>
<td>Yes</td>
</tr>
<tr>
<td>(2021)(102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PATENT-2</td>
<td>396</td>
<td>49±16 80% 49±16 80%</td>
<td>Riociguat</td>
<td>LPHQ</td>
<td>Primary 6MWD Secondary CWE, NTProBNP, WHO FC, haemodynamics</td>
<td>Yes</td>
</tr>
<tr>
<td>(2015)</td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D-5L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(109,148)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PATENT-1</td>
<td>443</td>
<td>51±17 78% 51±17 79%</td>
<td>Riociguat</td>
<td>LPHQ</td>
<td>Primary 6MWD Secondary CWE, NTProBNP, WHO FC, haemodynamics</td>
<td>Yes</td>
</tr>
<tr>
<td>(2013)(149)</td>
<td></td>
<td></td>
<td></td>
<td>EQ-5D-5L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIUMPH-I</td>
<td>235</td>
<td>52(18-75) 82% 55(20-75) 81%</td>
<td>Inhaled Treprostinil</td>
<td>MLWHF-PH</td>
<td>Primary 6MWD</td>
<td>No</td>
</tr>
<tr>
<td>(2010)(150)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EARLY(2008)</td>
<td>185</td>
<td>44±17 63% 45±18 76%</td>
<td>Bosentan</td>
<td>SF-36 (all)</td>
<td>Primary 6MWD, PVR Secondary CWE, NTProBNP, haemodynamics</td>
<td>No</td>
</tr>
<tr>
<td>(151)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIES2</td>
<td>394</td>
<td>51±14 68% 51±15 78%</td>
<td>Ambrisentan</td>
<td>SF-36 (physical)</td>
<td>Primary 6MWD Secondary CWE</td>
<td>Yes</td>
</tr>
<tr>
<td>(2008)(101)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACES(2008)</td>
<td>267</td>
<td>48±13 77% 48±13 82%</td>
<td>Sildenafil + IV Epoprostenol</td>
<td>SF-36 (all)</td>
<td>Primary 6MWD</td>
<td>No</td>
</tr>
<tr>
<td>(152)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AIR</td>
<td>203</td>
<td>53±12 68% 51±13 68%</td>
<td>Inhaled Iloprost</td>
<td>EQ-5D and EQ-VAS</td>
<td>Primary 6MWD Secondary NYHA, PVR, CWE</td>
<td>Yes</td>
</tr>
<tr>
<td>(2002)(153)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Characteristics of studies powered for HRQoL

6MWD six-minute walk distance, CPET cardiopulmonary exercise test, CWE clinical worsening events, EQ (EuroQoL)-5D-5L, EQ-VAS EuroQoL visual analogue scale, HRQoL health-related quality of life, NTProBNP N-terminal Pro-Brain Natriuretic Peptide, NYHA New York Heart Failure Association functional score, PROM patient-reported outcome measure, PVR pulmonary vascular resistance, WHO FC World Health Organisation Functional Class

†All secondary endpoints
Figure 1: HRQoL Instruments in PAH RCTs from systematic review categorised by ability to distinguish meaningful change in the PAH population.

20 of 43 trials selected an instrument with a MCID for HRQoL. 56 total instruments are included as 13 trials included more than one instrument, see supplementary table E2 for full details. No trials reported results in the context of MCID. Where available, statistical significance was reported as p<0.05.

BDI Beck’s depression inventory, CAMPHOR Cambridge Pulmonary Hypertension Outcome Review, E10 EmPHasis-10, EuroQol(EQ)-5D-5L, DFI dyspnoea fatigue index, FSS fatigue severity score, HADS hospital anxiety and depression questionnaire, HAP human activity profile, IPAQ International Physical Activity Questionnaire, KCCQ Kansas City Cardiomyopathy Questionnaire, LPHQ Living with Pulmonary Hypertension Questionnaire, MCID minimal clinically important difference. MLWHF Minnesota Living with Heart Failure, NHP Nottingham Health Profile, PAH-SYMPACT pulmonary arterial hypertension symptoms and impact questionnaire, PGA patient global assessment, SF-36 Short-Form-36, SGA subject global assessment.
Figure 2: Meta-analysis of HRQoL outcomes for riociguat

PROM instruments LPHQ (A) and EuroQoL (EQ)-5D-5L (B). 1.5mg dose in PATENT-1 for LPHQ was excluded as subgroup insufficiently powered. HRQoL health-related quality of life, LPHQ Living with Pulmonary Hypertension. Utility index score was not reported with EQ-5D-5L analysis. PROMs delivered at start and week 12 for PATENT-1, and every 2 weeks up to week 8 for PATENT-2 follow-on study. 12-month follow-up data for EQ-5D-5L from PATENT-2 not included. No imputation reported of missing data. 2.5mg riociguat 2013,(107) n= 254 (WHO FC III, n= 140 (55%) Vs WHO FC II, n = 108 (43%)** p>0.05; riociguat 1.5mg 2013, n =63 (WHO FC III, n= 39 (62%) Vs WHO FC II, n = 19 (30%)***p<0.0001); 2.5mg riociguat 2015,(104) n= 231 (WHO FC III, n= 127 (55%) Vs WHO FC II, n=97 (42%)** p>0.05; 1.5mg riociguat 2015, n = 56 (WHO FC III n= 35(63%) Vs WHO FC II, n =17 (30%)** p<0.005 all Fisher’s exact test.
Figure 3: (A) Conceptual framework for PAH HRQoL and (B) scope of PROMs used in PAH RCTs mapped onto the conceptual framework by professionals.

[A] Framework of patient-reported themes (n=6) and subthemes (n=25) identified by two independent reviewers on impact of pulmonary hypertension (majority group 1 PAH) on HRQoL [A]. Directly reported concepts are in bold (n= 8045 from supplementary table E6). Concepts may indirectly cross subthemes (cross loading), for example, treatment burden may impact on the EQ-5D-5L item ‘pain/discomfort’ as a reflection of treatment side effects however this is not included in analysis. Emphasis-10 (oval) is included as two RCTs are currently recruiting,(110) LPHQ is combined with MLWHF as instruments are identical. [B] Professional conceptual mapping of PROMs to HRQoL framework (6 PH consultants, 2 PH clinical fellows, 1 nurse specialist, 1 clinical psychologist, 1 physiotherapist). Emphasis-10 and LPHQ cover all main themes. Further work with patient perspectives required. LPHQ Living with Pulmonary Hypertension Questionnaire, MLWHF Minnesota Living with Heart Failure, SOB shortness of breath, SF-36 Short-Form 36.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting</td>
<td>Clinic &amp; Trial</td>
<td>Clinical</td>
<td>Clinical</td>
<td>NA</td>
</tr>
<tr>
<td>Completion</td>
<td>5-10 min</td>
<td>3-4 min</td>
<td>5-10 min</td>
<td>3-4 min</td>
</tr>
<tr>
<td>Number of Items / scales</td>
<td>21 Likert</td>
<td>10 Likert</td>
<td>36 Likert/ discrete</td>
<td>5 Likert</td>
</tr>
<tr>
<td>Recall period</td>
<td>One week</td>
<td>Recent experience</td>
<td>Varies with item</td>
<td>Today</td>
</tr>
<tr>
<td>Derivation PAH</td>
<td>US, Germany, France</td>
<td>UK &amp; Ireland</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Translations</td>
<td>English</td>
<td>&gt;20(134)</td>
<td>193</td>
<td>208</td>
</tr>
<tr>
<td>Cost***</td>
<td>Free(157)</td>
<td>Free(158)</td>
<td>Free(159)</td>
<td>Free(160)</td>
</tr>
<tr>
<td>Construct(s)</td>
<td>Total, Physical &amp; Emotional summary scores</td>
<td>Total score</td>
<td>8 Domains*, Physical &amp; Mental scores</td>
<td>Score indexed for population health state</td>
</tr>
<tr>
<td>Scoring / (Best HRQoL)</td>
<td>0-105 (0)</td>
<td>0-50 (0)</td>
<td>Varies</td>
<td>0-1 (1)</td>
</tr>
<tr>
<td>Quality of Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Internal Structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-retest</td>
<td>?</td>
<td>+</td>
<td>?</td>
<td>NA</td>
</tr>
<tr>
<td>Measurement error</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>NA</td>
</tr>
<tr>
<td>Criterion validity</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Other instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Response to intervention</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>Subgroups</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>?</td>
</tr>
<tr>
<td>Quality of Evidence Summary</td>
<td>Low</td>
<td>Low</td>
<td>Very Low</td>
<td>NA</td>
</tr>
<tr>
<td>Recommendation (A to C)</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
</tbody>
</table>

Table 2: Summary of PROM characteristics, measurement properties and evidence quality
sufficient (+), insufficient (−), or indeterminate (?), (NA) not available. Summary of evidence quality based on a modified GRADE approach. Properties with moderate to high evidence quality are shaded grey. Recommendations are made by three categories (A) PROM can be trusted for use with evidence for sufficient content validity and internal consistency, (B) potential to be recommended for use but not categorized as A or C or (C) PROMs with high quality evidence that a measurement property is insufficient and therefore should not be recommended for use.

***non-commercial use, *8 items: P physical, RP role physical, EF energy fatigue, SF social functioning, MH mental health, RE role emotional, GH general health, V vitality. Factor coefficients for mental and physical summary scores are held under copyright, reporting a total overall score is not recommended.(114,161) ‡inconsistencies with item functioning. Full summary of findings available in online supplement tables E7 and E8. GRADE Grading of Recommendations Assessment Development Evaluation, EuroQol-5D-5L (EQ-5D-5L), LPHQ Living with Pulmonary Hypertension Questionnaire, MLWHF Minnesota Living with Heart Failure, SF Short Form-36

Contributions
All authors contributed to the development of the manuscript and approved the final version

Acknowledgements
The authors would like to thank the patients from PHA UK and PHA UK representatives who supported the ratification of the conceptual framework. Thank you to Dr Charlie Eliot for ratifying the conceptual framework in addition to the authors.

Statement of Guarantor
N/A
## Contact emails and ORCID

<table>
<thead>
<tr>
<th>Email</th>
<th>ORCID</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="mailto:f.varian@sheffield.ac.uk">f.varian@sheffield.ac.uk</a></td>
<td>0000-0002-4644-8391</td>
</tr>
<tr>
<td><a href="mailto:joseph.newman@nhs.net">joseph.newman@nhs.net</a></td>
<td>0000-0001-5514-0754</td>
</tr>
<tr>
<td><a href="mailto:charlotte.pearson21@nhs.net">charlotte.pearson21@nhs.net</a></td>
<td>0009-0005-8120-9678</td>
</tr>
<tr>
<td><a href="mailto:R.Thompson@sheffield.ac.uk">R.Thompson@sheffield.ac.uk</a></td>
<td>0000-0002-0717-4551</td>
</tr>
<tr>
<td><a href="mailto:gregg.rawlings@sheffield.ac.uk">gregg.rawlings@sheffield.ac.uk</a></td>
<td>0000-0002-7916-9313.</td>
</tr>
<tr>
<td><a href="mailto:a.rothman@sheffield.ac.uk">a.rothman@sheffield.ac.uk</a></td>
<td><a href="https://orcid.org/0000-0002-7847-4500">https://orcid.org/0000-0002-7847-4500</a></td>
</tr>
<tr>
<td><a href="mailto:ciara.a.mccormack@mu.ie">ciara.a.mccormack@mu.ie</a></td>
<td><a href="https://orcid.org/0000-0003-4508-5522">https://orcid.org/0000-0003-4508-5522</a></td>
</tr>
<tr>
<td><a href="mailto:rebecca.burney@nhs.net">rebecca.burney@nhs.net</a></td>
<td>0009-0001-4984-6029</td>
</tr>
<tr>
<td><a href="mailto:T.Peasgood@sheffield.ac.uk">T.Peasgood@sheffield.ac.uk</a></td>
<td>0000-0001-8024-7801</td>
</tr>
<tr>
<td><a href="mailto:zmgo13@gmail.com">zmgo13@gmail.com</a></td>
<td>0000-0001-8912-8470</td>
</tr>
<tr>
<td><a href="mailto:j.carlton@sheffield.ac.uk">j.carlton@sheffield.ac.uk</a></td>
<td>0000-0002-9373-7663</td>
</tr>
</tbody>
</table>
References


16. United States Food and Drug Administration. PATIENT-FOCUSED DRUG DEVELOPMENT GUIDANCE PUBLIC WORKSHOP Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments [Internet]. 2018 [cited 2024 Jul 15]. Available from: https://www.fda.gov/media/116277/download


51. PHA UK. Helping people get the most from their PH treatments: Phoenix study proposal findings [Internet]. 2021 [cited 2023 Jul 18]. Available from: https://www.phauk.org/app/uploads/2021/04/Phoenix-study-research-report.pdf


111. Deliu N, DR, MA, SJ, DM, JR, WM, TM, VS. StratosPHere 2: A response-adaptive randomised placebo-controlled Phase II trial to evaluate hydroxychloroquine and phenylbutyrate in pulmonary arterial hypertension caused by mutations in BMPR2. pre-print. 2023;


146. National Institute for Health and Care Excellence. NICE health technology evaluations: the manual (PMG36) [Internet]. 2022 Jan [cited 2024 Jul 15]. Available from:


158. PHA UK. EmPHasis-10 Questionnaire [Internet]. 2024 [cited 2024 Aug 1]. Available from: https://www.phauk.org/pha-uk-resources/emphasis-10-questionnaire/#:~:text=The%20EmPHasis%20D10%20questionnaire%20is,patients%20during%20remote%20assessments%20only.


**Health-Related Quality of Life (HRQoL) Outcomes in Pulmonary Arterial Hypertension Trials**

**PROSPERO ID: CRD42024484021**

1. **Inclusion criteria**: Randomised Controlled Trials (RCTs) powered for minimal clinically important difference (MCID) of PROM

2. **COSMIN** evaluation of PROM psychometric properties

- **COSMIN**
  - A: Recommended
  - B: Further evaluation needed
  - C: Not recommended

**Systematic Review of Patient Reported Outcome Measures (PROM)**

- 90 RCTs
- 43 HRQoL endpoint
- 20 PAH MCID
- 8 Powered
- 1 Meta-Analysis
- 0 PPIE

**Consensus based standards for Selection of health-Measurement INstruments**

- 4 PROMs
  - EmPhasis -10
  - LPHQ
  - EQ-5D-5L
  - SF-36

- 2 (of 8) SF-36 domains reliable
- No PAH EQ-5D-5L psychometrics

- EQ-5D-5L less responsive than LPHQ
- For a valid HRQoL endpoint - select PROM with a PAH MCID
- PROM not yet suitable for quality adjusted life year (QALY) estimates

**CONCEPTUAL FRAMEWORK**

- 8045 patients identified
- Framework ratified

Ratified HRQoL framework used to map PROMs

**FUTURE FOCUS**

Map PROM to this framework from the perspectives of people living with PAH

**Themes**

- Sadness, Depression
  - Anxiety
  - Isolation
  - Fear and Uncertainty
  - Anger
  - Frustration

- Self-worth
  - Autonomy

- Sense of Loss
  - Relationships
- Physical
- Treatment burden
- Financial impact
- Activity

Most Commonly Reported

Not captured by EQ-5D-5L or SF-36

Less Commonly Reported

- Sexual Relationships
- Friends and Family
- Stigma/Hidden illness
- Clinical relationship
- Fatigue
- Breathlessness
- Dizziness
- Headache
- Sleep
- Daily variability
- Travel
- Hospital visits
- Work/Education
- Returning to activities
- Public places