SEE-ING WHAT'S IMPORTANT: ELICITING EXPERT JUDGEMENTS TO ESTIMATE THE MINIMALLY CLINICALLY IMPORTANT DIFFERENCE FOR SIX-MINUTE WALKING DISTANCE IN PH-ILD

Behr J[1], Howard L[2], Lewis D[3], Joseph A[3], Smith AB[3], Graziadio S[3], Green M[3], Funes DF[4], **Fernández Delgado M[4]**, Pujadas MB[4]

[1] University Hospital Munich, Munich, Germany; [2] Imperial College Healthcare NHS Trust, London, United Kingdom; [3] York Health Economics Consortium, York, United Kingdom; [4] Ferrer Internacional, Barcelona, Spain.

1. INTRODUCTION AND OBJECTIVE

Exercise testing is routinely recommended at baseline and follow-up assessments of pulmonary hypertension (PH) to evaluate disease severity, progression and response to therapy [1]. Cardiopulmonary measurements such as the six-minute walking distance (6MWD) test are frequently employed as primary endpoints in clinical trials. A challenge of examining patient outcomes in clinical trials, however, is determining if the observed differences constitute a clinically meaningful change for the patient. Prior work has established responder thresholds for pulmonary arterial hypertension (PAH); however, there is currently no estimate of the minimally clinically important difference (MCID) for the 6MWD of people with PH-ILD [2-4].

Where evidence is unavailable or uncertain, researchers may rely on expert judgements. However, these judgements are subject to uncertainty as experts have imperfect knowledge and are susceptible to cognitive and motivational biases [5]. Any analyses using uncertain evidence may give decision-makers a misleading view of the risks associated with their decision. This uncertainty must, therefore, be characterised. This study sought to employ a novel approach to MCID estimation, implementing a structured expert elicitation (SEE) workshop.

Objective: Implement an SEE workshop to elicit plausible estimates of the MCID for the 6MWD of people with PH-ILD.

2. METHODS

TRAINING AND PREPARATION

The Medical Research Council protocol and associated SEE resources (STEER) were implemented [6, 7]. The study received ethical approval from the University of York Research Governance Committee. A narrative review was conducted to inform an evidence dossier. The dossier was developed to contextualise questions and summarise available data surrounding the quantities of interest. The dossier was circulated one week before the workshop alongside training materials and an example judgement encoding task. The training materials clarified the research questions and defined concepts such as uncertainty (differentiating it from variability) and frequent types of bias. The example task responses were analysed and discussed at the beginning of the workshop, where misconceptions or clarifying questions were addressed. The protocol and materials were piloted by two experts independent of the workshop to ensure the procedure was understandable, the quantities were well-defined, and the training materials were sufficient.

IMPLEMENTING THE WORKSHOP

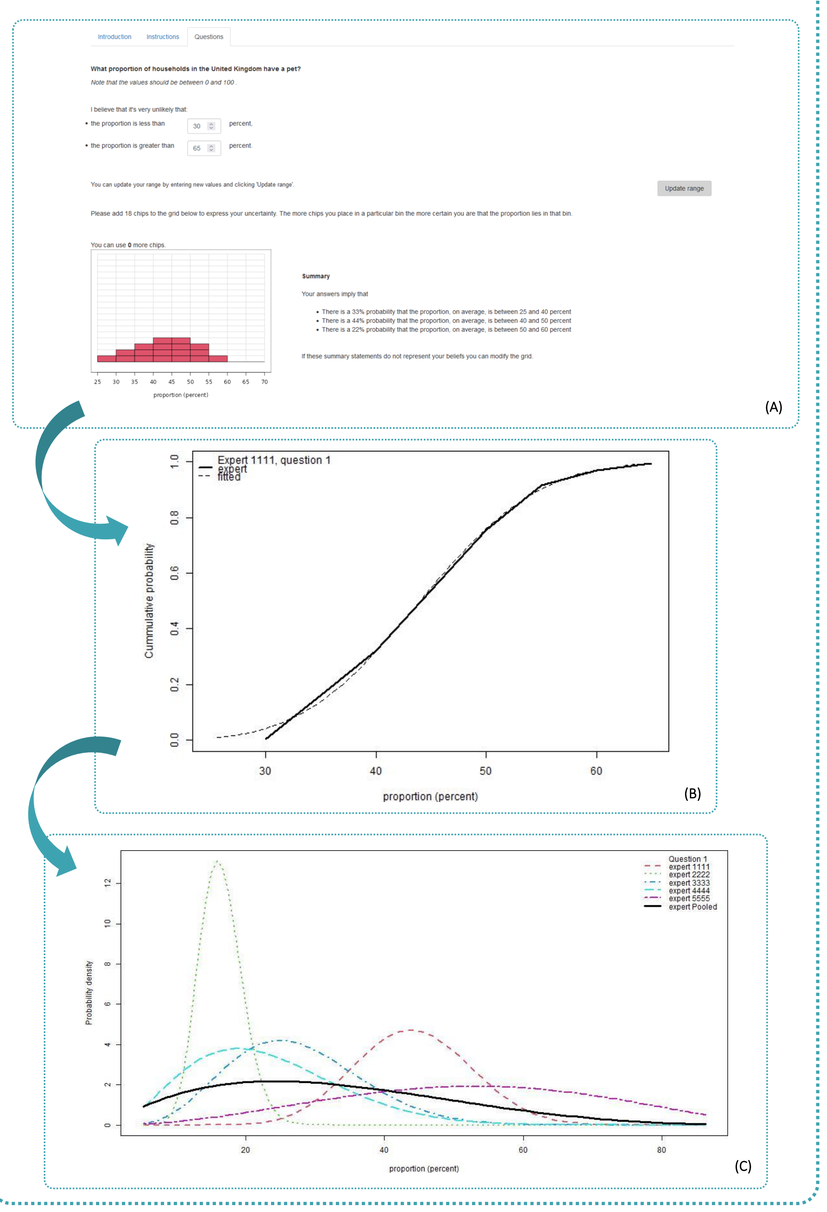
Pulmonologists and respiratory physicians (N = 5) from the UK and Germany were recruited. Facilitators delivered the workshop remotely using a web-based application. A fixed interval method was used. The experts were prompted to provide their plausible range by indicating limits representing 98% confidence intervals. Subsequently, they were presented with a grid dividing their plausible range into intervals, and a histogram was constructed by allocating a finite number of 'chips'. Each chip represented a fixed probability, and the number of chips placed into an interval indicated the expert's certainty that the unknown value fell within it. The experts provided rationales alongside their responses. The fitted distributions were displayed during facilitated group discussions to examine between-expert variability. The experts were allowed to alter their distributions post-discussion.

FITTING AND AGGREGATION

The cumulative distribution function was derived for all responses using the quantiles and probability masses generated from the encoding tasks. Modelled distributions were generated by fitting a suitable parametric distribution to the bounds using a least squares estimation method. The best fitting distribution was determined by a post hoc analysis of the sum of squared errors. Linear opinion pooling was employed to generate an aggregate distribution. All quantities were expressed as deterministic point estimates reflecting the expected value of the pooled distributions.

FIGURE 1

(A) Example web-based judgement encoding task; (B) Fitted cumulative density function; (C) Individual and aggregate distributions.



3. RESULTS

The expected value of the fitted distributions for the pre-discussion responses ranged between 17.5 m (SD = 1.4 m) and 47.1 m (SD = 11.8 m) and was 31.9 m (SD = 12.4 m) for the pooled distribution. For the post-discussion task, the expected value ranged between 24.8 m (SD = 2.0 m) and 40.6 m (SD = 8.7 m) and was 31.0 m (SD = 12.4 m) for the pooled distribution.

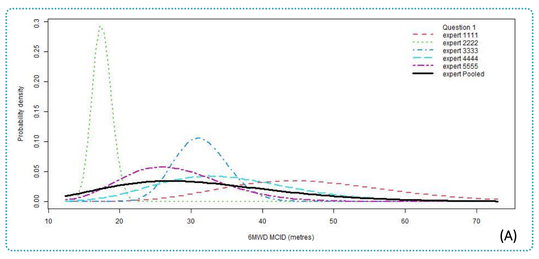
The group discussions and rationale statements revealed that the experts had based their responses on their clinical experience, what they perceived as clinically meaningful for patients, and existing data for PAH populations. Multiple experts indicated that the MCID for PH-ILD is likely moderately lower than that reported for PAH. When prompted for justification, they cited the comparatively reduced exercise capacity observed in this population. Expert `2222` maintained the highest degree of certainty in their estimate, noting that stability or even a small improvement was meaningful for patients.

One expert (`1111`) provided comparatively higher MCID estimates in both the pre- (mean = 47.1 m) and post-discussion (mean = 40.6 m) tasks. Their rationale statements and contributions during the group discussion indicated contradictory statements, such as failure to adjust their plausible limits. Due to task comprehension concerns, a scenario analysis was conducted, re-examining the responses to the second MCID elicitation task, excluding Expert `1111`. The expected value of the pooled distribution in the scenario was 28.6 m (SD = 6.9 m).

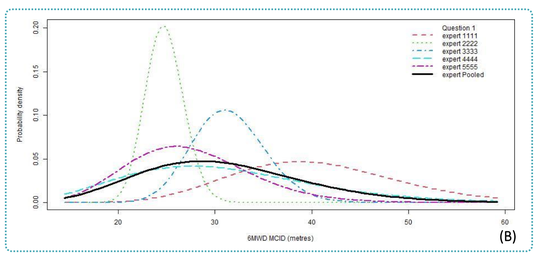
FIGURE 2

(A) Pre-discussion task; (B) Post-discussion task.

Expert	Mean	SD
1111	47.1	11.8
2222	17.5	1.4
3333	31.5	3.8
4444	35.4	9.8
5555	27.8	7.1
Pooled	31.9	12.4



Expert	Mean	SD
1111	40.6	8.7
2222	24.8	2.0
3333	31.5	3.8
4444	30.9	10.0
5555	27.4	6.4
Pooled	31.0	8.7



4. CONCLUSIONS

This work comprises one component of an ongoing study examining data from the INCREASE trial (NCT02630316) and hopes to facilitate the interpretation of clinical trials for PH-ILD-specific therapies. Multiple plausible values were generated using validated methods of SEE. The pooled estimates ranged between 28.6 m and 31.0 m, falling below published values for PAH populations. The findings provide a foundation for further work employing traditional estimation approaches, particularly anchor-based methods employing validated patient-reported outcome measures.

5. REFERENCES

(1) Gabler NB., et al. Validation of 6-minute walk distance as a surrogate end point in pulmonary arterial hypertension trials. Circulation. 2012 Jul 17;126(3):349-56.
(2) Gilbert C., et al. Estimating a minimally important difference in pulmonary arterial hypertension following treatment with sildenafil. Chest. 2009 Jan 1;135(1):137-42.
(3) Mathai SC., et al. The minimal important difference in the 6-minute walk test for patients with pulmonary arterial hypertension. American journal of respiratory and critical care medicine. 2012 Sep 1;186(5):428-33.
(4) Moutchia J., et al. Minimal clinically important difference in the 6-minute-walk distance for patients with pulmonary arterial hypertension. American journal of respiratory and critical care medicine. 2023 Apr 15;207(8):1070-9.
(5) Montibeller G, Von Winterfeldt D. Cognitive and motivational biases in decision and risk analysis. Risk analysis. 2015 Jul;35(7):1230-51.
(6) Bojke L., et al. Developing a reference protocol for structured expert elicitation in health-care decision-making: a mixed-methods study. Health Technology Assessment (Winchester, England). 2021 Jun;25(37):1.
(7) Centre for Health Economics. Structured expert elicitation resources (STEER). 2021.

Disclosure: Miriam Fernández Delgado, Diego Federico Funes and Mireia Bosch Pujadas are employees of Ferrer. Ferrer provided payment to York Health Economics Consortium (YHEC) to conduct the analyses and generate the publication. Damian Lewis, Andria Joseph, Adam B. Smith and Maisie Green are employed by YHEC.

Acknowledgements: The authors thank Colin Church, Colm McCabe, Laura Price, Marius Hoeper and Nazia Chaudhuri for their insight, expertise and feedback while participating in the workshop and reviewing the associated materials.

Presented at ISPOR EU 2024 Annual Meeting • 17 - 20 November 2024 • Barcelona, Spain