

Developing outcome measures for research

BY DEBORAH HALL

There are challenges in developing outcomes measures; **Professor Hall** presents five top pointers for making rapid progress in developing outcome measures for research purposes.

Anyone who has worked clinically with hearing loss will appreciate that every patient's experience is personal. This situation presents a practical challenge for how to measure treatment success. For one patient, treatment success may be about easier listening to the TV. For another, it may be about enhancing social experiences. Yet how can this diversity be addressed when an outcome measure for research must be the same for all participants?

Hearing loss affects about 300 million adults worldwide [1], and there is general consensus that hearing loss can have a negative impact on various aspects of an individual's quality of life. Difficulties in everyday life attributed to hearing loss vary considerably from person to person, and the degree of difficulties correlates poorly with audiometric profiles.

Probably the single most important factor in designing a clinical trial to test out whether a treatment is working for patients is the choice of outcome measure. This is crucial because in clinical trials, therapeutic benefit is determined using (usually) one measurement collected from all patients (i.e. treatment and control groups), at a pre-defined point in time relative to the treatment regime. This measurement is called the primary outcome measure and the data from this measure is used to

answer the main research question. The primary outcome is typically a variable relating to clinical efficacy, but could also relate to safety or tolerability of the treatment, or the patient's general quality of life. Generally speaking, the primary outcome should be the endpoint that is clinically relevant from the patients' perspective and relevant to healthcare providers in their everyday practice. In support of this, guidelines for good clinical practice state that: "The primary variable should be that variable capable of providing the most clinically relevant and convincing evidence directly related to the primary objective of the trial" [2].

Bearing in mind that the experience of hearing loss varies so widely across patients, there is no obvious outcome measure. Outcome reporting is typically based on a multi-item questionnaire that asks questions about a range of patient-reported complaints and generates a numerical score which indicates severity (Figure 1). The diversity of generic and hearing-specific complaints such as hearing disability, hearing handicap, quality of life, hearing aid benefit, communication and psychological outcomes perhaps helps to explain why so many different questionnaires have been developed to explore the impact of hearing loss. But these don't all ask questions about the same dimensions of complaint. There is no gold standard [3].

This situation presents the healthcare professional with two major practical challenges:

- Challenge 1 concerns how to comprehensively assess a patient for a precise clinical diagnosis.
- Challenge 2 concerns how to measure the therapeutic benefit for determining clinical efficacy (or for clinical practice audit).

With some degree of success, the first challenge has been resolved by creating multi-attribute, multi-item questionnaires

"We face two major challenges: 1. How to comprehensively assess a patient for a precise clinical diagnosis? 2. How to measure the therapeutic benefit for determining clinical efficacy?"



Figure 1. Outcome reporting typically relies on questionnaire that asks questions about a range of complaints.

Sample questions

1. Does a hearing problem cause you to use the phone less often than you would like? Yes Sometimes No

2. Does a hearing problem cause you to feel embarrassed when meeting new people? Yes Sometimes No

...

24. Does a hearing problem cause you to feel uncomfortable when talking to friends? Yes Sometimes No

25. Does a hearing problem cause you to feel left out when you are with a group of people? Yes Sometimes No

Creating a composite score
 For each item, 'Yes'=4, 'Sometimes'=2, and 'No'=0 points.
 Add up the scores for each of the 25 items to create a composite score out of 100.

Interpreting the composite score
 0-16: No handicap
 17-42: Mild to moderate handicap
 43-100: Significant handicap

Figure 2. Sample questions, scoring instructions and guidance for interpretation for the Hearing Handicap Inventory for the Elderly.

whose composite score can be used to discriminate between individuals. For example, the Hearing Handicap Inventory for the Elderly (HHIE) asks 25 questions about the emotional consequences of hearing impairment, social and situational effects, with pre-defined cut-offs of the overall score for determining 'no handicap', 'mild to moderate handicap' and 'significant handicap' (Figure 2). However, the solution to the first challenge tends to be incompatible with evaluating therapeutic benefit. This is because questionnaire items that discriminate well between different patients at the diagnostic appointment are not necessarily sensitive to evaluating changes over time within the same patient.

There are no easy solutions to this second challenge at present, but at least there are clear pointers as to where future research is likely to make the most rapid progress. Here are my top five recommendations:



Figure 3. There are so many different questionnaires and other outcome measures to choose from it can get bewildering at times.

(i) Future research studies should be reducing the diversity of outcome instruments.

Recent systematic reviews of instruments for measuring outcomes in audiological research have confirmed unacceptable heterogeneity. For example, there are more than 200 different measurement tools for assessing hearing loss, and more than 100 for tinnitus (Figure 3). This is not only confusing for investigators, but it also makes comparisons between studies impossible. Solutions can be found to identify those instruments that have good construct validity, are feasible to administer and have good statistical properties, but recommendations will have widespread impact only by working together to create and follow consensus guidelines [4].

(ii) Future research studies should clearly state the outcome of interest, and how it will be measured.

For trials of clinical efficacy, investigators clearly specify what they expect their treatment to change. Hearing-specific examples include 'effects on personal relationships' or 'conversation with a group of talkers'. These specific domain(s) of potential treatment-related change should define the clinical trial outcome(s). A measurement tool can then be selected for each outcome. And it is preferable to provide some explanation about why that particular tool was selected. Preferably the chosen tool should also have been validated for responsiveness

“Think critically about exactly what it is you are trying to measure. Don't select a questionnaire simply on the basis of its popularity or accessibility.”



to treatment-related change in the appropriate patient population. Helpfully, there are clear guidelines to choose the most appropriate tool [5].

(iii) Future research studies should recognise the difference between instruments developed to discriminate between patients and instruments developed to detect treatment-related change.

It is difficult to design a questionnaire instrument that is both discriminative and evaluative. To illustrate this with an example, communication difficulties and performance at work might both discriminate one patient from another, but only one of these might be responsive to treatment (e.g. hearing aids should improve communication difficulties, but might not necessarily affect working in an occupational setting). Averaging the benefit scores for these two different outcome domains could therefore compromise the sensitivity of the composite score for measuring treatment-related change. As a general rule, questionnaire instruments that successfully measure therapeutic benefit tend to be those with good statistical properties that enable the clinician or investigator to interpret specific complaints, rather than a global non-specific construct like 'severity' or 'handicap' [5].

(iv) Future research studies should interpret whether the outcome is clinically meaningful to the patient.

Just because a trial finds a statistically significant benefit does not necessarily mean that this is clinically meaningful. Statistics is all about numbers. It's possible for a trial to demonstrate a significant improvement, even when the size of that change is very small. For example, if a trial enrolls a large number of patients (high power) who have a low variability in scores (low variance). What is most important is whether the size of the change is meaningful from a clinical



Figure 4. Working out what magnitude of change on an outcome measure constitutes a clinically meaningful change should incorporate the patient perspective.

perspective. Various research methods do exist to assign a numerical value to clinical meaning often called the Minimal Important Difference (MID) or Minimal Clinically Important Difference (MCID). The most reliable methods take into account the patient perspective (Figure 4). How many hearing-related questionnaires do you know which have defined the MID / MCID for interpreting the change scores? Very few, I expect.

(v) Future research studies should describe both positive and negative outcomes.

All too often there is a bias towards highlighting the positive benefits of an intervention, at the expense of describing any negative or null findings. One way to minimise such source of bias is to always consider outcomes relating to patient safety since these are just as important as those relating to clinical efficacy. These can include, but are not restricted to 'adverse events', 'treatment-related adverse reactions', 'treatment adherence', and 'withdrawals from the study'. Another way to minimise such source of bias is to publish an outline of the clinical trial design and statistical analysis plan, while the data are still being collected. This includes defining the primary outcome and stating how it will be analysed. There are numerous public registries that provide such a publishing service (e.g. <https://clinicaltrials.gov/>, <https://www.isrctn.com/>, etc). And it is becoming increasingly common to publish more detailed clinical trial protocols in peer-reviewed journals (e.g. *BMJ Open*, *Trials*, *Pilot and Feasibility Studies*, etc).

There are no easy solutions to developing good outcome measures, but these five recommendations can guide future research studies and overcome current limitations.

- (i) Reduce the diversity of outcome instruments.
- (ii) Clearly state the outcome of interest, and how it will be measured.
- (iii) Recognise the difference between instruments developed to discriminate between patients and those developed to detect treatment-related change.
- (iv) Interpret whether the outcome is clinically meaningful to the patient.
- (v) Describe both positive and negative outcomes.

References

1. World Health Organization. Deafness and hearing loss fact sheet. 2015; <http://www.who.int/mediacentre/factsheets/fs300/en/> Last accessed 24 August 2017.
2. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. Harmonised Tripartite Guideline. Statistical Principles for Clinical Trials - E9. 1998; http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf Last accessed 24 August 2017.
3. Granberg S, Dahlstrom J, Moller C, Kahari K, Danermark B. The ICF Core Sets for hearing loss – researcher perspective. Part I: Systematic review of outcome measures identified in audiological research. *International Journal of Audiology* 2014;53(2):65-76.
4. COMIT'ID. Sharing your views to help improve the future of tinnitus research. https://www.youtube.com/watch?v=42X_mp59cKk&t=13s Last accessed 23 August 2017.
5. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials* 2016;17(1):449.



AUTHOR



Prof Deborah Hall,

Deputy Director NIHR Nottingham Biomedical Research Centre (BRC), Research Lead (Outcome measures for evidence synthesis and clinical trials), National Institute for Health Research Nottingham Biomedical Research Centre, Ropewalk House, 113 The Ropewalk, Nottingham, NG1 5DU, UK.

**E: deborah.hall@nottingham.ac.uk
www.hearing.nihr.ac.uk
[@hearingnihr](https://twitter.com/hearingnihr)**

Deborah Hall is the Deputy Director of the NIHR Nottingham Biomedical Research Centre and the Research Lead for 'Outcome measures for evidence synthesis and clinical trials'. Deborah graduated from Brunel University in Psychology and completed a PhD in Cognitive Neuropsychology at Birmingham University. Deborah has since worked at MRC Institute of Hearing Research, Nottingham Trent University, and is currently based at the University of Nottingham.

Declaration of Competing Interests

Prof Hall's research is supported by the National Institute for Health Research (NIHR). The views expressed are those of the author and not necessarily those of the NHS, the NIHR or the Department of Health. The author has made all reasonable efforts to ensure that the information provided in this review is based on available scientific evidence.

