

Mode of data elicitation, acquisition and response to surveys: a systematic review

K Hood,^{1*} M Robling,¹ D Ingledew,²
D Gillespie,¹ G Greene,¹ R Ivins,¹ I Russell,³
A Sayers,⁴ C Shaw⁵ and J Williams³

¹School of Medicine, Cardiff University, Cardiff, UK

²School of Psychology, Bangor University, Bangor, UK

³School of Medicine, Swansea University, Swansea, UK

⁴Faculty of Medicine and Dentistry, University of Bristol, Bristol, UK

⁵Faculty of Health, Sport and Science, University of Glamorgan,
Ponytypridd, UK

*Corresponding author

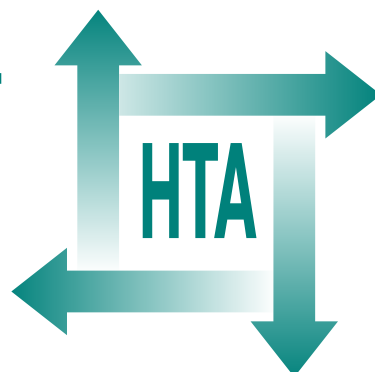


Executive summary

Health Technology Assessment 2012; Vol. 16: No. 27

DOI: 10.3310/hta16270

Health Technology Assessment
NIHR HTA programme
www.hta.ac.uk



Executive summary

Background

Many studies in health sciences research rely on collecting participant-reported outcomes. Although some of these are participant reports of factual information, such as adherence to drug regimes, that could be objectively validated, there is an increasing recognition of the importance of subjective measures such as attitude to, and perceptions of, health and services provision. Alongside the exponential increase in health-related literature devoted to participant-reported outcomes, attention is being paid to the method or mode of data collection. Much of this has been driven by the rapid development of new technologies, which can lead to increased ease, speed and efficiency of data capture alongside an increasing drive for maximising response rates. Survey methodologies (e.g. in the business, marketing, social and political sciences) have a literature base of their own, covering theory to practice, much of which has been only slowly recognised in the health arena. Few health-related outcome development papers indicate a theoretical approach to eliciting survey response and the focus in choosing a mode for a study is often based predominantly on improving response rates and minimising cost. The impact on the validity of response is not generally a consideration. In addition to this, in order to gain as complete a data set as possible, many studies are using multiple modes either to enhance participants' choice (e.g. opting for web- or paper-based surveys) or to improve follow-up rates (e.g. non-responders getting telephone data collection). Although for practical reasons these choices are entirely justifiable, consideration needs to be given to the validity of response via different modes and the impact that the choice of mode or modes might have on the conclusions from a study.

Objectives

- To provide an overview of the theoretical models of survey response and how they relate to health research.
- To review all studies comparing two modes of administration for subjective outcomes and assess the impact of mode of administration on response quality.
- To explore the impact of findings for key identified health-related measures.
- To create an accessible resource for health science researchers, which will advise on the impact of the selection of different modes of data collection on response.
- To inform the analysis of multimode studies.

Methods

In order to inform the systematic review of mode comparison studies, a review of the theoretical models and how they relate to the health domain was undertaken. This clarified the need to focus on features of mode rather than crude modes per se in order to understand the way in which responses to subjective outcomes could be affected. From this, a theoretical model based on Tourangeau was proposed with four main features: administration (interviewer or self), use of the telephone, use of the computer and sensory stimuli (audio, visual or both). Additional features were proposed that may belong in a model of response as well as potential mediating factors, such as cognitive challenge of questions. This approach was used to define the data extraction and coding classifications for studies.

Owing to the large body of literature relating to survey methodology which is published outside the health research arena, all studies that incorporate a mode comparison were included, regardless of setting. This led to a broad search strategy covering a wide range of disciplines. In order to target methodological studies, some innovations in search strategy that separate out the process from traditional reviews of the effectiveness of interventions were undertaken.

Identifying the literature

For a study to be included in the review it needed to:

1. provide evidence of a comparison between two modes of data collection of either the same question or the same set of questions referring to the same theoretical construct
2. compare a construct that is subjective and cannot be externally validated
3. explicitly reference a comparison in the analysis
4. collect quantitative data, i.e. use structured questions and answers.

Studies were excluded from the review if they involved:

1. a comparison between a quantitative measure and one or more qualitative data collection methods/analyses (e.g. unstructured interviews, focus groups)
2. a comparator derived from routine clinical records – unless explicit reference to specific self-reported construct is made within those records
3. a comparison between the response of two different judges, i.e. comparing a response from an individual to that made by someone other than the responder, for example a clinician providing a diagnosis.

A broad range of databases (for example EMBASE, PsychINFO, MEDLINE, EconLit, SPORTDiscus, etc.) were searched with no restrictions on start date or language. Searches were conducted up until the end of 2004. A matrix-based research strategy was developed and tested, searching for combinations of terms that would imply a mode comparison study.

Review process

The abstracts (and titles only for some foreign-language papers with no English abstract) were reviewed against the inclusion/exclusion criteria. Full papers were retrieved for all selected abstracts and then screened again using more detailed inclusion criteria related to the measures used. Papers that were still included were reviewed in full and detailed data extracted. At each stage, abstracts or papers were reviewed by a single reviewer after a period of training. Training for each stage included an assessment of reliability and sensitivity.

In order to assess the quality of the evidence contributing to this review, each paper was assessed for methodological quality. Assessing the quality of evidence becomes particularly challenging in reviews of studies having diverse methodologies. In this particular review, randomised controlled trials were not necessarily expected and so a more generic quality assessment tool was needed. A new tool was developed from two existing tools and tested.

Evidence synthesis

An overview of the studies identified is presented descriptively, highlighting the different mode features identified in the theory review. Those with appropriate data are subjected to quantitative methods of synthesis using exploratory metaregression to identify the association between mode features and differences in response. The primary analysis is based on three key summary statistics calculated for each comparison. These are the absolute difference between the means (standardised) of the two modes, the ratio of the largest to the smallest variance of the two modes and the effect size (ES; absolute mean difference/standard deviation) between two modes.

Between- and within-subject studies were analysed together, controlling for the study design. Analysis was conducted at two levels to account for clustering of comparisons within a study. This allowed for study-level characteristics, measure characteristics and mode features to be considered in a single model. The modelling approach assessed the four main mode features from the theoretical review, then tested the addition of other candidate features and then assessed model fit including other possible moderators of effect and identified interaction.

The two most frequently occurring outcomes – the Short Form questionnaire-36 items (SF-36) and the Minnesota Multiphasic Personality Inventory (MMPI) – are analysed in more depth using Mantel–Haenszel for between-group studies and Bland and Altman limits of agreements for within-group studies.

Results

The search strategy identified 39,253 unique references, of which 2156 were considered as full papers. Of these, 597 progressed to data extraction, with 381 finally included in the review. The most common reason (44%) for exclusion once the full paper was considered was that there was no actual mode comparison in the study. The majority of included studies were from North America (62%), with only 10% being from the UK.

Study designs were relatively evenly divided into between- and within-person studies (52% and 47%, respectively), with only 39% using some form of randomisation (random allocation for between-person studies and random ordering for within-person studies). In terms of quality assessment, most studies described their hypotheses and study design well, and drew appropriate conclusions (89%, 83% and 81% – good, respectively), but the description of participants, group allocation, potential impact of timing of data collection and presenting of variances was less good (22%, 50%, 27% and 35% – poor, respectively).

The 381 studies provided descriptions on 1282 outcome measures, of which 57% were health related. The most frequently reported outcomes were the SF-36 (17 studies) and the MMPI (9 studies). Thirty per cent of studies considered only a single outcome in their mode comparison, but most considered more (ranging from 1 to 21 outcomes). These studies also described a number of mode comparisons, giving in total 1522 comparisons between modes on multiple outcomes for analysis. Of these, 977 reported enough data to be included in the analysis of absolute mean differences, 910 in the analysis of the ratio of variances and 912 in the analysis of the ES.

Two features of mode were clearly associated with bias in response; however, none of the features of mode was associated with changes in precision. How the measure was administered, by an interviewer or by the person themselves, was highly significantly associated with bias ($p < 0.001$). A difference in sensory stimuli was also significant ($p = 0.03$). When both of these were present the average overall bias was < 1 point on a percentage scale. In terms of mediating factors, there was some suggestion that there was an interaction between both telephone and computer for data collection and date of publication, supporting the theory that differences disappear as new technologies become commonplace. Single-item measures were also related to greater degrees of bias than multi-item scales ($p = 0.01$).

Individual analysis of the SF-36 and MMPI showed a varied pattern across the different subscales, with conflicting results between the two types of study. None of the MMPI measures used to

detect deviant responding showed a relationship with the mode features tested. The limits of agreement analysis showed how variable measures were between modes at an individual rather than at a group mean level.

Conclusions

Implications for researchers

Researchers need to be aware of the different mode features that could have an impact on their results when selecting a mode of data collection for subjective outcomes. If researchers use a mixture of modes within their study (commonly a change in mode if there is poor or non-response), then consideration needs to be given to ameliorating potential biases consequent on this and controlling for them in analysis.

The potential does exist for there to be simple correction factors developed; however, these are likely to be measure specific. In analysis of current mixed-mode studies, researchers cannot just assume that results are comparable where a difference in administration or sensory stimuli exists and they need either to undertake sensitivity analyses or to formally control for mode in the analysis.

Recommendations for future research (in priority order)

There are already numerous studies considering a large number of outcome measures. However, these need to be reported in a standardised way to allow researchers to be able to make informed decisions about choice of mode with a particular outcome in a population. The development of reporting standards akin to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses), STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) or CONSORT (Consolidated Standards of Reporting Trials) for mode comparison studies is urgently needed and could build on the quality assessment tool developed here.

Further mode comparison studies are required, but these need to be experimentally designed to manipulate mode features and directly assess the impact. This is preferable to more studies comparing two modes at a relatively pragmatic level without consideration of those features. Studies need to give consideration to evaluation and direct testing of the impact of some of the mediators of mode effects, as the lack of data presented in papers in this review limited our ability to analyse this component.

Further primary studies need to be done to evaluate the impact of mode features over time. There was a suggestion across studies that this occurred for 'new' technologies for data collection (telephone and computer), but the 'learning effect' for any mode over time will be important to evaluate further in order to inform studies with long-term follow-up over multiple time points. The potential biasing impact of this 'learning effect' over time could be seen in single-mode studies as well as mixed-mode ones.

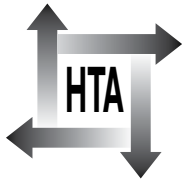
The focus of this review has been on measurement for research purposes and, therefore, has focused predominantly on the impact of mode features on estimated effects at a group level. However, the increasing use of subjective patient-reported outcomes in clinical practice means that considerable further work is required to consider measurement equivalence and reliability of assessment for individuals rather than groups.

Funding

Funding for this study was provided by the Health Technology Assessment programme of the National Institute for Health Research.

Publication

Hood K, Robling M, Ingledeu D, Gillespie D, Greene G, Ivins R, *et al.* Mode of data elicitation, acquisition and response to surveys: a systematic review. *Health Technol Assess* 2012;**16**(27).



How to obtain copies of this and other HTA programme reports

An electronic version of this title, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (www.hta.ac.uk). A fully searchable DVD is also available (see below).

Printed copies of HTA journal series issues cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our despatch agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per issue and for the rest of the world £3 per issue.

How to order:

- fax (with **credit card details**)
- post (with **credit card details** or **cheque**)
- phone during office hours (**credit card** only).

Additionally the HTA website allows you to either print out your order or download a blank order form.

Contact details are as follows:

Synergie UK (HTA Department)
Digital House, The Loddon Centre
Wade Road
Basingstoke
Hants RG24 8QW

Email: orders@hta.ac.uk

Tel: 0845 812 4000 – ask for ‘HTA Payment Services’
(out-of-hours answer-phone service)

Fax: 0845 812 4001 – put ‘HTA Order’ on the fax header

Payment methods

Paying by cheque

If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *University of Southampton* and drawn on a bank with a UK address.

Paying by credit card

You can order using your credit card by phone, fax or post.

Subscriptions

NHS libraries can subscribe free of charge. Public libraries can subscribe at a reduced cost of £100 for each volume (normally comprising 40–50 titles). The commercial subscription rate is £400 per volume (addresses within the UK) and £600 per volume (addresses outside the UK). Please see our website for details. Subscriptions can be purchased only for the current or forthcoming volume.

How do I get a copy of HTA on DVD?

Please use the form on the HTA website (www.hta.ac.uk/htacd/index.shtml). *HTA on DVD* is currently free of charge worldwide.

The website also provides information about the HTA programme and lists the membership of the various committees.

NIHR Health Technology Assessment programme

The Health Technology Assessment (HTA) programme, part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the effectiveness, costs and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined as all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care.

The research findings from the HTA programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA programme is needs led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, from the public and consumer groups and from professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA programme then commissions the research by competitive tender.

Second, the HTA programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Third, through its Technology Assessment Report (TAR) call-off contract, the HTA programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer reviewed by a number of independent expert referees before publication in the widely read journal series *Health Technology Assessment*.

Criteria for inclusion in the HTA journal series

Reports are published in the HTA journal series if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The research reported in this issue of the journal was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formally transferred to the HTA programme in April 2007 under the newly established NIHR Methodology Panel. The HTA programme project number is 06/91/07. The contractual start date was in June 2005. The draft report began editorial review in September 2010 and was accepted for publication in March 2011. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA programme or the Department of Health.

Editor-in-Chief: Professor Tom Walley CBE
Series Editors: Dr Martin Ashton-Key, Professor Aileen Clarke, Dr Peter Davidson, Dr Tom Marshall, Professor John Powell, Dr Rob Riemsma and Professor Ken Stein
Editorial Contact: edit@southampton.ac.uk

ISSN 1366-5278 (Print)

ISSN 2046-4924 (Online)

ISSN 2046-4932 (DVD)

© Queen's Printer and Controller of HMSO 2012. This work was produced by Hood *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health.

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (<http://www.publicationethics.org/>).

This journal may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising. Applications for commercial reproduction should be addressed to: NETSCC, Health Technology Assessment, Alpha House, University of Southampton Science Park, Southampton SO16 7NS, UK.

Published by Prepress Projects Ltd, Perth, Scotland (www.prepress-projects.co.uk), on behalf of NETSCC, HTA.

Printed on acid-free paper in the UK by Charlesworth Press.