

Resource-Use Measurement Based on Patient Recall: Issues and Challenges for Economic Evaluation

Joanna C. Thorn · Joanna Coast · David Cohen · William Hollingworth ·
Martin Knapp · Sian M. Noble · Colin Ridyard · Sarah Wordsworth · Dyfrig Hughes

Published online: 26 March 2013
© Springer International Publishing Switzerland 2013

Abstract Accurate resource-use measurement is challenging within an economic evaluation, but is a fundamental requirement for estimating efficiency. Considerable research effort has been concentrated on the appropriate measurement of outcomes and the policy implications of economic evaluation, while methods for resource-use measurement have been relatively neglected. Recently, the Database of Instruments for Resource Use Measurement (DIRUM) was set up at <http://www.dirum.org> to provide a repository where researchers can share resource-use measures and methods. A workshop to discuss the issues was held at the University of Birmingham in October 2011. Based on material presented at the workshop, this article highlights the state of the art of UK instruments for

resource-use data collection based on patient recall. We consider methodological issues in the design and analysis of resource-use instruments, and the challenges associated with designing new questionnaires. We suggest a method of developing a good practice guideline, and identify some areas for future research. Consensus amongst health economists has yet to be reached on many aspects of resource-use measurement. We argue that researchers should now afford costing methodologies the same attention as outcome measurement, and we hope that this Current Opinion article will stimulate a debate on methods of resource-use data collection and establish a research agenda to improve the precision and accuracy of resource-use estimates.

J. C. Thorn (✉) · W. Hollingworth · S. M. Noble
MRC ConDuCT Hub, School of Social and Community
Medicine, University of Bristol, Canynge Hall,
39 Whatley Road, Bristol, BS8 2PS, UK
e-mail: joanna.thorn@bristol.ac.uk

J. Coast
Health Economics Unit, School of Health and Population
Sciences, College of Medical and Dental Sciences, University
of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

D. Cohen
Health Economics and Policy Research Unit, Faculty of Health,
Sport and Science, University of Glamorgan, Pontypridd,
CF37 1DL, UK

M. Knapp
Personal Social Services Research Unit, LSE Health and Social
Care, London School of Economics and Political Science,
Houghton Street, London, WC2A 2AE, UK

M. Knapp
Centre for the Economics of Mental and Physical Health,
Institute of Psychiatry, King's College London,
De Crespigny Park, London, SE5 8AF, UK

C. Ridyard · D. Hughes
Centre for Health Economics & Medicines Evaluation, Institute
of Medical and Social Care Research, Bangor University,
Dean Street, Bangor, LL57 1UT, UK

S. Wordsworth
Health Economics Research Centre, University of Oxford,
Oxford, OX3 7AE, UK

Key Points for Decision Makers

- Methods for resource-use measurement based on patient recall have been relatively neglected compared with outcome measurement; a workshop discussing the issues highlighted the importance of addressing this imbalance.
- Topics identified for a future research agenda included measurement of total versus condition-related resource use, the implications of a lack of gold standard, validation issues, recall periods, patient burden and the standardisation of resource-use measures.
- We believe that costing methodologies must be afforded the same consideration as is devoted to outcome measures.

1 Introduction

1.1 Background

Resource-use measurement is one of the most challenging parts of an economic evaluation to get right, and is correspondingly time consuming. Despite this, in comparison with outcomes research and other areas of health economics, resource-use measurement is the subject of very limited research efforts. In 1999 a review of methods for ‘assessing the costs of healthcare technologies in clinical trials’ suggested that funders should establish an archive of data collection instruments and referred to the need to determine appropriate recall periods, develop standard questionnaires and determine validity and reliability of questionnaires [1]. Thirteen years later, there has been some progress on all the issues raised, but there remains considerable room for improvement.

Despite the opportunities afforded by increasing access to electronically stored routine healthcare data in the UK, patient-reported resource-use measures (RUMs) are likely to be necessary in economic evaluation for the foreseeable future, and it is therefore important to ensure that they are of high quality. RUMs are developed in many different formats, and might include questionnaires, diaries or interviews, administered by post, in person or by telephone. Currently, researchers frequently duplicate effort by independently reinventing RUMs. Funded by a grant from the Medical Research Council (MRC) Network of Hubs for Trials Methodology Research, the Database of Instruments for Resource-Use Measurement (DIRUM) was created at <http://www.dirum.org> to provide a repository where researchers can share RUMs and methods [2]. The database now houses over 50 instruments, and is rapidly growing as

more RUMs are submitted. Researchers are able to download and use the RUMs within the parameters of the permissions granted by the instrument developer. DIRUM also provides a preliminary archive of previous methodological work on resource-use measurement, which can be used as a platform for future research.

1.2 Workshop on Resource-Use Measurement Based on Patient Recall: Issues, Challenges and DIRUM

Following the development of DIRUM, a related workshop on resource-use measurement based on patient recall was held at the University of Birmingham, UK in October 2011. Of the 42 attendees, the majority came from UK universities but representatives from the UK National Health Service (NHS) and the pharmaceutical industry were also present. Speakers were drawn from a pool of UK academics with experience of resource-use measurement, and the workshop offered an opportunity to meet and discuss the issues, and to stimulate methodological research. The subjects were chosen as priority areas by speakers, and therefore represent a selected rather than systematic summary of topics. As the participants were primarily practising health economists, no introductory material on measuring costs alongside clinical trials was included; there are, however, many textbooks available which cover the subject (see Morris *et al.*, for example [3]). The presentation slides are available on the DIRUM website (<http://www.dirum.org>).

Themes presented and discussed during the workshop were in the context of the UK, and included descriptions of the state of the art, the challenges associated with developing RUMs, means of improving practice and key items for a research agenda. In this Current Opinion article, we summarise these themes with the aim of stimulating methodological research into resource-use measurement and thereby improving the accuracy of resource-use estimates in economic evaluation. Based on the workshop content, we argue that insufficient research resource is currently directed towards resolving resource-use measurement issues, and that this area should be more highly prioritised for research.

2 Resource-Use Measurement: the State of the Art

2.1 Good Practice in Questionnaire Design

A recent review showed that patient self-reported questionnaires are widely used, but have a significant amount of missing data [4]. A number of practices can be identified that improve questionnaire design and, in theory, minimise missing RUM data. A Cochrane review and meta-analysis

of methods to increase response rates to postal and electronic questionnaires identified over a hundred different strategies that had been evaluated through RCTs [5]. Although none of these referred specifically to patient-completed RUMs, some of the results can be generalised. For example, in terms of the administration of RUMs, a personalised approach identifying the originating institution and assuring confidentiality is most effective, while first-class mailing results in higher response rates than second class. Questionnaires should be as short as possible to identify the main costs from the chosen perspective [6], and filter questions should be used to direct completers past sections that are not relevant to them. The most relevant questions should be placed first; for RUMs, these would concern the main cost drivers.

A recent study aimed to determine whether the use of resource-use logs (diaries given to patients for recording resource use in real time) could reduce the amount of missing data in a subsequent patient-completed RUM [7]. Overall, the differences between trial arms were mostly negligible. However, for community-based NHS services there was significant improvement in completion rates with the log. It was concluded that a simplified resource-use log designed primarily using tick boxes has the potential to be a useful tool in reducing the amount of missing data from RUMs.

2.2 Experiences with Existing Questionnaires

Although many different RUMs have been developed, few have been as comprehensively tested as the Client Service Receipt Inventory (CSRI) [8]. The questionnaire was originally developed in the UK in the mid-1980s for evaluations of the closure of long-stay hospitals (particularly psychiatric and ‘mental handicap’ hospitals). The patient group had little experience of filling in forms and many of them used a wide range of health and other services. Taking approximately 20 minutes to administer, the questionnaire covers areas including the individual’s background, accommodation and living situation, employment history, social security benefits, health service usage and support from informal carers. There are now believed to be more than 200 different versions of the CSRI adapted to cater, for example, for different disease areas, healthcare systems, languages and modes of administration. The validity and reliability of the instrument are considered good, with consistency generally fairly well demonstrated [9–12].

Another established instrument is the Annotated Patient Cost Questionnaire (APCQ) [13]. In 1998, a need was identified by UK Health Economists’ Study Group (HESG) members for a standard questionnaire that could be used in a prospective economic evaluation to measure costs relating directly to patients and informal caregivers. A working

party was set up, with broad aims of defining the items of resource use that could reasonably be obtained directly from patients. A menu of questions was developed for different contexts for researchers to develop patient-completed questionnaires for costs, reducing the ongoing *ad hoc* approach by individual researchers producing their own questionnaires from scratch. It was hoped that this more standardised approach would improve the transferability of results across different studies.

An empirical study was conducted to test how well the APCQ performed in practice [14]. From the main version, a derived questionnaire was readily developed for dialysis patients, covering travel costs to and from the dialysis unit, time spent receiving dialysis and travelling and other healthcare facility costs, and was found to be easy for patients to complete. The performance of the dialysis questionnaire was examined in several ways, with internal consistency tests showing good agreement, and test–retest reliability studies demonstrating consistent answers.

Experiences from developing these questionnaires have suggested that a slimline approach in which only the dominant cost drivers are sought may be appropriate. Standardisation should be based on broad principles rather than being an attempt to create identical RUMs for all occasions. The effort expended and attention to detail should be proportionate to the expected contribution of different services or resources to total costs; for example, an inpatient episode counts much more highly than an infrequently accessed therapy group in total cost, and it is clearly more important to accurately assess use of the former. Questions must also be sensitive to the needs of groups who may have cognitive or addiction problems, and sensitive to the topic under consideration (crime and an expectation of loss of benefits, for example, may be aspects of resource use that provoke particular concerns amongst respondents).

3 Challenges Associated with Developing and Applying RUMs

The first challenge to arise is whether it is more appropriate to use an available questionnaire (such as those described above), or to develop a new one. Apart from being less costly, existing questionnaires have the advantage of providing consistency across studies, but even the most appropriate existing questionnaire may not represent the best way to capture key information for any given study. Researchers need to consider whether the marginal benefits of developing and testing a bespoke questionnaire will outweigh the marginal costs.

The design of new questionnaires requires consideration of whether to capture condition-related or total resource

use. Whilst in principle only those resources relating to the condition in question are relevant, the definition of ‘condition related’ may not always be obvious. For example, a study on blood glucose control that included only diabetes-related resource use might regard a hospital admission for a fractured leg as not relevant, when the accident may have been due to an episode of hypoglycaemia.

There are also a number of important issues that apply regardless of whether a new or existing questionnaire is used. For example, while it may seem obvious that ‘patient recall’ data will be obtained from patients, this will not be possible with children or cognitively impaired adults and the literature on preference elicitation indicates that in many circumstances the question of who to ask is not straightforward [15, 16]. Similarly, the decision whether or not to capture participants’ resource use prior to the start of a study has to be taken in context. In large trials, randomisation should ensure an even distribution between study arms of those who routinely are high resource users but, in smaller studies, it may be necessary to capture pre-study use of resources as a covariate [17]. The level of detail requested can also be problematic as patients’ lack of knowledge of health service structures may lead them to misclassify resource usage. The cost implications of inputs by consultants versus junior doctors are significant but respondents are unlikely to know the grade of doctor who treated them or maybe even whether the health professional in question was a doctor or a nurse.

3.1 Validation Issues

The validation of cost questionnaires represents another area of considerable uncertainty. Validity assessment includes considering the extent to which an instrument measures what it purports to represent, and the degree to which it might be helpful in answering a particular question. New questionnaires require validation to ensure the appropriateness of the response categories, the clarity of the instructions, and the layout, format and length of the questionnaire. However, the absence of a gold standard can make this problematic.

Only about 30 % of studies funded by the UK Health Technology Assessment (HTA) programme described an RUM validation process [18], which compares poorly with validation of patient-reported quality of life measures. However, these outcome measures are based on patient opinion rather than provable information, may apply to many cases, and represent a snapshot of health at a particular time; in contrast, RUMs are based on verifiable events, are completely specific to the context of the trial being undertaken, and cover a time period spanning the complete trial. Owing to the cognitive processes involved in responding to questionnaires and the nature of the data

being collected, there may be some intrinsic conceptual challenges associated with validating RUMs. Based on the cognitive response model in which comprehension of the question is followed by retrieval of the relevant data before a judgment can be made and a final response given [19, 20], much internal debate concerning a patient-reported quality of life measure takes place at the judgment stage, whilst for a resource-use question the internal debate may well arise at the retrieval stage. Given the different cognitive processes involved, it may not be reasonable to consider validity for patient-reported quality of life measures and costs in the same way.

Validity measurement can take several forms [21], although the empirical literature for RUMs has tended to focus on content and criterion validity. Some examples of content validity are expressed in terms of comparison with content in previously used questionnaires; however, these may not be adequate as evidence of validity—the argument that an instrument is validated because of a comparison to another (unvalidated) instrument is not necessarily logical. Criterion validity considers whether a measure performs well in relation to an identified gold standard; however, although a gold standard for costs may be provided by routine databases, the hypothesis that these systems are accurate and complete is questionable.

Construct validity, commonly studied in outcomes measurement, focuses on expected relationships: an instrument is said to have construct validity if relationships with other measurements are as expected given the construct being assessed. However, for resource-use data the absolute quantities are likely to be just as important as any expected correlations and may be much more difficult to assess.

Some early content validation could potentially be carried out by patient representatives on trial steering committees, patient groups and via qualitative interviews. Ideally, questionnaires should also be piloted to pinpoint where poor response exists and allow reduction in the length if resources are not consumed (provided the pilot is truly representative of the study population). However, it may not be feasible to validate every combination of questionnaire, condition, context and adaptation, and doing so might not represent a good use of research resources; validating individual questions, types of question or methods may be more appropriate. Currently, there is no consensus on whether a specific measure, or a general approach to measurement, should be validated.

4 Improving Practice

Workshop participants made many suggestions for improving the design and administration of RUMs. For example, sources of unit costs should be identified at the

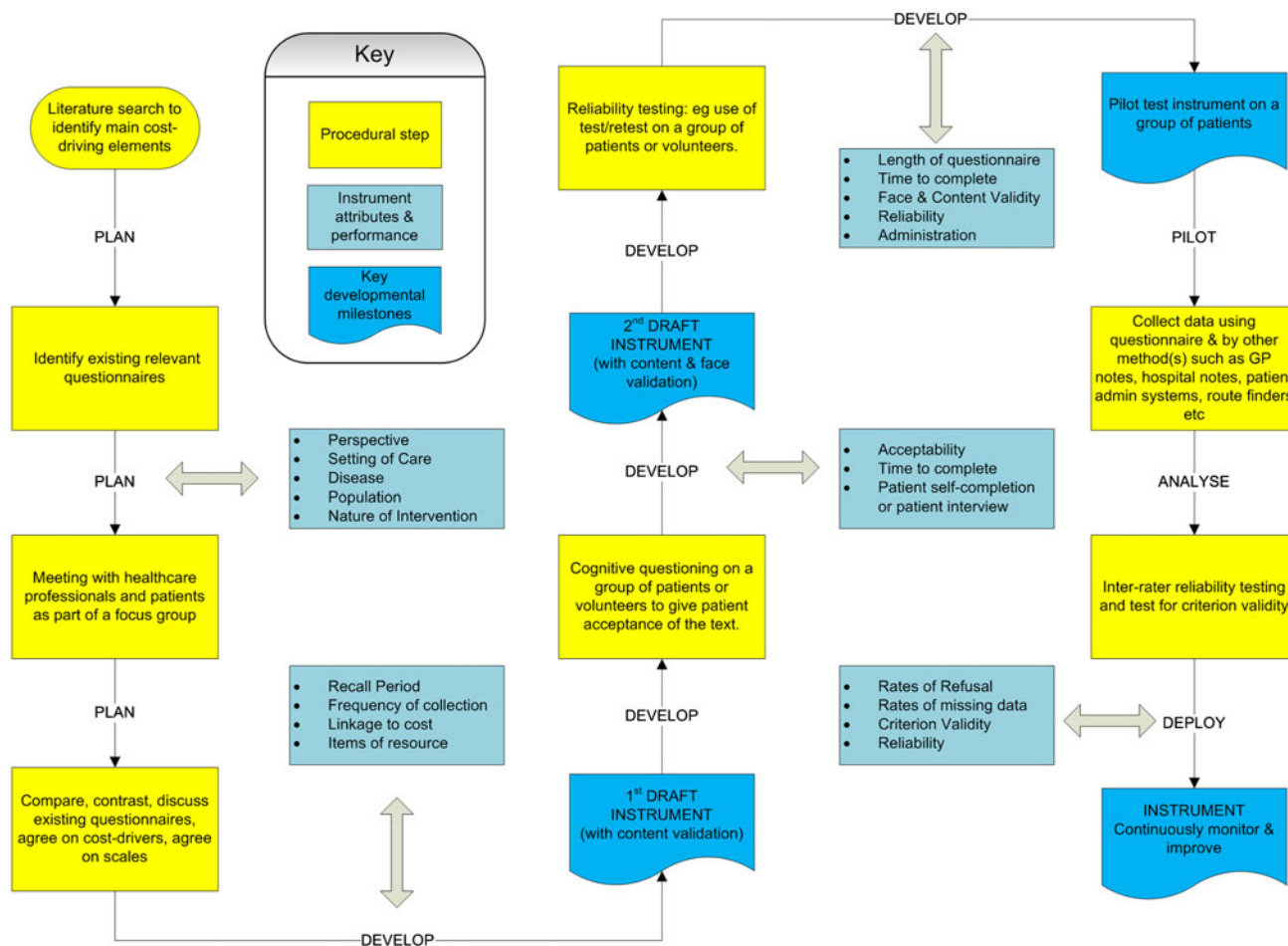


Fig. 1 Flowchart describing the processes for developing resource-use measures

outset of a study to help with specifying questions with the appropriate level of detail in the RUM, which makes many RUMs inherently country-specific as the granularity of unit costs varies between settings. Pilot or modelling studies should be used more extensively to identify important cost drivers where greater care should be taken to elicit valid responses. The patient should be treated as an essential source of information about the relevance of particular questions; different cultural or ethnic populations may differ in their resource use and the questionnaire might need to be adapted accordingly. DIRUM may have a role in allowing future trialists to identify and pilot relevant RUMs, before deciding whether to develop a new instrument.

Good-practice guidelines are required to improve methodological consistency. A recent review of the methods of resource-use data collection employed by clinical trials funded by the UK HTA programme [18] found that the majority of trials collected patient-level data and, of these, most used questionnaires, forms or interviews. Only a small proportion showed explicit evidence of having identified items of resource use at the planning stage

through consultation with healthcare professionals or conducting a review of published economic literature; however, it is likely that some form of analysis will have been undertaken even if not explicitly stated. From the described validation procedures within this review, a good practice checklist was developed, covering perspective, identification of resource-use items, planning for data collection and analysis, resource-use data collection, piloting, validation, non-trial estimates of resource use, methods for costing and standardisation of a reporting format [18]. This has further been developed into a flow diagram detailing an ideal way of approaching instrument development, encompassing planning, development, piloting, analysis and deployment of instruments (Fig. 1).

5 Research Agenda

One of the aims of the workshop was to identify a future research agenda. A number of challenges affect any potential research; for example, there is little funding

Table 1 Potential topics identified for developing into a research agenda

Measurement issues	<ul style="list-style-type: none"> • Validation when there is no alternative source of data or gold standard • Whether whole questionnaires, individual questions, types of question, methods, principles or approaches should be validated • Whether total or relevant resource use should be captured • The adequacy of expert panels for assessing costs
Questionnaire design	<ul style="list-style-type: none"> • Measuring the effect of attributes such as question order or recall period on completion rates • Improving aspects of current RUMs that typically are not covered well, such as travel and productivity • Identifying which resource-use questions are most prone to missing data • Assessing how much detail is required to obtain good medication data • Determining whether standardisation is better targeted at a modular level (applicable to specific diseases or populations, for example) or over a whole questionnaire
Patient burden	<ul style="list-style-type: none"> • Identifying whether it is better to collect a short-term snapshot of resource use and extrapolate to a longer period or to ask questions covering the whole period, which potentially increases recall bias • Assessing comprehension of medical terminology (e.g. psychologist/psychiatrist) • Dealing with the weakness of numerical skills required to multiply daily usage up to monthly consumption • Identifying the best approach to involving patients in the development of questionnaires

RUMs resource-use measures

available specifically for methodological research on measuring resource use. The health economic community does not have a culture of validation for resource-use measurement, and lacks experience of the specialist psychometric techniques required to conduct validation studies effectively; however, learning from other disciplines, such as market research or psychology, is a possible route to tackling these problems.

The workshop concluded with an open discussion of potential topics for future research. We have summarised each topic in Table 1 with the hope of stimulating such research. However, this list is not intended to be exhaustive, nor is it prioritised in order of importance.

6 Conclusion

While there are many aspects of resource-use measurement for which the health economics community has not yet reached consensus, one area on which participants appeared to agree was that there is a paucity of evidence regarding the collection of resource-use data and that there is an urgent need for this deficiency to be addressed. We believe it is time to build on the enthusiasm shown at the workshop, and that the many methodological challenges raised in the course of the workshop should now be addressed in order to generate firm recommendations for researchers. We hope that this Current Opinion article will stimulate debate and, ultimately, galvanise the health economics community into action, resulting in better tools with which to generate the cost-effectiveness estimates that ultimately determine the interventions that patients will and will not be entitled to receive.

Acknowledgments We thank the workshop attendees for their enthusiastic participation and constructive suggestions, and the manuscript's reviewers for their helpful comments. This work was funded by the MRC Network of Hubs for Trials Methodology Research.

Author contributions JT co-ordinated and wrote all drafts of the paper, and is the guarantor for the overall content. All other authors contributed verbal presentations to the workshop, revised the drafts for intellectual content and read and approved the final manuscript. WH conceived the idea for the paper.

References

1. Johnston K, Buxton MJ, Jones DR, et al. Assessing the costs of healthcare technologies in clinical trials. *Health Technology Assessment*. 1999;3(6):1–76.
2. Ridyard CH, Hughes DA. DIRUM team. Development of a database of instruments for resource use measurement: purpose, feasibility and design. *Value in Health*. 2012;15:650–5.
3. Morris S, Devlin N, Parkin D. *Economic analysis in health care*. New York: Wiley; 2007.
4. Noble SM, Hollingworth W, Tilling K. Missing data in trial-based cost-effectiveness analysis: The current state of play. *Health economics*. 2012;21(2):187–200.
5. Edwards PJ, Roberts I, Clarke MJ, et al. Methods to increase response to postal and electronic questionnaires. *Cochrane Database Syst Rev* 2009; (3): MR000008.
6. Rolstad S, Adler J, Rydén A. Response burden and questionnaire length: Is shorter better? A review and meta-analysis. *Value in Health*. 2011;14(8):1101–8.
7. Marques E, Johnson E, Goberman-Hill RJS, et al. Using resource use logs to reduce the amount of missing data in economic evaluations alongside trials. *Value Health*. 2013;16(1): 195–201.
8. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, Brewin C, Wing J, editors. *Measuring mental health needs*. London: Gaskell; 1992. p. 179–90.

9. Simpson S, Corney R, Fitzgerald P, et al. A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression. *Health Technol Assess (Winchester, England)* 2000; 4(36): 1–83.
10. Patel A, Rendu A, Moran P, et al. A comparison of two methods of collecting economic data in primary care. *Fam Pract.* 2005;22(3):323.
11. Byford S, Leese M, Knapp M, et al. Comparison of alternative methods of collection of service use data for the economic evaluation of health care interventions. *Health Econ.* 2007;16(5):531–6.
12. Mirandola M, Bisoffi G, Bonizzato P, et al. Collecting psychiatric resources utilisation data to calculate costs of care: a comparison between a service receipt interview and a case register. *Social Psychiatry Psychiatr Epidemiol.* 1999;34(10):541–7.
13. Thompson S, Wordsworth S. An annotated cost questionnaire for completion by patients. HERU Discussion Paper 03/01; 2001.
14. Wordsworth S. Improving the transferability of costing results in economic evaluation: an application to dialysis therapy for end-stage renal disease. UK: University of Aberdeen; 2004.
15. Tarride JE, Burke N, Bischof M, et al. A review of health utilities across conditions common in paediatric and adult populations. *Health Qual Life Outcomes.* 2010;8:12.
16. Hounsome N, Orrell M, Tudor Edwards R. Eq-5d as a quality of life measure in people with dementia and their carers: evidence and key issues. *Value Health.* 2011;14(2):390–9.
17. van Asselt AD, van Mastrigt GA, Dirksen CD, et al. How to deal with cost differences at baseline. *Pharmacoeconomics.* 2009;27(6):519–28.
18. Ridyard CH, Hughes DA. Methods for the collection of resource use data within clinical trials: a systematic review of studies funded by the UK health technology assessment program. *Value Health.* 2010;13(8):867–72.
19. Tourangeau R. Cognitive sciences and survey methods. In: Jabine TB, Straf ML, Tanur JM, Tourangeau R, editors. *Cognitive aspects of survey methodology: building a bridge between disciplines.* Washington: National Academic Press; 1984. p. 73–100.
20. Willis GB. *Cognitive interviewing: a tool for improving questionnaire design.* Thousand Oaks: Sage Publications, Inc.; 2005.
21. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use.* 3rd ed. Oxford: Oxford University Press; 2008.