P23: USING QUALITATIVE METHODS FOR ATTRIBUTE DEVELOPMENT FOR DISCRETE CHOICE EXPERIMENTS: ISSUES AND RECOMMENDATIONS

Joanna Coast,
Health Economics Unit, University of Birmingham
Hareth Al-Janabi,
Health Economics Unit, University of Birmingham
Eileen Sutton,
Department of Social and Policy Sciences, University of Bath
Sue Horrocks,
Faculty of Health and Life Sciences, University of West of England, Bristol
Jane Vosper,
Department of Social Medicine, University of Bristol
Dawn Swancutt,
Primary Care Clinical Sciences, University of Birmingham
Terry Flynn,
CenSoC, University of Technology Sydney

Address for correspondence:
Joanna Coast
Professor of Health Economics
Health Economics Unit
School of Health & Population Sciences
Public Health Building
University of Birmingham
Birmingham
B15 2TT
tel: 0121 414 3056
E-mail: j.coast@bham.ac.uk

Acknowledgements: We would like to thank the two main collaborators in the ICEPOP programme: Tim Peters and Jordan Louviere. We would like to thank Ini Grewal and Jane Lewis for their roles in the qualitative work on the quality of life index for older people; Karen Forbes, Julian Abel, Hilary Jennings, Alison Rich, Sue Perry for their

collaboration on the work on end of life care; David Kessler, Glyn Lewis, Mel Chalder, Katrina Turner, Anne Haase for assistance with mental health research collaborations; Chris Salisbury, David de Berker, Alison Noble, for their assistance in collecting dermatology data for use in methodological work; Sheila Greenfield, Sue Wilson for their supervision of the coloscopy work. We would like to thank the advisory group for ICEPOP who have commented freely on a number of these issues: Richard Smith (chair), John Brazier, David Cohen, John Gladman, Jeremy Horwood, Richard Huxtable, Liz Lloyd, Colette Reid, Mandy Ryan and Phyllis Watkins, and Steve Chaffey for administrative assistance. Mostly, we would like to thank all those informants and research participants who have freely given of their time to assist us in this research. The ICEPOP programme was supported by the MRC Health Services Research Collaboration.

ABSTRACT

Background: Attribute generation for discrete choice experiments (DCEs) is often poorly reported and it is unclear whether this element of research is generally conducted rigorously.

Aims: To explore general issues associated with the development of attributes for DCEs; to contrast different qualitative approaches to attribute development; to generate recommendations for using qualitative approaches to attribute development.

Data: The paper draws on data from seven studies. Three developed attributes for measures, four developed attributes for study specific questions. Five used interviews (in-depth or semi-structured), one used focus groups and one combined meta-ethnography with interviews.

Issues: Issues that have become apparent through these studies include: the conceptual framework for Random Utility Theory and the need for attributes that are neither too close to the latent construct nor too intrinsic to people's personality; the need to conceptualise attribute development as a two stage process involving conceptual development followed by development of language to evoke meaning; difficulty in resolving tensions inherent in the reductiveness of drawing precise terms from complex and nuanced qualitative findings.

Qualitative approaches: The paper presents comparisons of the advantages and disadvantages of alternative qualitative methods focusing on the need for qualitative research skills, time requirements, the 'richness' of attributes, sensitivity of the subject and potential for mis-specification.

Recommendations / discussion: We would like to draw on the experience of HESG members in terms of whether the issues we describe are resonant with their own experiences, and discuss the recommendations we suggest for future practice and reporting.

INTRODUCTION

Over the past ten years, discrete choice experiments (DCEs) have become an increasingly popular method of valuing outcome for economic studies of health and social care interventions. DCEs have been used as a means of valuation both for studies developing measures for use across different economic evaluations,[1] and for studies developing specific outcomes of interest for the particular evaluation in question.[2;3] DCEs form a particularly useful method of valuation for these latter studies where the outcomes of interest may not be entirely health related and where the methods typically used in (extra-welfarist) evaluation may be seen as inappropriate, involving, as they do, comparisons against death.

In the field of health economics, the general methods for DCEs (or as they were often referred to at the time, conjoint analysis studies) have been laid out by Ryan.[4;5] In these descriptions of the methodology of DCE, Ryan described the first two stages as being concerned with attribute development.[4] These stages are concerned with developing attributes that are the basis for the decision and then defining levels of those attributes for the experiment.[4] The information about what an attribute should look like is relatively limited, with the advice being limited to its being important to patients/policy makers, 'plausible' and capable of being traded.[4]

Nevertheless, and despite the lack of clarity about how attributes should look in health economic studies, a variety of means have been used to actually obtain attributes for empirical studies. These include literature review;[6-10] theoretical arguments from the literature;[11;12] existing health outcome measures;[13;14] professional recommendations;[8] focus groups, interviews or other consultation with staff or patients;[6;9;11;15;16] surveys with patients;[15] and expert review.[6;7] In one study attributes were described as having been defined by the policy questions[3] and in others[17;18] they were:

> *…chosen on the basis that there were statistically significant differences in the levels between the surgical and medical arm in the randomized trial. (Ryan & Hughes,[18] p.263)*

It is, however, highly recommended by one of the main experts in the field that qualitative work is conducted during attribute development:

> *We cannot overemphasise how important it is to conduct this kind of qualitative, exploratory work to guide subsequent phases of the stated preference study. (Louviere et al,[19] p.257)*
>
> *Study team members should keep in mind that consumers are always right about the way they think about products… they buy the product, after all! Hence the study team should endeavour to understand the dimensions… along which the product is evaluated by consumers and how specific levels of these dimensions*

are expressed… The primary objective is to be able to express the characteristics of the product to decision makers in the terms they employ. (Louviere et al,[19] pp.257-258)

Although these practitioners are enthusiastic about the potential for use of qualitative methods, it is not clear to what extent qualitative methods can or should be used relative to other methods. It is also not clear what particular qualitative data collection and analytical methods are most appropriate for the generation of attributes,[20] with Louviere and colleagues providing little guidance, beyond a suggestion that focus groups can be particularly useful[19] (which may be related to the marketing context in which they work, given that the focus groups are seen as a particularly useful rapid appraisal tool within this context).

There are also concerns about the reporting of attribute development, with the information provided generally being excessively brief, as the following quotes indicate:

The need to detect gastrointestinal malignancy early was discussed among four consultant gastroenterologists in one teaching hospital followed by wider consultation with nursing and clerical outpatient staff. Four key attributes were identified that could be altered to improve outpatient services… (Moayyedi,[15] p.430)

…derived from the literature on patient satisfaction with out of hours care, and also reflected the different models of care available across the country. (Scott,[10] p.388)[10]

Attribute selection focused on the parameters on which data were collected in the context of a clinical trial. (Bryan et al,[17] p.596)

To identify the key attributes of the programme, six focus groups were conducted with individuals who had used the bowel scan test kit. (Salkeld et al,[16] p.267)

It should be noted, however, that the empirical studies do not report any detail about how the information was distilled from the review or transcripts to form the chosen attributes. As has been observed previously,[20] where qualitative methods are used, no information is provided about the sampling, recording and transcription, analytical methods or reflexivity (the idea that the researcher should reflect on his/her role in the research and upon what is being found). Where other methods are used information is also in scant supply. For literature reviews: what search terms were used? How many abstracts were scrutinised? What inclusion and exclusion criteria were used? What methods of data extraction? What methods of data synthesis? For studies basing attributes on policy questions or trial results, how, for example, were policy options narrowed down, or trial outcomes selected? How were decisions made about the wording of attributes?

Clearly, although many DCEs are being conducted, there are still a number of questions around attribute development that remain to be resolved. Here four main questions have been identified as:

- What should attributes look like?
- Should qualitative methods be used, and to what extent?
- What are the advantages and disadvantages of different qualitative methods for developing attributes?
- How should attributes be reported?

This paper draws on the experience of an MRC research programme[21] in formulating answers to each of these questions. The paper begins by describing the seven DCE studies that were involved in this programme, each of which used qualitative methods to develop attributes. It then goes on to examine each of the four research questions out lined above, and then ends with a general discussion, that focuses particularly on recommendations for the future in terms of practice and research.

DESCRIPTION OF STUDIES

This paper draws on the experience of conducting seven DCEs in a variety of areas some of which are complete and fully published, others of which are ongoing. Three of the studies have been concerned with the development of measures for long-term use in economic evaluation and four have been concerned with more *ad hoc* policy-specific questions. Each of the studies is described briefly below and given a brief title to aid further discussion related to the individual studies later in the paper.

Studies developing measures

The capability study (2004-2008)

Forty in-depth interviews with older people, analysed using a framework approach, were used to develop attributes for an index measure of capability for older people. Five conceptual attributes were developed (attachment, security, role, enjoyment, control). A second stage of semi-structured interviews (n=19) analysed using constant comparison was used to confirm the attributes, develop meaningful wording, set levels and conduct a small amount of cognitive interviewing. Full details of the attribute development are published elsewhere.[1;22]

The carer study (2005-2009)

Meta-ethnography, a method for synthesising qualitative findings,[23] was used to combine the findings of qualitative studies on caring and determine attributes for the caring experience. Semi-structured interviews with 16 carers were then conducted iteratively (findings of early interviews influencing the conduct of subsequent interviews) and analysed using framework to confirm and refine attributes, develop meaningful language and set the number and wording of levels. The final measure contained six attributes (getting on, organisational assistance, social support, activities, control and fulfilment). Full details of the attribute development are published elsewhere.[24]

The end of life study (2006-2009)

Twenty-three in-depth interviews were conducted with older people from the general population, care homes and receiving palliative care, and analysed using constant comparison. These were used to develop seven conceptual attributes (Dignity, Autonomy, Physical Suffering, Emotional Suffering, Affection, Support, Completion) for an index measure to evaluate outcomes in end of life care. A second stage of 12 further in-depth interviews, also analysed using constant comparison, was used to confirm attributes and develop meaningful wording. This study is nearing completion.

Policy-specific studies

The dermatology study (2003-2007)

This study aimed to quantify patient preferences for access to services alongside other important factors in the context of a trial[25] comparing outpatient care for dermatology patients with care provided by a GP with special interests.[2] For development of attributes, semi-structured interviews were conducted with dermatology patients, purposively selected to obtain the views and experiences of people pre- or post- service and with an acute or chronic diagnosis. Using a combination of constant comparative and content analytical methods, analysis was conducted to develop and confirm conceptual attributes, and to develop meaningful language for these attributes. Interviews were carried out iteratively in three phases and four attributes (time waited, expertise, convenience, individualised care) were developed. Full details of the attribute development are published elsewhere.[20]

The colposcopy study (2005-2008)

This study aimed to identify women's preferences for aspects of their colposcopy appointments. In-depth interviews were conducted with 18 women attending two clinics in the West Midlands. The framework method of analysis was used to ascertain important factors in women's experience of colposcopy. Six attributes were identified: gender of the colposcopist, viewing the procedure on a monitor, staff attitude, waiting time, information provision and being rushed during the appointment. Precise wording of the DCE attributes was piloted on women within the screening age group (25-64).

The depression exercise study (2007-)

This study aims to assess patient preferences for attributes of a physical activity intervention for the treatment of depression. Semi-structured interviews were conducted with 12 patients participating in a trial investigating the use of a physical activity intervention for depression. Eight staff members (physical activity facilitators and GPs) associated with that trial were also interviewed. Interviews were conducted iteratively and analysed thematically. Four attributes resulted (Support, goals, time, choice/cost of activity). The study is ongoing.

The depression drug study (2007-)

This study aims to assess the preferences for drug treatment of depression. Five focus groups were conducted with 33 depression sufferers attending self-help groups or recruited by their GP. Focus groups were conducted iteratively, analysed thematically and the results interpreted in the light of the current literature on effectiveness and adverse side-effects of drug treatment for depression. Seven attributes resulted including adverse effects (weight gain, sexual dysfunction, drowsiness/sleep problems, anxiety/agitation, nausea/dizziness/headaches), effectiveness and class of medicine. The study is ongoing.

WHAT SHOULD ATTRIBUTES LOOK LIKE?

The first issue is that it has become apparent that attribute development studies need to address what criteria these conceptual attributes should fill. Current advice that attributes should be important to patients and/or policy makers, 'plausible' and capable of being traded,[4] clearly provides an important starting point, but is not sufficient in providing a complete description of the criteria that attributes should fill.

Four additional criteria that need to be fulfilled are associated with the conceptual framework for Random Utility Theory[26;27] such that this theory, which forms the psychological basis for DCEs is not violated. First, the attributes selected should include all those attributes that might be important for an individual in coming to a decision. If important attributes are ignored in the DCE this is likely to bias the findings obtained.

Second, the chosen attributes should not be too close to the latent construct that the DCE is designed to investigate. 'Too close' should be clear from common sense and piloting of the DCE. Thus, where utility is the underlying latent construct, common sense dictates not including an attribute within the DCE that is either labelled as utility or that expresses overall happiness or enjoyment with the good or service being valued. Essentially, where this problem is not avoided, the researcher would be including both the underlying construct in the measure and factors that contribute separately to that underlying construct. Since the framework of DCEs is consistent with Lancaster's characteristics theory,[28] this can have the effect of causing the attribute in question to dominate all others. Essentially the model becomes an identity relationship, telling the researcher nothing about the other attributes. Such problems can generally be identified during piloting of the DCE if they are not picked up during the initial attribute development process: the attribute in question will exhibit extremely large, significant, coefficients whilst those of all other attributes will be small and non-significant.

This issue, of the attribute being too close to the latent construct, is perhaps more likely to occur among the development of attributes for generic measures of the type developed here for the capability, carer and end of life studies. For development of such measures it is thus important to have a clear idea of what the underlying construct is, to avoid the potential for inadvertently including attributes that are too close to this underlying construct. For example, for the end of life study, the aim was to develop a measure of a good death, and thus the latent construct of interest here was that of a scale of 'good death'. It is also important to avoid attributes that are too close to this latent construct. With the end of life study there was some concern that an early theme that seemed to emerge as being important to informants, that they could continue 'to be myself' as they moved towards death was too closely related to the latent construct of a good death, in that, by having high levels of all the other attributes that had emerged, a person would inevitably be able to 'be myself' (or, put another way, not 'being myself' would not realistically allow high levels of the other attributes). Further, within an area such as end of life care, areas such as dignity and identity may be very difficult to separate from the latent construct, but potentially increase the acceptability of a measure for use by clinical colleagues.

Third, single attributes should not have such a large impact on decisions that respondents essentially make no errors in decision-making. Such attributes can cause individual's choices to become deterministic rather than stochastic and, again, violate the theory of random utility. A DCE will not, therefore, provide meaningful welfare estimates or trade-offs. Attributes exhibiting these characteristics may be possible to identify during qualitative work but again,

they can also be identified at the piloting stage – in this case, large coefficient estimates and non-convergence of the regression model is likely. Even when sample level estimates appear intuitive, subgroups of respondents might exhibit such preferences: it should be noted that serious mistakes in QALY ranking studies have been made by researchers who have averaged over people who do and don't conform to RUT.[29]

Fourth, the chosen attributes should not be too intrinsic to a person's personality and therefore aspects of a person that are not experimentally manipulable by policy or intervention should not be chosen as attributes. An example should help to make this clearer. One theme that was identified during the qualitative work for the depression-drug study was that of aversion to medication: some sufferers were particularly averse to taking medication, not just for their depression, but for any illness. Whilst this was an important theme within the qualitative work, general aversion to taking medicines is arguably not something that can be experimentally manipulated, but instead is a variable associated with the person that might affect their decision-making. It might therefore be useful information to collect from people to aid in analysing sub-groups of patients, but it does not form an attribute. Clearly, however, there will always be some debate about the extent to which some characteristics are manipulable or not, and some may hold the view that there are few characteristics that cannot be ameliorated if the benefits sufficiently outweigh the costs.

EXTENT OF USE OF QUALITATIVE METHODS

It has become very clear during this programme of work that attribute development is a two stage process, whether these two stages are, or are not, explicitly distinguished. The two stages of the process of attribute development are, first, conceptual development and second, development of language to evoke meaning. A third stage,* not discussed here, is the development of levels of attributes.

Conceptual development involves generating the attributes or dimensions or factors that together provide information about the latent construct. At this conceptual development stage academic terms are often used to describe the attribute. A good example of this is the capability study,[22] where five conceptual attributes were generated with headings of attachment, security, role, enjoyment and control, and descriptors that were relatively long and all encompassing were used to capture the meaning of these concepts for an academic audience. The conceptual

* The second stage according to Ryan.[4]

levels were not, however, particularly meaningful in describing the attributes to older people or to policy makers, and thus for the purpose of the valuation task (and the subsequent measure) a second stage of attribute development was required, around the development of language that more clearly evoked, for a general audience, the meanings behind the conceptual labels. During this second stage conceptual labels need to be 'translated' into lay language, ensuring that the language used conjures up similar factors for a lay audience as the conceptual labels conjure up for an academic audience whilst retaining the mutually exclusive nature of the conceptual attributes.[1]

Although these two stages have been clearly defined in some of the studies in this programme (for example, the capability and carer studies) and less clearly defined in other studies (for example, the dermatology study), it is plain that these two stages occur in all studies. For example, in the dermatology study where there were three iterations of qualitative work and the different attributes developed at different rates, each attribute was first defined and wording to capture the essence of that attribute was then generated. Given that these two stages are required in all attribute development studies, it is worth thinking about them separately in terms of the extent to which qualitative work might be desirable in the generation of attributes.

Conceptual development

It is clear that there are a number of methods that might be used for conceptual development of attributes, of which qualitative methods are just one type. Other options include the generation of attributes through literature review, the use of expert or researcher opinion, and basing attributes directly on the policy question of interest.

Strengths of using qualitative work

The particular strength of using qualitative work for conceptual development of attributes is the ability to draw on the views of others outside of the research process, in particular, those directly affected by a decision, in a manner that is directly relevant to the topic. The attributes developed through qualitative research are often 'richer' than those that can be generated through alternative methods, based as they are on more complex and nuanced data. Qualitative methods also enable sensitive subjects to be discussed and attributes thus captured, and reduce the potential for mis-specification of attributes through over-reliance on the views of experts or researchers, or on the basis of theoretical axioms. Through the process of rigorous qualitative analysis, by scrutinising transcripts for the identification of patterns and themes, developing initial categories and sub-categories, and constantly modifying and extending these categories in a hierarchical manner, it is possible to arrive at a mutually exclusive, comprehensive set of attributes.[30;31] This is in contrast to a less comprehensive use of qualitative methods, to generate possible lists

of questions, followed by the use of statistical techniques to reduce these to a coherent set of dimensions – see Brazier et al for details of this alternative approach.[32]

An alternative to using qualitative methods, or a means of side-stepping their use, is to allow attributes to be developed directly in response to the policy question. Such an approach, for example, was used by van den Berg et al[33] who specified just three attributes of informal care (time spent caring, caring task, and money to enable derivation of a willingness to accept value for carer time). It is clear, however, that this approach may lead to important attributes being omitted (in this case, for example, whether the individual was able to engage in paid employment outside caring), and may not, in itself, provide sufficient reason for avoiding other methods, such as qualitative analysis, to ensure that all important attributes are captured.

Literature review may also be seen as an alternative to qualitative methods. A strength of using qualitative methods, however, is that through selection of a sample tailored to the specific question, it enables data to be collected that are directly related to the question under consideration (it is, on the other hand, seldom the case that literature is obtained that contains precisely the desired information).

Challenges in using qualitative work

Some of the challenges associated with using qualitative work to generate attributes are not distinct from those of using qualitative methods more generally within a health economics setting: the acceptability of the methods used for an economic audience (particularly in relation to small sample sizes and the use of a reflexive approach in which the influence of the researcher is acknowledged), the difficulty in maintaining a different epistemological position in which theory development (attribute development) is inductive rather than deductive, and the practical problems associated with the opportunity cost of generating qualitative research skills.[34] In particular, qualitative research may be seen as time consuming and relatively difficult to undertake for researchers who are not experienced in these methods. The alternatives of literature review, expert/researcher opinion and policy driven, do not require the development of these specialised research skills, and can be conducted in a much shorter time. These challenges tend particularly to face economists, in that they are associated with a potential lack of acceptability of the use of qualitative methods for generating attributes. These challenges may, however, be less difficult to deal with than those facing qualitative researchers, that is, the potential lack of acceptability to these researchers of using qualitative methods for this particular purpose.

The programme of work that forms the basis for this paper includes qualitative work conducted both by those who would primarily consider themselves economists but who are trained in qualitative methodology, and by those who would primarily consider themselves qualitative researchers (ethnographers, anthropologists and so on). For all these researchers, but particularly for those who primarily consider themselves to be qualitative researchers, there is

a clear tension, particularly at the analysis stage, between the complex and nuanced findings produced by qualitative research and the reductiveness associated with attribute development. The process of attribute development thus contrasts greatly with the usual processes of qualitative analysis. Generally, qualitative analysis aims to be inductive and expansive, looking at the co-dependencies and inter-relationships across the rich data obtained. For attribute development, the process involves drawing precise terms for conceptual attribute development, describing key concepts in as few distinct attributes as possible,[2] and then determining precise wording to describe these attributes, particularly where language may be interpreted more broadly. The conduct of the qualitative research opens the researcher's eyes to the complexities and nuances of their data and can leave the researcher feeling that 'the brief descriptions of attributes and their levels could never do justice to the complexity of individuals' preferences' (Coast & Horrocks, p.29[20]). There is clearly the potential for criticism from pure qualitative researchers as new qualitative techniques such as meta-ethnography are harnessed by economists for the purpose of attribute development.

Generating meaningful language

For the second stage of attribute development, associated with generating meaningful language, qualitative work, using iterative techniques to constantly refine the language used, seems to have been particularly successful in this programme of work. In particular, it has (i) ensured that the meaning desired is that evoked amongst potential survey respondents, and (ii) avoided the use of terminology that, although generated in a brainstorm amongst the research team, does not, when used in a questionnaire setting with lay people, suggest the wished for meaning. For example, the carer study, which generated conceptual attributes using meta-ethnography developed a conceptual attribute on relationships that was concerned with the ability of the carer to communicate with the person cared for and avoid arguments. The terminology of 'relationships', however, led some informants to consideration of their marriage vows. The terminology of 'getting on' with the cared for person was therefore introduced and was found to better reflect the underlying attribute for informants. Precisely because this second stage of the attribute development process is directly concerned with language and meaning, it is difficult to see how the use of literature or quantitative methods might provide a more efficient means of achieving this second stage.

CHOICE OF QUALITATIVE METHODS

Once a decision has been made to use qualitative methods, there is then a decision about the precise nature of the methods that should be used for attribute development for DCEs, and this next section of the paper presents, in Table 1, a comparison of the advantages and disadvantages of alternative qualitative data collection methods used within this programme of research before providing some examples. On the basis of this experience, each of interviews, focus groups and meta-ethnography is judged in terms of the level of qualitative research skills required, time requirements, the 'richness' of the resulting attributes and the potential for mis-specification of attributes through the approach. Brief comments are also provided for each method.

Clearly, each of the techniques outlined in Table 1 requires some element of qualitative research skill. These skills include those in both data collection (conducting interviews; facilitating focus groups) and data analysis (with different types of analysis requiring different skill levels). It is worth noting that in this programme of work qualitative analysis was successfully conducted by those trained primarily in economics who received additional training in qualitative methods, as well as by those trained primarily in qualitative methods; the different backgrounds of researchers led to different challenges in the use of qualitative methods to develop attributes as indicated above, but it was certainly feasible for economists to develop the required level of skill in conducting qualitative research.

The time required to conduct the qualitative research tended to be higher for those studies where the ultimate aim was the generation of a measure, rather than informing a particular policy question. This higher time involvement seemed to be largely related to the breadth of the topic. Those areas where the aim was to generate a measure, on the whole, covered broader topics than those where the aim of the DCE was more tightly focused on a specific policy issue. For example, interviews for the capability study involved discussion of informants' entire lives and interviews for the end of life study involved discussions of informants' views of death and dying, whereas, for example, interviews for the dermatology and colposcopy studies were much more tightly focused on the specific conditions. On the whole, therefore, data collection took longer, and the subsequent analysis was more complex and time-consuming. Given, however, that the intention is that measures should be used repeatedly, this greater investment in their development can be justified.

From the experience here, attributes resulting from interviews, particularly in-depth interviews, tended to be slightly richer in their content than those arising from focus groups. This can be explained by the greater time spent with individual informants during interviews than focus groups and the greater opportunity for in-depth exploration of particular issues and concepts. Inevitably, the opportunities for individuals to speak in-depth are reduced during focus groups where there are a number of people whose views all need to be heard. Compared with meta-

ethnography, interviews also appeared to perform well in terms of the richness of attributes although there was only one meta-ethnography performed in this programme of work. For meta-ethnography, it appears that the richness of the attributes generated will be heavily dependent on the material that is available for inclusion. The ability to include a much broader

Table 1—Characteristics of different qualitative methods for generating attributes

Qualitative approach	Qualitative research skills	Time	'Richness' of attributes	Potential for attribute mis-specification	Other comments
Interviews	Medium, high if in-depth	High for both data collection & analysis	High	Some, depending on sample	May be particularly useful for sensitive subjects that people are unwilling to discuss in a group
Focus groups	Medium	Medium for data collection (but requires presence of more than one researcher), can be high for analysis	Medium	Some, depending on sample	Can be difficult to analyse and to link views across individuals; may generate topics through discussion that the researcher was previously unaware of.
Meta-ethnography	Medium	Low for data collection, medium for analysis	Depends (see comment)	Depends (see comment)	Very dependent on the quality and richness of the material available for inclusion in the meta-ethnography.

range of experiences and informants than would ever be feasible in a primary study, suggests the possibility for richer attributes than from a primary focus group or interview study, but the necessary use of secondary, less detailed published data, rather than having access to a full set of primary data suggests the possibility that attributes

will be less rich or potentially less relevant for the particular issue under consideration. Of course, any meta-ethnography will also be heavily dependent on the extent and richness of the source material.

Even with qualitative methods, there will always be some potential for the mis-specification of attributes, depending on the sample included within the qualitative work. In well-designed qualitative work, such as the studies included in this programme, that uses purposive sampling with the aim of obtaining the full range of views about the issue under consideration, however, this potential should be minimised. It would certainly be much lower than would be expected based on non-qualitative methods such as, for example, expert or researcher opinion.

It is worth noting that the choice between interviews, focus groups and meta-ethnography may ultimately come down to practical considerations such as the sensitivity of the topic area, current knowledge of the topic amongst researchers and the extent of literature currently available in the area. For example, in the end of life study, focus groups were not considered appropriate because of the extreme sensitivity of the subject and the potential for distress to be caused during discussion, and the extent of available literature on the subject of a good death was relatively limited both generally, and specifically within the UK context. For the depression-drugs study, the particular benefit of focus groups was the opportunity to allow depression sufferers to discuss amongst themselves the adverse effects that they had experienced through medication for depression, enabling topics that the researcher might not have considered to be discussed.

REPORTING ATTRIBUTES

In the introduction to this paper, a number of quotes were given from previous studies, that outline the brevity with which the development of attributes in DCEs has, in the past, been reported. The problem with the lack of detail given in this reporting is that it is unclear whether the work has been conducted rigorously but has not been reported well, or whether it has not been conducted rigorously. There are a number of concerns with such brief reporting of the details of attribute development. First, where economists are conducting qualitative work it is not always clear that they have the necessary experience to do this work well, and so the lack of good reporting conventions makes it less likely that this important element of any DCE will be conducted rigorously. Second, where researchers entering this area want to draw on the experience of others who have previously conducted DCEs, there will be little information about attribute development, with each new study essentially ‘re-inventing the wheel’. Third, and as a

direct consequence of this constant 're-invention' the lack of reporting does not allow debate and subsequent methodological development.

For all these reasons, better reporting conventions are required for these studies. Such reporting conventions are required both to ensure better reporting in studies submitted for publication and to assist peer-reviewers of these DCE studies: clearly those most likely to be asked to review DCE studies in health care will be health economists and these peer-reviewers themselves may not feel qualified to comment on the qualitative element of the work and so may be happy with the one or two lines of explanation generally currently provided.

It is suggested here that, as a minimum, the following information should be reported:

- A rationale for the method used to develop attributes for the DCE;
- The nature of the sample included in the qualitative work (patients, clients, staff, public, etc.);
- Details of how the sample was obtained (sampling that is purposive, snowball, maximum variation, random etc.);
- Brief details of the topic schedule (questions asked);
- Details of who conducted the interviews and in what setting;
- Information about whether the interviews were transcribed;
- Details of the type of analysis conducted (for example, content analysis, framework, thematic analysis, constant comparison etc.);
- A brief description of the results obtained, including information about how many potential attributes were generated and how these were manipulated to become the final set of attributes;
- Whether any attribute might prove problematic in the DCE, for example, by being potentially 'too' important to any subgroups of respondents (and therefore likely to lead to deterministic decision-making).

The level of detail in reporting need not be the same for all studies. Where studies aim to produce instruments for future use, a paper in its own right outlining all these issues is of particular importance so that future users of the instrument can be clear about how the measure was generated. Such papers include those produced for the capability study[22] and carer study[24] within this programme. Where studies are more *ad hoc* and are conducted within the context of a particular policy issue such a specific paper may not be required (although such papers can be useful in informing debate, particularly where they outline new suggestions or describe methods previously not described, as for example with the dermatology study[20]) but instead a small amount of information can be placed in

the main study. This was found to be feasible within current word-limits for the main paper from the dermatology study.[2]

DISCUSSION & RECOMMENDATIONS

This paper draws on a set of seven studies that have all used comprehensive qualitative studies as an integral part of the development of a DCE for the valuation of particular outcomes. It should be reiterated here that the attributes included in these studies, and how they are worded, are fundamentally important for obtaining valuable end-results from DCEs and need, in general, to be given greater attention by health economists developing and conducting these studies.

There are clearly a number of limitations to what has been presented here. First, the paper relies on work conducted through one research programme. Although this research programme had a number of strands, the research team for all the different projects had common elements, thus there may be issues that have been faced by other teams of researchers working in other contexts that this team has not come across. In particular, this team has focused on the use of qualitative methods and there may be other perspectives on attribute generation from those who have used methods such as literature review or the use of expert opinion that are not included here. Second, even within this research team, greater experience has been generated in the use of interviews, rather than focus groups or meta-ethnography. Additional experience in these latter two areas may generate further insights.

There are a number of recommendations following on from the work. First, it is important for researchers to be clear about the latent construct that they are working with and aware of the possibility of generating attributes that are perhaps too close to this latent construct. Second, it is vital for researchers to be clear about the need for attributes to be manipulable in policy, such that attributes that are essentially personal characteristics are not included. Instead, these may be useful to identify during qualitative work and then collect information about in the subsequent valuation work for use in detailed sub-group analysis. In dealing with both these issues, it is valuable for researchers to repeatedly generate mock-ups and conduct pilot work to ensure that such attributes do not 'sneak into' the final DCE. Third, researchers should explicitly consider the two stages of attribute development – conceptual development and the generation of meaningful language – and how they will be dealt with, even where the two stages will run concurrently. Fourth, researchers should be clear about the methodological choices that they have made in relation to attribute development and the reasoning behind them, giving consideration to issues such as

their skills, the time commitment, the extent of current knowledge/literature in the area, the sensitivity of the topic and so on. Finally, researchers should adhere to the reporting conventions outlined in the section above.

In terms of further research in this area, direct comparisons of the attributes generated through different methods would be extremely interesting, although there are clearly a number of practical limitations on the development of such studies. There is also potential for using qualitative methods to explore the existence of interactions between attributes but precisely how this might be done needs further work: questions such as *Would having less on attribute A make attribute B more important to you?* would potentially inform this issue but could make interviews excessively long and the cognitive burden imposed by these questions might be greater than some respondents can cope with. Further potential for qualitative methods lies in the possibility of determining 'market segments' within health-related DCEs, that is, those groups that are likely to express different preferences.[19] It is important to identify such sub-groups to ensure that sufficient numbers are sampled across these different segments as a basis for later analysis. Research is required to determine whether qualitative methods can successfully identify such sub-groups. Similarly, research also needs to be conducted to determine whether qualitative methods can identify sources of variance (scale) heterogeneity. Identifying such sources is important because failure to properly model variance (scale) heterogeneity on the latent scale leads to biased DCE estimates.[35] This would involve establishing the success of qualitative methods both in identifying sub-groups who do not engage well with the DCE task and in identifying attributes that are more vulnerable to misunderstanding by respondents.*

This paper draws together the experiences from a relatively large number of DCE studies, but, as we suggest above, there may be other issues that we have not covered. In the HESG discussion, we would particularly like to draw on the experience of HESG members in relation to the following issues:

- Are the issues we describe resonant with others' experiences?
- Are there other aspects of qualitative analysis in relation to attribute development that we should cover?
- Is there agreement that the reporting around attribute development is poor?
- Are the recommendations we suggest for future reporting sufficiently comprehensive? Do you agree with this approach?

* if the variance of the random utility components associated with each of the various attributes is not constant, then not only will parameter estimates be biased, but they cannot be corrected by transforming into willingness to pay space. [louviere, ERE 2006].

REFERENCES

(1) Coast J, Flynn TN, Natarajan L, Sproston K, Lewis J, Louviere JJ et al. Valuing the ICECAP capability index for older people. Social Science and Medicine 2008; 67:874-882.
(2) Coast J, Salisbury C, de Berker D, Noble A, Horrocks S, Peters TJ et al. Preferences for aspects of a dermatology consultation. British Journal of Dermatology 2006; 155:387-392.
(3) Ryan M, McIntosh E, Dean T, Old P. Trade-offs between location and waiting times in the provision of health care: the case of elective surgery on the Isle of Wight. J Public Health Med 2000; 22:202-210.
(4) Ryan M. Using consumer preferences in health care decision making. The application of conjoint analysis. London: Office of Health Economics; 1996.
(5) Ryan M. Discrete choice experiments in health care. Br Med J 2004; 328:360-361.
(6) Phillips KA, Maddala T, Johnson FR. Measuring preferences for health care interventions using conjoint analysis: an application to HIV testing. Health Services Research 2002; 37:1681-1705.
(7) Hall J, Kenny P, King M, Louviere J, Viney R, Yeoh A. Using stated preference discrete choice modelling to evaluate the introduction of varicella vaccination. Health Economics 2002; 11:457-465.
(8) Hundley V, Ryan M, Graham W. Assessing women's preferences for intrapartum care. Birth 2001; 28:254-263.
(9) Ubach C, Scott A, French F, Awramenko M, Needham G. What do hospital consultants value about their jobs? A discrete choice experiment. Br Med J 2003; 326:1432-1438.
(10) Scott A. Identifying and analysing dominant preferences in discrete choice experiments: an application in health care. Journal of Economic Psychology 2002; 23:383-398.
(11) Bech M. Politicians' and hospital managers' trade-offs in the choice of reimbursement scheme: a discrete choice experiment. Health Policy 2003; 66:261-275.
(12) Ratcliffe J. Public preferences for the allocation of donor liver grafts for transplantation. Health Economics 2000; 9:137-148.
(13) McKenzie L, Cairns J, Osman L. Symptom-based outcome measures for asthma: the use of discrete choice methods to assess patient preferences. Health Policy 2001; 57:193-204.
(14) Johnson FR, Banzhaf MR, Desvousges WH. Willingness to pay for improved respiratory and cardiovascular health: a multiple-format stated-preference approach. Health Economics 2000; 9:295-317.
(15) Moayyedi P, Wardman M, Toner J, Ryan M, Duffett S. Establishing patient preferences for gastroenterology clinic reorganization using conjoint analysis. European Journal of Gastroenterology & Hepatology 2002; 14:429-433.
(16) Salkeld G, Ryan M, Short L. The veil of experience: do consumers prefer what they know best? Health Economics 2000; 9:267-270.
(17) Bryan S, Buxton M, Sheldon R, Grant A. The use of magnetic resonance imaging for the investigation of knee injuries: an investigation of preferences. Health Economics 1998; 7:595-604.
(18) Ryan M, Hughes J. Using conjoint analysis to assess women's preferences for miscarriage management. Health Economics 2004; 6:261-273.
(19) Louviere JJ, Hensher DA, Swait JD. Stated choice methods - analysis and application. Cambridge: Cambridge University Press; 2000.
(20) Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. J Health Serv Res Pol 2007; 12:25-30.
(21) Coast J, Flynn N, Sutton E, Al-Janabi H, Vosper J, Lavender S et al. Investigating choice experiments for perferences of older people (ICEPOP): Evaluative spaces in health economics. J Health Serv Res Pol 2008; 13(Suppl 3):31-37.

(22) Grewal I, Lewis J, Flynn TN, Brown J, Bond J, Coast J. Developing attributes for a generic quality of life measure for older people: preferences or capabilities? Social Science and Medicine 2006; 62:1891-1901.
(23) Noblit G, Hare D. Meta-ethnography: synthesizing qualitative studies. Sage; 1988.
(24) Al-Janabi H, Coast J, Flynn TN. What do people value when they provide unpaid care to an older person? A meta-ethnography with interview follow-up. (In press.). Social Science and Medicine 2008;1-11.
(25) Salisbury C, Noble A, Horrocks S, Crosby Z, Harrison V, Coast J et al. Evaluation of a General Practitioner with Special Interest service for dermatology: randomised controlled trial. Br Med J 2005; 331:1441-1446.
(26) Thurstone LL. A law of comparative judgement. Psychological Review 1927; 34:273-286.
(27) McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, editor. Frontiers in econometrics. New York: Academic Press; 1974. 105-142.
(28) Lancaster K. A new approach to consumer theory. Journal of Political Economy 1966; 74:134-157.
(29) Flynn TN, Louviere JJ, Marley AAJ, Coast J, Peters TJ. Rescaling quality of life tariffs from discrete choice experiments for use as QALYs: a cautionary tale. Population Health Metrics 2008; 6(6).
(30) Glaser BG, Strauss AL. The discovery of grounded theory: strategies for qualitative research. London: Weidenfeld and Nicolson; 1968.
(31) Strauss A, Corbin J. Basics of qualitative research. Grounded theory procedures and techniques. London: Sage; 1990.
(32) Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. Measuring and valuing health benefits for economic evaluation. Oxford: Oxford University Press; 2007.
(33) van den Berg B, Maiwenn A, Brouwer W, van Exel J, Koopmanschap MA. Economic valuation of informal care: the conjoint measurement method applied to informal caregiving. Social Science and Medicine 2005; 61:1342-1355.
(34) Coast J, McDonald R, Baker R. Issues arising from the use of qualitative methods in health economics. Journal of Health Services Research and Policy 9, 171-176. 2004.
(35) Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinomial logit models. Journal of Marketing Research 1993; 30:305-314.