Health Economics, Policy and Law (2010), **5**, 91–111
© Cambridge University Press 2009 doi:10.1017/S1744133109990016

Debate

Contingent valuation: what needs to be done?

RICHARD D. SMITH*
Health Policy Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK
TRACEY H. SACH
School of Chemical Sciences and Pharmacy and Health Economics Group, University of East Anglia, Norwich, UK

Abstract: Contingent valuation (CV) has been argued to have theoretical advantages over other approaches for benefit valuation used by health economists. Yet, in reality, the technique appears not to have realised these advantages when applied to health-care issues, such that its influence in decision-making at national levels has been non-existent within the health sector. This is not a result of a lack of methodological work in the area, which has continued to flourish. Rather, it is a result of such activities being undertaken in a rather uncoordinated and unsystematic fashion, leading CV to be akin to a 'ship without a sail'. This paper utilises a systematic review of the CV literature in health to illustrate some important points concerning the conduct of CV studies, before providing a comment on what the remaining policy and research priorities are for the technique, and proposing a guideline for such studies. It is hoped that this will initiate some wider and rigorous debate on the future of the CV technique in order to make it seaworthy, give it direction and provide the right momentum.

Contingent valuation: a ship without a sail?

Contingent valuation (CV) provides a survey-based method for estimating the monetary benefits of non-marketed goods and services (see Mitchell and Carson (1989) and Bateman *et al*. (2002) for fuller description of the method). Although CV has been applied in the environmental and transport literature since the early 1960s, and was applied in the health area in the mid 1970s (Acton, 1973, 1976), it remained neglected by health economists until the late 1980s and early 1990s (see Diener *et al*. (1998), Klose (1999) and Olsen and Smith (2001)). Since that time, the number of CV studies in health care has increased rapidly

*Correspondence to: Richard D. Smith, Health Policy Unit, Department of Public Health and Policy, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK. Email: Richard.Smith@lshtm.ac.uk

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

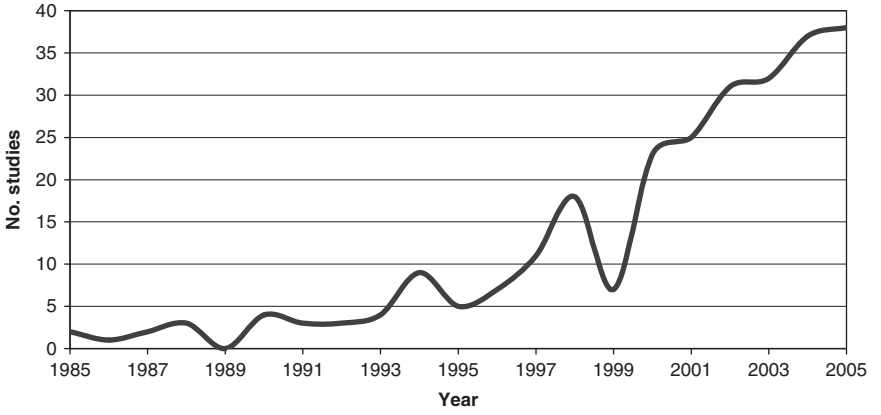


Figure 1. Number of CV studies per year (1985–2005)

(see Diener *et al.* (1998), Klose (1999), Olsen and Smith (2001), Smith (2003) and Sach *et al.* (2007)), covering a wide variety of disease areas and countries; although this number remains small compared to the number of cost-effectiveness and cost–utility studies conducted over the same period (for instance, the OHE (Office of Health Economics) Health Economic Evaluation Database holds 35,000 plus references (http://www.ohe-heed.com/)).

Using data from a CV database compiled by the authors following on-going systematic reviews of the literature, the methods of which have been described extensively elsewhere (Olsen and Smith, 2001; Smith, 2003; Sach *et al.*, 2007), Figure 1 illustrates that in the 5 years, from 2001 to 2005, 163 CV studies were published.

Table 1 shows that the USA and UK remain the largest producers of CV studies in health care (Sweden contributed a number of studies during the 1990s, but activity has declined recently). However, the most obvious change in the last few years has been the application of CV in countries other than the USA, UK, Canada and Sweden (the main, or only, players in the 1980s and 1990s). From 2000, at least one CV study in health care had been conducted in 35 other countries. CV thus appears to be making inroads in a number of different health systems and cultures, and especially so in some developing countries where often these studies are used to estimate possible demand and co-payment rates, and thus have a very specific policy focus.

Table 2 shows that many of the early studies were focussed on cardiovascular disease – predominantly by Johannesson in Sweden – but the wider geographical use of CV illustrated above coincides with a broadening of clinical area from the mid-1990s.

Table 3 shows that a significant, but nonetheless a minority, of interventions evaluated have been pharmaceutical products. However, since 2000, there have been more studies focussed on screening for various diseases – perhaps because a

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 1. Country of origin by year

Year	UK	USA	Canada	Sweden	Other	Total
1985	0	2	0	0	0	2
1986	0	1	0	0	0	1
1987	0	1	0	0	1	2
1988	1	2	0	0	0	3
1989	0	0	0	0	0	0
1990	0	2	1	0	0	3
1991	0	1	1	2	0	4
1992	1	1	0	1	0	3
1993	0	1	0	2	1	4
1994	0	4	1	1	3	9
1995	1	1	2	1	0	5
1996	1	1	0	3	2	7
1997	6	3	0	2	0	11
1998	3	10	0	2	3	18
1999	1	4	0	0	2	7
2000	3	9	2	1	8	23
2001	5	14	1	1	4	25
2002	7	8	2	0	14	31
2003	7	8	5	0	12	32
2004	8	11	1	1	16	37
2005	0	11	7	0	20	38
Total	44	95	23	17	86	265

significant component part of screening concerns the value of information, not usually captured in other evaluation methods.

It can be seen, therefore, that there has been an expansion in total number, breadth of geographical coverage (and hence health systems and cultures), clinical areas, and interventions. The popularity of CV thus shows no signs of abatement – indeed, quite the opposite – and this makes it timely to critically review the development of CV and raise future research directions.

Despite the technique's growing popularity, there remain a number of critical unresolved issues. These may be categorised as concerning: (i) whether or not there are (or could be) substantial advantages to the use of the technique over alternative outcome measurement strategies; and (ii) the appropriate methodology for conducting CV generally, and in its application in health economics more specifically (Cookson, 2003; Hanley *et al.*, 2003).

Numerous *specific* CV research questions could be raised as important in a paper such as this. For instance, what the 'threshold' income level is for impacting on sensitivity, the impact of prominent numbers on valuations for a wide range of goods, and the impact of response acquiescence or 'yeah-saying'. These have been considered in more detail, including by the authors, elsewhere (e.g. Smith, 2001, 2004, 2005, 2007a, 2007b, 2008; Yeung *et al.*, 2003, 2006;

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 2. Clinical area/diagnosis over time

Year	Cardiovascular	Cancer	Obstetrics and gynecology	Infection	Mental illness	Arthritis	Other*	Total
1985	0	0	0	0	0	0	2	2
1986	0	0	0	0	0	1	0	1
1987	0	0	1	0	0	0	1	2
1988	1	0	0	0	0	0	2	3
1989	0	0	0	0	0	0	0	0
1990	0	0	0	0	0	0	3	3
1991	2	0	0	0	0	0	2	4
1992	1	0	0	0	0	0	2	3
1993	3	0	0	0	0	0	1	4
1994	1	0	2	1	0	0	5	9
1995	0	0	2	0	1	0	2	5
1996	2	0	1	1	0	0	3	7
1997	1	0	3	0	0	0	7	11
1998	0	1	3	4	0	1	9	18
1999	0	0	0	1	1	0	5	7
2000	0	2	1	4	0	2	14	23
2001	2	5	1	6	2	0	9	25
2002	1	1	0	12	1	1	15	31
2003	2	5	0	5	3	1	16	32
2004	2	4	2	6	0	2	21	37
2005	0	5	0	13	2	1	17	38
Total	18	23	16	53	10	9	136	265

* Other includes, for example, different locations for care, irradiation of food, asthma and respiratory illness, blood products, and information from pharmacists on medicine use.

Covey and Smith, 2006; Whynes *et al*., 2007). However, focus on such questions, important as they may be for methodological advancement, would miss more significant and fundamental issues facing the technique. The discussion contained in this paper therefore invites health economists to consider: (i) whether there is a role for CV in health economics; and (ii), if yes (or perhaps a weaker not sure), how can CV development progress to a stage to realise the desired potential, in light of developments over the last 20 years of research in the area?

At present, we would suggest that CV is like a ship without a sail – it is adrift in the sea of economic evaluation methodologies, unsure of a direction and without the ability to move in a purposeful manner to achieve its destination. Throughout the paper the metaphor of the ship is used, since for a boat to float it first needs to be built correctly, which requires an in depth understanding of the dynamics and dimensions of a successful boat; that all components are present to make it move; and that once moving, an experienced captain oversees the direction and speed of travel. This is a useful metaphor for CV because uncertainties remain about the exact dynamics and dimensions of the technique,

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 3. Intervention by year

Year	Drug	Surgery	Screening	Other*	Total
1985	0	0	0	2	2
1986	1	0	0	0	1
1987	0	0	0	2	2
1988	0	0	2	1	3
1989	0	0	0	0	0
1990	0	0	0	3	3
1991	3	0	0	1	4
1992	1	0	0	2	3
1993	2	0	0	2	4
1994	1	0	0	8	9
1995	1	1	1	2	5
1996	2	0	0	5	7
1997	1	1	2	7	11
1998	8	0	1	9	18
1999	1	0	2	4	7
2000	6	3	4	10	23
2001	7	1	5	12	25
2002	15	0	3	13	31
2003	8	1	4	19	32
2004	10	2	3	22	37
2005	20	1	1	16	38
Total	87	10	28	140	265

* Other includes, for example, information provisions, unspecified movement between health states, and irradiation.

and earlier research has proceeded in a rather ad hoc and uncoordinated manner that has resulted in the technique having limited influence in decision-making in health care.

Is CV seaworthy?

The first major question to ask is whether the good ship CV is ready to be taken to the sea. Is the technique worth further research, time, and money being invested into it so that it can form part of the health economists' toolkit? This depends on whether there is a realisable advantage to using CV over other techniques. If the answer to this question is no then CV should be abandoned as unseaworthy, and efforts put in to developing other ships.

In this respect, Olsen and Smith (2001) advanced three arguments why CV may offer real advantage over other outcome measures. These included its foundations in welfare economics, that it does not restrict outcomes assessed to health outcomes only, and that monetary valuation of benefits and costs is required for decisions about allocative efficiency (that is, it allows a full cost–benefit

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

analysis (CBA) to be undertaken). The first of these arguments was largely refuted as unimportant on the basis that it is more important for a method to fit with societal value judgements, which may be expected to differ across countries and health systems. For instance, one could argue that the failure by the National Institute for Health and Clinical Excellence (NICE), in the UK, to adopt CV as an approach in its methods manual is one indication that UK society does not share the values of CV within health care (NICE, 2004). That this may reflect the situation that health care is free at the point of use, does not seem a logical argument since NICE still has to make a decision about societal willingness to pay for a quality-adjusted life year (QALY; as discussed below); although, removing this judgement from the individual level may in some way make it appear more acceptable.

In terms of the second argument, Olsen and Smith (2001) categorised the benefits measured by CV as: health outcomes, where three dimensions were identified including change in health state, duration, and probability (which are those assessed by the QALY); process utility, which may also derive from the use of health care; and non-health outcomes, covering option value and caring externalities. Their review however, found that, in practice, few studies made use of this potential advantage, mainly through poor scenario specification.

In respect of the third argument in favour of CV – that it can inform allocative efficiency decisions through application in CBA – Olsen and Smith (2001) found that most studies reviewed were partial evaluations, valuing one programme alone. Together with the fact that most also did not compare CV values to the cost of the programmes, this suggests that few CV studies were making use of their theoretical advantage. Such was the evidence back in 2001, that Olsen and Smith had 'the distinct feeling of a huge mismatch between the theoretical glory of WTP and the usefulness for public health policy of the majority of surveys which have applied this method' (Olsen and Smith, 2001: 47). Updating this review to see if there has been a change over time finds that the same conclusion can largely be reaffirmed.

Looking first at health outcomes, one can see from Table 4 that the majority of studies do not explicitly address the valuation of health outcomes comprehensively. Table 4 also suggests that this situation has not improved, although there are fewer who provide no details, as a proportion, or for whom health is not the focus.

Table 5 provides a breakdown of non-outcome attributes or 'process utility', and provides a comparison between the two time periods.

In terms of non-current use attributes – reflecting option value and externalities – a more complex picture of the interaction between the scenario specification and the respondent arises. That is, which values are activated in this respect depends upon the role in which respondents are placed. To assess the extent to which CV studies had actually sought to include these elements of value, each study was examined for three types of respondents: users, convenience sample, and the general population. Users are most often patients,

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 4. Health outcomes description

	No. of studies		
Dimension	1985–1998	1999–2005	1985–2005
Not the focus	19	16	35
None	19	50	69
Health status	6	66	72
Probability	10	29	39
Health status and duration	5	5	10
Health status and probability	4	21	25
Duration and probability	1	0	1
Health status, duration and probability	7	7	14
Total	71	194	265

Table 5. Non-outcome descriptions

	No. of studies		
Dimension	1985–1998	1999–2005	1985–2005
Not the focus	23	82	105
Not provided	7	58	65
Information	10	20	30
Caring	5	25	30
Other	13	13	26
Information and caring	1	5	6
Caring and other	2	0	2
Information and caring and other	1	0	1
Total	71	194	265

but sometimes their relatives. Convenience samples are most often groups of students or seminar participants. General population samples are those who are supposed to be representative of the population. Table 6 shows that most surveys included only use-value, reflecting the fact that most surveys asked patients only. Only 24% of studies have asked the general population. When users may have expressed option value or externalities, this refers to preventive programmes. Table 6 also illustrates that this picture has changed little over the two time periods considered.

Given that CV studies in health care are overwhelmingly constructed to elicit use-value alone, the question that arises therefore is whether CV studies in health care are misspecified. Empirical research suggests that, if these results are more widely replicable, most CV studies in health care may indeed be misspecified, as a significant element of the value of the good in question is not

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 6. Types of values by type of respondents

	Use value	Option value	Externalities	Option and externalities	Total
Users					
1985–1999	42	2	1	1	46
1999–2005	68	18	9	12	107
Total	110	20	10	13	153
Convenience sample					
1985–1999	8	2	0	2	12
1999–2005	23	0	0	1	24
Total	31	2	0	3	36
General population					
1985–1999	5	1	0	3	9
1999–2005	41	7	2	4	54
Total	46	8	2	7	63
Users and convenience sample					
1985–1999	0	1	0	0	1
1999–2005	5	1	1	0	7
Total	5	2	1	0	8
Users and general population					
1985–1999	0	0	0	1	1
1999–2005	0	3	0	1	3
Total	0	3	0	1	4
Convenience and general population					
1985–1999	1	0	0	0	1
1999–2005	0	0	0	0	0
Total	1	0	0	0	1
Total					
1985–1999	56	6	2	7	71
1999–2005	137	29	11	17	194
Total	193	35	13	24	265

being captured (Smith, 2007). This is important as, ultimately, the cost–benefit calculus and policy recommendation would be biased against the good in question, and in a comparison between goods. Of course, whether studies actually *are* misspecified depends upon the policy context within which the CV study is conducted. For instance, it may be argued that the most appropriate policy context is often one that is *ex ante*, and thus that the *only* source of relevant value is option value.

Updating the contribution that CV can make to CBA and allocative efficiency decisions reveals that in terms of the number of partial evaluations (163 or 62% of studies) this does seem to have improved slightly over time, but the number comparing CV values to the cost of interventions was not much higher (29 studies). The challenge here is that, given the failure of CV to achieve its

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

theoretical advantages, can other outcome approaches devise ways to incorporate these advantages. If they can then clearly CV might become redundant.

Of course, one possible alternative to conducting *individual* CV studies is to maintain the current practice of conducting cost–utility analysis (CUA), but have a 'translation' factor to convert these in to CBA through use of a statistic or formula representing social willingness-to-pay (WTP) for a QALY (Adler, 2006; Gafni and Birch, 2006). Without such a 'translation' factor, there is no way in which the cost effectiveness of health interventions can be compared with that of interventions in other sectors, such that an inefficient allocation across sectors is a distinct possibility (noting that, of course, even if such a translation factor were used, basing this on a QALY would still only represent the WTP for health benefits, thus still excluding some of the claimed advantages of CV over other techniques).

Currently, bodies such as NICE, have to make an implicit judgement about whether the QALY gains are 'worth' what they will cost (Devlin and Parkin, 2004). However, the 'threshold' values adopted (such as £20,000–£30,000 per QALY above or below which a new therapy will be rejected or recommended for adoption in England and Wales, and the commonly-mentioned €18,000 in the Netherlands) are more or less 'arbitrary' (NICE, 2004; Rawlins and Culyer, 2004), and NICE has recently funded work looking at the assessment of a money value (Devlin and Parkin, 2004). Similar work is being undertaken elsewhere, for instance, Australia (George *et al*., 2001). There are also a growing number of papers that have either explicitly called for, or at least implied the need for, estimates of the monetary value of a QALY, with empirical estimates ranging from around £1,000 to over £500,000 per QALY, depending on assumptions used and whether values have been imputed from the value of a prevented fatality or a specific survey looking at WTP for changes in health states (Phelps and Mushlin, 1991; Johannesson, 1995; Garber and Phelps, 1997; Johnson *et al*., 1997; Bleichrodt and Quiggin, 1999; Hirth *et al*., 2000; Dolan and Edlin, 2002; Hammitt, 2002; Gyrd-Hansen, 2003; Devlin and Parkin, 2004; Eichler *et al*., 2004; Buxton, 2005; Byrne *et al*., 2005; Franica *et al*., 2005; King *et al*., 2005; Haninger and Hammitt, 2006; Adler, 2006; Van Houtven *et al*., 2006).

However, although such a calculation is appealing, and interest in generating such a value increasing, it is not as straightforward as it initially appears and there are a range of issues that need to be considered when applying WTP at a *societal* level. These have been considered in depth elsewhere (Richardson and Smith, 2004; Smith and Richardson, 2005), but in brief concern whether the sum of individual WTP equals social WTP, whether individual WTP maps directly into social WTP, whether personal income is the appropriate budget constraint, and whether WTP should be adjusted for ability to pay. The degree to which these considerations can be overcome will determine the degree to which CV may be unnecessary. The recently completed SVQ (Social Value of a

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

QALY) project in the UK has considered a number of these issues (Donaldson *et al*., 2003) and will be extended, on a European level, through the EuroVaQ (European Value of a QALY) project (Donaldson *et al*., 2006). These projects are designed to provide information concerning whether the values/weights attached to gains in quality and length of life should vary according to the characteristics of the patients receiving them, and how much the health system should be spending (at the margin) to achieve such gains. This involves testing alternative procedures in order to: (i) determine whether it is feasible to derive a baseline value of a QALY sufficiently robust to be of use to decision makers; (ii) explore the issues involved in incorporating 'citizen' concerns into a WTP framework; and (iii) identify which method(s) (if any) appear to be the most promising candidate(s) for national decision-making (and, in EuroVaQ, look at cross-country differences).

Of course, despite all this effort, whether a 'social value of a QALY' is obtainable, or even desirable, is a moot point. It could simply be that economists are trying to combine conceptually incomparable measures – that the conditions for equivalence are excessively restrictive and unrealistic (e.g. Dolan and Edlin, 2002). Given that we are almost certainly looking for a non-linear function, rather than a universal constant, WTP for money value of a QALY, it might well be that in trying to value a QALY we simply support the replacement of the QALY with WTP, and thus the replacement of CUA with CBA (Van Houtven *et al*., 2006).

Therefore, if a 'social value of a QALY' is not feasible, or perhaps desirable, then we are left in the position of having to address the issues discussed in this and previous reviews of CV (Diener *et al*., 1998; Klose, 1999; Olsen and Smith, 2001, Smith 2003, Sach *et al*., 2007). In this respect, the need for a coordinated research effort and the specification of some form of reference case and/or guidelines for conduct of CV in health economics appears unavoidable, and perhaps even more important. That is, if CV is deemed seaworthy, then a sail needs to be constructed.

To summarise, CV does not seem to be capitalising on the major advantages it holds as a technique over other outcome methodologies. Overall, it is unclear why and how a decision maker could use *existing* CV studies in resource allocation decision-making. This failure to capitalise on the potential advantages of CV by researchers could suggest that training in the technique is inadequate, the technique is fundamentally flawed, the costs of doing CV appropriately are astronomical, or simply that there are inadequate guidelines available to guide researchers. The detachment between the doers and users of this research may also mean that interested individuals do the studies for reasons of personal interest rather than with a policy use in mind. It is quite likely that the answer contains elements of all these arguments. However, these authors would not suggest that the technique is fundamentally flawed; rather, it has not been appropriately developed and applied in health economics.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

What sort of sail?

If the foundations of CV are seaworthy, decisions about the construction of CV, or to use our metaphor the construction of the sail, need to be made. Historically, the key specifications have focused on scenario development and presentation, payment vehicle, respondent type, the incorporation of risk or uncertainty, the time period over which WTP/WTA is estimated and the mode of administration (Smith, 2003). In order to see what the specifications of the sail should be, it is important to reflect back to see if, over time, a consensus has been reached on any of these design aspects, whether there has been any convergence towards a common agreed standard, as has been the case in other areas of economic evaluation (NICE, 2004). Using the CV database used in previous reviews of CV (see Olsen and Smith (2001), Smith (2003) and Sach *et al*. (2007) for details) it is possible to address this question by looking at a few examples.

To know if the methodology used in CV studies has improved, one would need a standard against which to judge. Within health economics, there has been no formal guidelines established or validated, hence many studies claim to have followed guidelines established within environmental economics, although very much in an ad hoc fashion. Assuming this is acceptable for the moment, we can illustrate change over time with reference to the National Oceanographic and Atmospheric Administration (NOAA) guidelines (Arrow *et al*., 1993). Taking an example using question-elicitation format first, the NOAA Panel recommended that dichotomous-choice questions should be used in preference to open-ended questions, as the latter are 'biased, erratic and unreliable' (Arrow *et al*., 1993). The dichotomous-choice format was developed to attempt to resolve three problems that had been encountered in using the open, bidding, and payment-card formats: low response rates; strategic bias; and starting point (for bidding games) or range (for payment card approaches) bias. This format was argued to have three advantages over other techniques: it more closely resembles consumer choice in an actual market situation; no one respondent can unduly influence the overall WTP value; and it avoids starting point (and range) bias since respondents are only presented with one bid (Smith, 2000).

One would expect such a significant recommendation to spill over to the conduct of CV studies in health care. Table 7 demonstrates that between 1990 and 1996, dichotomous-choice questions represented *fewer* than any other single format, yet for the period from 1996 to 1999 this format represented *more* studies than were undertaken with any other single format. However, Table 7 also illustrates that the majority of studies (80% across the full period) were still not undertaken using the dichotomous-choice format. In more recent years, since 2000, the payment card and bidding approaches have, especially, been more widely reported, and critical attention has only been paid to the use of different question formats *within health economics* in the last few years (Frew *et al*., 2003, Whynes *et al*., 2003, 2004).

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Table 7. Questionnaire format over time

Year	Open	Bidding	Payment card	Dichotomous-choice single	Dichotomous-choice double	Not available	Multiple format	Total
1985	1	0	1	0	0	0	0	2
1986	0	0	0	0	0	0	1	1
1987	1	0	0	0	0	1	0	2
1988	0	0	2	0	0	1	0	3
1989	0	0	0	0	0	0	0	0
1990	0	1	1	0	1	0	0	3
1991	1	1	0	1	0	0	1	4
1992	1	0	1	0	0	1	0	3
1993	0	1	1	1	0	1	0	4
1994	1	1	2	1	0	2	2	9
1995	2	0	1	0	0	0	2	5
1996	1	2	1	2	0	0	1	7
1997	1	2	2	5	0	0	1	11
1998	1	5	4	4	2	0	2	18
1999	1	1	1	0	0	0	4	7
2000	2	4	6	0	1	5	5	23
2001	5	2	6	1	2	5	4	25
2002	3	2	9	3	3	7	4	31
2003	7	4	7	3	1	3	7	32
2004	3	10	6	8	3	2	5	37
2005	3	7	6	5	7	6	4	38
Total	34	43	57	34	20	34	43	265

As a second example, the method of survey administration recommended by the NOAA Panel is that of face-to-face interviews. Smith (2003) reported that 41 studies (37%) collected data from face-to-face interview. Of the remaining 70 (63%) who did not undertake face-to-face interview, 12 (11%) used telephone interview, 23 (21%) used postal self-administration and 29 (26%), other self-administration. The technique was not available for six studies. Of the 154 studies published subsequently, the corresponding figures for studies collecting data from face-to-face interviews, telephone, postal, and other self-administration were 45% (70/154), 16% (25/154), 14% (21/154) and 15% (23/154), with the technique not being available in 10% (15/154) of cases. This would seem to indicate some improvement – there are more studies using face-to-face and telephone interview, and less using postal or other self-complete forms of administration, which should improve the quality of studies in this latter period. There is, therefore, some hope, but the degree of change has appeared non-universal across different aspects of design, perhaps indicative of a lack of coordination or direction.

Of course, it may not be appropriate or desirable to apply the NOAA guidelines, which were developed within a very specific context – valuing

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

environmental damage caused by pollution for use in legal settlement. By its very nature, environmental economics is mostly concerned with 'non-use' valuation of unfamiliar commodities, such as wildlife, areas of natural beauty or air quality, and irreversible effects. In contrast, CV surveys in health care are valuing a 'good' that is generally more familiar, the context (health system) is familiar, the good (usually) providing direct 'use-value', and the good itself being of a private (i.e. rival and excludable, and therefore tradable) rather than public nature. Thus, for instance, there may be very justifiable reasons why health economics may not wish to adopt NOAA recommendations for elicitation formats, given the differences in what is being valued (Smith, 2006). However, the critical issue is that there is no systematic exploration of this justification, or application of alternative, health-specific, guidelines adopted in health economics. As with other economic methodologies, there is a need for readily identifiable criteria by which the quality of CV studies may be judged.

At the most fundamental level, this has meant that CV studies are not currently comparable, as health-specific CV studies lack agreement and adherence to a specific design choice for each dimension (including, *inter alia*, how to develop CV studies appropriately, what and how information should be provided in the scenario, elicitation format, respondent type, payment vehicle, time frame for valuation, partiality, administration mode, and how risk and uncertainty should be incorporated). The result of this is an inability to compare across CV studies because they use incomparable methods on two or more dimensions. This leads to an inability to look at the existing CV literature to try to establish the impact of different design choices, as it is not clear which design choice may have impacted the result across studies, even for studies of the same or similar interventions, as they are intrinsically confounded. It therefore seems critical that more systematic work be undertaken in order to consider how key dimensions should be handled in these studies, and thus enable us to craft a sail that is ready to drive the ship.

Setting sail: which direction?

This paper has thus far argued that applications of CV have failed to deliver on the theoretical advantage of the technique, which, at least in part, has reflected the absence of agreed guidelines in health economics about how to conduct CV studies. The consequence has been a methods-policy gap where the technique has had limited influence in decision-making processes. If the further pursuit of CV is seen as a worthy cause then there is some work needed to make CV useful. In this section, we make some tentative suggestions as to some ways forward, and whilst readers may disagree with the detail, the key point being made is the need to make some decisions on where to set our sail, otherwise the direction of travel will never be determined and the good ship CV will continue to drift about, not really being of any use to anyone.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Box 1. Checklist of what should be reported in published CV studies

CV development and context

- Country where the CV survey has been conducted and health care financing details
- Focus – methodological or policy
- Specificity of questionnaire (part of wider survey)
- Details of other measures of QoL incorporated
- Scenario development
- Welfare measure (WTP or WTA)

CV scenario description

- Intervention(s)
- Partiality (single good or close substitutes)
- Outcomes (health status, probability and time)
- Non-outcomes (information, care, other)
- Payment vehicle
- Presentation of uncertainty/risk
- Survey period
- Time period for WTP
- Question/elicitation format

CV reporting and results

- Method of data collection
- Type of respondent
- Sample size
- Response rate
- Type of outcomes incorporated (Use, option or externality value)
- Duration of interview/length of questionnaire
- WTP values (results of the studies)
- Transformation of values from one context/time to another
- Price year
- Currency
- Cost of intervention
- Cost–benefit ratio
- Time period used in analysis

CV validity and reliability

- Tests for bias – order effect, starting point, range, interviewer, strategic
- Statistical analysis performed
- Assessment of zero/high bids
- Distributional issues considere
    - Income assessed
    - Income adjustment
- Validity tests
- Reliability tests

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

For this, there has to be a common standard in the reporting of CV studies in the published peer-reviewed literature. For example, Sach *et al.* (2007) found important design details missing in CV studies of pharmaceutical interventions published since 2000. Here, 25% failed to report the elicitation method used and 23% failed to report the price year. Both of these details are required in order to interpret and compare the results in a useful manner. An obvious and easily implemented solution to this problem would be the adoption by health economists and journal editors of an agreed checklist, listing what details should be included in published CV studies. Box 1 suggests what such a checklist might include, covering all aspects from development, design and results, through to tests of reliability or validity undertaken (see also Yeung and Smith (2006)).

Given the issues highlighted in this paper, we suggest that guidelines need to be produced in a coordinated manner that can be used by authors, reviewers and policy makers alike, which should also include guidance concerning how to deviate from the guidelines appropriately when undertaking methodological work. For instance, isolating the change of interest, rather than confounding the issue with multiple simultaneous changes. The importance of guidelines is clearly demonstrated by looking at the application of CV in the environmental and transport safety literature where NOAA was so influential in shaping the development of the technique, and the production of government guidelines for CV conduct in these areas. Given the methodological uncertainties that remain in CV as applied to health care, any guidelines developed need to be formed on the basis of some consensus so that they are adhered to. The simplest approach might, in the first instance, be to adapt the NOAA guidelines in health (outlined in brief in Box 2).

Importantly, however, there has been little consideration of whether the *methodology* found in the application of CV in the environmental or transport safety literature is directly translatable to health care and, if so, whether this is even desirable (in most areas of the application of economics to health care, we consider 'health' to be sufficiently 'different' to warrant specific adaptations to theory and/or practice (Mooney, 1986, 1994), so why not here?). It would seem sensible for research to first address the applicability of CV guidelines in other branches of economics before developing and coordinating work to create new ones in health. The main point to note is that there is no point tinkering if there is no common purpose. A coordinated research effort, that ensures research is reported appropriately and isolates the one aspect being varied, would go some way towards making CV useful to policy.

Others might fundamentally disagree with using NOAA as a starting point on the basis that health is sufficiently different from the environment and that it makes no logical sense to start from that point. Such an argument could be seen as quite persuasive in terms of the misspecification issue raised earlier in the paper, for instance. Because health CV values extend beyond non-use and irreversible decisions, there could be a recommendation that the default assumption be that all elements of value are of importance such that researchers

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Box 2. NOAA guidelines (adapted from Arrow *et al.*, 1993)

Development:

- Pilot work and pretesting should be conducted and described

Design choices:

- Willingness to pay format should be used
- Probability sampling
- Define the budget constraint
- State the payment vehicle
- Face-to-face interviews or telephone interviews
- Dichotomous-choice question format, with a no answer option such respondents should be asked to explain their choice
- Adequate information about that which is being valued should be included (this should include the change in health state, the probability and duration)
- Respondents need to be reminded of other commodities directly and that their expenditure on other private goods will be reduced as a result of the WTP question
- The survey should be employed across different time points with independent samples to test for time trends
- The survey should include questions about other factors which might explain responses, for instance:
    - Income
    - Prior knowledge of the issue
    - Prior interest in the issue
    - Attitudes towards health
    - Attitudes toward the NHS
    - Relevance of issue
    - Understanding of the task
    - Belief in the scenarios
    - Ability/willingness to perform the task
- Tests of interviewer effects should be incorporated

Reporting, should include clear details on:

- The population sampled
- The sampling frame used
- The sample size
- The overall sample non-response rate, detailing refusals and item non-response
- The paper should reproduce the exact wording and sequence of the questionnaire and of other communications to respondents so that design choices are clear to see
- Data from the study should be retained and made available to those interested

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Quality assessment:
- High response rate
- Responsiveness to the scale of the good or service being valued
- Understanding of the task by respondents
- Belief in the scenario
- Yes or no votes should be explained in terms of cost and/or value of the good or service.

must explicitly consider the relevance of use, externality and/or option value in their design and discuss how this issue was tackled. Similarly, research around elicitation format in health by no means necessarily supports the use of the dichotomous-choice questions over other formats and, therefore, health economists may wish to agree on the use of payment scales or another format as their preferred base case option. The key point being made here is that this requires systematic discussion and application.

Clearly, guidelines will never produce a 'perfect' CV study. By their very nature, CV studies are flawed representations of the world. Their purpose is not to achieve perfection but rather to provide a structure to guide movement toward obtaining more accurate values, a belief presumably driving all forms of research into aspects of the valuation exercise, such as embedding effects, payment vehicles, or scenario specification. Indeed, it is perhaps somewhat ironic that not only despite, but perhaps because, of such high-profile guidelines, and the wealth of research undertaken into CV surveys in general, there is not greater consistency in the conduct of CV surveys. However, this paper indicates not only that CV studies in health care have performed poorly in the construction, specification and presentation of the contingent market, but also that there has been little, if any, improvement in this respect over the last 20 years.

Who will captain the ship?

Of course, setting direction and staying on course requires a captain. The paper has thus far considered the *what*-needs-to-be-done question, but we also need to consider *who* should, or could, captain the ship.

Health economists could choose to be proactive and coordinate themselves so as to develop the technique themselves or in collaboration with researchers from other disciplines. Alternatively, health economists could wait for provocation to do something. This might come in the form of a major health accident, akin to the impetus for the NOAA recommendations, or rather more likely as a policy agenda that forces change. From a policy point of view, if health economists were to focus narrowly on evaluating the value for money of curative technologies that only have health benefits, then arguably CV might be unnecessary, other than in helping determine the overall allocation of resources to health from the treasury perhaps. In reality, of course even clinical interventions often

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

have spillover effects onto family carers or children in terms of educational benefits for example. Such benefits are not captured in traditional cost per QALY analysis, which could be justified by organisations such as NICE in the UK, on the grounds that an NHS and social services perspective is taken. However, in 2005 the remit of NICE expanded to incorporate public health, as it took on the functions of the Health Development Agency, an area where the costs and benefits very clearly extend beyond the health care arena and have consequently challenged standard economic evaluation methods.

To deal with this, NICE has extended the scope of *costs* captured in economic evaluations and whilst *benefits* remain assessed by QALYs, a cost-consequent approach has also been adopted to take account of the wider benefits where possible (Drummond *et al*., 2007). Such an approach does not seem ideal since potentially non-health benefits could be attached less weight than the health effects, biasing decisions towards those with larger health gains. Potentially, this agenda could provide an impetus to develop CV methods in health care. This being the case then NICE might be an appropriate lead. However, the CV approach then risks not being captained anywhere, except to the bottom of the sea perhaps, because certain elements of the health economics community may not approve of the technique. This paper has shown that *on current evidence* such an opinion of the technique might actually be well founded. However, it is one thing to demonstrate that a technique has failed in the past but another to argue that it could never work in the future, which is what concerns us in this paper.

To conclude, it seems to us that now would seem an appropriate time to take stock of the history of CV in health economics, in part because the ad hoc application of CV can no longer be ignored and in part because new opportunities such as the public health economics agenda may renew the potential importance of the technique. This paper has attempted to be provocative in the hope of stimulating response and debate about the future of CV because the status quo seems clearly inadequate, and as a profession it is arguably our responsibility to debate whether this means it should be put in to port and never sailed again, or is given the ability to set sail and travel to new and exciting places!

Acknowledgements

The authors would like to thank Professor David Whynes for comments made prior to the starting of writing this paper. Tracey Sach is a member of the Public Health Interventions Advisory Committee at NICE. The comments and views expressed in this paper are the responsibility of the authors alone.

References

Acton, J. P. (1973), *Evaluating Public Programs to Save Lives: The Case of Heart Attacks*, Santa Monica: R-950-RC, RAND.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Acton, J. P. (1976), 'Measuring the monetary value of lifesaving programs', *Law and Contemporary Problems*, **40**: 46–72.

Adler, M. D. (2006), 'QALYs and policy evaluation: a new perspective', *Yale Journal of Health Policy, Law and Ethics*, **VI**: 1–92.

Arrow, K., R. Solow, P. R. Portney, E. E. Leamer, R. Radner and H. Schuman (1993), 'Report of the NOAA panel of contingent valuation', *Federal Register*, **58**: 4601–4614.

Bateman, I., R. T. Carson, B. Day, W. M. Hanemann, N. Hanley, T. Hett, A. Jones, G. Loomes, S. Mourato, E. Ozdemiroglu, D. W. Pearce, R. Sugden and J. Swanson (2002), *Economic Valuation with Stated Preferences Techniques. A Manual*, Cheltenham: Edward Elgar.

Bleichrodt, H. and J. Quiggin (1999), 'Life-cycle preferences over consumption and health: when is cost-effectiveness analysis equivalent to cost–benefit analysis?', *Journal of Health Economics*, **18**(6): 681–708.

Buxton, M. J. (2005), 'How much are health-care systems prepared to pay to produce a QALY?', *European Journal of Health Economics*, **6**: 285–287.

Byrne, M. M., K. O'malley and M. E. Suarez-Almazor (2005), 'Willingness to pay per quality-adjusted life year in a study of knee osteoarthritis', *Medical Decision Making*, **25**: 655–666.

Cookson, R. (2003), 'Willingness to pay methods in health care: a sceptical view', *Health Economics*, **12**: 891–894.

Covey, J. and R. D. Smith (2006), 'How common is the "prominence effect"? Additional evidence to Whynes *et al.*', *Health Economics*, **15**: 205–210.

Devlin, N. and D. Parkin (2004), 'Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis, *Health Economics*, **13**: 437–452.

Diener, A., B. O'Brien and A. Gafni (1998), 'Health care contingent valuation studies: a review and classification of the literature', *Health Economics*, **7**: 313–326.

Dolan, P. and R. Edlin (2002), 'Is it really possible to build a bridge between cost–benefit analysis and cost-effectiveness analysis?', *Journal of Health Economics*, **21**: 827–843.

Donaldson, C., I. Bateman, S. Chilton, M. Jones-Lee, G. Loomes, H. Metcalf, M. Mugford, A. Robinson, M. Ryan, P. Shackley, R. D. Smith, R. Sugden and J. Wildman (2003), *What is the Value to Society of a QALY?*, Department of Health. NCCRD, Birmingham, UK.

Donaldson, C., W. Brouwer, L. Gulacsi, D. Gyrd-Hansen, A. Kozierkiewicz, J. A. Olsen, A. Manca, A. Mataria, J. P. Moatti, U. Persson, J. L. Pinto Prades, M. Ryan and R. D. Smith (2006), 'European Value of a Quality Adjusted Life Year', *European Union Framework*, **6**. http://cordis.europa.eu/fetch?CALLER=FP6_PROJ&ACTION=D&DOC=1&CAT=PROJ&QUERY=0121eaa95d53:22fc:1d7ece4e&RCN=84063

Drummond, M., H. Weatherly, K. Claxton, R. Cookson, B. Ferguson, C. Godfrey, N. Rice, M. Sculpher and A. Sowden (2007), Assessing the Challenges of Applying Standard Methods of Economic Evaluation to Public Health Interventions. http://www.york.ac.uk/phrc/D1-05_FR.pdf [3 April 2008].

Eichler, H. G., S. X. Kong, W. C. Gert, P. Mavros and B. Jonsson (2004), 'Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge?', *Value in Health*, **7**: 518–528.

Franica, D. M., D. S. Pathakb and A. Gafni (2005), 'Quality-adjusted life years was a poor predictor of women's willingness to pay in acute and chronic conditions: results of a survey', *Journal of Clinical Epidemiology*, **58**: 291–303.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Frew, E. J., D. K. Whynes and J. L. Wolstenholme (2003), 'Eliciting willingness to pay: comparing closed-ended with open-ended and payment scale formats', *Medical Decision Making*, **23**: 150–159.

Gafni, A. and S. Birch (2006), 'Incremental cost-effectiveness ratios (ICERs): the silence of the lambda', *Social Science and Medicine*, **62**: 2091–2100.

Garber, A. M. and C. E. Phelps (1997), 'Economic foundations of cost-effectiveness analysis', *Journal of Health Economics*, **16**(1): 1–31.

George, B., A. Harris and A. Mitchell (2001), 'Cost-effectiveness analysis and the consistency of decision-making. Evidence from pharmaceutical reimbursement in Australia (1991–1996)', *PharmacoEconomics*, **19**: 1103–1109.

Gyrd-Hansen, D. (2003), 'Willingness to pay for a QALY', *Health Economics*, **12**: 1049–1060.

Hammitt, J. K. (2002), 'How Much is a QALY Worth? Admissible Utility Functions for Health and Wealth', Working Paper, Harvard Center for Risk Analysis, Harvard University.

Haninger, K. and J. K. Hammitt (2006), 'Willingness to Pay for Quality Adjusted Life Years: Empirical Inconsistency between Cost-effectiveness Analysis and Economic Welfare Theory', Working Paper, Harvard Center for Risk Analysis, Harvard University.

Hanley, N., M. Ryan and R. Wright (2003), 'Estimating the monetary value of health care: lessons from environmental economics', *Health Economics*, **12**: 3–16.

Hirth, R. A., M. E. Chernew, E. Miller, A. M. Fendrick and W. G. Weissert (2000), 'Willingness to pay for a quality-adjusted life year: in search of a standard', *Medical Decision Making*, **20**: 332–342.

Johannesson, M. (1995), 'The relationship between cost-effectiveness analysis and cost–benefit analysis', *Social Science and Medicine*, **41**: 483–489.

Johnson, F. R., E. E. Fries and H. S. Banzhaf (1997), 'Valuing morbidity: an integration of the willingness to pay and health status index literature', *Journal of Health Economics*, **16**: 641–665.

King, J. T., J. Tsevat, J. R. Lave and M. S. Roberts (2005), 'Willingness to pay for a quality-adjusted life year: implications for societal health care resource allocation', *Medical Decision Making*, **25**: 667–677.

Klose, T. (1999), 'The contingent valuation method in health care', *Health Policy*, **47**: 97–123.

Mitchell, R. and R. Carson (1989), *Using surveys to value public goods: The contingent valuation method*, Washington, DC: Resources for the Future.

Mooney, G. H. (1986), *Economics, Medicine and Health Care*, London: Harvester Wheatsheaf.

Mooney, G. H. (1994), *Key Issues in Health Economics*, New York: Harvester Wheatsheaf.

National Institute of Health and Clinical Excellence (2004), Guide to the Methods of Technology Appraisal. Accessed 1 July 2008. http://www.nice.org.uk/pdf/TAP_Methods.pdf

Olsen, J. A. and R. D. Smith (2001), 'Theory versus practice: A review of "willingness-to-pay" in health and health care', *Health Economics*, **10**: 39–52.

Phelps, C. E. and A. I. Mushlin (1991), 'On the (near) equivalence of cost-effectiveness and cost–benefit analyses', *International Journal of Technology Assessment in Health Care*, **7**(1): 12–21.

Rawlins, M. and A. Culyer (2004), 'National Institute for Clinical Excellence and its value judgements', *British Medical Journal*, **329**: 224–227.

Richardson, J. and R. D. Smith (2004), 'Calculating society's willingness-to-pay for a QALY: key questions for discussion', *Applied Health Economics and Health Policy*, **3**(3): 125–126.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press

Sach, T., R. D. Smith and D. Whynes (2007), 'A "league table" of contingent valuation results for pharmaceutical interventions: a hard pill to swallow?', *PharmacoEconomics*, **25**: 107–127.

Smith, R. D. (2000), 'The discrete choice willingness-to-pay question format in Health Economics: should we adopt environmental guidelines?', *Medical Decision Making*, **20**(2): 194–206.

Smith, R. D. (2001), 'The relative sensitivity of willingness-to-pay and time-trade-off to changes in health status: an empirical investigation', *Health Economics*, **10**: 487–497.

Smith, R. D. (2003), 'Construction of the contingent valuation market in health care: a critical assessment', *Health Economics*, **12**: 609–628.

Smith, R. D. (2004), 'The reliability of willingness to pay for changes in health status', *Applied Health Economics and Health Policy*, **3**(1): 35–38.

Smith, R. D. (2005), 'Sensitivity to scale in contingent valuation: the importance of the budget constraint', *Journal of Health Economics*, **24**: 515–529.

Smith, R. D. (2006), 'Its not just what you do, its the way that you do it: the effect of different payment card formats and survey administration on willingness to pay for health gain', *Health Economics*, **15**: 281–293.

Smith, R. D. and J. Richardson (2005), 'Can we estimate the "social" value of a QALY? Four core issues to resolve', *Health Policy*, **74**: 77–84.

Smith, R. D. (2007), 'Use, option and externality values: are contingent valuation studies in health care mis-specified?', *Health Economics*, **16**: 861–869.

Smith, R. D. (2007a), 'The relationship between reliability and size of willingness to pay values: a qualitative insight', *Health Economics*, **16**: 211–216.

Smith, R. D. (2007b), 'The role of "reference goods" in contingent valuation: should we help respondents to "construct" their willingness to pay?', *Health Economics*, **16**: 1319–1332.

Smith, R. D. (2008), 'Contingent valuation in health care: does it matter how the "good" is described?', *Health Economics*, **17**: 607–617.

Van Houtven, G., J. Powers, A. Jessup and J.-C. Yang (2006), 'Valuing avoided morbidity using meta-regression analysis: what can health status measures and QALYs tell us about WTP?', *Health Economics*, **15**: 775–795.

Whynes, D. K., E. J. Frew and J. L. Wolstenholme (2003), 'A comparison of two methods for eliciting contingent valuations of colorectal cancer screening', *Journal of Health Economics*, **22**: 555–574.

Whynes, D. K., E. J. Frew and J. L. Wolstenholme (2004), 'Comparing willingness-to-pay: bidding game format vs. open-ended and payment scale formats', *Health Policy*, **68**: 289–298.

Whynes, D., E. Frew, Z. Philips, J. Covey and R. D. Smith (2007), 'The "prominence effect" on contingent valuation', *Journal of Economic Psychology*, **28**: 462–476.

Yeung, R. Y. T. and R. D. Smith (2006), 'Can we use contingent valuation to assess the private demand for childhood immunization in developing countries? A systematic review of the literature', *Applied Health Economics and Health Policy*, **4**: 165–173.

Yeung, R. Y. T., R. D. Smith and S. M. McGhee (2003), 'Willingness to pay and size of health benefit: an integrated model to test for "sensitivity to scale"', *Health Economics*, **12**(9): 791–796.

Yeung, R. Y. T., R. D. Smith, L. M. Ho, J. Johnston and G. Leung (2006), 'Empirical implications of response acquiescence in discrete-choice contingent valuation', *Health Economics*, **15**: 1077–1089.

https://doi.org/10.1017/S1744133109990016 Published online by Cambridge University Press