

Using mixed methods to explore the issue of scope sensitivity in contingent valuation

Gillian Currie (University of Calgary)
Cam Donaldson (University of Newcastle upon Tyne)
Helen Lee (University of Calgary)

Work in progress

This draft: May 18, 2005

Please do not quote without permission of the authors

Draft for discussion at HESG meeting, Newcastle, June 29-July 1, 2005.

Author for Correspondence and contact details:

Gillian Currie, PhD
University of Calgary
(currie@ucalgary.ca)

Abstract:

Issue being addressed: This paper reports on the results of experimental work testing scope sensitivity in contingent valuation of health care programs. Theory suggests that 'rational' preferences should exhibit scope sensitivity, that is, greater willingness to pay for larger health benefits. This is not always observed in empirical applications.

Methods: Mixed methods are employed. Quantitative experimental tests of scope sensitivity using both the standard approach to the elicitation of willingness to pay values as well as the marginal approach were conducted. Both within sample and between sample tests of scope sensitivity were used. In addition, qualitative methods were used to generate hypotheses regarding the reasons behind responses to WTP questions, and why scope insensitivity may occur.

Results: There were mixed results, with some tests indicating sensitivity to scope and others not. The quantitative results suggest that there is merit to further exploration of the marginal approach to eliciting willingness to pay values. The qualitative results indicate that apparently 'irrational' responses may not, in fact, represent departures from economic theory.

Implications/discussion: The implications of both the quantitative and qualitative components for future conduct of contingent valuation studies will be discussed.

Introduction

Contingent valuation (CV) is a method for valuing the benefits of non-marketed goods and services that is being increasingly used in empirical applications in health economics. A key focus of the CV literature has been in assessing the validity of the technique. A key test of the validity of the method relates to scope sensitivity. Theory suggests that willingness to pay should increase with the ‘size’ of the benefits of the intervention or program being valued. This prediction of theory is not always verified in empirical tests – there is mixed results both in environmental economics and health economics literature. Thus, sensitivity to the size of benefit remains a concern within the contingent valuation literature. This paper reports on a study that explores sensitivity to scope employing mixed methods – a quantitative experimental design using both internal and external tests for sensitivity to scope alongside qualitative follow-up interviews exploring the reasons and rationales behind individual responses to CV questions and why they may or may not be sensitive to scope. The policy background to the paper is the use of willingness to pay to explore the values of members of the public with respect to disparate uses of health care resources as in the original case of helicopters, hearts and hips (Olsen and Donaldson, 1998). The requirement of such studies for multiple valuations from each respondent creates significant opportunities for exploring the validity of CV-based willingness to pay estimates, and it is one of these methodological aspects that forms the focus of this paper.

The remainder of the paper is organized as follows. In section 1, some background on the empirical investigation of scope sensitivity will be provided. Section 2 describes the methods used in this research, and outlines the experimental design employed to explore scope sensitivity. Section 3 outlines the methods employed in the analysis, and Section 4 presents the results of both the quantitative and qualitative analysis. Section 5 concludes and outlines issues for further research.

1. Background

One of the key validity tests in contingent valuation is to examine sensitivity of respondent’s stated WTP values to the size of the health benefit being valued. The ‘size’ characteristic may be measured in a number of different ways including number of patients treated, or larger benefits to those treated in terms of the magnitude of the health gain, the size of the risk reduction, or the range

of health benefits provided.¹ According to economic theory, one would expect that WTP would be greater for options that offer a greater benefit. Theory also suggests that the marginal utility or value of additional benefits is diminishing, thus we would not expect to see a proportional increase in the size of WTP with the size of the benefit, however defined. Nevertheless, theory does predict a positive correlation. This methodological issue is particularly important when using CV results in a health policy context. If WTP values fail to exhibit sensitivity to the size of the benefit, they may not be accurately reflecting underlying values and thus may lead to misallocations of resources.

In empirical work across the fields of environmental, safety and health economics, findings have been mixed however many studies find scope insensitivity. In the environmental economics literature, the overall conclusion is that while scope insensitivity is a common finding, it is often attributable to issues of survey design (Bateman et al 2002, Carson et al 2001). In the safety literature, tests of scope insensitivity also are an active area of current research with scope sensitivity to risk reduction a particular area of concern (Beattie et al, 1998, Carthy et al, 1998). The issue of scope sensitivity in environmental economics and safety economics remains unresolved and remains an active area of research.

It has been pointed out that findings from literature in other areas may not be relevant in health economics for a variety of reasons (Smith 2005). Within health economics specifically, recent empirical evidence regarding scope shows some positive results (Smith 2001, Yeung et al, 2003), some mixed (Smith 2005) and some negative results (Olsen et al 2004). However, to date, relatively little research has been undertaken on the sensitivity of WTP values to the size of the benefit in health care. Given the mixed empirical results to date further research is needed to address this issue.

The research reported on in this paper represents a component of a broader study conducted in Alberta investigating the validity of the CV method employing mixed quantitative and qualitative methods. The overall study investigated issues raised in EuroWill (Donaldson, 1999) and thus has been dubbed “AlbertaWill”. This paper will focus on the scope sensitivity issue that was examined in a large CV study of 640 participants in Calgary, Canada. Both external (between-sample) and internal (between-subject) scope tests were conducted. In addition, an experiment was designed to compare scope sensitivity between two different methods of eliciting the WTP values from

¹ Smith 2005 distinguishes between scope and scale sensitivity where the latter refers to the amount of the good and the former refers to the range of the benefits.

respondents. A split sample approach was taken where half of respondents received a survey that posed the WTP question via the ‘standard’ approach and the other half received a survey that elicited WTP values via the ‘marginal approach’. In addition, increasingly authors have called for further understanding of the reasons behind potentially anomalous or inconsistent quantitative results in preference elicitation studies (Olsen et al 2005, San Miguel et al, 2005). This study therefore also incorporated a significant qualitative component whereby a sub-sample of respondents participated in an in-depth qualitative follow-up interview following the completion of the CV survey.

2. *Methods*

2.1 *The valuation scenarios*

Each valuation exercise involved respondents valuing potential expansions to three different health care programs by expressing their willingness to pay in terms of extra taxation for the three programs. There were four potential health program expansions that were valued: ambulance service, heart program, cancer program A, and cancer program B. Each version of the survey only contained three of the programs. The programs that were valued in each version of the survey are shown in Table 1. Detailed program descriptions are shown in Appendix A.

2.2 *The ‘standard’ approach versus the ‘marginal’ approach*

There were two different approaches taken to eliciting the WTP values. In the standard approach, absolute WTP values were elicited for each of the three program expansions. An alternative approach that has been used in an exploration of preference order consistency (Shackley and Donaldson 2002) was adopted in this test of scope sensitivity. In the marginal approach, absolute WTP was elicited for the respondent’s lowest ranked program expansion. Following that, they were asked how much they would be willing to pay for their next highest ranked program *relative to the first* and then their WTP for the top ranked program *relative to the second*. It was hypothesized that the marginal approach should lead to increased sensitivity to scope, as it frames the WTP question in an incremental fashion that would serve to highlight the relative differences between the programs being valued. Table 1 indicates that Survey Version 1 (A&B) used the standard approach, while survey Version 2 (A&B) used the marginal approach.

Version	Elicitation Approach	Program expansion			
		Ambulance	Heart	Cancer A	Cancer B
	Standard (n = 160)	X	X	X	
	Standard (n = 160)	X		X	X
	Marginal (n = 160)	X	X	X	
	Marginal (n = 160)	X		X	X

2.3 Scope tests

There were two different types of scope tests incorporated in the quantitative component of the study. There was an internal (within-subject) scope test where each respondent was asked to value two different health program expansions that differed only in terms of the number of patients treated. Cancer program B was identical in all respects to Cancer program A, except that program B involved treatment of 450 additional patients and program A involved treatment of 300 additional patients. There was also an external (between-sample) scope test where a comparison was made between the valuation between the sub-samples of the WTP for two cancer programs which differed in the number of patients treated.

There were two internal scope tests comparing WTP for Cancer A and Cancer B.: one within respondents of survey 1B using the standard approach and one within respondents of survey 2B using the marginal approach. There were also two external scope tests. Within the standard version, there was an external scope test comparing WTP for Cancer A in Survey 1A with WTP for Cancer B in Survey 1B. Within the marginal version, there was an external scope test comparing WTP for Cancer A in Survey 2A and WTP for Cancer B in Survey 2B.

2.4 Design of the questionnaire

The questionnaire was based upon a design that has been extensively pre-tested and applied in EuroWill (Olsen et al 2005, Shackley and Donaldson 2004). In addition, the questionnaire was piloted on a sample of 80 respondents in Calgary, Canada prior to being implemented for the full study. Section 1 of the survey contained introductory information. The purpose of the study – to establish people’s values for three different potential health care program expansions – was explained. They were told to imagine that these three program expansions were competing with each other for implementation, and that not all of them could be funded, and that the information

they (and other members of the public) provide about their values is important in setting priorities in the health region. They were also told that they would be asked about their maximum WTP in the form of extra taxation for the program expansions, and that the dollar values they gave would be taken as a measure of their value of each program.

Section 2 asked respondents about their own (or family) experience of the types of health programs that they were going to be asked about (cancer, heart disease and use of emergency services), as well as their perceptions of their risk of needing these types of services in future.

Section 3 introduced the detail on the program expansions that were to be valued. The program descriptions included information about how many additional patients would be treated in the program expansions, how many were currently being treated, the age/gender characteristics of the patients, and what the health outcome of the treatment would be. Then respondents were asked to rank the programs in terms of their importance (they were told that ties were allowed).

The next section involved WTP elicitation. In the standard version, the respondent was presented each of the three programs in random order. For each program, respondents were asked if they would be willing to pay (in terms of extra taxation) for the program expansion and the reasons why (or why not). They were explicitly told to consider only one program at a time, in their valuations. Respondents were also reminded of their budget constraints in that they were asked to remember that their contributions would reduce the money they had available to spend on other things. If people said no to extra taxation, they were also asked the question in terms of a voluntary donation in order to capture the possibility of a protest against extra taxation. For those who indicated they would be willing to contribute (in either form), respondents were asked to indicate their maximum WTP using a payment card format (with an option to indicate a value outside the range of the scale). As previously noted, the marginal version of the questionnaire followed a different method. The lowest ranked program was presented first and the procedure outlined above for the standard version was followed. Then, for the second ranked program, respondents were asked if they would be willing to pay for that program expansion and the reasons why or why not. If they said they were willing, they were reminded of how much they were willing to pay for the lower ranked program and then asked how much more than that they would be willing to pay for the higher ranked program.

The final section collected demographic information on the respondents including gender, number of persons in the household, number of children in the household, marital status, age, self rated health status, health care (doctor) visits in the last year, education level, employment status, and income.

2.5 *Interview procedure*

2.5.1 *Quantitative contingent valuation survey*

A random sample of the general public in Calgary was sought. A market research company was employed to undertake the recruitment and a random sample of 640 Calgarians agreed to participate in the study. Participants were invited to come to a central location for an in-person interview conducted by trained interviewers. The interviews took place between July and September 2003.

2.5.2 *Qualitative follow-up interview*

In addition to the main survey, a sub-sample of respondents were also invited to participate in an in-person follow-up interview. For the pilot study, all 80 respondents also participated in the qualitative follow-up interview. For the main study, 43 respondents participated in the qualitative follow-up interview. The full set of 123 respondents was used for the qualitative analysis. These interviews were conducted by one of the authors (GC). The interviews were digitally recorded. A semi-structured open-ended interview explored the reasons behind people's answers to the CV survey. The factors underlying rankings as well as the WTP values were explored. Reasons for ranking the larger program as more important (or not) were explored. Reasons for being WTP more for the larger program (or not) were also explored. Also, questions probing the respondents understanding of the CV exercise, difficulties with the questions, additional information wanted and difficulty arriving at dollar values for the WTP questions were asked. Finally, respondents were asked their views on if and how decision-makers in health care should consider public values. In this paper, we will focus on the questions pertaining to scope sensitivity.

3. *Analysis*

3.1 *Quantitative analysis*

The goal of the quantitative analysis was to compare WTP for different sized programs across and within samples. The first step was to identify 'protest zeros' from reasons for not being WTP. In addition, we needed to test for heterogeneity of the sub-samples, that is, test for differences in mean values (or proportions) for demographic characteristics of the sub-samples.

For each of the standard and marginal versions, hypothesis testing was conducted to examine the within and between sample scope tests. For the within sample scope test, this involved a test of differences in mean WTP for cancer program A and cancer program B in sub-sample who did Version 1B (2B). It was expected that WTP for the larger cancer program B would be larger than that for the smaller cancer program A. For the between sample scope test, this involved a test of differences in mean WTP for cancer program A from Version 1A (2A) and WTP for cancer program B from Version 1B (2B). It was expected that WTP should be larger for the larger cancer program B. Another between sample test involves examining the difference in mean WTP for cancer program A from Version 1A(2A) with WTP for cancer program A from Version 1B(2B). It was expected that the WTP for these same sized programs should not be different between samples. In addition, if there was evidence of sample heterogeneity, we repeat between-sample scope tests using predicted WTP values that were derived from regression analysis of the WTP values on the demographic predictors. In addition, we examined scope sensitivity at the individual level, and whether the marginal version had an impact on this as hypothesized.

3.2 Qualitative analysis

The digital recordings of the follow-up interviews were transcribed verbatim. The qualitative data was then reviewed and coded for reasons for (in)sensitivity to scope by one of the co-authors (GC). The coded manuscripts were then subject to thematic analysis where the reasons were then grouped into themes describing sets of similar reasons.

4. Results

4.1 Quantitative: Scope sensitivity at the aggregate level

The demographic characteristics of the samples for the standard versions of the survey (Versions 1A and 1B) are reported in Table 2. There are no significant differences in the demographic characteristics at the 5% level. Table 3 shows the WTP comparisons relevant for the between sample and within sample scope tests for the standard version. The results of the hypothesis tests for scope sensitivity are shown in Table 4. The third column shows the results of the t-test for the within-sample scope test (ie, whether WTP for the larger cancer program is different from that for the smaller cancer program in Version 1B). These two WTP values were not significantly different at the five percent level. However, the WTP value is larger for the larger program and the difference is significant at the 10% level. The second column shows the results of the t-test for the between-sample scope test (ie, whether WTP for the larger cancer program in Version 1B is different from

WTP for the smaller program in Version 1A). The two WTP values were not significantly different; thus the between-sample test fails to show scope sensitivity. Finally, we would also expect that WTP for the small cancer program in Versions 1A and Versions 1B are not significantly different. The results for this hypothesis test is shown in column 1 and indicates that the WTP *is* significantly different for the same program between the two sub-samples in Version 1A and 1B. Overall, all indicators of scope sensitivity failed for the standard version of the survey, when testing at the 5% level of significance. Small caveats on this conclusion, however, might be that the direction of WTP differences from the within-sample test was as expected whilst the WTP values for the small cancer program in version 1B were significantly lower than those for 1A. Given this lower initial starting point, it may not be surprising that the between-sample test failed.

Table 5 shows the demographic characteristics of the sample for the marginal version of the survey. There are no significant differences between groups in terms of these characteristics. Table 6 shows the WTP comparisons relevant for the between sample and within sample scope tests for the marginal version. The results of the hypothesis tests for scope sensitivity are shown in Table 7. The second column shows the results of the t-test for the between-sample scope test. The two WTP values were significantly different from each other, however they go in the wrong direction. The WTP value for the smaller program is significantly higher in Version 2A than the WTP for the larger program in Version 2B. Thus the between-sample test fails the scope sensitivity test. However, as noted in column 1, the WTP for the smaller program in Version 2A is significantly larger than WTP for the larger program in Version 2B, a more extreme version of the caveats listed above vis-à-vis the results from using the standard versions of the questionnaire. The third column shows the results of the t-test for the within-sample scope test. These two WTP values were significantly different, thus the within-sample scope test indicates that there is scope sensitivity.

We suspected that there may be heterogeneity in the sub-sample populations that were affecting the between sample scope tests. Expanded significance tests of the full set of characteristics of the sample were conducted. Regression analysis of the dependent variables that could affect WTP was performed, and adjusted or predicted WTP values were computed assuming the samples were identical. The hypothesis tests for the between sample tests were then repeated examining the adjusted WTP values.

For Versions 1A and 1B (the standard version), family experience, gender, number of children less than 16, health status, type of employer, and income were significantly different at the 10% level.

The hypothesis tests were repeated for the between sample scope tests. If all independent variables are included, there are no significant differences. If only those independent variables that were significant in the regression analysis are included, the differences in WTP across samples is not significant at the 5% level, but both are significant at the 10% level.² In other words, at the 5% level of significance, WTP for the larger program is not different from that for the smaller program but neither is WTP for the smaller programs different. At the 10% level of significance, WTP for the larger program is larger in sub-sample 1B but so is WTP for the same sized program. Either way, this is mixed news for the between sample scope tests in the standard version.

For Versions 2A and 2B (the marginal version), family experience and employment status were significantly different between the samples. Again, if all independent variables are included, there are no significant differences. In the adjusted WTP hypothesis tests (including only those independent variables that were significant predictors of WTP), the significant differences for the between sample WTP values disappeared at the 5% level. The WTP for the larger program is not significantly different than the smaller program, and neither is the WTP for the same program. Both differences are still significant at the 10% level.³ Overall, there is mixed results in the between sample scope sensitivity tests for the marginal version of the survey.

Given that there were significant differences in the mean WTP for the same program in the two sub-samples for both the standard and the marginal version of the survey, we also investigated whether there were outlier values that were driving these differences. For the standard version, when outliers with values over \$1000 were removed, the findings were unchanged. When values over \$500 were removed, the significant difference between WTP for the same program in the two sub-samples disappeared. In the marginal version, there were seven outliers in the sub-sample 2A whose WTP was greater than the maximum WTP in sub-sample 2B. In the marginal version, the significant difference between WTP for the same sized programs remains. But looking at the scope test, where previously the WTP for the smaller program was significantly larger in sub-sample 2A, now there is no significant difference. The between sample scope test still fails, but perhaps not as dramatically.

² For the test of the difference between the predicted WTP for the larger program versus the smaller program across the samples, the p-value was 0.0646. For the test of the difference between the predicted WTP for the same size program across the two samples, the p-value was .0600).

³ For the test of the difference between the predicted WTP for the larger program versus the smaller program across the samples, the p-value was 0.0594. For the test of the difference between the predicted WTP for the same size program across the two samples, the p-value was .0456).

4.2 *Quantitative: Scope sensitivity at the individual level*

Focusing on the sub-samples that contained the within-sample scope tests (versions 1B and 2B), we can examine scope sensitivity at the individual level. Table 8 shows the distribution of respondents' relative increase in WTP as the size of the program increased by 1.5. Some respondents in both versions of the survey were willing to pay less for the larger program, which clearly appears inconsistent. The primary difference between the standard and marginal versions was in terms of the number of respondents who gave the same WTP for the two programs – 52% of respondents indicated that they would be willing to pay the same amount for the two programs in the standard version compared to only 23% in the marginal version. Thus, the marginal approach to eliciting WTP values is associated with a reduction in this apparent discrepancy.

4.3 *Qualitative results*

The qualitative follow-up interviews were conducted immediately following the participant's completion of the CV survey. The interviewer had the responses to the CV survey prior to conducting the follow-up interview, and the questions asked were determined by whether the respondent had exhibited scope sensitivity or not. They were asked first if they had noted a difference between the two cancer program, and if so to explain what that was. When respondents were willing to pay more for the larger program, they were asked to explain the reasons behind this. When they were not willing to pay more for the larger program, they were asked to explain why.

The vast majority of respondents did recognize that there was a difference between the two cancer programs, and identified the difference. If an individual did not notice a difference between the two programs, they were not probed further as to the reasons behind their WTP values. This was only the case for a couple of respondents.

4.3.1 *Scope-sensitive respondents*

The explanations behind why people were willing to pay more for the larger program expansion were as expected. In fact, it was viewed as a bit of a strange question to ask as the answer seemed to obvious to participants. They indicated that they were willing to pay more for the larger program because more people were helped so the program was more valuable:

“Just because...being able to help more people”

Raising some concern, although noted by only one respondent, was a perception that the marginal approach forced them to answer in this way:

Using mixed methods to explore scope sensitivity in contingent valuation

“...I felt obligated to pay more because of the process...”

However, this respondent went on to say that he probably would have answered that way in any case. He pointed out though that:

“I suppose that in my warped way of thinking I would have you know I kind of would have been thinking that I'd pay the same amount for both and I would hope that they would go ahead and use the one that would help more people.”

But in the end, he concluded:

“...And you know the answer to your original questions would be I would pay more for it because more people were affected.”

This was raised by only one respondent out of 320, plus 40 pilot interviews. Nevertheless, it was reported here as it raises the issue that perhaps respondents were pressured into answering in a way that was scope sensitive by the marginal approach.

4.3.2 Scope-insensitive respondents

When respondents were not willing to pay more for the larger health program expansion, they were first asked whether or not they had noted a difference in the programs. Most people had in fact noticed a difference, thus their scope insensitivity was not as a result of failing to note that the one program was larger. There were several types of reasons given for the difference:

#1: “Contributions to the cause/Did not see them as different programs”: Respondents were willing to pay a token amount towards the cancer program expansions but did not distinguish between them. Perhaps the difference between the programs was not “big enough” to induce them to be scope sensitive.

“...I think just because, ah, even though one helped more or was a larger program whichever, um, just the equal, I guess the disease - the impact of disease -to me was the same..”

“..because I didn't really see it as larger, I just saw it as the same thing happening..”

“...just on this, this little bit of information that I have here and all, I'll put the same amounts..”

“..I was thinking they were doing the same thing and the number wasn't that as important..”

#2. “Budget constraint kicked in”: Some individual's did distinguish between the smaller and larger programs in their rankings but did not distinguish in their willingness to pay values. Another

explanation for this that came up in people's responses is that their budget constraint did not allow them to distinguish in their WTP values.

“Oh, I think because, um, I sort of had a fixed a dollar amount and that was kind of the maximum dollar amount that I would give for [that] program...I kind of just saw 75 and thought okay that's really my cap. So I kind of grouped those two together..”

“I'm a student I don't have a lot of money...”

#3. “Inferring differences in programs”: Respondents inferred quality related attributes of the program not specified in the descriptions that were presented. Thus they were responding in a way that was consistent with economic theory, given the extra information they were inferring.

One of the possibilities is that a quality advantage was present in the smaller program.

“Because, I thought they could get better care, um, for the amount of money that was available... instead of treating maybe 1000 people mediocre, or poorly or fairly... you could maybe be cured [in a program treating less].

If it's true or not I don't know, but that's what I was thinking...”

Another type of difference assumed is that additional information about the programs. For example, in referring to the larger health program one respondent indicated:

“ um, yah, in the second one the amount of success or pain didn't seem to be as good, as high... yah, like removing the pain from the patient wasn't as successful...”

This individual had indicated he would be willing to pay less for the larger program, a response that would seem 'irrational' in that it's not sensitive to scope. However, given that he had inferred a difference in the success of pain relief between the two, his WTP was 'rational' given his belief.

Another possibility is that the smaller program represented an inefficient program and thus it could actually achieve the same numbers treated as the larger program if they used the resources more efficiently. Thus the two programs were not seen as different.

“I considered that inefficiency in the system. I think there is lots of money if it was distributed properly.”

Overall the qualitative results yielded some interesting insights into respondents' thinking behind their responses to the CV survey.

5. *Discussion*

The standard versions of the survey failed to show scope sensitivity in all tests – both the within sample and between sample tests. The marginal version of the survey also failed to show scope sensitivity for the between sample tests, but did show sensitivity for the within sample scope tests. This indicates that there is some potential for improvement with the marginal approach to eliciting WTP values and this should be further explored. One could argue that the marginal approach ‘forces’ rationality. However, others have argued that this approach gives the method its best chance at consistency with theory and that failure in that case indicates a serious problem. Not all respondents did indicate a higher WTP value for the larger program, so there is some evidence that respondents still perceived the ability to answer according to their preferences. Further analysis of whether there were any differences between those who exhibited scope sensitivity and those who did not may shed some light on this issue.

Looking at the individual level responsiveness to scope confirms that the marginal approach results arise primarily as a result of affecting the number of people who give the same monetary values for the two differently sized programs. Olsen et al (2004) found that the majority of respondents (62%) gave the same monetary values for differently sized programs. Our results also found this for the standard version, although fewer gave this answer (52%). However, the marginal version dramatically reduces this response to 23%. A further important attribute of the data set is that it gives the ability to explore individuals simple ranking of the programs with rankings implied by their WTP values. This will be the subject of further work within this project.

Key finding of the qualitative work looking at scope sensitivity suggests that many people were not being ‘irrational’ in their responses to the CV questions even when they were not willing to pay more for the larger programs. In fact, a common finding was that though they recognized the scope difference, it was not perceived as significantly large to distinguish in the WTP values. Another finding was that respondents had ‘filled in’ information not provided in the survey to infer differences in the two health programs that were not stated. Thus, given their assumptions, their quantitative responses were not inconsistent with economic theory.

These findings suggest further investigation of the marginal approach is warranted. Qualitative work alongside this is crucial to understand the reasons behind the findings of the empirical work. In

Using mixed methods to explore scope sensitivity in contingent valuation

the meantime, and given that this the first exploration of this data set, the authors would appreciate comments on, amongst other things, matters of:

- Interpretation – what do such results mean for willingness to pay elicited through such CV exercises?
- How to make more of the qualitative data that have been collected?
- The quantitative analytic strategy; in particular, how to account for lower initial values for the small cancer programs given by groups 1B and 2B, and, indeed, how to interpret these results.
- Implications of the qualitative results for drawing inferences about rationality and for what tests of rationality are important.
- Other issues for further research.

References

- Bateman IJ, Carson RT, Day B, Hanemann M, Hanley N, Hett T, Jones-Lee M, Loomes G, Mourato S, Ozdemiroglu E, Pearce DW, Sugden R, Swanson R, 2002. *Economic Valuation with Stated Preference Techniques: A Manual*. Edward Elgar, Cheltenham, pp. 296-342.
- Beattie et al, On the contingent valuation of safety and the safety of contingent valuation: part 1-caveat investigator. *J Risk Uncertainty* 1998; 17(1): 5-26.
- Carson RT, Flores NE, Meade NF, Contingent valuation: controversies and evidence. *Environmental and Resource Economics* 2001 19, 173-210.
- Donaldson C. *Developing the method of 'willingness to pay' for assessment of community preferences for health care*. Final report to Biomed 2 Programme (PL950832) of the European Commission. Health Economics Research Unit, University of Aberdeen and Departments of Economics and Community Health Sciences, University of Calgary.
- Hanley N, Ryan M, Wright R. Estimating the monetary value of health care: lessons from environmental economics. *Health Economics* 2002, 12: 3-16.
- Olsen JA, Donaldson C. Helicopters, hearts and hips: using willingness to pay to set priorities for new public sector health care programmes. *Social Science and Medicine* 1998 46, 1-12.
- Olsen JA, Kidholm K, Donaldson C, Shackley P. Willingness to pay for public health care: a comparison of two approaches. *Health Policy* 2004, 70: 217-228.
- Olsen JA, Donaldson C, Pereira J. The insensitivity of 'willingness-to-pay' to the size of the good: New evidence for health care. *Journal of Economic Psychology*. 2005 25: 445-460.
- San Miguel F, Ryan M, Amaya-Amaya M, 'Irrational' stated preferences: a quantitative and qualitative investigation. *Health Economics* 2005 14: 307-322.
- Shackley P, Donaldson C. Should we use willingness to pay to elicit community preferences for health care? New evidence from using a 'marginal' approach. *Journal of Health Economics* 2002, 21: 971-991.
- Smith RD, The relative sensitivity of willingness-to-pay and time-trade-off to changes in health status: an empirical investigation. *Health Economics* 2001 10: 487-497.
- Smith RD. Sensitivity to scale in contingent valuation: the importance of the budget constraint. *Journal of Health Economics* 2005, 24: 515-529.
- Yeung R, Smith RD, McGee SM. Willingness to pay and size of health benefit: an integrated model to test for 'sensitivity to scale'. *Health Economics* 2003, 12: 791-796.

Appendix A: Program descriptions

EXPANSION IN CANCER TREATMENTS A

300 more patients with advanced cancer could have pain relief with radiation therapy, in addition to the 600 patients in Calgary who are currently getting this treatment.

Without the radiation treatment, the 300 patients would get pain reducing medicine. With just this medicine, however, many patients will still not get satisfactory pain relief, while others will have unacceptable side effects in the form of tiredness and poor quality of life.

Radiation therapy can provide 75% of these patients with good pain relieving effects and most patients will experience improved functioning. The treatment has few side effects. On average, the patients will have better quality of life from this treatment at the end of their life, but the treatment will not actually prolong most patients' lives.

There are patients in every age group, but the average age is 60 years old. Men and women are affected in equal numbers.

EXPANSION IN CANCER TREATMENTS B

450 more patients with advanced cancer could have pain relief with radiation therapy, in addition to the 600 patients in Calgary who are currently getting this treatment.

Without the radiation treatment, the 300 patients would get pain reducing medicine. With just this medicine, however, many patients will still not get satisfactory pain relief, while others will have unacceptable side effects in the form of tiredness and poor quality of life.

Radiation therapy can provide 75% of these patients with good pain relieving effects and most patients will experience improved functioning. The treatment has few side effects. On average, the patients will have better quality of life from this treatment at the end of their life, but the treatment will not actually prolong most patients' lives.

There are patients in every age group, but the average age is 60 years old. Men and women are affected in equal numbers.

EXPANSION TO HEART PROGRAM

250 more heart operations could be provided each year, in addition to the 1000 which are currently performed in Calgary.

Most of these extra heart patients are men aged 60-70 years. They have chest pain and breathe heavily when strained.

The operation will make 75% of the patients completely free from pain with the rest experiencing a reduction in pain. Without the operation the patients are expected to live 8-10 years. With the operation the average patient will live for an extra 1 year on top of this.

The operation mortality risk is 1% (i.e. 1 in 100 people die while being operated on).

EXPANSION TO AMBULANCE SERVICE

Each year in Calgary about 1 in 80 people use the ambulance service following an accident (e.g. car accident, work, accident, fall, etc.).

With increased funding to this program, it would be possible to reach those who have had an accident in less time (particularly those that suffer an accident in more remote areas), and to increase the quality of early emergency care.

Most people who use ambulances following an accident are aged between 15 and 40 years.

Well-equipped ambulances with qualified personnel improve the prospects for recovery (i.e. clinical prognosis) and probability of survival of those who have had an accident.

Table 2

Demographic characteristics of Samples for Standard Versions 1A and 1B

	<i>Version 1A</i> (n = 160)	<i>Version 1B</i> (n = 160)
Median age	41-45	41-45
Male gender (%)	45%	36.9%
Median income (\$)	60,001 – 70,000	60,001 – 70,000
High School or below (%)	31 (15.6%)	25 (15.6%)
Married (%)	79 (49.4%)	79 (49.4%)

Table 3

WTP comparisons for the standard versions

	Small Cancer Program (300 patients)		Large Cancer Program (450 patients)
	Version 1A N=160	Version 1B N=160	Version 1B N=160
	Number of Protest Zeros	9	5
Number of True Zeros	11	7	4
Mean WTP	\$120.94	\$88.89	\$105.8
Mean WTP (exclude protest)	\$128.15	\$91.75	\$110.64
Confidence Interval	\$98 - 159	\$77 – 107	\$88 – 134
Median	100	65	75
Interquartile range	\$35 – 150	\$25 – 100	\$ 40 – 150

Table 4

Hypothesis testing for scope sensitivity for standard versions

	<i>Between sample comparisons</i>		<i>Within sample comparison</i>
	WTP for small program in 1A and 1B	WTP for small program in 1A with large program in 1B	WTP for small program in 1A with large program in 1B
T-test (p-value)	1.9337 (0.0270)	0.8137 (0.2082)	1.2428 (0.1074)
Excluded protest zeros			
T-test (p-value)	2.1267 (0.0171)	0.9053 (0.1830)	1.3533 (0.0885)

Table 5

Demographic characteristics of Samples for Marginal Versions 2A and 2B

	<i>Version 2A</i>	<i>Version 2B</i>
	(n = 160)	(n = 160)
Median age	41-45	41-45
Male gender (%)	39	36.9%
Median income (\$)	70,001 – 80,000	60,001 – 70,000
High School or below (%)	28 (17.5%)	31 (19.4%)
Married (%)	81 (50.6%)	79 (45.0%)

Table 6**WTP comparisons for marginal versions**

	Small Cancer Program (300 patients)		Large Program (450 patients)
	Version 2A N=160	Version 2B N=160	Version 2B N=160
	Number of Protest Zeros	7	1
Number of true zeros	4	21	14
Mean WTP	\$164.42	\$93.48	\$ 124.36
Mean WTP (without protest)	\$171.94	\$94.06	\$ 125.93
Confidence Interval	\$130 – 214	\$77 –111	\$106 – 146
Median	100	60	100
Interquartile range	\$50 – 100	\$25 – 100	\$ 50 – 150

Table 7**Hypothesis testing for scope sensitivity in marginal versions**

	<i>Between Sample comparisons</i>		<i>Within sample comparison</i>
	WTP for small program in 2A and 2B	WTP for small program in 2A with large program in 2B	WTP for small program in 2A with large program in 2B
T-test (p-value)	3.1826 (0.0008)	1.7477 (0.0407)	2.3296 (0.0102)
Excluded protest zeros T-test (p-value)	3.4336 (0.0003)	1.9672 (0.0250)	2.3922 (0.0087)

Table 8: The distribution of respondents relative increase in WTP as the size of the effect increased by 1.5

Relative increase in WTP	Regular version	Marginal version
'true zeros'	3	9
>0 – <1	10	13
1	79	37
>1 – <1.5	20	36
1.5	14	16
>1.5 – <2	7	10
2	11	13
>2 – <3	4	7
≥3	5	17
Total	153	158