

# Assessing the impact of deliberative processes on the views of participants: is it ‘in one ear and out the other’?

Tania Stafinski PhD\*, Devidas Menon PhD† and Yutaka Yasui PhD†

\*Associate Director, Health Technology & Policy Unit, School of Public Health, University of Alberta, Edmonton, AB, Canada,

†Professor, Public Health Sciences, School of Public Health, University of Alberta, Edmonton, AB, Canada

## Abstract

### Correspondence

Dr Tania Stafinski PhD  
3201 Research Transition Facility  
University of Alberta  
8308-114 Street  
Edmonton, AB T6G 2V2  
Canada  
E-mail: tancias@ualberta.ca.

### Accepted for publication

6 October 2011

**Keywords:** citizens' juries, deliberative processes, evaluation, social values

**Background** Interest in citizens' juries for eliciting the views of the public to inform coverage decisions on new health technologies has grown. However, evaluative information, particularly regarding their short- and/or longer-term impact on participants' views is limited. As citizens' juries can be resource intensive, such information is required to make ‘evidence-based’ decisions about their use.

**Objectives** To assess the impact of citizens' juries on participants' preferences for the distribution of health care across populations over time.

**Setting and participants** Two citizens' juries, involving a different representative sample of the public, were held. Participants completed identical questionnaires before (T1), directly after (T2) and 6 weeks following the jury (T3). Questionnaires comprised rating, ranking and choice-based questions related to four characteristics of competing patient populations (age, current health, life expectancy without treatment and health gain resulting from an intervention). Semi-structured telephone interviews were also conducted to explore the impact of the jury on participants' distributive preferences. Changes in responses to the self-administered survey over the three time points were assessed quantitatively, while interview questions were analysed using qualitative techniques.

**Results** No significant differences in responses to rating questions were observed. Pre/post-jury changes in the rankings of two factors were statistically significant in one of the juries. However, in both juries, T1–T2 changes in responses to several of the choice-based questions reached statistical significance. The number was lower between T2 and T3, suggesting that jurors retained their views. According to findings from the interviews, jurors' views changed or were clarified through participation in the jury.

**Conclusions** There appears to be evidence suggesting that the views of individuals who participate in citizens' juries change as a result of the experience, and those ‘informed’ views are sustained.

## Introduction

With increased recognition that resource allocation decisions in health care are both complex and value-laden, interest in deliberative processes for eliciting the views of the public to inform such decisions has grown.<sup>1–3</sup> Deliberative processes aim to gather input from an 'informed citizenry' (i.e. one who has had an opportunity to hear all perspectives related to an issue, discuss and debate options in a non-coercive fashion and arrive at a collective decision).<sup>4–7</sup> Thus, they have been used to seek the views of the public around complex policy issues in and outside of health care.<sup>4,8,9</sup> In recent years, one such process, citizens' juries, has received considerable interest from decision makers.<sup>1,10,11</sup> Citizens' juries bring together 12–16 individuals selected to be broadly representative of their community. Over a 2- to 4-day period, they learn about a relevant issue, hear from expert 'witnesses' who offer different perspectives, engage in deliberations among themselves and arrive at a common ground answer.<sup>4</sup> To date, evaluations of citizens' juries, while positive, have, for the most part, been limited to feedback questionnaires examining jurors' experiences and qualitative analyses of deliberations to assess jury competence and rationality.<sup>4,12,13</sup> There appears to be little information regarding their short- and/or longer-term impact on individual jurors' opinions. It has been proposed that citizens' juries may also serve as a mechanism for managing public expectations by facilitating a shift in attitudes from more self-interested to more socially aware ones.<sup>14,15</sup> Therefore, pre/post-jury assessments of participants' views are needed to determine the broader value of citizens' juries.

## Objectives

The purpose of this paper was to assess the impact of citizens' juries on jurors' preferences for the distribution of health care across populations. Specifically, it aimed to examine whether jurors' views on the importance of factors/patient characteristics that may be consid-

ered during resource allocation decision making for new health technologies changed following participation in the jury and, if they did, whether such views were retained over time.

## Methods

To assess the impact of citizens' juries on jurors' views, a mixed methods approach was used, involving pre/post-jury administration of a common questionnaire and individual telephone interviews with participants.

### The citizens' jury

Two citizens' juries [Northern Alberta Citizens' Jury (NA Jury) and Southern Alberta Citizens' Jury (SA jury)] were held over two and a half days using similar methods (i.e., the same facilitators, presentations, 'witnesses' and decision simulation exercises) but with different samples of the public to elicit distributive preferences for the allocation of resources across competing patient populations (T Stafinski, D Menon, Y Yutaka, unpublished data).<sup>16</sup> Each jury consisted of 16 participants, recruited to comprise a broadly representative sample of citizens residing in northern or southern Alberta. To accomplish this, personalized study information packages were first sent to 1500 individuals from each region. The names were obtained through random sampling of a commercial database of billing addresses for land and mobile telephone numbers (Survey Sampling International® Survey Sampling International, Markham, ON, Canada). Study packages invited respondents to participate in telephone screening interviews designed to collect information needed to assemble two juries that reflected the socio-demographic profiles of the two regions. Information on socio-demographics included age, gender, ethnicity, employment status and household income. Respondents were also asked about (i) potential affiliations with health-related special interest groups (e.g. patient advocacy groups) and (ii) employment as a health-care professional or with government. As the purpose of the juries was to obtain the views of 'ordinary

citizens' with no particular axe to grind, or whose voices might otherwise not be heard, those who met either of these two criteria were excluded from further consideration. To reduce the potential for volunteer bias, packages indicated that participants selected to comprise the jury would each receive a \$400 honorarium (plus reimbursement of jury-related expenses). To select 16 jurors for each jury, a combination of purposive and stratified random sampling was used. Eligible respondents were first grouped by gender and age. They were then stratified by level of education and household income (before taxes) and purposively selected to match the distribution of the population of the region. Finally, random sampling was used to select jurors when several individuals with the same set of characteristics had been identified.

In each jury, participants engaged in a series of increasingly complex trade-off exercises, which involved simultaneous consideration of multiple patient-related factors/characteristics. Based on their responses, preference statements, reflecting the extent to which jurors' choices depended on the presence of certain factors, were generated (see Appendix I).

#### Data collection

##### *Development of pre/post-jury and follow-up questionnaire*

For each jury, a separate self-administered survey comprising three different types of questions commonly used in studies designed to elicit social values or distributive preferences of the public for health care across the population was constructed.<sup>11</sup> They included the following: (i) ranking, (ii) rating (Likert) and (iii) choice-based (choice-based conjoint analysis) questions. Each type of question incorporated the following four factors/patient characteristics around which distributive preferences of the public have frequently been sought: (i) current health or severity of illness, (ii) imminence of death, (iii) age and (iv) health improvement or gain with the technology (Appendix II). Regarding the ranking question, jurors were asked to rank the four factors from most to

least important. The four Likert questions asked jurors to rate the importance of each factor on a five point scale (from very important to not important at all). The choice-based questions asked jurors to choose between two unique patient populations characterized by a different combination of categories or levels within individual factors, 'divide funds equally between populations' or 'let someone else decide' (Appendix II). Unique populations and pair-wise comparisons were generated using SPSS (IBM Canada, Markham, ON, Canada) Orthoplan, which creates 'scenarios' (unique populations) and choice sets (pair-wise comparisons) to obtain the maximum amount of information through the fewest possible comparisons. Comparisons were checked for plausibility and 'level balance' (i.e. the number of times each level or category of a factor appears should be approximately the same) before adding them to the questionnaire. Sixteen such comparisons were included, a number representing the upper limit of manageable cognitive burden.<sup>17</sup> In addition, one duplicate question was included to assess the reliability of jurors' responses. While ranking and rating questions were identical across the two questionnaires, the 16 choice-based questions differed (one set per jury).

##### *Administration of the questionnaire*

All jurors completed the same self-administered questionnaire three times: (i) at the beginning of the jury; (ii) at the end of the jury; and (iii) 6 weeks following the jury. The first two were completed in the room in which jury sessions were held. The third was mailed to jurors, along with a self-addressed, postage paid envelope and a cover letter, which repeated instructions for completing the questionnaire and included a reminder to answer questions independently. The cover letter also provided contact information for the researchers, should jurors require further clarification. The purpose of the 6-week follow-up questionnaire was to assess the stability of jurors' preferences over time. The 6-week time period was selected based upon findings from a review of behavioural psychology

literature in which studies examining the permanence of attitude changes to various types of information employed periods ranging from 2 days to 12 months (median: 1 month) and the need to minimize losses to follow-up.<sup>18–20</sup>

#### *Telephone interviews*

Upon receipt of completed follow-up questionnaires, semi-structured telephone interviews, each 15 min in length, were conducted. Jurors were asked about their overall experience participating in the jury and any impact it might have had on their perceptions of resource allocation decision making for new health technologies in the province. They were also asked questions related to their individual views/distributive preferences, including whether or not they felt they had changed as a result of the jury and, if yes, whether they felt they had changed again over the 6 weeks since the jury was held. They were then reminded of the final set of preference statements arrived at by the jury and asked whether they felt it reflected their individual preferences prior to the jury and whether they still agreed with its contents. Lastly, jurors were given the opportunity to provide any additional comments. To minimize interviewer bias, all interviews were conducted by the same researcher using a pre-tested interview guide.<sup>21</sup>

#### Analysis of questionnaires and interviews

##### *Questionnaires*

To assess the extent to which jurors' views changed immediately following the jury, responses to the first questionnaire (pre-jury survey) were compared with those to the second questionnaire (post-jury survey). To assess the extent to which jurors retained their views following the jury, responses to the second questionnaire were compared with those of the third questionnaire (follow-up survey). Both comparative analyses were performed using the following statistical tests. To identify statistically significant differences in jurors' responses to the ranking and rating questions, the Wilcoxon matched-pairs signed-rank test was used.<sup>22</sup> This test analyses the differences between paired measurements for

each study subject, generating a *P*-value that answers the question, 'If the median difference is actually zero, what is the likelihood that random sampling would have resulted in the median difference observed in this study?' If the *P*-value is small (i.e. <0.05), the possibility that the difference may be a coincidence is rejected.

For each of the choice-based questions, the number of jurors whose responses differed between the pre- and post-jury questionnaires and between the post-jury and follow-up questionnaires was counted. This value was then used to calculate the proportion of jurors who changed their minds on an individual question. It was assumed that the probability of a juror changing his/her mind on an individual question followed a binomial distribution (change or no change). Confidence intervals were then calculated around the proportion of jurors whose responses differed on each question.<sup>23,24</sup> For the purposes of this paper, a meaningful value for the proportion of jurors who changed their minds was set at 0.20. Therefore, if a confidence interval around a proportion fell entirely above 0.20 (20%) that value was considered to be statistically significant.<sup>25</sup> Lastly, the probability of a Type I error was set as 0.05, corresponding to 1/20. As there were 16 choice-based questions, the maximum number expected to change by chance, alone, was, at most, one.

##### *Interviews*

Interviews were digitally recorded, transcribed and analysed using content analytic and constant comparison techniques.<sup>12,21</sup> Data (chunks of information) were sorted, arranged and coded using qualitative data management software (NVivo®, QSR, Cambridge, MA, USA). All transcripts were analysed by the same researcher who conducted the interviews. To assess observer bias, a second researcher independently analysed transcripts from two, randomly selected interviews. Findings from both researchers were subsequently compared.<sup>26</sup>

Lastly, responses to the questionnaire were compared with those collected through interviews to determine the extent to which they could be triangulated.

The study received ethics approval from the University of Alberta Health Research Ethics Board.

## Results

Socio-demographic characteristics of the 16 individuals who comprised each jury are presented in Table 1. Slight variations in the distribution of age, education and household income between juries were found, but none achieved statistical significance (all *P*-values > 0.5).

All 16 jurors of the Northern Alberta Citizens' Jury completed the three questionnaires. In the Southern Alberta Citizens' Jury, one participant was not able to stay for the entire jury session. Therefore, only 15 of the 16 jurors completed all three questionnaires and were included in the analyses.

Pre/post-jury comparisons of responses to common questionnaire

### Ranking question

Findings from a comparative analysis of the Northern Alberta jurors' rankings of the four factors/patient characteristics prior to and directly after the jury are presented in Table 2. Statistically significant differences in the rankings of 'current health/severity of illness' and 'health gain/improvement' were observed. 'Current health/severity of illness' decreased in rank, while 'health gain/improvement' increased, moving from a median rank of third to first. Jurors considered health gain first, consistently favouring patient populations if the available health gain could at least bring them to sufficient functioning.

In the Southern Alberta Jury, no statistically significant differences in the rankings of jurors prior to and directly following the session were found.

### Rating questions

For both juries, no statistically significant differences in jurors' ratings of the importance of each factor/patient characteristic were detected (Table 3).

**Table 1** Comparison of socio-demographic profiles of the two juries

Characteristic	Number of jurors (%)		<i>P</i> -value*
	Southern Alberta Jury	Northern Alberta Jury	
Gender			
Male	8 (50)	8 (50)	0.64
Female	8 (50)	8 (50)	
Age			
18–24	2 (13)	2 (13)	1.00
25–34	2 (13)	2 (13)	
35–44	2 (13)	2 (13)	
45–54	4 (26)	3 (19)	
55–64	3 (19)	3 (19)	
65–74	2 (13)	2 (13)	
> 74	1 (6)	2 (13)	
Education (highest level)			
< High school	1 (6)	1 (6)	1.00
High school	4 (25)	5 (31)	
Post-secondary diploma	4 (25)	4 (25)	
Undergraduate degree	4 (25)	4 (25)	
Graduate degree	3 (19)	2 (13)	
Annual household income (\$ Cdn, before taxes)			
< \$25 000	2 (13)	3 (19)	1.00
\$25 000–\$45 000	4 (25)	4 (25)	
\$46 000–\$70 000	3 (19)	3 (19)	
\$71 000–\$100 000	4 (25)	3 (19)	
> \$100 000	3 (19)	3 (19)	
Employment status			
Employed	12 (75)	12 (75)	1.00
Unemployed	2 (13)	2 (13)	
Retired	2 (13)	2 (13)	
Ethnicity			
Asian	2 (13)	1 (6)	0.60
Caucasian	13 (81)	13 (81)	
First nations (aboriginal)	1 (6)	2 (13)	
Geographic location			
Urban	12 (75)	12 (75)	0.66
Rural	4 (25)	4 (25)	

\*Fisher's Exact Test.

### Choice-based questions

In the Northern Alberta Jury, the percentage of jurors whose responses to an individual question differed post-jury from those pre-jury ranged from 31 to 63% (Table 4). There were no questions to which all jurors answered the same way both times. Further, for six of the questions, the percentage of jurors who changed their minds

**Table 2** Comparison of pre-, post- and follow-up jury questionnaire responses to ranking question for both juries (rank of 1 = highest)

Factor/patient characteristic	Jury	Median score			Pre- to post-jury P-value*	Post-jury to follow-up P-value*
		Pre-Jury	Post-Jury	Follow-up		
Age	Northern Alberta	2.5	3	4	0.68	0.61
	Southern Alberta	3	2	3	0.52	0.50
Current health/severity of illness	Northern Alberta	2.5	3	3	0.039	0.43
	Southern Alberta	3	3	2	0.56	0.45
Imminence of death	Northern Alberta	2	2.5	2	0.25	0.16
	Southern Alberta	2	3	3	0.12	0.94
Health gain/improvement	Northern Alberta	3	1	2	0.016	0.16
	Southern Alberta	3	3	2	0.75	1.0

\*Wilcoxon signed rank test.

reached statistical significance. As observed in the ranking question, responses provided to the post-jury questionnaire were consistent with the jury's collective preference statements. Most jurors selected patient populations who could at least achieve sufficient functioning. When health gains were the same across populations, there was a preference for funding those who were 'worse-off' or facing imminent death without treatment. Lastly, few jurors selected the 'divide resources equally' option, instead they chose one of the two patient populations.

In the Southern Alberta Jury, values were slightly lower, ranging from 14 to 53%, and the percentage of jurors who changed their minds was statistically significant in five of the questions. As in the Northern Alberta Jury, such 'changes' were consistent with the jury's collective preference statements. Patient populations facing imminent death or considered the most severely ill were only favoured when the available health gain was enough to restore them to sufficient functioning.

Importantly, all jurors in each jury answered duplicate questions consistently (i.e. the same way), suggesting that their responses were reliable.

Post-jury and follow-up comparisons of responses to common questionnaire

#### *Ranking question*

For both juries, no statistically significant differences in jurors' rankings of the four fac-

tors/patient characteristics at the end of the jury session and 6 weeks later were observed (Table 2).

#### *Rating questions*

As in the pre/post-jury comparison, no statistically significant differences in jurors' views of the importance of each factor were found 6 weeks after either jury (Table 3).

#### *Choice-based questions*

In both the Northern and Southern Alberta Juries, the range of percentages of jurors whose responses differed between the post-jury and follow-up jury questionnaires was lower than that found for the pre- and post-jury comparisons (NA Jury: 13–50%; SA Jury: 7–40%) (Table 4). Additionally, the number of questions in which the proportion of jurors who changed their minds reached statistical significance was lower (NA Jury: 2; SA Jury: 0). In the Northern Alberta Jury, the two questions both involved trade-offs between populations in which neither could be returned to sufficient functioning. None of the preference statements generated through the actual jury session addressed such circumstances. Therefore, a comparison of the two results was not possible. Again, all jurors in each jury answered duplicate questions consistently.

#### *Findings from interviews*

All 31 jurors indicated that the jury had affected their views in some way. Approximately

**Table 3** Comparison of pre-, post- and follow-up jury questionnaire responses to rating Likert questions for both juries

Questions	Jury	Times	Frequencies of answers (row %)						Total	Medians (pre- to post-jury)	P-value*	Medians (post- to follow-up jury)	P-value*
			Very important	Important	Moderately important	Of little importance	Not important at all						
Age	Northern Alberta	Pre-jury	7 (44)	3 (19)	4 (25)	2 (13)	0	16 (100)	2 to 2	0.96	2 to 2	0.47	
		Post-jury	5 (31)	7 (44)	3 (19)	0	1 (6)	16 (100)					
		Follow-up	7 (44)	4 (25)	3 (19)	1 (6)	1 (6)	16 (100)					
	Southern Alberta	Pre-jury	3 (20)	5 (33)	5 (33)	1 (7)	0	15 (100)	2 to 1	0.08	1 to 2	0.34	
		Post-jury	8 (53)	2 (13)	5 (33)	0	0	15 (100)					
		Follow-up	5 (33)	4 (27)	6 (40)	0	0	15 (100)					
Current health / severity of illness	Northern Alberta	Pre-jury	5 (31)	7 (44)	4 (25)	0	0	16 (100)	2 to 2	0.48	2 to 2	0.32	
		Post-jury	4 (25)	7 (44)	4 (25)	1 (6)	0	16 (100)					
		Follow-up	3 (19)	11 (69)	2 (13)	0	0	16 (100)					
	Southern Alberta	Pre-jury	3 (20)	10 (67)	2 (13)	0	0	15 (100)	2 to 2	0.56	2 to 2	0.61	
		Post-jury	7 (47)	6 (40)	1 (7)	1 (7)	0	15 (100)					
		Follow-up	6 (40)	6 (40)	2 (13)	1 (7)	0	15 (100)					
Imminence of death	Northern Alberta	Pre-jury	7 (44)	7 (44)	2 (13)	0	0	16 (100)	2 to 2	0.45	2 to 2	0.78	
		Post-jury	6 (38)	6 (38)	4 (25)	0	0	16 (100)					
		Follow-up	7 (44)	6 (38)	4 (25)	0	0	16 (100)					
	Southern Alberta	Pre-jury	6 (40)	6 (40)	2 (13)	1 (7)	0	15 (100)	2 to 2	0.83	2 to 2	0.71	
		Post-jury	5 (33)	7 (46)	2 (13)	1 (7)	0	15 (100)					
		Follow-up	5 (33)	7 (47)	3 (20)	0	0	15 (100)					
Health gain / improvement	Northern Alberta	Pre-jury	9 (56)	5 (31)	2 (13)	0	0	16 (100)	1 to 1	0.08	1 to 1	0.32	
		Post-jury	13 (81)	3 (19)	0	0	0	16 (100)					
		Follow-up	12 (75)	3 (19)	3 (19)	0	0	16 (100)					
	Southern Alberta	Pre-jury	7 (47)	6 (40)	2 (13)	0	0	15 (100)	2 to 1	0.10	1 to 1	0.66	
		Post-jury	9 (60)	6 (40)	0	0	0	15 (100)					
		Follow-up	9 (60)	5 (33)	1 (7)	0	0	15 (100)					

\*Wilcoxon signed rank test.

**Table 4** Comparison of pre-, post- and follow-up jury questionnaire responses to choice-based questions in both juries

Pair-wise questions		Pre- to post-jury		Post- to follow-up	
Patient population A	Patient population B	Proportion of jurors who changed their responses	95% confidence interval†	Proportion of jurors who changed their responses	95% confidence interval†
<b>Northern Alberta Jury</b>					
Y, Mol, weeks, FF	Y, Mil, 1 year, SF	6/16 (0.375)	0.152–0.646	4/16 (0.250)	0.073–0.524
O, Mil, weeks, SF	Y, Mil, 1 year, FF	6/16 (0.313)	0.152–0.646	3/16 (0.188)	0.041–0.457
O, Mil, weeks, IF	O, Mol, 5 years, SF	4/16 (0.250)	0.072–0.524	3/16 (0.188)	0.041–0.457
Y, Mil, weeks, FF	Y, SI, 1 year, SF	6/16 (0.375)	0.152–0.646	3/16 (0.188)	0.041–0.457
Y, SI, 5 years, SF	O, Mil, 1 year, FF	5/16 (0.313)	0.110–0.587	5/16 (0.313)	0.110–0.587
Y, Mol, weeks, IF	O, SI, weeks, FF	9/16 (0.563)	0.299–0.803*	7/16 (0.438)	0.198–0.701
O, Mol, 5 years, SF	Y, Mol, weeks, SF	5/16 (0.313)	0.110–0.587	6/16 (0.375)	0.152–0.646
Y, Mol, 5 years, IF	Y, SI, 5 years, IF	8/16 (0.500)	0.247–0.754*	9/16 (0.563)	0.299–0.803*
O, Mol, weeks, FF	O, Mol, 1 year, SF	6/16 (0.375)	0.152–0.646	4/16 (0.250)	0.073–0.524
Y, Mil, weeks, IF	Y, Mol, 1 year, IF	7/16 (0.438)	0.198–0.701	7/16 (0.438)	0.198–0.701
O, SI, weeks, FF	O, Mil, 1 year, SF	7/16 (0.438)	0.198–0.701	4/16 (0.250)	0.073–0.524
Y, Mil, 1 year, IF	Y, Mil, 5 years, FF	6/16 (0.375)	0.152–0.646	2/16 (0.125)	0.016–0.384
Y, SI, weeks, IF	O, Mol, 5 years, FF	9/16 (0.563)	0.299–0.803*	2/16 (0.125)	0.016–0.384
O, SI, weeks, IF	Y, SI, 1 year, IF	8/16 (0.500)	0.247–0.754*	9/16 (0.563)	0.299–0.803*
O, SI, 5 years, SF	Y, Mil, 5 years, IF	8/16 (0.500)	0.247–0.754*	4/16 (0.250)	0.073–0.524
O, SI, weeks, IF	O, Mol, 5 years, IF	9/16 (0.563)	0.299–0.803*	4/16 (0.250)	0.073–0.524
<b>Southern Alberta Jury</b>					
Y, Mol, weeks, FF	Y, Mil, 1 year, SF	7/15 (0.467)	0.213–0.734*	4/15 (0.267)	0.078–0.551
O, Mol, weeks, SF	Y, Mil, 1 year, FF	8/15 (0.533)	0.267–0.787*	5/15 (0.333)	0.118–0.616
Y, Mil, weeks, FF	O, SI, 1 year, FF	6/15 (0.400)	0.163–0.617	4/15 (0.267)	0.078–0.551
Y, SI, 5 years, FF	O, Mil, 1 year, FF	7/15 (0.467)	0.213–0.734*	5/15 (0.333)	0.118–0.616
Y, Mol, weeks, IF	O, Mil, weeks, FF	2/15 (0.133)	0.017–0.405	2/15 (0.133)	0.017–0.405
O, Mil, 5 years, SF	Y, Mol, weeks, SF	5/15 (0.333)	0.118–0.616	1/15 (0.067)	0.002–0.320
O, Mol, weeks, FF	O, Mil, 1 year, SF	3/15 (0.200)	0.017–0.405	3/15 (0.200)	0.043–0.481
O, SI, weeks, FF	O, Mil, 1 year, SF	5/15 (0.333)	0.118–0.616	3/15 (0.200)	0.043–0.481
O, SI, weeks, IF	Y, SI, weeks, SF	7/15 (0.467)	0.213–0.734*	5/15 (0.333)	0.118–0.616
Y, SI, 5 years, SF	O, Mil, 1 year, SF	8/15 (0.533)	0.267–0.787*	5/15 (0.333)	0.118–0.616
O, SI, 5 years, SF	O, SI, 1 year, FF	2/15 (0.133)	0.043–0.481	2/15 (0.133)	0.017–0.405
Y, Mol, 5 years, FF	O, Mol, 1 year, FF	7/15 (0.467)	0.213–0.734*	6/15 (0.400)	0.163–0.677
Y, SI, 5 years, IF	O, SI, 5 years, FF	8/15 (0.533)	0.267–0.787*	6/15 (0.400)	0.163–0.677
Y, Mol, 1 year, IF	Y, Mol, 5 years, FF	3/15 (0.200)	0.017–0.405	3/15 (0.200)	0.043–0.481
O, SI, 5 years, IF	Y, SI, 1 year, IF	4/15 (0.267)	0.078–0.551	2/15 (0.133)	0.017–0.405
Y, Mol, 1 year, IF	Y, Mol, 5 years, FF	3/15 (0.200)	0.017–0.405	1/15 (0.067)	0.002–0.320

\* $P < 0.05$  for  $H_0$ : Proportion changed = 0.20, † Based on a binomial distribution.

Y, young; O, old; SI, severely ill; Mol, moderately ill; Mil, mildly ill; FF, full functioning; SF, sufficient functioning; IF, insufficient functioning.

one-third felt it had helped to clarify their views ('Yeah, I now say 'it depends' a lot more, and I am okay with that'; 'I couldn't get my head around how I felt before the jury. Now at least I know how I am going to think about things'; and 'I learned so much – not just about the healthcare system but about myself, what I believe in'; and 'I am not sure I had a clue before'). Approximately one-third thought it

had actually changed their views ('Holy smokes, this stuff is really complicated... I mean now I realize it is never as easy as just helping the worst off people...you gotta think about so much more, like what are we really gettin' out of it'; 'I think, hope, I am less judgemental now about, like, people who do unhealthy things. You know, when you think about trading off people, it forces you to think twice about whether you

want to use that against them'; and 'I totally changed my mind, 180°...it's about more than [the] greatest good for the greatest number saying'). Over half of the jurors noted the complexities involved in making resource allocation decisions and expressed empathy for those charged with such a task ('I wish all Albertans could have a chance to participate in one of these. I have a way better handle on how hard these decisions really are'; 'Gosh, I wouldn't trade places with the health minister any day of the week...[he] has a tough job'; 'To have to say 'no' to a family, I just know I couldn't do it, yet it has to be done'; and 'I appreciate how difficult a job it is now'). Just over one quarter mentioned feeling hopeful or reassured by the views of their fellow jurors ('For me, the experience was reassuring. I think we can be proud of what we came up with, and there is no way that we could have thought that way together without something like this'; 'I was happy to see that people weren't just automatically going to throw old people like me under the bus'; 'I was surprised how easy it was for us to make some decisions – like we were on the same page even though we were pretty different'; and 'I think we did a great job – move over [health minister's name removed]').

All but one of the jurors thought their views had not changed since the jury ('Nope, my memory isn't that bad yet'; 'After all those exercises and discussions, things will be stuck in there for a while'; and 'No, but I think about how much I think society should be willing to give up for things a lot more, like that Zamboni procedure'). The single juror who stated otherwise mentioned that (s)he had changed her personal views around 'last chance' therapies which may offer important but small health gains. While (s)he felt (s)he had 'written them off' during the jury, (s)he realized afterwards that there may be value to loved ones, which (s)he hadn't considered at the time ('I never voted for the small improvement, even when they [patients] were pretty ill. There is a psychosocial benefit that I didn't pay much attention to and should have').

When asked about the set of collective preference statements formulated by each jury, all 31

jurors stated that it still reflected their views ('Yes, I am still pretty comfortable with them'; 'I still like them and even tell others about them'; and 'I think so, I mean, when I read them again I still feel okay about them'). In contrast, only 13 of the 31 jurors thought that the set accurately captured their views prior to the jury ('Can't say would have thought of these things beforehand, but I do think they are what I believe now'; 'Nope, no way I would have thought that way'; 'I wish I could say that they did, but no, I am afraid not'; and 'Values are so hard to explain to people. Until the jury, I hadn't thought about them as distributive preferences – that helped me out a whole bunch').

## Discussion

This study assessed the immediate and longer-term impact of citizens' juries on the distributive preferences/views of participants regarding resource allocation decision making for new health technologies. To our knowledge, it represents the first attempt to evaluate citizens' juries using a mixed methods approach, combining traditional feedback interviews and repeated administration of an identical questionnaire before, immediately following and 6 weeks after the jury. According to the results of qualitative analyses of the telephone interviews, the views of jurors changed as a result of the citizens' jury and were retained. Findings from quantitative analyses of questionnaire data were less clear, but appeared to be similar to the results of the jury sessions (i.e. collective preference statements). For example, during such sessions, both juries indicated a preference for funding health technologies that could return patient populations to at least 'sufficient functioning', regardless of the characteristics of that patient population (e.g. severity of illness). Individual juror's responses to choice-based questions in the post-jury surveys demonstrated a consistent willingness to 'trade-off' more severely ill populations who, with access to a particular health technology, could not achieve enough health gain to restore sufficient functioning, for less ill populations who could

achieve such functioning. Further, their preferences did not vary with age of the competing populations. In one of the juries, changes in the ranking of two of the four factors from before the jury to immediately after the jury were statistically significant. However, in both juries, no statistically significant differences in responses to the rating questions over time were observed. The number of choice-based questions in which the proportion of jurors whose responses changed reached statistical significance and the magnitude of the change (number of jurors with differing responses) were less for the post-jury follow-up comparison than for the pre/post-jury comparison. Thus, 'change' appeared to depend on the type of question. The lack of differences in responses to the rating questions might be explained by the fact that such questions do not require a trading-off of 'goods'. In other words, jurors may have realized that they were not losing or 'giving up' anything by simply providing an immediate response. While this may have also been the case for the ranking questions, they do require consideration of the *relative* value of 'goods', as there is a single-ranking position to which only one of several 'goods' can be assigned.<sup>27</sup> Regarding the choice-based questions, the set of 16 differed between the two juries. Therefore, it was not possible to compare responses across juries. Although each set was generated using the same approach, questions may not have represented equivalent 'difficulty' levels. Nonetheless, in both juries, the majority of responses to the post-jury and follow-up questionnaires were consistent with the final preference statements of the jury session, suggesting that jurors may have adopted and retained a more societal view as a result of the experience. According to findings from the interviews, the jury appeared to play a role in shaping the views of participants. However, there are several limitations of the study. First, the permanence of jurors' 'informed' views was based on a 6-week follow-up only. While this time period has frequently been used to assess the permanence of attitude changes in the behavioural sciences, the extent to which it may be adequate in the context of social values is not

clear. Also, the three questionnaires were not administered in the same setting (i.e. the third was mailed to jurors' homes). Although jurors were asked to complete questionnaires independently, it was not possible to ensure compliance with the request. Thus, respondent bias may have been introduced. Finally, the findings are based on a limited type and number of questions contained in questionnaire.

## Conclusions

Based on the qualitative and quantitative results of this study, there appears to be some evidence suggesting that the views of individuals who participate in citizens' juries change as a result of the experience, and those 'informed' views are retained a minimum follow-up period of several weeks.

## Acknowledgements

This study was made possible through funding of a New Emerging Team grant from the Canadian Institutes of Health Research.

## Conflict of interest

The authors have no conflicts of interest to declare.

## References

- 1 Government of Alberta. *A Foundation for Alberta's Health System*. Report of the Minister's Advisory Committee on Health: A New Legislative Framework for Health. Edmonton, AB, Canada: Alberta Health and Wellness, 2010. Available at: <http://www.health.alberta.ca/documents/MACH-Final-Report-2010-01-20.pdf>.
- 2 Stafinski T, Menon D, McCabe C. To fund or not to fund: development of a decision-making framework for the allocation of resources to new health technologies. *Pharmacoeconomics*, 2011; **29**: 1–10.
- 3 Stafinski T, Menon D, Philippon D, McCabe C. Health technology funding decision making processes around the world: the same yet different. *Pharmacoeconomics*, 2011; **29**: 475–495.
- 4 Coote A, Lenaghan J. *Citizens' Juries: Theory into Practice*. London, UK: Institute for Public Policy Research, 1997.

- 5 Lenaghan J. Involving the public in rationing decisions. The experience of citizens juries. *Health Policy*, 1999; **49**: 45–61.
- 6 Werner P. Israeli lay persons' views on priority-setting criteria for Alzheimer's disease. *Health Expectations*, 2009; **12**: 187–196.
- 7 Elster J (ed.). *Deliberative Democracy* [Cambridge Studies in the Theory of Democracy]. Cambridge, UK: Cambridge University Press, 1998.
- 8 Armour A. The citizens' jury model of public participation. In: Renn O, Webler T, Wiedemann P (eds) *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse*. Dordrecht, The Netherlands: Kluwer, 1996: 175–188.
- 9 Aldred J, Jacobs M. Citizens and wetlands: evaluating the Ely citizens' jury. *Ecological Economics*, 2000; **34**: 217–232.
- 10 Ryan M, Scott DA, Reeves C *et al.* Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment*, 2001; **5**: 1–186.
- 11 Stafinski T, Menon D, Marshall D, Caulfield T. Societal values in the allocation of health care resources: is it all about the health gain? *The Patient*, 2011; **4**: 1–19. (forthcoming).
- 12 Renn O, Webler T, Wiedemann P (eds). *Fairness and Competence in Citizen Participation: Evaluating Models for Environmental Discourse* [Risk, Governance and Society], Vol. 10. New York: Springer, 1995.
- 13 Edmonton G, Mercer D. Scientific literacy and the jury: reconsidering jury 'competence'. *Public Understanding of Science*, 1997; **6**: 327–359.
- 14 Barnes JA. *Building a Deliberative Democracy: An Evaluation of Two Citizens Juries*. London, UK: Institute for Public Policy Research, 1999.
- 15 McIver S. *Health Debate? An Independent Evaluation of Citizens' Juries in Health Settings*. London, UK: King's Fund Publishing, 1998.
- 16 Stafinski TD. Social values and their role in allocating resources for new health technologies, dissertation. Edmonton, AB, Canada: University of Alberta, 2010.
- 17 Louviere J, Hensher D, Swait J. *Stated Choice Methods: Analysis and Applications*. Cambridge, UK: Cambridge University Press, 2000.
- 18 Boninger DS, Brock TC, Cook TD, Gruder CL, Romer D. Discovery of reliable attitude change persistence resulting from a transmitter tuning set. *Psychological Science*, 1990; **1**: 268–271.
- 19 Conner MF, Sheeran PF, Norman PF, Armitage CJ. Temporal stability as a moderator of relationships in the theory of planned behaviour. *British Journal of Social Psychology*, 2000; **39**: 469–493.
- 20 Sengupta J, Goodstein RC, Boninger DS. All cues are not created equal: obtaining attitude persistence under low-involvement conditions. *Journal of Consumer Research*, 1997; **23**: 351–361.
- 21 Crabtree BF, Miller WL (eds). *Doing Qualitative Research* [Research Methods for Primary Care], 2nd edn. Thousand Oaks, CA, USA: Sage Publications, 1999.
- 22 Wilcoxon F. Individual comparisons by ranking methods. *Biometrics*, 1945; **1**: 80–83. Available at: <http://www.jstor.org/stable/pdfplus/3001968.pdf?acceptTC=true>, accessed 27 October 2011.
- 23 Pezzullo JC. Exact binomial and poisson confidence intervals, 2009. Available at: <http://statpages.org/confint.html>, accessed 25 October 2011.
- 24 Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. *Statistics in Medicine*, 1998; **17**: 857–872.
- 25 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. New Jersey: Lawrence Erlbaum Associates, 1988.
- 26 Pape C, Mays N. Rigour in qualitative research. In: Pope C, Mays N (eds) *Qualitative Research in Health Care*, 3rd edn. Hoboken, NJ, USA: Wiley-Blackwell, 2006.
- 27 Alwin DF, Krosnick JA. The measurement of values in surveys a comparison of ratings and rankings. *Public Opinion Quarterly*, 1985; **49**: 535–552. Available at: <http://poq.oxfordjournals.org/content/49/4/535.full.pdf+html>, accessed 25 October 2011.

## Appendix I

Preference statements emerging from the two juries

1. There is a preference for funding health technologies that could return patient populations to at least sufficient functioning, regardless of the characteristics of that patient population.

The remaining five statements assume that the technology can restore patient populations to at least sufficient functioning:

1. If patient populations are otherwise the same, there is a preference for funding technologies that could benefit those who are the 'worst off' or most severely ill;

2. In general, there is a preference for funding health technologies to patient populations who are facing imminent death;

3. If the number of patients in each population is the same, there is a preference for funding health technologies that could benefit young populations, except when the older populations are facing imminent death and, with treatment, receive at least the same individual health gain as those in the young population; and

4. There is a preference for funding health technologies that could benefit the greatest number of patients (i.e. large populations over small populations) regardless of age, unless the individual health gain achieved through funding health technologies for a smaller number of patients is considered substantial. The amount of health gain needed depends upon age. A larger amount is required for small, older populations than for small, younger populations.

## Appendix II: Sample questions from pre/post-jury survey

Please imagine that you are the provincial Minister of Health. Your job is to decide which new health services the province should pay for. These health technologies may help different groups/populations of patients, and there is not enough money to fund all of them. When deciding which on to fund, you might consider:

Age	What is the average age of patients in the group who might benefit from the health technology?
Current health or severity of illness	How ill are the patients in the group who might benefit from the health technology?
Imminence of death	What is the life expectancy of patients in the group if the technology is not funded?
Health gain or improvement	How much will the health of patients in the group improve if the health technology is funded?

### Part A: Sample question

Please answer the following questions by placing a check mark in the box that best describes your views:

1. How important is it to think about the age of patients when deciding which new health services to fund?

- Very important   
 Important   
 Moderately important   
 Of little importance   
 Not important at all

### Part B: Sample question

Please rank the following four factors that might be used to help determine the priority of different patient groups for health care in order of importance to you, starting with the most important one:

Age  
 Current health  
 Imminence of death  
 Health gain

Most important	1.
↓	2.
	3.
Least important	4.

## Part C: Sample question

Please imagine that you only have enough money to fund one of two technologies. Technology A is used to treat illness A (patient population A), while technology B is used to treat illness B (patient population B). The two groups of patients are different.

The below table shows the possible characteristics of the patient groups.

Characteristic	Categories	Description
Age	Young (Y)	25 years old
	Old (O)	65 years old
Current health	Severely ill (SI)	Unable to perform daily activities (working, family or leisure); in extreme pain or discomfort; depressed
	Moderately ill (Mol)	Able to perform some daily activities; in moderate pain; mildly depressed
	Mildly ill (Mil)	Able to carry out daily activities; in mild pain
Imminence of death*	Will die within a few weeks (weeks)	
	Will die in 1 year (1 year)	
	Will die in 5 years (5 years)	
Health gain or improvement with treatment	Health returns to normal (what it was before the illness) [i.e. full functioning (FF)]	
	Health does not return to normal (pre-illness), but patients are able to perform most daily activities [i.e. sufficient functioning (SF)]	
	Health does not return to normal and patients are not able to carry out most daily activities on their own [i.e. insufficient functioning (IF)]. However, patients still improve a little	

\*Some disease progress slowly and others progress very quickly.

The next set of questions asks you to make a choice between funding a technology for Group A or Group B. Alternatively, you can choose to fund both groups equally or indicate that you would prefer that someone else makes the choice. For each question, please place a check mark in the box that best describes your view.

## Question 1

	Group A	Group B
Age	Young	Old
Current health or severity of illness	Severely ill	Mildly ill
Imminence of death without the technology	Will die in 5 years	Will die in 1 year
Health gain or improvement with the technology	Health returns to normal (as it was prior to the illness)	Health returns to normal (as it was prior to the illness)

Please check one of the following boxes:

Fund Group A

Fund Group B

Divide funds equally between the groups

Let someone else decide