Can a moral reasoning exercise improve response quality to surveys of healthcare priorities?

M Johri,1 L J Damschroder,2 B J Zikmund-Fisher,3 S Y H Kim,4 P A Ubel5

ABSTRACT

Objective: To determine whether a moral reasoning exercise can improve response quality to surveys of healthcare priorities.

Methods: A randomised internet survey focussing on patient age in healthcare allocation was repeated twice. From 2674 internet panel members from the USA and Canada, 2020 (79%) completed the baseline survey and 1247 (62%) completed the follow-up. We elicited respondent preferences for age via five allocation scenarios. In each scenario, a hypothetical health planner made a decision to fund one of two programmes identical except for average patient age (25 vs 65 years). Half of the respondents (intervention group) were randomly assigned to receive an additional moral reasoning exercise. Responses were elicited again 7 weeks later. Numerical scores ranging from −5 (strongest preference for younger patients) to +5 (strongest preference for older patients) were assigned to receive an additional moral reasoning exercise. Responses were elicited again 7 weeks later. Numerical scores ranging from −5 (strongest preference for younger patients) to +5 (strongest preference for older patients) were assigned to younger patients; 0 indicates no age preference.

Conclusions: Although people generally prefer allocating scarce resources to younger patients in all clinical contexts, this preference for younger patients was weaker in the intervention group. Indicators of response quality favoured the intervention group.

As a complement to assessments of safety, efficacy and economic efficiency, decision-makers seeking to allocate healthcare resources increasingly demand information on the values, preferences, ethical principles and beliefs of the constituencies they serve. In Canada, values-based decision-making and public engagement in the priority-setting process were recently identified as national priority areas for research.1 In the UK, both the National Institute for Health and Clinical Excellence (NICE) and the National Health Service (NHS) now view the values of the British public as vital to fulfilment of their democratic mandates.1 2

But what is the best way to identify the public’s values? This paper introduces a novel technique to investigate social values, evaluates its performance via a randomised experiment and assesses its potential contribution to strengthening values-based decision-making and public engagement. We evaluate our technique by exploring people’s attitudes toward the role of patient age in healthcare resource allocation. This topic was chosen because it has been researched extensively in the scientific literature13 14 and by NICE,15 which has assumed a leadership role in exploring social value judgments related to priority setting.2 We first motivate the enquiry with a brief review of current methods and findings.

Large-scale surveys have been widely used to explore social value judgements such as the role of age in healthcare allocation. Despite their ability to elicit information from a large cross-section of the public rapidly and at relatively low cost, these approaches suffer from at least four important shortcomings. (1) People may require time to form a considered judgement on an issue, while such surveys prompt responses based on initial reactions with little time to reflect.14 15 (2) Responses can be shaped inappropriately by subjective factors and framing effects.13 14 16 17 (3) The perspective of survey designers may also unintentionally skew results. Large-scale surveys have found extensive evidence of public preferences favouring allocation to younger age groups18 19 but health economists may systematically have neglected dimensions where preferences for age-based priority setting are neutral across age groups or where they favour the elderly.20

Deliberation is a form of discussion focussing on careful weighting of reasons for and against some proposition.21 22 Proponents of deliberative approaches argue that, compared with large-scale surveys, methods such as focus groups, citizen’s panels and citizen’s juries increase the quality of reflection contained in individual responses and reduce prejudicial responses.16 18 23 The most important initiative currently used by NICE to investigate public values is the Citizen’s Council: a 30-member panel reflecting the age, gender, socio-economic status and ethnicity of the people of England and Wales. The Council’s role is to engage in deliberation to help develop the broad social values that NICE should adopt in preparing its guidance.16 24 25 Reflections on age and priority setting in Citizen’s Councils have proven more nuanced and less conclusive than survey responses.26

Deliberative approaches have important strengths: chief among them is that they enhance response quality as respondents reflect on and
rewards the criteria used to adjudicate scenarios through conversation.14 15 Although superior in certain respects, deliberative approaches have drawbacks. For example, the number of participants who can meaningfully deliberate at one time is small. Deliberative debates, therefore, become exclusive processes restricted to a small group of citizens.20 This raises concerns about issues such as (1) participant selection; (2) adequate representation of community views;21 (3) the high cost per respondent; (4) potential for dominance of group discussions by strong personalities; and (5) a lack of scientific generalisability due to inherent difficulties in replicating focus group results.14

We hypothesised that it is possible to draw upon some of the strengths of deliberation in an affordable manner that would enable widespread public participation. We conducted a web-based survey of social value judgments concerning healthcare allocation among age groups and used a randomised design to assess whether an intervention to improve moral reasoning influences the relative value people place on age. We also measured the impact of the moral reasoning intervention on response quality, decision-making characteristics and quality of moral reasoning.

METHODS

Participants
Study subjects were members of a panel of over one million internet users (the “Survey Spot Internet panel”) residing in the United States or Canada who voluntarily agreed to participate in research surveys.23 Subjects were recruited via email; participation was encouraged via entry into a sweepstakes upon survey completion. Web-based surveys are a cost-effective way to reach large numbers of respondents and are increasingly used for survey research.23 24 25 26 We employed a stratified sampling strategy mirroring the age and sex composition of the census population of the participant’s country of residence and an experimental survey design using computer-generated randomisation and blinding of participants to group assignment.

Questionnaire design
Respondents were presented with five allocation scenarios (Appendix 1; see online supplementary material). All were designed from the perspective of a health system planner faced (due to budgetary limitations) with a decision to fund one of two programmes. Programmes were identical except for patient age. Scenarios elicited preferences to fund a programme targeted towards younger patients (average age 65 years). They included three life-saving programmes (liver and lung transplants and coronary bypass surgery), depression treatment and palliative care. Allocation preferences were expressed on a sliding scale from −5 (target younger patients) to +5 (target older patients), with 0 indicating no preference between age groups. After responding to the five scenarios, subjects were asked about emotions evoked while responding to the survey, whether they wished to change their responses (although they were not permitted to do so) and to provide demographic information.

Study procedures
We randomised half of the respondents to a survey with an embedded moral reasoning exercise (intervention group) and half to a survey without (control group). Two survey versions reversing scenario order (liver transplant, palliative care, depression treatment, coronary bypass and lung transplant) were created and randomly assigned within both control and intervention groups. Liver and lung transplant scenarios were identical except for the organ specified. At no point were subjects able to refer to their previous responses. The questionnaire was administered at two time points: baseline and 7 weeks later (follow-up). Everyone received the same questionnaire at follow-up as at baseline (fig 1).

Intervention
For subjects in the intervention group, the questionnaire included a moral reasoning exercise. For each of the five scenarios, intervention group members were asked to read the scenario description, perform a moral reasoning exercise and provide allocation preferences. The exercise asked participants to select which 3 of 10 possible allocation principles they deemed most important for the scenario under study. The 10 principles appeared in random order for each participant; the order once presented was consistent within subject across the scenarios and the two time points. The list of 10 principles was developed initially from accounts of allocation decisions given by participants in a prior study on age and priority setting.23 We coded the rationales, supplemented and refined the list through a review of the literature in ethics and health economics, and condensed it to produce an inventory of principles pertinent to discussions of age and resource allocation (table 1). The intervention aimed to remind people of principles that they might not spontaneously have considered and to prompt them to reflect on their applicability prior to making a decision, as often happens in conversation with others. Moral reasoning can be viewed as a species of practical reasoning; that is, as a type of reasoning directed towards deciding what to do.23 The assigned exercise aims to enhance participants’ ability to determine a morally appropriate course of action in response to a specific scenario. We hypothesised that this moral reasoning exercise would improve response quality.

Participants in the control group were asked to respond to the five choice scenarios without undertaking the moral reasoning exercise. However, to gain insight into their moral reasoning, control group members were presented with the moral reasoning exercise once, after the last scenario (either liver or lung transplant, depending on the scenario order to which they were randomised).

Outcome measures
In order to assess respondents’ age-related allocation preferences, we compared response patterns between intervention and control groups on the −5 to +5 numerical response scale. We also compared the three principles selected by intervention and control group members as most important in the moral reasoning exercise for either liver or lung transplant (depending on survey version received). To gain insight into moral reasoning for the intervention group, we analysed their responses to the moral reasoning exercise for all scenarios.

We used four criteria, defined prior to fielding the study, to evaluate response quality:

1. Rates of “no preference” and “extreme” responses On our response scale, a value of 0 signals neutrality between age groups, while −5 (+5) reflects the strongest preference to allocate resources to younger (older) patients. Based on the responses of an earlier study, we expected that participants would exhibit higher rates of “no preference” responses for depression treatment and palliative care, but maintain strong preferences towards younger groups for the three

---

**Note:** The content continues with more information on the study design, data analysis, and results related to the intervention and outcome measures. However, the provided excerpt focuses on the methodology and key aspects of the study design, ensuring clarity and coherence in the text representation. Further details are necessary to complete the full understanding of the research outcomes and implications.**
Figure 1 Flowchart of trial. The study was originally designed as two identical experiments fielded in parallel: one in Canada and one in the USA. As analyses showed that preferences did not differ significantly between Canada and the USA, we pooled data in the results and present a consolidated design in this flowchart. We adjusted for country as appropriate in modelled analyses. All respondents who completed baseline analyses were invited to repeat the survey 2 weeks later. Experimental group assignments were preserved during follow-up analyses.
Economic productivity

Duration of benefit

Patient need

Relief from suffering

Capacity to benefit/best outcomes

Maximise number helped

Family responsibilities

Guarantee chance for "full life"

Equal treatment

Patient need

Relief from suffering

Duration of benefit

Capacity to benefit/best outcomes

Maximise number helped

Family responsibilities

Guarantee chance for "full life"

Equal treatment

Patient need

Relief from suffering

Duration of benefit

Capacity to benefit/best outcomes

Maximise number helped

Family responsibilities

Guarantee chance for "full life"

Allocation principle Sample comment* Scenario Liver transplant Palliative care Depression treatment Coronary bypass Lung transplant Overall

"All patients deserve the best medical care, regardless"

"Give it to the group that needs it most"

"We should always relieve pain when we can"

"Give the treatment to patients who will benefit most"

"Fund the programme that helps the most people"

"At 35, people are likely to be raising families"

"Give the younger patients a chance for full life"

"Giving the treatment to the younger group makes sense since they will enjoy it longer"

"Society shouldn’t pay to treat people who haven’t taken care of themselves"

"Help those who are most likely to have jobs they have to do"

46 55 50 44 47 48

42 54 50 47 41 47

34 68 33 34 37 41

35 24 33 35 37 33

31 34 36 28 31 32

38 17 31 33 36 31

22 26 36 28 23 27

24 12 15 21 21 19

16 7 8 22 17 14

11 3 8 9 9 8

*Analysis for intervention group at baseline only, n = 942. A similar analysis performed at follow-up is available upon request. The follow-up analysis showed no differences compared with the baseline analysis with the exception of a statistically significant reduction in selection of “family responsibilities” as a relevant allocation principle; for each scenario was introduced, intervention group respondents were asked to perform a moral reasoning exercise asking them to select 3 of a possible 10 allocation principles (reproduced here). They were asked to select principles relevant to the scenario under consideration, in response to the question: “How important is each principle to you for making a decision as to which programme to fund?” indicates a statistically significant difference (p < 0.05) based on logistic regression models adjusting for age and male, taking liver transplant as the reference scenario; several sample comments were provided for each allocation principle—these are illustrative only.

Table 1 Proportion (%) of respondents* selecting each allocation principle† by scenario‡

Statistics

We compared demographic characteristics across experimental groups using Student’s t test for continuous variables (that is, age) and χ² tests for categorical variables.

Numerical preference scores were assessed as continuous variables on the −5 to +5 response scale. We compared mean outcomes and frequencies of neutral and extreme responses across experimental groups using the Student’s t test. For each of the 10 allocation principles, we used logistic regression to model the odds of its occurrence using dummy variables for scenario. We assessed the impact of the intervention and survey timing on preference scores using cross-sectional time-series random-effects regression models adjusted for demographic characteristics (country, health status and education level) and

Statistical analyses

We compared demographic characteristics across experimental groups using Student’s t test for continuous variables (that is, age) and χ² tests for categorical variables.

Numerical preference scores were assessed as continuous variables on the −5 to +5 response scale. We compared mean outcomes and frequencies of neutral and extreme responses across experimental groups using the Student’s t test. For each of the 10 allocation principles, we used logistic regression to model the odds of its occurrence using dummy variables for scenario. We assessed the impact of the intervention and survey timing on preference scores using cross-sectional time-series random-effects regression models adjusted for demographic characteristics (country, health status and education level) and
clustered by participant. Results were confirmed using ordered logistic regression.

Analyses were performed using Stata (v.9) on all participants for whom reliable information on outcome measures was available (per protocol analysis).

Additional information on the methods can be found in the online supplementary material.

RESULTS

Sample characteristics

From April to June 2005, 2574 (8%) of internet users who received an email invitation responded by clicking onto the survey (fig 1). Of those who responded, 2020 (79%) completed the baseline survey and were invited to return and 1247 (62%) completed the follow-up 7 weeks later. Participants who admitted that they gave "intentionally wrong answers" were excluded from the analyses (11 individuals at baseline and 2 at follow-up). Overall, of the people who initially responded to the invitation, 78% were included in the baseline analyses (1239 receiving the intervention and 1335 controls) and 48% in the follow-up analyses (574 receiving the intervention and 671 controls).

Characteristics of subjects at baseline are described in table 2. Participants in experimental groups had similar characteristics.

Outcome measures

Analyses showed that preferences did not differ significantly between Canada and the United States. As recruiting mechanisms were identical, concurrent and from the same source, we pooled data and adjusted for country in regression analyses.

Numerical preference scores

Participants in both experimental groups preferred offering scarce resources to young patients over older patients in all

Table 2  Respondent characteristics at baseline*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Intervention group</th>
<th>Control group</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>942 (47%)</td>
<td>1067 (53%)</td>
<td>2009 (100%)</td>
</tr>
<tr>
<td>Age (years)</td>
<td>45 (SD 15)</td>
<td>46 (SD 15)</td>
<td>45 (SD 15) (14 to 95)</td>
</tr>
<tr>
<td>Male sex</td>
<td>457 (47)</td>
<td>497 (47)</td>
<td>48 (954)</td>
</tr>
<tr>
<td>Country of residence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USA</td>
<td>466 (49)</td>
<td>552 (52)</td>
<td>1018 (51)</td>
</tr>
<tr>
<td>Canada</td>
<td>476 (51)</td>
<td>515 (48)</td>
<td>991 (49)</td>
</tr>
<tr>
<td>Minority† (US only)</td>
<td>103 (22)</td>
<td>135 (23)</td>
<td>238 (23)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secondary school or less*</td>
<td>189 (20)</td>
<td>255 (24)</td>
<td>444 (22)</td>
</tr>
<tr>
<td>Some post-secondary†</td>
<td>483 (51)</td>
<td>499 (47)</td>
<td>982 (49)</td>
</tr>
<tr>
<td>Bachelors degree</td>
<td>182 (19)</td>
<td>231 (22)</td>
<td>413 (21)</td>
</tr>
<tr>
<td>Masters degree or more**</td>
<td>87 (9)</td>
<td>82 (8)</td>
<td>169 (8)</td>
</tr>
<tr>
<td>Self-rated health†‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excellent</td>
<td>112 (12)</td>
<td>114 (11)</td>
<td>226 (11)</td>
</tr>
<tr>
<td>Very good</td>
<td>349 (37)</td>
<td>400 (38)</td>
<td>749 (37)</td>
</tr>
<tr>
<td>Good</td>
<td>356 (37)</td>
<td>385 (36)</td>
<td>740 (37)</td>
</tr>
<tr>
<td>Fair</td>
<td>108 (11)</td>
<td>136 (13)</td>
<td>244 (12)</td>
</tr>
<tr>
<td>Poor</td>
<td>23 (2)</td>
<td>31 (3)</td>
<td>54 (3)</td>
</tr>
<tr>
<td>Employment status†‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full-time employment</td>
<td>470 (50)</td>
<td>518 (49)</td>
<td>999 (49)</td>
</tr>
<tr>
<td>Part-time employment</td>
<td>148 (16)</td>
<td>179 (17)</td>
<td>327 (16)</td>
</tr>
<tr>
<td>Not employed</td>
<td>324 (34)</td>
<td>370 (35)</td>
<td>624 (35)</td>
</tr>
<tr>
<td>Civil status‡</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single/never married</td>
<td>216 (23)</td>
<td>212 (20)</td>
<td>428 (21)</td>
</tr>
<tr>
<td>Married</td>
<td>496 (53)</td>
<td>582 (55)</td>
<td>1078 (54)</td>
</tr>
<tr>
<td>Divorced/separated/widowed</td>
<td>152 (16)</td>
<td>203 (19)</td>
<td>355 (18)</td>
</tr>
<tr>
<td>Living with partner</td>
<td>77 (8)</td>
<td>69 (6)</td>
<td>146 (7)</td>
</tr>
<tr>
<td>Annual income ($) (n = 1822):‡†</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;$20 000</td>
<td>121 (14)</td>
<td>164 (17)</td>
<td>285 (16)</td>
</tr>
<tr>
<td>20 000–49 999</td>
<td>360 (43)</td>
<td>417 (43)</td>
<td>777 (43)</td>
</tr>
<tr>
<td>50 000–74 999</td>
<td>211 (25)</td>
<td>210 (22)</td>
<td>421 (23)</td>
</tr>
<tr>
<td>75 000–99 999</td>
<td>84 (10)</td>
<td>102 (11)</td>
<td>186 (10)</td>
</tr>
<tr>
<td>100 000–150 000</td>
<td>48 (6)</td>
<td>59 (6)</td>
<td>107 (6)</td>
</tr>
<tr>
<td>&gt;$150 000</td>
<td>18 (2)</td>
<td>18 (2)</td>
<td>36 (2)</td>
</tr>
<tr>
<td>Time to complete survey (minutes)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>17 (SD 19)</td>
<td>13 (SD 16)</td>
<td>15 (SD 18)††</td>
</tr>
<tr>
<td>Follow-up</td>
<td>11 (SD 10)</td>
<td>9 (SD 7)</td>
<td>10 (SD 8)††</td>
</tr>
</tbody>
</table>

*Variables reported as frequency (per cent) unless otherwise stated. Sample size (n) is given at top of column unless otherwise indicated. Percentages may not sum to 100 due to rounding; †the denominator is 2009 unless otherwise indicated. Numbers may not sum to 2009 due to missing values; ‡information on race/ethnicity was sought only for US participants. Minority status is defined as "underrepresented minority" by the US Federal Office of Management and Budget (OMB) for the health professions, science and engineering (http://www.whitehouse.gov/omb/fedreg/race.ethnicity.html (accessed 27 October 2008)) and was defined as those who self-identified as black, Hispanic, American Indian/Alaskan Native or Pacific Islander; ‡denotes all those whose maximum educational attainment is successful completion of 12 years of formal schooling (high-school diploma); ††defined as those with a high-school diploma who attended trade school or attended some college. Those who completed a 2-year degree (for example, AA, AS) are included in this category; ‡‡"masters’ degree or more" refers to those respondents holding a masters degree, a doctoral degree or a professional degree such as an MD; ††the denominator for these variables is 2020. Numbers may not sum to 2020 due to missing values; ‡‡annual household incomes reported in nominal dollars; ††difference between intervention and control groups statistically significant at the p<0.001 level.
Table 3 Mean preference scores* for intervention and control groups at baseline and follow-up†

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Baseline</th>
<th>Follow-up</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intervention</td>
<td>Control</td>
</tr>
<tr>
<td>Liver transplant</td>
<td>−1.35</td>
<td>−1.89‡</td>
</tr>
<tr>
<td>Palliative care</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>Depression treatment</td>
<td>−0.71</td>
<td>−1.04‡</td>
</tr>
<tr>
<td>Coronary bypass</td>
<td>−0.83</td>
<td>−1.22‡</td>
</tr>
<tr>
<td>Lung transplant</td>
<td>−1.24</td>
<td>−1.86‡</td>
</tr>
</tbody>
</table>

*Scores range from −5 (preference for programme serving 35-year-olds) to +5 (preference for programme serving 65-year-olds); † indicates no preference between age groups; ‡ all comparisons based on unadjusted scores using the t test to assess the equality of means for independent populations with equal variances; †† difference between experimental groups (intervention baseline vs control baseline; intervention follow-up vs control follow-up) statistically significant at the p<0.001 level; * difference between experimental groups (intervention baseline vs control baseline; intervention follow-up vs control follow-up) statistically significant at the p<0.05 level; ††† difference within experimental groups (intervention baseline vs intervention follow-up; control baseline vs control follow-up) statistically significant at the p<0.001 level.

Assessment of “prejudicial reasons”
The coronary bypass scenario deliberately included a prejudicial factor stressing the importance of maintaining a healthy lifestyle. Intervention group members identified lifestyle as an important consideration more frequently in the bypass scenario than in any other scenario (table 1, 22% vs 16–17% for liver and lung and 7–8% for palliative care and depression), suggesting that our scenario description was effective at prompting people to think about this factor.

As predicted, however, the moral reasoning intervention reduced the weight people placed on this lifestyle factor. For the bypass scenario, the mean preference score for those in the intervention group is closer to 0 than for controls (table 1, p<0.001).

Temporal stability
While both groups modified their responses over time to reflect less age preference, control group responses were less stable (table 5). Responses for the control group showed a significant shift towards neutrality in preferences between age groups from baseline to follow-up for three of the five scenarios (liver transplant (p<0.001), lung transplant (p<0.05) and palliative care (p<0.05)). Responses for the intervention group showed a significant shift toward neutrality from baseline to follow-up for only one of the five scenarios (liver transplant; p<0.05).

DISCUSSION
Principal finding
Can a moral reasoning exercise improve response quality to surveys of healthcare priorities? We used an experimental design and elicited preferences for patient age via choice scenarios to study this question. Consistent with previous work, for both experimental groups, we found that age preferences favour younger patients,*2 23 26–28 and differ in strength according to the clinical context.*2 23 People displayed the strongest preference for the younger group in lifesaving scenarios when resources are especially scarce, such as transplants, but showed no age preference in the palliative care scenario.

The principal finding of this study is that age preferences were modified by the moral reasoning exercise. Specifically, for
all scenarios except palliative care, the strength of preferences favouring the young was significantly diminished among those receiving the moral reasoning intervention, although it was not eliminated. This change was accompanied by an improvement in response quality.

Strengths and weaknesses

This study has several important strengths, including a randomised design and a large and heterogeneous pool of respondents drawn from two countries. Among social value surveys of age, it is unique in assessing response quality and innovation are of lesser concern. However, future research is, to our knowledge, the first social value survey to assess the temporal stability of responses in this context.

Three issues merit close consideration:

1. Although we stratified the invitation list to mirror the United States and Canadian populations with respect to age and gender, concerns can be raised regarding the sample composition. Our strong 79% completion rate notwithstanding, given the initial 6% response rate for the baseline survey overall, only 6% of people in the United States and 15% in Canada invited to participate in the baseline survey actually completed it. Moreover, the sample is self-selected to include a predominance of internet users with higher income and higher educational attainment. These problems are common to internet surveys based on opt-in panels. In a review of internet surveys, Couper cites response rates ranging from 8–60% and notes the potential for response bias. Our intent was to conduct a randomised study to elicit responses from a heterogeneous sample of subjects so as to assess the potential of a novel intervention to improve response quality. We succeeded in recruiting a large and diverse pool of subjects to this task and do not intend to generalise the specific age preference values obtained to the general population. Given our objectives, issues of sample selection are of lesser concern. However, future research might usefully consider departing from opt-in panels by employing methods designed to represent general populations such as proportional sampling based on random digit dialling.

2. Although intention-to-treat (ITT) analysis is considered an ideal for randomised clinical trials, our study took a per-protocol approach. An ITT approach, which analyses all participants according to original experimental group assignment regardless of subsequent events, would have required us to impute missing age preference scores for all participants who failed to complete the survey. We judged this task inappropriate to the goals of the study. ITT analysis is legitimately defended as the gold standard when the goal is to assess the effect of a treatment or policy in a general population that will include deviations from protocol and non-compliant individuals. The question we are asking is a distinct one, oriented towards establishing the potential for an intervention effect in a large and heterogeneous population. As specified in the original protocol, our analyses included all participants who completed a survey and reported giving truthful answers. We feel that this approach is fully suited to our study question.

3. While the intervention serves to modify response patterns, does it improve response quality? Our study traced the impact of the moral reasoning intervention on response quality in four dimensions of which two were selected to permit comparison with findings from focus groups. Although there is no gold standard for the public’s age preferences, focus groups are commonly acknowledged to yield better-quality responses than standard large-scale surveys. In keeping with focus groups results suggesting that participants become more nuanced in their judgments and attenuate the strength of their age preferences through deliberation, we found that rates of no preference responses were greater and rates of extreme responses favouring the young were lower for the intervention group for all scenarios. Results from the bypass scenario are consistent with the intervention group giving less weight to a prejudicial statement, as documented in focus groups. However, these results are also consistent with the overall pattern of attenuation of age preferences. The remaining two criteria, internal consistency and temporal stability, are formal criteria of good response quality. Experimental groups were found to be equally consistent but the intervention group generally gave more stable responses over time. Temporal stability was assessed by comparing consistency in numerical responses between experimental groups at baseline and follow-up. While both groups modified their responses over time to reflect less age preference, control group responses were considerably less stable than those from the intervention group, showing significant short-term tracing that five scenarios (vs one of five for the intervention group).

Together, these results provide plausible evidence of improved response quality for the intervention group.

Interpretation of findings

The internal validity of the study is high: the fundamental finding that the moral reasoning exercise attenuated age preferences is unlikely to be due to bias in the design and conduct of this randomised study. In fact, four of the ten allocation principles presented in the moral reasoning exercise gave explicit rationales favouring younger patients; the rest made no reference to age. Due to issues of sample selection, specific age preference values from our study should not be attributed to the general populations of Canada and the United States. This caution notwithstanding, the qualitative insights generated from these results have a remarkable convergence with findings from other studies on age in priority settings conducted using different study designs, populations and methodologies.

The moral reasoning exercise generated valuable insights in its own right. Despite the diversity of scenarios, and although several principles explicitly favoured younger patients for all scenarios, the three principles selected as most important were equality of treatment, meeting patient needs and relief of pain and suffering. These social values do not reflect age preferences and are not aligned with the allocation principles (namely maximisation of best outcomes, duration of benefit, maximisation of the number helped) privileged in cost-effectiveness analysis (CEA).

CONCLUSIONS

Age preferences favouring the young exist, vary with clinical context and are significantly reduced when participants are encouraged to reflect on a wide range of moral principles in a moral reasoning exercise. These findings have several implications. Concerning the debate on the relevance of patient age for healthcare priority setting, they show that claims that the public favours age-based rationing on equity grounds have a much weaker empirical basis than previously thought, under-cutting calls for age weighting of health benefits in CEA. Findings from standard surveys using choice scenarios to elicit age preferences should be interpreted with caution. In the
absence of a moral reasoning exercise similar to our own, they are likely to overstate age preferences.39

Results from the moral reasoning exercise showed that the values held by the general public are in tension with those privileged in economic evaluations. Participants favoured allocation principles that supported procedural justice and humanitarian response to suffering over the consequentialist approach privileged in CEA methods and welfare econom-
ics.14 23 36 40 Since public values and CEA values are demonstrably distinct, public consultation is essential for democratic, respon-
sive setting of health priorities. The results of CEA should be interpreted and contextualised in light of broader public values related to equity and social justice.

Large-scale surveys have a unique contribution to make to the process of public consultation. In contrast to face-to-face deliberative methods, surveys extend the possibility of engaging a large and diverse population of respondents at a relatively low cost, enhancing external validity and replication of results.

Incorporation of a moral reasoning exercise into large-scale surveys of social values is a promising new strategy for engaging the public in the process of priority setting. It should be investigated as a tool for policy setting bodies seeking to incorporate social values in decisions concerning allocation of health resources.

Acknowledgements: We thank Dr Howard Bergman and François Bélond of the Saldanha Research Group for invaluable guidance. Funding: The study was supported by the Department of Veterans Affairs, Veterans Health Administration, Ann Arbor Veteran Affairs Healthcare System R&D Center of Excellence, and by grants from the Canadian Institutes of Health Research (CIHR), (Project number 43817) and the US National Institutes of Health (NIH) (R01-HD38963). Dr Johri is supported by a career development award from the American Cancer Society (MRSG 06-130-S1-CPPB). The funding agreements ensured the authors’ full independence in designing the study, interpreting the data, and writing and publishing the report.

Competing interests: None.

Ethics approval: Exemption from ethics approval was granted by the Institutional Review Boards of the University of Michigan Medical School (IRRBMED).

REFERENCES


6. Lewis PA, Charney M. Which of two individuals do you treat when only their ages are different and you can’t treat both? J Med Ethics 1985; 11:26–34.


