

NOT FOR CITATION WITHOUT PERMISSION

**SOLICITING PUBLIC PREFERENCES
FOR HEALTH CARE PRIORITIES:
A CRITICAL REVIEW OF METHODOLOGIES**

Anita Kothari

Please address correspondence to:

Anita Kothari
McMaster University
Centre for Health Economics and Policy Analysis
1200 Main Street West
Hamilton, Ontario, CANADA
L8N 3Z5
kotharar@fhs.mcmaster.ca

Cite as: Kothari A: Soliciting Public Preferences for Health Care Priorities: A Critical Review of Methodologies, McMaster University Centre for Health Economics and Policy Analysis Research Working Paper # 99-07, June 1999.

ABSTRACT

Various health care reforms have included a mandate to solicit local preferences for health care priorities. There is little consensus with respect to defining the participants, outlining the types of decisions which are appropriate for public input or establishing the way in which public preferences can be incorporated into decision-making. This paper critically reviews the methodological approaches that have been used to solicit public preferences for health care priorities in publicly funded systems. Exercises to solicit public preferences were carried out in the United States, England, New Zealand and the Netherlands. These experiences are critically reviewed using a framework which outlines requirements for internal validity, generalizability, reliability and objectivity. While the process of soliciting public preferences is still in its infancy, the analysis demonstrated that some aspects of internal validity, generalizability, reliability and objectivity were being given due attention. Key to developing a methodologically sound process is an *a priori* definition of the public (e.g., citizens or advocacy groups), from which appropriate data collection, sampling and analytical techniques follow. The potential instability of public preferences supports the use of multiple methods of data collection to corroborate findings. There is a need to share the research data, analysis and findings with the community.

INTRODUCTION

The call to include the public in decisions around the allocation of health care resources has come from many camps, at various points in time, to satisfy various perceived needs and obligations. The health care reforms of the mid-eighties and nineties have been the most recent drivers of public inclusion; these reforms, occurring worldwide, are generally aimed at increasing efficiency while containing costs. The mechanisms used to implement the goals of reform include such things as managed competition, decentralization, devolution and their variants and they are usually accompanied with a mandate for soliciting local input into health care priorities.

This mandate has stimulated debate around the extent and form of the public's inclusion in decision-making. There is little consensus with respect to defining the participants, outlining the types of decisions which are appropriate for public input or establishing the way in which public preferences can be incorporated into decision-making (1). The purpose of this paper is to clarify one aspect of this conundrum -- to critically review the methodological approaches that have been used to solicit public preferences for health care priorities in publicly funded systems. Preferences are defined here as those choices made on the basis of desires, which may or may not reflect actual health need. Presumably, preferences are influenced by such things as values, attitudes and beliefs.

Preferences for the distribution of health care resources can be solicited at various policy-making levels. At the clinical level, the health care provider may involve the patient in decision-making to determine the best treatment option in accordance with the patient's personal context and values. At an institutional level, hospital policies ensure patients receive care which is both effective and affordable; such institutional policies could be developed with information about the community's preferences for treatment alternatives.

Unlike these micro levels of decision-making, this paper is focused on macro levels of decision-making in the health care sector (i.e., public policy). Our focus is on decisions made by a state or regional authority, to determine which services will be offered, in what quantity and to whom. Other decisions at this level may include the size of the annual health care budget, and the amounts within the budget that will be allocated to various sectors. These decisions imply a trade-off between value sets, and consequently the public's values and preferences are seen by some to be just as or more legitimate than the politicians' (2,3). This legitimacy is based on the rights of community members as taxpayers, as potential patients and as citizens.

The processes used to solicit community preferences will be examined. More specifically, the scientific quality or soundness of these methods will be assessed using the principles of internal validity, generalizability, reliability and objectivity. Of less concern is whether the process achieved the intended goals, the scope of the issues considered, the extent to which the information about public preferences was actually used for decision-making or whether the process was appropriate given the specific context. Other researchers have discussed these issues; Kathlene and Martin (4), and Renn (2), consider participation within broader stages of public policy development (e.g., problem definition, information gathering and analysis, development of policy, policy outcomes), whereas our framework only considers the soundness of one of these stages (information gathering & analysis). In that we use an *a priori* developed framework to analyze previous experiences, our approach is more prescriptive than Lomas' (5).

The next section describes the search strategy that was used to identify examples of public participation in the context of current health care reforms, followed by the conceptual framework used to evaluate the various experiences. The framework is then applied to the cases of public participation in New Zealand, England, the United States and the Netherlands, followed by a discussion based on the analysis.

METHODS

A primary literature search was conducted using a combination of the MESH terms “social values”, “public policy”, “decision making”, “resource allocation” and “health care rationing” and the key terms “community participation”, “priority setting” and “public participation” with the databases Sociofile (1974-) and Medline (1980-), and the LCSH terms “citizen participation”, “decision making”, “health policy”, “policy sciences” and “resource allocation” and the key terms “community participation”, “public participation” and “priority setting” with the databases EconLit (1969-), Humanities Index (1984-), Social Sciences Index (1983-) and Sociofile (1974-). The reference lists of relevant articles were searched for additional references. Other sources of information included articles from personal files and referrals from researchers in this area.

A secondary strategy evolved after it became clear which jurisdictions would be applicable for this analysis. Medline (1985-) was searched again with the name of the jurisdiction and the MESH terms “resource allocation” and the key terms “public participation” and “health care rationing”. Additional articles were found using this strategy.

Articles were restricted to those written in English and in the published literature (i.e., secondary sources) as opposed to government documents and other grey literature. The articles were also limited to those describing developed countries which share contexts (e.g., health status profiles, general standards of living, available resources for health care, political stability) which differ considerably from those in under-developed or developing countries.

CONCEPTUAL FRAMEWORK FOR ANALYSIS

The framework which follows was developed to examine the soundness or methodological rigour of the processes that have been used to solicit public preferences. For policy-makers, this approach can be attractive for the very reason that it attempts to separate the politics of prioritizing health care services from the information used to influence the politics, making the process more credible to constituents. Attention to methodology represents a thoughtful and planned approach, one that produces research information rather than policy-makers' impressions of local preferences. Furthermore, this approach is in accordance with recent calls for more evidence-based decision-making in health care.

Our review of the literature revealed little if any direction in terms of what a methodologically sound process might encompass. Bowling writes, "The methodology for eliciting the public's priorities for health care is in its infancy." (6, p. 851). Donovan and Coast concur: "Determining the most appropriate methodology specifically for obtaining public views as part of a priority-setting exercise is not easy and only a limited amount of advice is available." (7, p. 215). For reasons of credibility and accountability, governments and the public might want the process to be conducted in a systematic and reproducible fashion. The process of scientific research is also characterized as systematic and reproducible, and thus serves as a reasonable starting point. Four principles key to an investigation are its internal validity, generalizability, reliability and objectivity. The underlying meanings of these principles were reviewed, along with possible design "threats" to their integrity. These principles were then applied to the context of a public preferences exercise. The framework developed for this analysis is presented in Table 1 and is adapted from the works of qualitative and quantitative researchers (8,9).

In the realm of soliciting preferences, internal validity is concerned with ensuring that the findings are true to the actual preferences¹ the community holds at that period in time. The ability to solicit true preferences may be hampered by the participants' misunderstanding of the issue or a focus on an unrelated issue. A strategy of "informed preferences" guards against this threat by ensuring that participants have a common understanding of the issue at hand. Another way to enhance internal validity is through the use of multiple data collection methods (e.g., surveys,

¹ Of course, we can never be sure that the "true preferences" are equivalent to the preferences that participants wish to reveal. For the purposes of this paper we will treat them as the same.

focus groups) to explore various domains of the issue and to confirm common emergent themes. Internal validity may be clouded by ambiguous findings, which may represent a true preference or an “error” in the research process. To clarify ambiguities, the process could include a probing opportunity during or after an initial stage of the research.

The “truth” of the solicited preferences may be subject to criticism if inconsistencies or a mismatch existed between the definition of community and the methods used to collect those preferences (see Table 2). Community preferences might refer to the wishes and underlying values that individuals within a geographical area possess. Individuals then serve as an appropriate source of information for preferences. Alternatively, community preferences may be perceived as something other than individual preferences. In this situation preferences are seen as collective in nature and may be measured at a geographic level or at a sub-group level, where groups are organized by member’s interests or affiliations (e.g., advocacy groups, groups of health professionals). If a collective measure of preferences is sought, than the method of summing up individual responses or outcomes to derive a grand level measure is flawed on the grounds that individual responses or attitudes are inherently “different” than a collective, group-level measure (10,11). To achieve such a measure, McIver argues that participants ought to have a forum for education and debate in order for a collective opinion to prevail. In other words, aggregate-level preferences infer a consensus of sorts resulting from discussion among group members.

The last indicator of internal validity is interpretation check, defined here as a technique of feeding back research results to the study participants to ensure that their views were interpreted correctly -- an attempt to establish face validity of results.

Having established the internal validity or “truth” of the findings, the next concern is the generalizability of those findings. Generalizability or external validity refers to the ability to infer that the preferences from the participants represent the views of the greater community (however defined), or that the preferences can be applied to a similar context in another jurisdiction. If quantitative research methods were employed, we would ask if an attempt was made to establish that participants were representative of the community. If qualitative research methods were used, it would be more appropriate to ask if multiple sources were tapped through purposive sampling to obtain a breadth of preferences.

Reliability refers to the reproducibility of the process, all else being equal. One indicator of reliability is replication -- was the process repeated and results compared? Another

method of establishing reliability is to ensure a transparent process, where the research logs are of sufficient quality and are available for public scrutiny, similar to the publication of scientific research for peer review. A transparent process allows the reader to assess the consistency of the process (i.e., that the process was not erratic or irregular). We could also look for stable membership on the research team and common findings attributable to at least two members of the research team who interpreted the data independently. It is unlikely, however, that this level of detail would be reported in the literature.

Objectivity in the research process is closely related to internal validity. The researcher's personal opinions or needs ought to be kept distinct from the community's preferences. The availability of research records, methods of interpretation and results of findings for an examination of possible external influences or biases (e.g., "were the findings supported by the data?") can help to establish objectivity.²

Another indicator that can be used to inspect objectivity is a statement from the author indicating the research's funding source or other potential sources of conflict. Thus, establishing objectivity requires the reader to assess the research process for possible biases. To accomplish this the framework asks whether there are opportunities to conduct this assessment.

Details about how individual data collection methods, such as surveys or interviews, can be conducted in a rigorous fashion (i.e., that the four principles are upheld) are well established in the literature (e.g., 12,13). The framework in this paper was developed to assess the process of soliciting public preferences, which in some cases entails a combination of data collection methods. In these situations the framework was applied to *the combination* of methods since this collection comprised the "process" which yielded the desired information.

² This "scrutiny" indicator was previously enlisted to determine reliability. The way in which the indicator is used to determine objectivity differs from the way it is used to determine reliability, and we can imagine cases which satisfy one but not the other principle. For example, research documents from a public preferences exercise may reveal that a questionnaire was shortened upon its second administration. This may indicate an inconsistency in the process (reliability) but does not suggest a bias influencing the process.

RESULTS & ANALYSIS

The framework was applied in two ways. First, the framework was used to examine methodological quality *within* each public preference case. The cases are presented by thematic cluster -- those using single methods, those using multiple methods for a single preference source and those using multiple methods to solicit multiple preferences -- followed by tables (Tables 3a-c) detailing the analysis. The framework was then used to examine general aspects of validity, generalizability, reliability and objectivity *across* cases. These findings are described in the discussion section. Table 4 presents an overview of the methods that were used to determine preferences in each case.

Single Methods

There were four jurisdictions which employed a single data collection method to solicit preferences from the public: England (national level) and District Health Authorities within England -- Pimlico, Bath and Solihull.

In England, a face-to-face interview survey was used to obtain a national perspective on health care service priorities. Respondents were asked to rank health services and treatments for specific groups of people (e.g., the elderly), posed attitudinal questions, asked who should participate in priority setting and questioned around budget allocation. A random sample of 2,005 people (originally selected for the Office of Population Census and Surveys) resulted in a response rate of 75%. Sampling was based on postcode address files, stratified by region, housing tenure and socioeconomic group. The sample population was generally representative of the population in terms of age and sex, but there was an under representation of those with a longstanding illness. The data set is held on archive at the University of Essex (14).

In Pimlico, a semi-structured interview format was used in 1994 to understand how, by exploring their underlying assumptions, people approach the problem of ranking priorities. The sampling frame was an inner city general practice which served a range of ethnic groups; approximately 25% of the practice is poor. There is an annual 30% patient turnover rate in the practice. In qualitative fashion, sampling was purposive to ensure a heterogenous participant group. Of the 20 people selected, 16 people attended the interviews and were diverse in social class, ethnicity and age (20-73 years), reflecting a range rather than representation of the practice. Pre-interview information was distributed with the initial invitation and described the objectives of the exercise and provided participants with a ranked list of nine health care services. The analysis employed a technique which allowed the patient's themes to emerge rather

than the researcher's. Participants began by using their own priorities as a reference point but gradually found it easier to use the community's needs to justify ranking services (15).

Postal surveys were used in Bath and Solihull. The Bath District Health Authority explored their residents' priorities, knowledge and views on the public's role in decision-making around health care priorities. About 1,500 residents were identified from electoral registries using a systematic procedure (further details are not provided). A second mailing was not sent on the grounds that anonymity was to be maintained, but pre-paid reply envelopes were included. After adjustments for residents who had moved, a 49.2% response rate was achieved and the authors concede that response and sampling bias may be present. Representativeness was reported on the basis of age and sex (16).

The Solihull Health Authority was interested in the public's values around health services and states. In 1994, members of the public were asked to value 16 different health states through the use of a visual analogue scale. Participants were also asked to rank services, which were accompanied with relative cost information. A broad range of 11 services were posed under 2 different scenarios (if you had more money, if you had less money). This section of the questionnaire was found to be internally consistent (17). The sampling frame was the Family Health Services Authority register, and a random sample of 630 adults, stratified by neighbourhood, age and sex, received the postal questionnaire. Three reminders were sent, after which non-respondents were contacted by phone or in person. After follow-up the response rate was 68%. The respondents were representative of the population (18).

Multiple Methods for Single Preferences

The cases in this cluster are characterized as primarily tapping into one preference source using a group discussion method with the addition of individual questionnaires to confirm the collective consensus.

The Somerset District Health Authority in England has implemented a process whereby eight health panels are held three times a year to determine the community's values around specific health issues that the Authority is facing at the time (e.g., should the Authority pay for treatment outside of the area?). About 12 people are invited to attend each panel, and recruitment is conducted by a professional using quota sampling "to ensure a spread of characteristics" (p. 1155). Efforts are made to encourage participation when inviting potential sample members. To cover costs, each participant receives £10. Participants are asked to serve for a year, and a rolling membership ensures four new and eight old members at every session. Based on the five

meetings which have occurred to date, the participants have been representative of the community in terms of sex, age, housing status, occupation, having a child under 15 years of age, car ownership and a hospital visit in the last year. Those working in health care have been slightly over represented (18).

Prior to the meeting, participants were given a paragraph briefly describing the issue at hand. At the meeting more information was provided about the nature of the problem, the number of people affected, the costs and other aspects of the issue. Participants were asked to take the community's preferences into account in their decision-making. A professional moderator conducted the sessions, which were tape recorded and qualitatively analysed. The qualitative analysis revealed that a breadth of arguments had been explored, that consistency in issues and decisions across focus groups had been achieved and that a community's perspective was used in discussions. In addition, each participant was asked to score priorities via an anonymous survey at the end of the focus group. A report is available to the public (18).

The Institute for Public Policy Research and the Cambridge & Huntingdon Health Authority in England organized a pilot citizens' jury to gather information about how priorities should be set, the criteria that ought to be used, and to what extent the public should be involved. A professional recruiter used a stratified random sampling strategy to invite participants who represented the socioeconomic characteristics of their community (characteristics were identified). No information is provided about refusal rates. For 4 days, 16 jurors listened to "experts" present research information using clinical examples to illuminate priority setting challenges. The jurors were able to clarify points presented by the expert witnesses, and they held further discussions among themselves. The process was tape recorded for qualitative analysis and quantitative information was derived from surveys that jurors filled out before and after the proceedings. Jurors were reimbursed £250 for the four days, and attendance was almost 100%. A professional moderator ensured that the agenda and points of interest were followed and that all members of the panel had an opportunity to express themselves (19).

Jurors were asked to evaluate the process. They indicated that future jurors should receive more education prior to the proceedings. There was also some concern about how to handle "biases" that experts may inadvertently present, either in their material or during the questioning period. A report is being submitted to the Health Authority (19).

In the Netherlands, a qualitative panel was used to explore the arguments underlying priorities in health care, and five panels with about nine people each were purposively chosen

from five parties concerned with health policy: patients (representing national patient groups), the public (university students and civil servants not working in health), general practitioners, specialists and health insurers. Overall, the panel members were well educated and the authors note that purposive sampling is not intended to obtain a representative sample (20).

The panels were asked to play the role of a parliamentary committee charged with translating a 1/3 reduction in the health care budget among 10 representative services. Researchers looked at the criteria they used to ration services and the underlying reasons for the criteria. Participants were given a variety of background information before coming to the panel. They were provided with selection criteria outlined in the government's Dunning Report, along with a precis about each of the 10 services (e.g., number of people affected, cost). They also received literature written by those supporting and opposing the public funding of health care (20).

Each panel took five hours. One panel member led the discussion and a facilitator was present to guide the discussions, which were taped and transcribed. At the end of the panel, members were asked to individually complete a questionnaire to confirm the unanimity of the group decisions (20).

Multiple Methods for Multiple Preferences

The remaining cases share the feature of soliciting multiple preferences using multiple data collection methods.

The City and Hackney District Health Authority in England wanted to determine the values (around health services) held by their general population and held by health professionals. They used a common questionnaire asking respondents to rank 16 services from the perspective of the needs of the local people (6,21).

The questionnaire was piloted extensively with various community groups. The final questionnaire was given to 359 people and administered either at routine meetings held by each of the community groups or at locations heavily populated by group members. The questionnaire was filled out by 350 people, accounting for 27 different groups (e.g., elderly or Jewish association). Afterward, a semi-structured discussion was held with each group. The respondents were comparable to the population in terms of sex, age and race/ethnicity but were over represented in terms of those who had had additional training after leaving school (6,21).

A random sample of the public was the second source of information. The sampling frame was drawn from 454 people registered with general practitioners in the area, and selected participants received the questionnaire by mail. Reminder post cards were sent four times. Non responders were followed up with personal interviews, which increased the response rate by 11% to a total of 78%. The sample mirrored the local population in terms of sex, ethnicity/race, further education or training, and proportion with a longstanding illness. Those over 30 years of age were over represented in the sample (6,21).

The third source of information was general practitioners, consultants and public health doctors. After four mailings the general practitioner and consultants' response rate was approximately 67% (86% of public health doctors responded) (6,21).

The remaining two cases describe national level efforts. Public consultation about New Zealand's health care reforms was extensive, comprised of public meetings, public forums, consensus conferences and publication of documents for public comment. The first such document was the New Zealand Department of Health's 1991 "Your Health and the Public Health", commonly known as the Green and White paper. The resulting 1,586 submissions to this document were attributed to 821 professional organizations, volunteer organizations and local health groups (22).

The National Advisory Committee on Core Health and Disability Support Services (Core Service Committee) was formed to provide guidelines for the four Regional Health Authorities' provision of services. The Core Service Committee held 19 public meetings to establish priority areas, and as a result of the findings, decided that "core" services would be based on historical patterns but explicit guidelines would also be developed to reflect the clinical contexts for appropriate treatments (23).

To develop these treatment protocols or boundary guidelines, the Core Service Committee began holding consensus conferences with consumers and health professionals in 1992. Consumers were composed of patients and those representing advocacy groups. The conferences focused on high cost, high volume services, such as major joint replacement and hormone replacement therapy, which fell under the previously established priority areas (22,24).

Particular access issues were identified in these conferences and put to the public by the Core Service Committee in the 1993 document "Seeking Consensus" (22).

There was also a series of seven workshops, conducted over a two month period in 1993, to explore ethical and justice issues in health care rationing. Only a range of community groups -- those not normally heard from (the elderly, people with disabilities, rural residents, low income urban residents, Maori and those from the Pacific Island, and high school students) - - were invited to the workshops as “individuals with particular perspectives”. About 25-50 participants attended each workshop, each lasting about five hours in length. The format was composed of discussions, game playing and small group tasks (24).

As the Core Services Committee continued their task of trying to define core services, they developed an evaluative framework which was presented to the public in the 1993/94 document ‘The Best of Health 2: How We Decide on the Health and Disability Support Services We Value Most’ (22).

Finally, a panel discussion was organized in 1994, and randomly selected members of the public and health professionals were invited to determine which social factors should be taken into account when selecting candidates for cataract surgery, coronary artery bypass graft and angioplasty (22).

A similarly complex exercise was carried out in the United States. The Oregon Health Services Commission (OHSC) was formed to devise a means of expanding health care coverage to a broader population than currently served. They envisioned a list of services ranked by priority which could be used to determine a package of core health services based on budgetary restraints. To accomplish this, the OHSC tried to use a cost-utility approach using a quality of well-being measure. The cost and outcomes for condition/treatment pairs were derived from the literature and administrative records. As well, 54 panels of health care providers were assembled to determine their estimation of probability of duration of illness and effectiveness of treatment (25-27).

In 1990, 1001 state residents from the general population were telephone surveyed to determine values for symptoms and levels of functional impairment associated with illness. Respondents were asked to rate 31 health states from 0 to 100, and this information was used in combination with outcomes in the cost-utility algorithms. The completion rate was 23.3%, the refusal rate was 23.3% and 53.4% of non-responders were attributed to things such as answering machines, telephone difficulties, language difficulties, no answer, etc. The majority of respondents were white and had a household income above the poverty line (28).

Due to unreliable cost and outcomes data, this method was discarded and never submitted for federal approval. In 1991 a second attempt began with the development of disease categories within which the condition/treatment pairs from the first attempt would be slotted. The idea was to first rank the categories and then rank the pairs within the categories. The OHSC required some measure of social values to proceed with ranking the categories (28).

To obtain these measures, 47 community meetings and public hearings, comprising 1,048 participants, were held across the state. An introductory presentation was followed by small group discussions, the results of which were brought to the larger group. To stimulate discussions, participants were asked to prioritize nine types of health care services. Participants were then asked “why is this health care service important to us” and asked to consider the “common good”. The values that were expressed were ranked according to frequency of response. Despite the door-to-door canvassing method employed by grassroots organization members to encourage attendance, the majority of participants were white, college educated and financially comfortable, with about 50% working in the health field. Only 9.4% of them were uninsured (compared to 16% of the Oregon population) (26,28-30).

The OHSC used these resulting social values to rank the 13 categories, and then they ranked the condition/treatment pairs within each category using a net benefit approach which emphasized duration of benefit and quality of well-being (based on the results of the 1990 telephone survey). The final list was adjusted by the OHSC’s subjective judgement. The Bush government stated that the list violated the Americans with Disabilities Act and the Commission was directed to omit any reference to quality of well-being (26,28-30).

DISCUSSION

A conceptual framework designed to assess the methodological rigour of soliciting community preferences was applied to a variety of processes. These processes differed in the types of preferences that were being sought; jurisdictions were interested in preferences for health states, for medical treatments and for decision-making roles. This application has uncovered some issues common to all processes and preferences being sought.

Before discussing these issues, we note that the analysis may be perceived as limited in that aspects of ensuring internal validity, generalizability, reliability and objectivity might have actually been carried out during the research process but were not *reported* in the literature. Nonetheless, the literature serves as the public record of the event and as such remains a fair primary source of information for outside investigators.

Internal Validity: In some cases, informed preferences might have been accomplished through the distribution of educational material. But Fishkin (31) stresses that informed preferences require both education and deliberation among participants in order for the participants to be able to reflect on the arguments at hand. A direct application of the deliberative approach demonstrated that participants were able to provide in-depth comments on a complex issue (32), and a recent study provides further evidence that people’s views change after given the opportunity to debate the topic at hand (33). Hence, the distribution of educational materials without a discussion does not ensure that participants understood the issue. All the cases which utilized multiple methods to obtain preferences from one source incorporated both an educational and deliberative component in their process. Some aspects of the New Zealand and Oregon experiences also shared this feature.

The clusters demonstrated a strong effort to use multiple methods at two different levels: within one source of preference (e.g., group preferences) and across sources of preferences (e.g., group and individual preferences). In terms of clarifying ambiguous findings, those processes which included a discussion format with a moderator were more likely to have opportunities to question participants about their remarks. None of the “closed” formats, such as surveys, followed-up on ambiguous results with the use of focus groups or other methods.

The internal validity indicator requiring further attention by researchers is consistency or a match between the particular definition of the public and the selection of appropriately corresponding data collection and analytical methods. In some cases, when a group-level preference

was sought, participants were able to interact in order to come to a aggregate-level decision, and when an individual-level preference was sought the individual became the sole source of information. But in many cases it was not clearly stated at the start whose preference was of interest, and in these cases the definition of the public fell, by default, to the way in which the data was collected. For example, when public government documents are released with requests for the public to respond, highly-organized interest groups with available resources are more likely to submit an opinion than will members of the general public (34). This weakness was especially prevalent among cases using multiple methods to seek multiple preferences.

In all cases, details about methods used to analyse the data were not provided. This was particularly apparent in those cases where both individual and group-level preferences were solicited in the same forum, leaving one to wonder whether one type of preference was weighted more heavily in the analysis.

None of the cases reported any type of interpretation check of findings with participants before a formal report was prepared. Improvements could be made on this front.

According to the framework, focus groups and panels demonstrated a high level of internal validity, followed by semi-structured interviews, jury format and the multiple method, evolving process in New Zealand.

Generalizability: The issue of generalizability was closely linked to the way in which the public was defined and the data collected. If an individual-level preference was of interest, then the cases demonstrated that samples were selected on the basis of sociodemographic characteristics to represent the larger community. Marmor and Morone call this descriptive representation, and note that one of the challenges is identifying characteristics which are most key to the purposes of the exercise (an issue not addressed by the cases reviewed). For example, how does a researcher decide whether health preferences identified by those representative of neighbourhoods are more or less informative than those identified by representatives of racial/ethnic groups? In situations where small samples sizes are called for, Marmor and Morone point out the practical difficulty in trying to select a limited number of people to reflect “everyone” in the community. They also call attention to the fallacy of assuming that just because people have the same demographic features they necessarily hold the same preferences (34).

Group-level preferences, conveyed by a member of the collective (substantive representation), also bows to the logical inconsistency of “shared interest or affiliation therefore shared preferences”. Other difficulties include knowing which groups to invite (another issue not addressed in the cases), and balancing the preferences of the group with the greater good of the jurisdiction. The advantage of this type of representation is that, at least where interest groups are involved, representatives are held accountable for their actions by their group (34). Those cases focussing on collective-level preferences usually used a qualitative method of data collection and consequently relied on purposive sampling to obtain a variety or extreme sample of preferences without specifically considering which groups ought to be invited because of particular health care preferences. An exception was the Netherlands, where panels representing “distinct parties in health policy” were selected.

Related to the issue of representation is the action of *explicitly telling* the participants the perspective expected of them. Often this was obvious from the way in which the questions were posed, but in one interesting case (Pimlico), participants were individually selected (with a corresponding data collection method) but not asked to represent any particular point of view. During the interview process respondents shied away from answering as individuals and preferred to take the perspective of the community at large (although there was no forum to establish what the collective might think or want).

Overall, there was an attempt to select participants on the basis of a few sociodemographic characteristics (mostly age, sex and neighbourhood) or to select purposively. Single data collection methods and those cases employing multiple methods for a single preference demonstrated better generalizability; cases of multiple methods for multiple preferences were ambiguous with respect to who was to be represented. More attention is needed around the definition of the public and implications for representativeness and generalizability of findings.

Reliability: In assessing reliability we ask if the same results would be found if the process was repeated, or, if the research logs were scrutinized, whether we would come to similar conclusions. Some cases incorporated replication into their design (e.g., repeated focus groups or town hall meetings), allowing an analysis to be conducted across as well as within sessions. In those cases using quantitative surveys, statistical indicators describing the reliability of the instrument could be reported (but never were). Only in two cases (England national survey and Pimlico) are research logs available for inspection of consistencies. In most cases, only a final report was available. Even at that, in many instances it was not clear whether the report was ever disseminated to the public or whether only the health authority was the recipient.

Oregon was a case in the extreme. So much has been written about the process that details have been inconsistently reported.

One of the underlying assumptions of these exercises is that preferences can be measured in the first place. At issue is the stability of these preferences. McIver (11) summarizes research on public values as they relate to lay views and beliefs. The author describes how framing effects (e.g., changing the wording of a question), contexts (e.g., asking about one's own experiences versus hypothetical experiences) or additional knowledge can change lay beliefs.

When speaking about the stability of preferences for health outcomes, health economists use different paradigms to conceptualize the process. Some ascribe to the notion that individuals hold consistent preferences over health outcomes, and that the process of tapping into them actually helps to shape the preference. Consequently, different methods may result in different results. Other economists feel that individuals have strong opinions for a few items about which they are well informed. In this situation the action of soliciting preferences may form a preference which was not there initially. The third paradigm describes preferences as present but not fully expressed, and the process of eliciting and debate helps the individual articulate his or her preference (35).

The stability of preferences needs to be better understood if they are going to be used to inform health care rationing decisions. At the very least, the preceding discussion underlines the importance of using multiple methods within preferences to corroborate findings.

Overall, the framework revealed that moderate levels of reliability were demonstrated by those complex cases of New Zealand and Oregon, where public meetings were repeated and reports of results were available to the public. Similarly, the multiple focus group design in Somerset allowed for consistency checks across groups, and reports were available to the public.

Objectivity: As noted above, more effort is required to ensure that the research protocol, data, analysis and results are made available to the local community and to outside investigators. In terms of communicating potential researcher biases, in half of the cases the author(s) declared their funding source. A few authors stated their conflict of interest as “none” (probably reflecting a journal's editorial policy for declaration). While objectivity *per se* is determined by the reader, opportunities to carry this out were provided by the England (national survey), Pimlico and Somerset cases.

The Framework: Strengths and Limitations

Although an evaluation of the framework was not an explicit objective of this paper, the analysis has stimulated some preliminary thoughts about its strengths and limitations. In terms of limitations, the framework treats the solicitation of different types of preferences (e.g., health states or treatments) as common endeavours; whether this is appropriate or whether it is in effect comparing “apples and oranges” needs further investigation. Additionally, the framework treats each indicator as equally important; other researchers may wish to weight particular indicators as more vital than others.

The framework is novel, which in itself presents both strengths and weaknesses. Further work is required to refine the indicators and to determine the framework’s utility and appropriateness given the larger political milieu. For example, methodological soundness implies “better quality information” (however this is defined), but the current analysis makes no attempt to determine whether sounder processes did in fact lead to “better information”.

In that the framework is novel it contributes to the literature in this area. Of importance is that some of the indicators, which were developed from a methodological perspective, concur with recommendations for better processes based on a democratic ideals (e.g., representativeness). Our perspective is somewhat more neutral than others’ in that we deliberately leave the definition of the public to the jurisdiction involved but recognize that the definition has implications for other aspects of the process.

The framework was able to differentiate between the cases presented and provides guidance for researchers involved in this process.

CONCLUDING REMARKS

We refrain from recommending a “winner” based on a simple tally of methodological pluses and minuses. More importantly, the analysis indicated that each individual method had particular strengths but could be further improved with the incorporation of methodological details outlined in the framework. For example, postal surveys could be supplemented with educational and deliberative efforts and follow-up strategies with a sub-sample of the respondents. Or, focus groups could be improved by the addition of interpretation checks.

For those embarking on a public preferences exercise, this analysis points to key issues to consider. First, in order to plan an internally valid and generalizable exercise, our analysis suggests a need for a clear, *a priori* definition of the public in terms of their composition (stakeholders? citizens? both?) and whether individual or collective level preferences are desired. This definition of the public will in turn influence the data collection method (interaction among participants?), sampling (which larger group do we wish to infer the results to, or who are the relevant groups?) and analysis (will responses be weighted equally?). The inclusion of multiple data collection methods may enhance internal validity and reliability by providing some safeguard against the possibility of instable preferences. The availability of research logs provides an opportunity for outsiders to assess the reliability and objectivity of the exercise.

In terms of feasibility, there are possible trade-offs inherent in adopting the current framework for future processes. Adherence to the framework’s principles may require more time and financial resources than originally planned. As well, the focus on methodology may distract from other questions around involving the public, such as whether it is appropriate to include them at all or at what stage in the policy process ought they be included. It is not our intent to push these important questions into the background. But the use of this framework can lend to the credibility of the entire exercise, and may forge favourable links between politicians and their communities.

While the indicators of internal validity, generalizability, reliability and objectivity have not been previously applied to this context, they were useful in differentiating among the cases presented and allowed some general evaluative comments to be made. Other aspects of the process also deserve attention. These may include questions around whether moral or ethical standards were upheld (was the process fair?) (36). Alternatively, one could compare the costs or cost-effectiveness of various processes (did we get our money’s worth?). An extended evaluation could incorporate these elements into the framework used in the analysis. The framework described in this paper represents an initial attempt to assess the methodological aspects of the processes that have been used to solicit public preferences for health care rationing decisions.

REFERENCES

1. Eyles J. The role of the citizen in health-care decision-making. Centre for Health Economics and Policy Analysis Working Paper C93-1, 1993.
2. Renn O, Webler T, Rakel H, Dienel P, Johnson B. Public participation in decision making: A three-step procedure. *Policy Sciences* 1993;26:189-214.
3. Coast J, Donovan J. Conflict, complexity and confusion: The context for priority setting. In: J. Coast, J. Donovan and S. Frankel, eds. *Priority setting: The health care debate*. Chichester: John Wiley; 1996. p. 3-34.
4. Kathlene L, Martin JA. Enhancing citizen participation: Panel designs, perspectives, and policy formation. *Journal of Policy Analysis and Management* 1991;10(1):46-63.
5. Lomas J. Reluctant rationers: Public input to health care priorities. *Journal of Health Services Research and Policy* 1997;2(2):103-111, 1997.
6. Bowling A, Jacobson B, Southgate L. Explorations in consultations of the public and health professionals on priority setting in an inner London health district. *Soc. Sci. Med.* 1993;37(7):851-857.
7. Donovan J, Coast J. Public participation in priority setting: commitment or illusion? In: J. Coast, J. Donovan and S. Frankel, eds. *Priority setting: The health care debate*. Chichester: John Wiley; 1996. p. 203-234.
8. Luba Y, Guba EG. *Naturalistic inquiry*. Newbury Park (CA): Sage; 1985.
9. Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia (PA): Lippincott-Raven Publishers; 1998.
10. McIver S. Information for public choice. *British Medical Bulletin* 1995;51(4):900-913.
11. Shiell A, Hawe P. Health promotion community development and the tyranny of individualism. *Health Economics* 1996;5:241-247.
12. Bowling A. *Research methods in health: Investigating health and health services*. Philadelphia (PA): Open University Press; 1997.
13. Cresswell JW. *Research design: Qualitative and quantitative approaches*. Thousand Oaks (CA): Sage; 1994.
14. Bowling A. Health care rationing: The public's debate. *BMJ* 1996;312:670-4.

15. Dicker A, Armstrong D. Patients' views of priority setting in health care: An interview survey in one practice. *BMJ* 1995;311:1137-9.
16. Richardson A, Charny M, Hanmer-Lloyd S. Public opinion and purchasing. *BMJ* 1992;304:680-2.
17. Szczepura A, Sidhu K, Cobb B, Cooper R, Geller R. Setting priorities in UK health care and the new NHS R&D strategy. In: M Malek, ed. *Setting priorities in health care*. Chichester: John Wiley; 1994. p. 243-261.
18. Bowie C, Richardson A, Sykes W. Consulting the public about health service priorities. *BMJ* 1995;311:1155-8.
19. Lenaghan J, New B, Mitchell E. Setting priorities: Is there a role for citizens' juries? *BMJ* 1996;312:1591-3.
20. Stronks K, Strijbis A, Wendte JF, Gunning-Schepers LJG. Who should decide? Qualitative analysis of panel data from public, patients, healthcare professionals, and insurers on priorities in health care. *BMJ* 1997;315:92-96.
21. Jacobson B, Bowling A. Involving the public: Practical and ethical issues. *British Medical Bulletin* 1995;51(4):869-875.
22. Coast J. Core services: Pluralistic bargaining in New Zealand. In: J. Coast, J. Donovan and S. Frankel, eds. *Priority setting: The health care debate*. Chichester: John Wiley; 1996. p. 65-82.
23. Cooper MH. Core services and the New Zealand health reforms. *British Medical Bulletin* 1995;51(4):799-807.
24. Campbell AV. Defining core health services: The New Zealand experience. *Bioethics* 1995;9(3/4):252-258.
25. Honigsbaum F, Calltorp J, Ham C, Holmstrom S. *Priority setting processes for healthcare*. Oxford: Radcliffe Medical Press; 1995.
26. Tengs T. An evaluation of Oregon's Medicaid rationing algorithms. *Health Economics* 1996;5:171-181.
27. Bowling A. Setting priorities in health: the Oregon experiment (part 2). *Nursing Standard* 1992;6(38):28-30.
28. Coast J. The Oregon plan: Technical priority setting in the USA. In: J. Coast, J. Donovan and S. Frankel, eds. *Priority setting: The health care debate*. Chichester: John Wiley; 1996. p. 37-63.

29. Daniels N. Is the Oregon rationing plan fair. *JAMA* 1991;265(17):2232-2235.
30. Bowling A. Setting priorities in health: the Oregon experiment (part 1). *Nursing Standard* 1992;6(37):29-32.
31. Fishkin JS. *Democracy and deliberation*. New Haven: Yale University Press; 1991.
32. Abelson J, Lomas J, Eyles J, Birch , Veenstra G. Does the community want devolved authority? Results of deliberative polling in Ontario. *CMAJ* 1995;153(4):403-412.
33. Dolan P, Cookson R, Ferguson B. Effect of discussion and deliberation on the public's views of priority setting in health care: focus group study. *BMJ* 1999;318:916-919.
34. Marmor TR, Morone JA. Representing consumer interests: Imbalanced markets, health planning, and the HSAs. *Milbank Memorial Fund Quarterly* 1980;58(1):126-65.
35. Dolan P. The nature of individual preferences: A prologue to Johannesson, Jonsson and Karlsson. *Health Economics* 1997;6:91-93.
36. Goold SD. Allocating health care: Cost-utility analysis, informed democratic decision making, or the veil of ignorance? *Journal of Health Politics, Policy and Law* 1996;21(1):69-98.

TABLE 1
PRINCIPLES OF RIGOUR IN A PUBLIC SOLICITATION PROCESS

TABLE 2
PREFERENCES AND THE COMMUNITY

TABLE 3a
ANALYSIS OF PARTICIPATION PROCESSES -- SINGLE METHODS

TABLE 3b
ANALYSIS OF PARTICIPATION PROCESSES
-- MULTIPLE METHODS FOR SINGLE PREFERENCES

TABLE 3c
ANALYSIS OF PARTICIPATION PROCESSES
-- MULTIPLE METHODS FOR MULTIPLE PREFERENCES

Table 4
METHODS THAT WERE USED IN SOLICITING PUBLIC PREFERENCES

NOT FOR CITATION WITHOUT PERMISSION

Paper 99-07

**SOLICITING PUBLIC PREFERENCES
FOR HEALTH CARE PRIORITIES:
A CRITICAL REVIEW OF METHODOLOGIES**

Anita Kothari

