Choosing health technology assessment and systematic review topics: The development of priority-setting criteria for patients’ and consumers’ interests

Hilda Bastian
National Institutes of Health

Fülöp Scheibler, Marco Knelangen, Beate Zschorlich
Institute for Quality and Efficiency in Health Care, Germany

Mona Nasser
Universities of Exeter and Plymouth

Andreas Waltering
Institute for Quality and Efficiency in Health Care, Germany

Background: The Institute for Quality and Efficiency in Health Care (IQWiG) was established in 2003 by the German parliament. Its legislative responsibilities are health technology assessment, mostly to support policy making and reimbursement decisions. It also has a mandate to serve patients’ interests directly, by assessing and communicating evidence for the general public.

Objectives: To develop a priority-setting framework based on the interests of patients and the general public.

Methods: A theoretical framework for priority setting from a patient/consumer perspective was developed. The process of development began with a poll to determine level of lay and health professional interest in the conclusions of 124 systematic reviews (194 responses). Data sources to identify patients’ and consumers’ information needs and interests were identified.

Results: IQWiG’s theoretical framework encompasses criteria for quality of evidence and interest, as well as being explicit about editorial considerations, including potential for harm. Dimensions of “patient interest” were identified, such as patients’ concerns, information seeking, and use. Rather than being a single item capable of measurement by
One year’s worth of Cochrane reviews (352 new reviews beginning with issue 3 of 2004) were used to gauge the key steps in priority setting for HTA, constitutes something of a “black box,” despite its major influence on topic selection (14). Yet effective and transparent priority setting is essential for the use of IQWiG’s scarce resources. We estimate that we can only aim to keep at the most 1,000 topics evidence-based and up-to-date in the long term. Evolving objective and transparent methodologies that are genuinely stringent and severely limit eligibility has, therefore, been a focus of the developmental stage of the Institute (5).

METHODS
Overview
IQWiG’s process for selecting topics for its patient/consumer health information was developed in iterative steps, combining data gathering and analyses with reflective editorial practice. An opinion poll was conducted in 2006 on level of interest in statements drawn on conclusions of Cochrane reviews. The methods for polling are described in detail below. One central editor (H.B.) made final editorial decisions throughout 2006, evolving an explicit categorization of the multiple criteria routinely influencing decisions to reject or proceed with a topic. A theoretical framework and potential data sources for the topic selection process were developed using the poll results and the principles of multi-criteria decision analysis priority setting (2). With each decision, the framework and list of criteria was re-appraised and modified, until decisions no longer led to further modifications.

From this framework, specific criteria and data sources practical for routine daily use in scanning new and updated systematic reviews and HTAs and assessing their eligibility were chosen. Selection processes were then operationalized (through 2008 and 2009). Finally, in 2010, methods for monitoring the apparent effectiveness of the topic selection process were developed and incorporated into the procedures for information updating. Plans were developed for adjusting the overall portfolio of information in the light of consumer need and reader response.

Polling Methods
One year’s worth of Cochrane reviews (352 new reviews beginning with issue 3 of 2004) were used to gauge the...
interestingness of statements based on review conclusions. These Cochrane reviews were filtered by two people (H.B., B.Z.), according to eligibility criteria which had been determined by IQWiG. These eligibility criteria were largely to be consistent with the standards applied for HTAs within the Institute: (i) The authors of the systematic review determined at least one randomized controlled trial to be adequate quality, with data on at least one patient-relevant outcome; (ii) The review/HTA was not sponsored by a manufacturer of a reviewed product; (iii) A search for studies was conducted within the past 5 years; and (iv) Enough evidence was found to support a conclusive statement.

A total of 124 reviews were determined to be eligible (35 percent). To determine whether rigid application of the final criterion might result in excluding many highly interesting reviews, we also included all noneligible reviews from one issue of the Cochrane Library (issue 1 of 2005, adding a further forty-five reviews to the sample). The subset from issue 1 of 2005 thus included seventy-four eligible and noneligible reviews in total.

Results of these reviews were summarized in one to two sentences, in both German and English. Depending on the evidence in the review, these summative statements might be conclusive or inconclusive, for example:

- Conclusive statement: “People who reduced their salt intake by an average of 4.5 g per day for an average duration of 4 to 6 weeks, experienced a reduction in blood pressure”;
- Inconclusive statement: “Radiotherapy for age-related macular degeneration (vision loss) might be able to improve outcomes, but research is inconsistent and the treatment might damage eyes.”

Two Web-based surveys were conducted, giving respondents a 5-point scale (from “not interesting” to “extremely interesting,” with the center point being neutral). The following groups were invited to participate:

- All employees of IQWiG (including nonscientific staff);
- Members of the Federal Joint Committee’s patient information committee (including several patient representatives) and representatives of the Federal Ministry of Health;
- The national “authorized” organizations representing patients and consumers in Germany (authorized by the Federal Ministry of Health in consultation with national patient and consumer organizations); and
- Members of an international evidence-based medicine mailing list and the Cochrane Collaboration’s Consumer Network email list.

The mailouts resulted in 194 completed surveys, with almost half of the respondents choosing the description “health professional” (45 percent). A total of 95 completed surveys came from Germany (49 percent) and 60 percent of the respondents were female.

**Table 1. Theoretical Framework for Priority-Setting Criteria**

<table>
<thead>
<tr>
<th>Healthcare evidence</th>
<th>Editorial considerations</th>
<th>Patients’/consumers’ interests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systematic reviews of the effects of health and self care</td>
<td>Balanced mixture of topics</td>
<td>Patients’/consumers’ expressed concerns</td>
</tr>
<tr>
<td>Evidence of impact of information as an intervention in topic area</td>
<td>Currently topical issues</td>
<td>Information sought by patients/consumers</td>
</tr>
<tr>
<td>Potential adverse effects of publishing information</td>
<td>Topics that attract readers’ interests</td>
<td></td>
</tr>
<tr>
<td>Priorities of commissioning bodies (payers)</td>
<td>Unmet information needs</td>
<td></td>
</tr>
<tr>
<td>Workload and other resource issues</td>
<td>Burden of disease</td>
<td></td>
</tr>
<tr>
<td>What experts believe patients/consumers need or want to know</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean interest scores (with 95 percent confidence intervals) were calculated for each review summary statement. If the lower end of the confidence interval (CI) was higher than the middle (neutral) category, the review was rated as “significantly interesting.” If the upper end of the CI was lower than the middle category, the review was rated as “significantly uninteresting.” All other reviews were rated as “not significant.” Two of the authors (H.B. and A.W.) coded all conditions in the summary statements as either “common” (1 percent lifetime prevalence or relatively high prevalence within a major population age group or gender) or “not common” for a post-hoc analysis.

**RESULTS**

**Poll Results**

The key poll results are shown in Supplementary Tables 1 and 2, which can be viewed online at www.journals.cambridge.org/thc2011026. The sample was not sufficient to identify any significant differences in ratings between the health professionals and non-health professionals. Although Germans rated the interestingness of the messages more highly in general, roughly the same reviews were regarded as interesting and uninteresting.

Very few statements were rated as significantly interesting (8 percent), while around half of all statements were significantly uninteresting (Supplementary Table 1). Based on the planned subset analysis, reviews without enough
evidence to draw a conclusion were less likely to be significantly interesting (12 percent versus 7 percent). However, the rigid application of the criteria would have resulted in two of seven significantly interesting reviews being excluded.

A post-hoc analysis was conducted because there appeared to be a discriminative criterion for “uninteresting” statements: the combination of “not enough evidence” and a condition that is not common. That analysis showed that of fifteen reviews in uncommon conditions where there was not enough evidence to support a conclusive statement, all but one were significantly uninteresting (93 percent) (Supplementary Table 2). All of the significantly interesting statements in the subset from Issue 1 of 2005 were for common conditions (100 percent), and all but two had enough evidence for a conclusive statement (71 percent). The poll confirmed the general value of the eligibility criterion “enough evidence was found to support a conclusive statement,” with exceptions in practice for common conditions or commonly used interventions.

### Theoretical Framework and Potential Data Sources

The framework for considering consumer-oriented priorities (Table 1) includes three major dimensions: (i) Healthcare evidence (both about interventions or technologies and the process of communication itself); (ii) Editorial considerations (including the potential for adverse effects); and (iii) Patients’/consumers’ interests.

The potential for adverse effects of publishing information requires further methodological and practical development, but is an explicit and important criterion for IQWiG. Our ultimate goal is to empower people, including providing support for them to set and achieve their own health goals. However, we also need to be concerned with adverse effects such as disempowering or debilitating people, causing despair, and worsening people’s health or well-being. Some messages could have unintended effects or potentially predictable adverse effects. For example, the potential adverse effects of communicating a beneficial effect of continuing to smoke (some weight control) had to be taken into consideration in the editorial decision on whether or not to communicate all or some of the results of a systematic review on weight gain and smoking cessation (16). This was more than a theoretical concern, as there was some suggestion that this information could contribute to decreased motivation to quit smoking (16). We published the information, trying to reduce the risk of harm while informing comprehensively by including particular emphasis on the value of quitting smoking. We do not know whether we were successful here, as we have yet to develop methods for determining the effects in our readers of what we do (5).

The process of developing the framework resulted in the construct of “interestingness” to/for patients/consumers becoming increasingly explicit and differentiated. Patients’ and consumers’ interests in the context of information are both what is interesting to them and what it might be in their interests to know. Therefore, the term interests here incorporates interesting to and importance for. Rather than being a single item capable of measurement or determination by a single method, the concept consists of multiple potentially measurable elements. Table 2 fleshes out aspects of these elements, by indicating potential ways that they could be assessed.
The processes ultimately chosen for assessing interesting to patients/consumers at IQWiG include literature searches for qualitative research addressing patients’ concerns and what affects their decision making. This is critical for our product development in areas where significant resources are invested or where we are particularly concerned about the potential effects of what we communicate or how we communicate. Published qualitative research can also be used to help inform our understanding of what topics should be addressed within a priority condition (such as stroke), for example.

Information seeking on the Internet (both within the search engine of our own Web site, and search query data reported by Google) as well as at the call centers operated by Germany’s statutory health insurers have emerged as vital sources for overall topic planning for the Institute. We plan to use a deeper analysis of these data to map out the major conditions which we must address to have a reasonably comprehensive information portfolio.

**IQWiG’s Routine Topic Selection Process**

Figure 1 provides an overview of the process we use for assessing a proposed topic: Figure 2 provides more detail of the assessment steps for relevance and evidence. This process could be triggered by one of two things: either beginning with a suggested topic, or a topic could be triggered by the publication of an eligible systematic review or HTA.

The Department of Health Information monitored the publication of new and updated systematic reviews and HTAs published in both English and German, primarily through the Cochrane Database of Systematic Reviews, the Database of Reviews of Effects, the INAHTA (International Network of Agencies for Health Technology Assessment) database of HTAs, and the McMaster Online Rating of Evidence system (10). The template for summarizing the topic assessment process is included as Supplementary Table 3, which can be viewed online at www.journals.cambridge.org/thc2011026. Routine sources for rapidly assessing priority included national routine data collections on reasons for doctor visits, hospitals and sickness-related work absence, data on enquiries to statutory health insurance call centers (supplied in confidence by major insurers), the McMaster ratings of clinical relevance (10), and national prescribing data (19). Prevalence and burden of disease information was obtained from the evidence sources, supplemented by data available in German clinical practice guidelines and the BMJ publication, Clinical Evidence.

**Monitoring Effectiveness of Topic Selection and Determining Whether to Disinvest in Topics**

Monitoring effectiveness of our methods for determining what is likely to be interesting is critical for the Web site’s impact. Any priority setting before production is essentially predictive and presumptive. Whether the resulting information actually proves to be interesting to the intended audience and whether it attracts readers requires evaluation, including monitoring behaviors that indicate interest. Assessing actual rather than predicted interest could contribute to improved topic selection methods. In addition, being able to determine actual interest would enable good decisions about whether to continue to invest in keeping a topic up-to-date, or to choose instead to disinvest and archive (or remove) a topic.

With an estimated 1,000 topics as the Institute’s goal for its information portfolio, errors in selection of topics are to
### Relevance assessment - topics

- Data on prevalence or incidence
- Clinical relevance by McMaster Online Rating of Evidence if available for one or more relevant systematic reviews
- Accessibility of the interventions evaluated

### Evidence assessment – systematic reviews and HTAs

- Addresses a question relevant to patient action, decision or critical knowledge
- Recent enough search for evidence (generally within 3 to 5 years)
- Assesses at least one patient-relevant outcome (mortality, morbidity or a quality of life outcome)
- Not sponsored by a manufacturer of the product evaluated in the review
- Adequately conducted systematic review (according to Oxman and Guyatt rating tool (15), addressing search strategy, assessment of the quality of primary studies and appropriate synthesis and interpretation of data)
- Data on gender and age of participants ideally included
- For a statement of effectiveness, at least one adequate quality trial

**Figure 2.** More detail of the assessment steps for relevance and evidence.

---

be viewed critically. Decisions that have been supported by excellent information and great expertise are not necessarily going to get it right. Informed expert opinion has been shown to be poorly predictive of which technologies will become important (8), for example. Medical journal editors’ decisions, too, on which items of research will interest journalists are far from universally right (3). Yet, despite the widespread use of priority setting in health, the issue of how to determine its effectiveness has received much less attention than mechanisms for choosing priorities. Each inadvertently unimportant/uninteresting topic selected by IQWiG not only has an initial opportunity cost: it continues to consume resources by means of the updating process. Correcting for topic selection errors, or changes in interest over time, could maximize the potential impact of the Web site.

### DISCUSSION

**Evaluating the Process and the Resulting Choices**

A major limitation of our developmental process was that our convenience polling did not involve a representative sample of our intended audience. Further more rigorous research to explore differences in actual rather than presumed interest of stakeholder groups appears warranted.

The appropriateness of our methods for topic selection, and the results of that process, were a key part of the evaluation conducted on our evidence assessment and information service by the World Health Organization in 2008 and 2009 (5). The WHO’s review team commended aspects of the
process, but encouraged further development and evaluation of the process as a priority.

Undertaking specific research with representative populations could provide us with a measure of the extent to which we have achieved our goal of providing information that is interesting to patients and consumers. Consultation with experts such as patient or consumer representatives was also suggested as a potential mechanism (5). Neither of these forms of evaluation has been done to date. We do, however, currently have several opportunities to at least roughly gauge the actual rather than predicted level of interest for consumers or patients in information we have produced. Before publication, whether or not the information is interesting is one of the aspects addressed by the independently conducted focus groups (of five participants) where every draft of our information is user-tested. This routine user-testing began in 2008 and has been described elsewhere (20).

After publication, the following items of routinely collected data may provide indications of actual interest: Number of visitors to the information (on our Web site or use by means of a syndicated provider); Requests for syndication or reproduction; Google position achieved for that information (collected routinely for selected subjects only); and Online user ratings of the information (an option to rate a piece of information on a 5-point scale at the end of the article, currently used by less than 1 percent of our readers).

None of these is in itself free of confounding or bias as a method of assessing actual interest in information. The number of visitors and Google position are affected by many factors, including the amount of competition for attention on that subject. Readers who rate an article online poorly are not necessarily specifically expressing a negative opinion on the topic itself: the fact of reading and rating already suggests some level of interest in the topic. However, their rating may reflect, at least in part, how interesting the message about the topic was. As the poll results suggested, the level of evidence can contribute to how interesting the message about a topic might be. For example, a conclusive statement that “Drug A causes heart attacks” may be interesting, while the statement “Not enough evidence of the effects of drug A” might be uninteresting.

Priority Setting for Disinvestment Decisions

While acknowledging the limitations of the data, we nevertheless used both visitor numbers and online ratings to help us determine whether or not to continue to invest in a published piece of information at the time that major updates were undertaken. In 2010, poor visitor numbers and ratings contributed significantly to the decision to archive several items of information. The Web site currently covers around 250 topics. As the Institute nears its ultimate limits for generating new information, these disinvestment decisions will play an increasingly important role. When no recently updated eligible systematic review of evidence for a topic exists, updating based on searching for and assessing primary studies becomes particularly resource-intensive and can exceed the resources required to simply create new information on another topic where up-to-date systematically reviewed evidence is available. Furthermore, updates are not a one-off activity: updates may need to be undertaken several more times in the following decade, amounting to a significant resource investment.

Thus, methods for priority setting for disinvestment will be at least as important as methods for priority setting for new topic selection. Some criteria are obvious and easy to apply (such as when an intervention is no longer available, or a disease such as avian influenza is no longer prevalent). However, relative value of one topic over another also involves more complex prioritization decisions. Developing an approach to considering whether or not information is having adverse effects other than opportunity costs is critical here.

Limitations of the Approach and Lessons for Others

The process we used to develop the criteria and other methods was very specific, being for a specific task (identifying evidence which would interest our intended users), within the specific constraints of IQWiG’s remit, general methods, and resources. As with many HTA agencies (9), the integrity of IQWiG’s scientific process and its independence required a transparent process that was as demonstrably objective as possible and not susceptible to special interest influence. The goal of consumer involvement in the specific form of advocate representation, therefore, presents particular challenges (9;17). It has been suggested that meaningful consumer involvement in research agenda setting requires well-networked consumers who are then provided with information, resources, and support and engaged in repeated facilitated debate (13). Although feasibility has been shown, effectiveness has not been demonstrated.

Even where processes for stakeholder involvement in priority setting exist, a considerable part of prioritization will always occur within a “black box” inside an agency, or involving competing subject areas where no identifiable practical proxy for affected communities is available. Interest for the public is a relevant priority-setting criterion for some HTA agencies (1;11). For the researchers and others involved in the process, a better understanding of what matters and is interesting to patients and the community would be valuable. Our work represents a first step down the path of increasing the ability of people working in HTA to make choices of interest to patients and the community, with or without the representation of consumers.

CONCLUSIONS

The best evidence to communicate to patients/consumers is right, relevant and likely to be considered interesting to and/or
important for the people affected. Methods for determining whether or not research findings are likely to be “right” are well-developed. Relevance for consumers and patients suggests primarily issues on which they may make decisions, or information which people need or would not know. That determination is relatively straightforward. However, determining what is interesting to patients and consumers themselves, rather than relying solely on opinions, is a methodological challenge.

Priority-setting criteria for national and local HTA activities often focus on areas imposing high costs on the healthcare system and where there is large uncertainty (2;12;14). This is unsurprising given that the context of HTA is often led by policymakers. Nevertheless, some agencies also are explicitly concerned with patients’ priorities. Our experience suggests that the interests of patients and other end users of information at the societal level may even be the polar opposite of what fundamentally concerns policy makers. At the community level, interest appears to focus more on knowledge that is at least somewhat certain and conditions which represent a common burden (whether or not they impose system cost burdens). Further methodological and evaluative work could clarify the characteristics of evidence and information that are interesting specifically to patients and consumers to increase their use of the information. Ultimately, methods able to determine the actual rather than predicted interest will be necessary to refine priority-setting processes in the future. Additional areas where methodological work would be helpful is exploring adverse effects of information use and methods for determining priorities for disinvestment in continued updating.

SUPPLEMENTARY MATERIAL
Supplementary Table 1
Supplementary Table 2
Supplementary Table 3
www.journals.cambridge.org/thc2011026

CONTACT INFORMATION
Hilda Bastian (hilda.bastian@nih.gov) National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, Maryland 20894

Fülöp Scheibler, PhD, Marco Knelangen, Diploma of Health Science (fuelop.scheibler@iqwig.de), Beate Zschorlich, Masters of Public Health (beate.zschorlich@iqwig.de), Institute for Quality and Efficiency in Health Care, Dillenbergerstr 27, Cologne 51105 Germany

Mona Nasser, Dr. dentistry, Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth, The John Bull Building, Tamar Science Park, PL6 8BU Plymouth, UK

Andreas Waltering, Dr. medicine (andreas.waltering@iqwig.de), Institute for Quality and Efficiency in Health Care, Dillenburgerstr 27, Cologne 51105 Germany

CONFLICT OF INTEREST
All authors report they have no potential conflicts of interest.

REFERENCES
12. Noorani HZ, Husereau DR, Boudreau R. Priority setting for health technology assessments: A systematic review of current

INTL. J. OF TECHNOLOGY ASSESSMENT IN HEALTH CARE 27:4, 2011 355


