CURRENT OPINION

Drug Saf 2012; 35 (11): 983-993
0114-5916/12/0011-0983/$49.95/0

Adis © 2012 Springer International Publishing AG. All rights reserved.

Good Decision Making Requires Good Communication

Baruch Fischhoff

Department of Social and Decision Sciences, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA, USA

Abstract

The methods used for regulatory decisions must facilitate three kinds of communication: (i) with individual experts who must translate their knowledge into usable form; (ii) among the experts whose pooled knowledge informs those choices; and (iii) between regulators and those affected by their choices. Decision-making methods vary in their reliance on expert judgement and computational methods and, hence, in their ability to meet the goals of sound decision making: breadth, depth, precision, neutrality, evaluability and transparency. An approach developed by the US FDA, the Benefit-Risk Framework, integrates judgement and computation, cognizant of their strengths and weaknesses. Its application both requires and facilitates good communication about risks and benefits.

Sound, credible regulatory decision-making faces three communication challenges. One is having methods and conclusions that are transparent to the many stakeholders concerned with the choices, including physicians, patients, patient advocates, pharmaceutical companies, pharmacists and politicians. The second is facilitating communication among an agency's expert reviewers, so that all issues are aired and resolved as far as the data allow. The third is communication with individual experts, so that they can express their knowledge in relevant terms. The methods available to meet these challenges vary in their reliance on expert judgement and computational methods. This article considers, in turn, the general strengths and weaknesses of judgement and computation, their specific expressions in regulatory decisions, and an integrated approach being developed by the US FDA.

In order to be approved for use, drugs must be shown to be effective and to have benefits that outweigh their risks. Regulators making those assessments consider both direct effects on patients who might take a drug and indirect effects on other public health concerns, such as promoting the development of treatment options for patients who do not tolerate or will not use existing products. In making these choices, regulators need the best available evidence regarding products' benefits and risk, including candid assessments of the uncertainties surrounding that evidence.[1-3]

Drug approval decisions often force regulators to balance diverse, complex, uncertain outcomes. The methods for that balancing can be described as falling along a continuum ranging from pure judgement to pure computation. At the one extreme, regulators can rely on their expert reviewers' experience to assess whether the apparent risks and benefits warrant making it available to patients, conditional on its other public health impacts. At the other extreme, regulators can rely

on formal models, such as those developed for cost-effectiveness analysis, to compute aggregate estimates of benefit and risk.[4,5] Many of those models seek to measure the relative importance of all outcomes in common units, such as Quality-Adjusted Life-Years (QALYs). An attraction of those models is offering explicit, replicable procedures for computing the overall attractiveness of treatment options.[6-8] A danger is obscuring the exercise of expert judgement involved in resolving matters of fact (i.e. analysing the evidence about key outcomes) and matters of value (i.e. determining those outcomes' relative importance).[9]

Section 1 draws on results from decision science[10,11] to characterize the strengths and weaknesses of approaches relying on judgement and computation in terms of six goals of sound decision making. Section 2 considers the specific challenges posed by regulatory benefit-risk decisions. Section 3 describes an approach being deployed by the FDA, designed to integrate judgement and computation in ways suited to those decisions. Briefly, that approach takes computation as far as it can go, then uses disciplined expert judgement to produce a fuller understanding of benefits and risks. Section 4 discusses how sound decision-making methods require sound communication.

1. Six Goals for Decision-Making Processes

1.1 Breadth

A decision-making process should address all issues relevant to the decision maker. To that end, judgement has a natural advantage, in that experts can draw on their broad knowledge to consider any issue set before them. In contrast, formal analysis can only address issues translated into its standard terms. As a result, expert judgement potentially has much greater breadth than does formal analysis. The value of that potential depends on how much breadth is needed and how well experts can provide it.

An important finding in decision science is that experts often overestimate their ability to think broadly, in the sense of considering many issues systematically.[12,13] Indeed, the research finds that experts often make no better decisions than do simple formulae (e.g. compare the number of good and bad outcomes) that rely on expert judgement to identify the key issues. One reason for this finding is that experts cannot keep everything in their heads. As a result, rather than trying to think broadly, they may do better by addressing every item in a standard set of issues, and then performing a simple calculation. Checklists are often useful for just the same reason. Achieving greater consistency compensates for restricting the set of issues.[14,15] Like checklists, simple computations are also transparent, unlike complex formulae whose results must be taken on faith. How these issues emerge in any specific decision-making setting is an empirical question, answered by studying how well a method predicts health outcomes and communicates its findings.[16]

However complex their computations might be, all models reflect the judgements of the experts who set their initial terms, revise them in order to incorporate new issues, assess the importance of the issues that they omit, and explain their logic. As a result, when seeking breadth of coverage, the best balance of judgement and computation involves computing the computable, then using expert judgement for insight into issues that the models ignore.

1.2 Depth

Analogous issues arise with achieving the depth needed to understand each element of the relevant evidence adequately. Formal methods can compute statistical summaries of identically collected observations. Such calculations avoid the imprecision of relying on mental arithmetic and the biases from relying on intuitive judgement (e.g. unduly weighting memorable observations).[11,17,18] However, they still require the expert judgement of statisticians to select methods, evaluate data quality and address problems (e.g. outliers, missing data, imperfect randomization).

Even more expert judgement is needed to combine evidence from diverse sources. That judgement might be used to construct formal models

Adis © 2012 Springer International Publishing AG. All rights reserved.

that estimate the likelihood of future events by combining statistics for their components (e.g. probabilistic risk assessments of new medical devices). Or, it might be elicited as holistic judgements that seek to accommodate all sources of uncertainty, including statistical variability.

As with breadth, the best way to achieve needed depth is to compute as much as possible, so as to avoid the limits to intuitive statistics, and then rely on expert judgement for the rest, so as to accommodate additional forms of evidence. These judgements should be elicited with methods found to reduce bias and subject to evaluation, so that users know how much to trust them.[3,10,19-23]

1.3 Precision

Formal models require users to define their issues precisely, thereby reducing the ambiguity that is part of much thinking and discourse. Doing so addresses the concern raised in Michael Faraday's famous quote, "If you cannot measure it, you cannot improve it." Precision also facilitates communicating decisions and evaluating their quality.[24]

For measures to achieve those ends, they must be communicated clearly to the experts who provide the inputs to the calculations and to the decision makers who use their outputs. Those inputs might include raw data (e.g. survey reports of how often medications are taken or adverse effects are experienced) and expert judgements of those data (e.g. the accuracy of such patient reports). The outputs of the calculations might be predictions of critical outcomes (e.g. adverse effect rates; the number of people who need to be treated in order to achieve a clinical outcome), analyses of potential antecedents (e.g. blood chemistry) or assessments of the uncertainty surrounding them.

Unfortunately, precision is easily overestimated. Even seemingly clear and simple terms, such as 'rain', 'likely', 'pain' or 'safe sex', can mean different things to different people – or to the same people in different contexts.[25,26] As a result, using such terms forces people to read between the lines, and guess at the intended meaning.[27] Faced by such ambiguity, people may not only guess wrong, but also question the motives of those who fail to make themselves clear. A problem specific to formal models is that analytically precise terms may have little intuitive meaning.[28] Such miscommunication can easily go unnoticed unless the degree of shared understanding is measured directly.[29]

In this light, the best way to balance judgement and computation in pursuit of precision is to use the most precise measures that people can comfortably understand, and then rely on results from behavioural research to extrapolate from there to the measures that decision makers need.[30,31] For example, if decision makers need more precise estimates of health behaviours (e.g. annual adverse effect rates) than people can give, one might ask about the last occurrence of an event, then infer the desired rate by projecting over the relevant time period, adjusting for telescoping (the tendency to exaggerate how recently events occurred). Asking people about distinctive periods (e.g. how did you feel the week before Thanksgiving – or finals?) can evoke incidental cues that aid memory, while possibly incurring bias (e.g. if those periods have atypically high or low rates).[32]

1.4 Neutrality

Decision makers' choices should reflect their organizations' values. Their experts' job is to summarize the relevant evidence so that decision makers can apply those values. Experts can undermine those decisions if their personal values affect their work. That can happen both with formal analyses and with expert judgement.

With formal analyses, values are inevitably embedded in their precisely defined terms.[10,33] For example, mortality risk can be measured (precisely) as the probability of premature death or as the expected number of life-years lost. The first measure treats all deaths as equal; the latter values young people more (as more years are lost when they die). There are ethical cases for either definition, but an analysis that uses just one ignores the other. Similar value-laden judgements accompany other terms that need precise definitions when computing expected impacts (e.g. whether

Adis © 2012 Springer International Publishing AG. All rights reserved.

to distinguish people by their sex, ethnicity or ability to provide informed consent for treatment).

With expert judgement, values can affect results both deliberately and inadvertently. On the one hand, experts might make explicitly 'conservative' estimates in order to induce a measure of caution. On the other hand, their judgement might be clouded by 'motivated cognition', as they exploit ambiguity in order to see what they hope to see and dismiss unwanted observations.[34,35]

The best way to balance judgement and computation in achieving neutrality is to have experts communicate the values embedded in their work, along with the implications of using alternative values (e.g. different definitions of mortality). Decision makers can then choose the perspective closest to their mandate. Such help is especially needed when people face novel choices, forcing them to 'construct' their preferences, by working through the implications of their basic values for specific choices.[36,37] To this end, experts must communicate the values embedded in their models (e.g. how they defined 'mortality') and their analyses (e.g. how conservative their estimates are; how much they have discounted future outcomes). Having experts disclose their values also offers an explicit way to express those concerns, reducing any influence on their judgements of the evidence.

1.5 Evaluability

Decision makers need to know how definitive analyses are, when basing choices on them. Formal models can address this question with sensitivity analyses: varying the inputs through the range of plausible values, allowing decision makers to see if their choices change (hence are 'sensitive' to which value is used). Statistical measures of variability (e.g. standard deviations, confidence intervals) can guide the choice of values for sensitivity analyses by capturing what has been observed (e.g. in clinical trials). However, judgement is needed to capture what might be observed (e.g. with actual use of a product or actual reporting of adverse events).

Structured elicitation of expert judgement is the standard approach to identifying plausible values. Like all measurements, such judgements can be evaluated in terms of consistency or accuracy.[38] Consistency is evaluated by asking the same question in different ways, and then seeing how compatible the answers are (e.g. comparing judgements elicited for different time periods).[39] Accuracy is evaluated by eliciting judgements precisely enough to compare them to subsequent events. For example, studies of the accuracy of probability-of-precipitation forecasts have found them to be well calibrated, in the sense that it rains about 70% of the time when forecasters assign that probability.[23]

The best way to balance judgement and computation when designing a decision-making process for evaluability is to compute sensitivity analyses, guided by expert judgements, taking into account any suspected bias. For example, research has found that judgements are sometimes anchored on the first value that people consider. As a result, the preferred way to elicit probability distributions begins by asking about extreme values – because asking first for a best guess would anchor subsequent judgements on that value, producing unduly narrow intervals.[18] If experts insist on starting with a best guess, then sensitivity analyses can use broader intervals than the ones that they provide. Such strategies acknowledge the possibility of bias without letting that threat disqualify all judgements.

1.6 Transparency

Good communication between experts and decision makers should be an emergent property of a well-designed decision-making process. That process should allow decision makers to communicate their information needs to experts, so that relevant analyses are performed and communicated comprehensibly. Formal analysis coupled with structured expert elicitation can characterize information needs in analytical terms.[36-38] Behavioural research can convey that information effectively, and then assess recipients' understanding.[40,41]

Understanding regulators' choices requires understanding both the evidence and the decision-making process. Which issues and stakeholders

Adis © 2012 Springer International Publishing AG. All rights reserved.

have standing? Which values are embedded in its terms? How much is invested in clarifying those values and separating them from assessments of fact? How diligently are uncertainties sought? How conscientiously are their implications explored? How candidly are they expressed? Effective communication need not be expensive, but does require evidence.[16] Sound decision-making methods should simplify communication, by focusing it on the few most critical issues and explicating the rationale for regulators' choices.[42]

2. Design for Decision: Approving Medical Treatments

Applying these general design principles to specific decisions requires considering the kinds of *options*, *outcomes*, and *evidence* that those decisions entail. For regulators' benefit-risk decisions about drugs, these three elements typically have the following properties, with the attendant design implications.

2.1 Options

Depending on their legal framework, regulators might have but two discrete options: approving or not approving a drug (perhaps revisiting that choice as evidence accumulates). In other cases, though, regulators can create new options, hoping to modify a drug's benefits or risks and improve its benefit-risk profile. For example, the FDA[43] can approve a drug conditional on a Risk Evaluation and Mitigation Strategy (REMS), requiring actions such as special provider training, patient education, or restrictions on drug dispensing.

With discrete options, regulators need to know only whether expected benefits exceed expected risks (conditional on other public health impacts). That determination should be straightforward when the treatment is a clear winner supported by strong data, but difficult when those conditions are not met. A sound decision-making procedure will allow regulators to adapt their level of effort to the difficulty of the choice – making quick, careful work of the easy decisions, while investing more effort in the hard ones.

When regulators can create options, they need more than just bottom-line estimates of expected benefits and risks. Rather, they also need to understand the processes determining those outcomes well enough to envision new options that might push a treatment over the threshold of acceptability. To that end, the decision-making process must reveal which issues make experts nervous (e.g. drug-drug interactions that could not be observed in clinical trials, given patient selection criteria) and where they see opportunities for improvement (e.g. checks integrated into pharmacy systems). That knowledge must then accompany their decision, as part of the rationale communicated to stakeholders.

2.2 Outcomes

Regulators' legal framework determines which outcomes have standing. Those may be just direct impacts on patients' health or other public health outcomes as well (e.g. the availability of novel dosing methods). That legal framework should also determine the decision rule to use in combining evaluations of a drug's relevant outcomes. That rule might be compensatory, if a drug's strengths can cancel out its weaknesses; conjunctive, if a drug must be acceptable on every outcome (e.g. with no more than a maximum tolerable risk and no less than a minimum useful benefit); or disjunctive, if it must have at least one exceptional property (e.g. a new mechanism of action), while being acceptable in other respects.

Healthcare analysts have developed various methods for comparing the potentially diverse outcomes that patients might experience.[7,8,22,31] One strategy, mentioned earlier, translates all outcomes into a common unit (e.g. QALYs). A second strategy elicits preferences among vectors of outcomes, hoping to reveal the relative weights assigned to each.[44]

One common criticism of these methods is that expressing all values on a single scale loses distinctions essential to making, and then communicating, decisions. A second criticism is that these methods obscure value questions (e.g. how should one define mortality or deal with effects distributed over time). A third criticism questions whether

Adis © 2012 Springer International Publishing AG. All rights reserved.

people can articulate their preferences well enough to provide the stable judgements that these methods require.[36,37] A fourth criticism argues that these value questions should be resolved through vigorous public debate, rather than being buried in analytical details.[11]

These criticisms apply to the 'traditional' use of these methods in cost-effectiveness analysis, where the benefits of making many, relatively similar decisions consistently might compensate for ignoring the differences among those decisions.[12,13] Regulatory approval decisions, though, pose unique choices with complex implications for decision made by potentially heterogeneous and uncertain patient populations. Analysts can often shed some light on any issue (e.g. a putative value for the theoretical possibility that a treatment would slow the development of bacterial resistance), as long as their work is properly qualified and does not displace needed discussion.

A method that informs decision makers' judgement, rather than replacing it with computed solutions, must provide the alternative perspectives that they need to articulate their preferences.[37] Those perspectives might include calculations made under different assumptions, as long as decision makers recognize that the more they rely on others' calculations the more difficult it will be for them to communicate the rationale for their choices.

2.3 Evidence (and Uncertainties)

Decision makers need three kinds of knowledge about the evidence regarding a treatment: (i) statistical summaries of the relevant data; (ii) critical evaluations of quality of the studies producing those data; and (iii) an understanding of the strength of the underlying science. Here, too, they must get the best combination of judgement and computation.

Given the limits to intuitive statistics, computation is obviously better than judgement for producing statistical summaries. As mentioned, those calculations require the expert judgement of statisticians (e.g. to choose right tests, handle missing observations and outliers). Once done, they must be communicated to users who may need help in avoiding common pitfalls (e.g. confusing absolute and relative risk, confusing practical and statistical significance).[11,18]

Expert judgement is also essential for the other two assessments that decision makers need: the quality of the studies and the strength of the science underlying them. Ascertaining the quality of the studies is the central work of regulatory reviewers, who have the strength of the science in the back of their minds.[2,3] Reviewers know that even the most careful studies in new scientific areas can provide a shaky foundation for decision making, as can weak studies in mature areas. Computation can inform these judgements, as when a formal meta-analysis assesses the overall strength of the benefit or risk signal lurking in a set of studies[1] or when meta-methodology studies estimate the size of the errors introduced by design flaws (e.g. not randomizing; not blinding participants).[35]

Regulators need to know about these uncertainties in order to know how much confidence to place in their decisions, what additional data to require, what precautions to take and how vigilant to be for surprises (perhaps requiring them to revisit a decision). Moreover, they need to receive that knowledge in a form that allows them to design better options, construct thoughtful preferences and communicate their conclusions. Meeting those conditions is the challenge facing those who design regulatory decision-making processes.

3. A Design for Decision: US FDA's Benefit-Risk Framework

Since 2009, the FDA has been developing an approach to benefit-risk decision making that seeks to balance judgement and computation. Eggers,[45] Jenkins[46] and Walker et al.[47] provide brief descriptions. The regulatory philosophy underlying the approach is expressed in discussions of specific decisions (e.g. Beasley et al.,[48] Parks and Rosebraugh[49]). The following is my personal view of the approach, cast in terms of the design principles described above. It does not obligate the FDA in any way. Alternative approaches, developed within other design constraints, can be found in Breckenridge,[50] Coplan et al.,[6] Eichler

Adis © 2012 Springer International Publishing AG. All rights reserved.

et al.,[51] European Medicines Agency,[52] Levitan et al.,[7] Mussen et al.[53] and Phillips et al.[8] They differ from the FDA's Framework primarily in requiring more formal (and more demanding) judgements from experts and in relying more heavily on computation.

A distinctive feature of the FDA's process in developing the framework has been extensive testing with its reviewers. That testing sought to ensure that the method addressed the three essential communication needs: (i) communicating with individual experts, so that they understand the method and can express themselves in its terms; (ii) allowing effective communication within review teams, so that members' expertise is fully shared; and (iii) facilitating stakeholders' understanding of how decisions are reached and ability to provide requested input. The framework evolved through that testing process, guided by the underlying decision science.[10,54]

3.1 The Framework

Table I shows the Benefit-Risk Framework. The five rows represent the factors that the FDA considers, in an order that tells the story of a decision: what condition a drug treats, what are the unmet medical needs in that therapeutic area, what appear to be the drug's clinical benefits, what are its estimated risks and what risk management activities could improve those outcomes. These rows express the FDA's regulatory philosophy. For example, they show that the FDA considers benefits and risks, but not monetary costs – leaving them to other decision makers. The first row shows the FDA's attention to defining the benefit precisely (e.g. just initial infections, just children aged over 6 years, just patients without the complications excluded in the proposed label). That description of the condition should communicate its severity well enough for readers to understand why the FDA decided that predicted benefits outweighed predicted risks – or not.

The Framework's two columns distinguish evidence and uncertainties (left) from conclusions and reasons (right). The former summarizes the submitted data, as analysed by the FDA's subject matter experts. The latter summarizes their opinions regarding the implications of that evidence for the pending regulatory decision. The former has judgements of fact; the latter has judgements of value. Having separate columns clearly distinguishes those two kinds of subjectivity. It also invites reviewers to express their thinking about policy issues, recognizing that senior officials with signatory authority ultimately make the decision. It might reduce any tendency for value judgements to influence scientific ones (e.g. injecting caution with 'conservative' estimates).

When the framework has been completed for a product, its cells summarize the issues that reviewers consider critical to the approval decision. These entries will include both calculations (e.g. confidence intervals for effect sizes) and judgements (e.g. concerns about data quality). The completed framework may note issues that once seemed relevant, but proved immaterial (e.g. manufacturing problems that were resolved, novel outcome measures that were rejected). Detailed support for each entry appears in traditional FDA review documents. Checklists (not shown) suggest issues to consider for each cell. The table is accompanied by a narrative summary of the decision and the reasoning behind it, thereby communicating how the FDA balanced benefits

Table I. US FDA's Benefit-Risk Framework

Decision factor	Evidence and uncertainties (summary of evidence)	Conclusions and reasons (implications for decision making)
Analysis of condition		
Unmet medical need		
Clinical benefit		
Risk		
Risk management		

Adis © 2012 Springer International Publishing AG. All rights reserved.

and risks, given its appraisal of the condition and the unmet medical need, and its specification of risk management measures. The Framework aids transparency by focusing attention on the main issues, keeping them from being hidden in plain sight.

3.2 Characterizing Uncertainty

The FDA's regulatory philosophy, as captured in the Benefit-Risk Framework and other public expressions,[3,45-49] recognizes that its decisions must reflect the quality of the available evidence. A method for assessing each form of uncertainty (variability, data quality, strength of science) must find a 'sweet spot' between judgement and computation, such that experts provide the most precise judgements they can without losing touch with what they are saying, as a result of expressing themselves in strange terms. What follows is a possible strategy based on basic decision science research. As with any method, its applicability to the FDA's regulatory decision making is an empirical question.

Variability: The effects observed with a medical treatment vary, reflecting variation in the procedure (e.g. consistency of administration), the condition (e.g. forms of a disease), patient responsiveness (e.g. due to genetics or medical history), other events (e.g. drug-drug interactions) and measurement error. That variation is conventionally represented in confidence intervals, computed for outcome measures with clinical significance. Choosing those measures requires expert judgement, as does communicating the statistical summaries (e.g. not confusing statistical significance with practical value; properly discounting *post hoc* tests).

Data quality: A sophisticated research community supports the clinical trials that inform regulatory decisions. These labours notwithstanding, all trials are imperfect – as they must be, treating real people in complex settings. As a result, the FDA's reviewers intensively analyse how well trials fared. Their analyses address both internal validity (how well was a trial conducted?) and external validity (how well can it predict experience in the postmarket setting?). One standard approach to such evaluations is Risk of Bias (RoB),[35] part of the Cochrane Collaboration's Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework for assessing the quality of evidence in healthcare. RoB research reviews clinical trials in order to identify methodological problems with the greatest impact on results. A formal RoB analysis applies correction factors, in order to estimate what clinical results might have been had trials not had biases (e.g. high attrition rates, poor randomization, lack of blinding). A less formal application has reviewers indicate whether a study was vulnerable to each major bias on the RoB list and, if so, how that might have affected the observed outcomes.

Strength of science: The confidence placed in any evidence depends on the quality of the basic science underlying it. A novel drug creates uncertainty if the science is weak for predicting its performance with other populations, over longer periods of time and in more varied settings. A complex condition creates uncertainty if it relies on surrogate measures without strong science linking them with health effects. One approach to characterizing the strength of the science is in terms of its 'pedigree'.[55,56] Full pedigree analyses evaluate a science along multiple dimensions (e.g. the directness of its measures, the precision of its models). Simple ones just rate the science on a single dimension anchored at 'strong' and 'weak'. Properties of strong science include measuring outcomes directly, rather than relying on surrogates (e.g. biomarkers); being supported by large rigorous experiments, rather than just statistical predictions (e.g. dose-response relationships estimated from noisy epidemiological data); and using established, widely accepted methods, rather than ones that vary by investigator. Clinical trial reviewers should have little trouble making such ratings, explaining their reasoning and interpreting the implications of those uncertainties.

Combining these three kinds of uncertainty requires expert judgement. From a decision science perspective, the standard summary is the credible interval, expressing experts' summary judgement as a range of plausible values (e.g. 'there is a 90% chance that the sustained remis-

Adis © 2012 Springer International Publishing AG. All rights reserved.

sion rate will be between X and Y, if the treatment is approved'). Statistical confidence intervals are a natural point of departure for such judgements. Credible intervals should be broader than confidence intervals when clinical trials have design flaws or weak scientific foundations. They should be narrower when trials disrupt normal processes (e.g. with intrusive tests), making trial experience noisier than that in everyday life.

Credible intervals are routinely elicited in some domains (e.g. forecasting, meteorology, probabilistic risk assessment[23,33,54]), whereas in other domains (e.g. intelligence analysis, climatology), some experts are uncomfortable producing them.[28,57,58] Elicitation procedures that address these experts' concerns might reduce that resistance: (i) elicit judgements for well-defined outcomes (e.g. adverse effect rates in specific populations), so that experts are not asked to give precise answers to vague questions; (ii) specify which issues have standing, so that experts are not forced to guess what 'uncertainty' entails; (iii) stress regulators' need for quantitative assessments of uncertainty, so that they are not forced to read experts' minds. If experts are still uncomfortable, a complete, clearly communicated uncertainty analysis (variability, data quality, strength of science) might provide decision makers with enough information to infer the credible intervals that the experts see, but hesitate to say.

3.3 Evaluating the Decision-Making Process

As mentioned, how well a decision-making process works is an empirical question. Evaluation requires defining appropriate measures and collecting the requisite data. Proper evaluation is essential to the continuous improvement that all organizations need. It protects against faulty intuitive evaluations, such as seeing some impacts (e.g. time spent in meetings), while missing others (e.g. time saved by avoiding problems or not analysing unimportant issues). It draws attention to organizational priorities, such as respecting the views of all staff members, demonstrating consistency in decision rules and creating a transparent record. A regulatory body making benefit-risk decisions might define a method's success by how well it does the following.

Capture the bases of the decisions: A sound decision-making method should address all issues relevant to pending regulatory actions, in ways that faithfully represent the evidence, as interpreted by agency experts. The method should record the agency's expectations precisely enough that they can be evaluated in the light of experience, making learning possible. It should express the rationale of the decisions in terms that allow the organization to assess and, if warranted, demonstrate the consistency of its regulatory philosophy.

Communicate how the agency reaches its decisions: A sound method should convey the rationale of regulatory decisions with a clear organizing narrative (e.g. 'We accelerated the review given the treatment's clear benefits and minimal risks.' 'We gave cautious approval, but mandated intensive tracking of adverse events among patients who chose to use the product.') The method should summarize the relevant evidence in comprehensible terms, with ready links to more detailed analyses. It should provide any context that readers need in order to interpret its conclusions (e.g. what the medical condition is like, what the agency's regulatory powers allow it to do, what prior agreements were reached with the sponsor).

Use staff time and energy well: For individual experts, a decision-making method should be clear, efficient and motivated. For the agency as a whole, the method should promote internal communication, recognize staff members' contributions and save time by focusing resources and anticipating problems. The FDA's framework addresses these goals by paying attention to all cells, respecting the expert judgement needed to fill them and soliciting experts' opinions about approval decisions, even if responsibility for them lies elsewhere.

4. Conclusions

Benefit-risk decisions about medical treatments must assess complex, uncertain, diverse outcomes, and then determine whether the benefits outweigh the risks. The success of these decisions

Adis © 2012 Springer International Publishing AG. All rights reserved.

requires properly balancing judgement and computation. Judgement alone is inadequate because of its inherent limits and potential biases. Computation alone is inadequate because it cannot capture all the issues that must inform sound benefit-risk decision making. The FDA's emerging Benefit-Risk Framework draws on decision science in seeking that balance. If successful, it will address the three communication challenges facing those decisions: (i) communication with the individual experts who must translate their knowledge into usable form; (ii) communication among the teams of experts who must pool their knowledge; and (iii) communication between regulators and those affected by their choices. A decision-making process has the greatest chances of success if it is grounded in the theory of decision science and evaluated with its methods, giving the process a sound foundation and the opportunity to learn from experience.

Acknowledgements

No sources of funding were used to prepare this manuscript. The author has served as Chair of the FDA Risk Communication Advisory Committee and as a consultant to the FDA. The views expressed here are his alone.

This article is part of a theme issue co-edited by Priya Bahri, European Medicines Agency, UK, and Mira Harrison-Woolrych, New Zealand Pharmacovigilance Centre, New Zealand. No external funding was used to support the publication of this theme issue.

References

1. Mussen F, Salek S, Walker S. Benefit-risk appraisal of medicines: a systematic approach to decision making. Chichester: Wiley, 2009
2. Temple R. How FDA currently makes decisions on clinical trials. Clin Tri 2005; 2: 278-81
3. Woodcock J, Temple R, Midthun K, et al. FDA senior management perspectives (roundtable). Clin Tri 2005; 2: 383-8
4. Drummond MF, Shculpher MJ, Torrance GW, et al. Methods for the economic evaluation of health care programmes. 3rd ed. Oxford: Oxford University Press, 2005
5. Gold MR, Siegel JE, Russell LB, et al., editors. Cost-effectiveness analysis in health and medicine. New York: Oxford University Press, 1996
6. Coplan PM, Noel RA, Levitan BS, et al. Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. Clin Pharmacol Ther 2011; 89: 312-25
7. Levitan B, Andrews E, Gilsenan A, et al. Application of the BRAT framework to case studies: observations and insights. Clin Pharmacol Ther 2011; 89: 217-24
8. Phillips LD, Fasolo B, Zafiropolous N, et al. Is quantitative benefit-risk modeling of drugs desirable or possible? Drug Discov Today Technol 2011; 8: e3-10
9. Temple R. Quantitative decision analysis: a work in progress. Clin Pharmacol Ther 2007; 82 (2): 127-30
10. Fischhoff B, Kadvany J. Risk: a very short introduction. Oxford: Oxford University Press, 2011
11. Kahneman D. Thinking, fast and slow. New York: Farrar, Straus & Giroux, 2011
12. Dawes RM. The robust beauty of improper linear models in decision making. Am Psychol 1979; 34: 571-82
13. Meehl PE. Clinical versus statistical prediction. Minneapolis (MN): University of Minnesota Press, 1954
14. Gawande A. No mistake. The New Yorker 1998 Mar 30; 74-81
15. Gawande A. The checklist manifesto. New York: Henry Holt, 2009
16. Fischhoff B, Brewer N, Downs JS, editors. Communicating risks and benefits: an evidence-based user's guide. Washington, DC: US FDA, 2011
17. Hastie R, Dawes RM. Rational choice in an uncertain world. 2nd ed. Thousand Oaks (CA): Sage, 2010
18. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. Science 1974; 185: 1124-31
19. Edwards W, Lindman HR, Savage LJ. Bayesian statistical inference for psychological research. Psychol Rev 1963; 70: 193-242
20. Fischhoff B, Beyth-Marom R. Hypothesis evaluation from a Bayesian perspective. Psychol Rev 1983; 90: 239-60
21. Phillips LD. Bayesian statistics for the social sciences. London: Thomas Nelson, 1974
22. Hunink M, Glasziou P, Siegel JE, et al. Decision making in health and medicine: integrating evidence and values. Cambridge: Cambridge University Press, 2011
23. O'Hagan A, Buck CE, Daneshkhah A, et al. Uncertain judgements: eliciting expert probabilities. Chichester: Wiley, 2006
24. Fischhoff B. What forecasts (seem to) mean. Int J Forecast 1994; 10: 387-403
25. Campbell P. Understanding the receivers and the reception of science's uncertain messages. Philos Trans Royal Soc 2011; 369: 4891-912
26. Murphy AH, Lichtenstein S, Fischhoff B, et al. Misinterpretations of precipitation probability forecasts. Bull Am Meteorolog Soc 1980; 61: 695-701
27. Schwarz N. Self reports. Am Psychol 1999; 54: 93-105
28. Budescu DV, Broomell S, Por H-H. Improving communication of uncertainty in reports of the IPCC. Psycholog Sci 2009; 20: 299-308
29. Downs JS. Evaluation. In: Fischhoff B, Brewer N, Downs JS, editors. Communicating risks and benefits: an evidence-based user's guide. Washington, DC: US FDA, 2011: 11-8
30. Schwartz A, Bergus G. Medical decision making. New York: Cambridge University Press, 2008
31. Sox HC, Blatt MA, Higgins MC, et al. Medical decision making. Philadelphia (PA): American College of Physicians, 2007

Adis © 2012 Springer International Publishing AG. All rights reserved.

32. Turner CF, Martin E, editors. Surveying subjective phenomena. Washington, DC: National Academy Press, 1984
33. Crouch EAC, Wilson R. Risk-benefit analysis. Cambridge (MA): Harvard University Press, 1981
34. Kunda Z. The case for motivated cognition. Psychol Bull 1990; 108: 480-98
35. Turner RM, Spiegelhalter DJ, Smith G, et al. Bias modelling in evidence synthesis. J R Stat Soc Series A 2009; 172 (1): 21-47
36. Fischhoff B, Manski C, editors. The elicitation of preferences. Boston (MA): Kluwer, 1999
37. Lichtenstein S, Slovic P, editors. Construction of preferences. New York: Cambridge University Press, 2006
38. Fischhoff B. Cognitive processes in stated preference methods. In: Mäler K-G, Vincent J, editors. Handbook of environmental economics. Amsterdam: Elsevier, 2005: 937-68
39. Bruine de Bruin W, Fischhoff B, Brilliant L, et al. Expert judgments of pandemic influenza. Glob Public Health 2006; 1 (2): 178-93
40. Löfstedt RE. Risk management in post-trust societies. Basingstoke: Palgrave/MacMillan, 2005
41. Löfstedt R, Bouder F, Chakraborty S. Transparency and the FDA: a quantitative study. J Health Commun. In press
42. Reyna VF. A theory of medical decision making and health: fuzzy-trace theory. Med Decis Making 2008; 28: 850-65
43. FDA. Guidance for industry: medication guides – distribution requirements and inclusion in risk evaluation and mitigation strategies (REMS). Washington, DC: FDA, 2011
44. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk tradeoffs of treatment alternatives. Med Decis Making 2009; 29: 104-15
45. Eggers S. FDA CDER's benefit-risk assessment framework. ISPOR International 2012 Annual Meeting; 2012 Jun 5; Washington, DC
46. Jenkins J. A United States regulator's perspective on risk benefit considerations [online]. Available from URL: www.fda.gov/downloads/aboutFDA/CentersOffices/CDER/UCM210155.pdf [Accessed 2012 Apr 14]
47. Walker S, McAuslane JN, Liberti L. Developing a common benefit-risk assessment methodology for medicines: a progress report. Scrip Regul Aff 2011 Dec; 18-21
48. Beasley BN, Unger EF, Temple R. Anticoagulant options: why the FDA approved a higher but not lower dose of dabigatran. N Engl J Med 2011; 364: 1788-90
49. Parks M, Rosebraugh C. Weighing risks and benefits of liraglutide: the FDA's review of a new antidiabetic therapy. N Engl J Med 2010; 362: 774-7
50. Breckenridge A. Regulatory challenges, reimbursement, and risk-benefit assessment. Clin Pharmacol Ther 2010; 88: 153-4
51. Eichler HG, Pignatti F, Flamion B, et al. Balancing early market access to new drugs with the need for benefit/risk data. Nat Rev Drug Discov 2008; 7: 818-26
52. European Medicines Agency. Benefit-risk methodology project. Brussels: European Medicines Agency, 2009
53. Mussen F, Salek S, Walker S, et al. A quantitative approach to benefit-risk assessment of medicines. Pharmacepidemiol Drug Saf 2007; 16: S2-41
54. Morgan MG, Henrion M. Uncertainty. New York: Cambridge University Press, 1990
55. Funtowicz SO, Ravetz JR. Uncertainty and quality in science for policy. Dordrecht: Kluwer, 1990
56. NUSAP.net: robust knowledge for sustainability [online]. Available from URL: http://www.nusap.net [Accessed 2012 Sep 7]
57. National Research Council. Intelligence analysis for the future. Washington, DC: National Academy Press, 2011
58. Fischhoff B. Communicating uncertainty: fulfilling the duty to inform. Issues Sci Technol 2012; 29: 63-70

Correspondence: Professor *Baruch Fischhoff*, Howard Heinz University Professor, Department of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15213-3890, USA.
E-mail: baruch@cmu.edu

Adis © 2012 Springer International Publishing AG. All rights reserved.