A Benefit-risk Model to Facilitate DMC-sponsor Communication and Decision Making

In drug development, independent data monitoring committees (DMCs) are routinely engaged by clinical trial sponsors to monitor patient safety. Transferring the obligation of interim data monitoring to an independent party helps ensure trial integrity by allowing the sponsor to remain blinded to emerging treatment results. But it also poses a dilemma: after each interim analysis, the DMC makes a recommendation to the sponsor regarding the continuation or modification of the ongoing trial, on the basis of which the sponsor must make a decision. Yet the unblinded results, which alone are the ideal basis for decision making, are only seen by the DMC and by design are withheld from the sponsor. We propose application of a multiattribute benefit-risk model to enable more informed decision making. The model provides an assessment of the emerging benefits and risks of active therapy versus control, based on an evaluation of attributes prespecified by the sponsor. While revealing no specific study outcomes, it allows an appreciation of the direction and magnitude of differentiation between treatment groups in terms of benefit and risk, enabling conversations around scenarios that would lead to this kind of assessment. The model also reflects uncertainty in the evaluation of benefits and risks, an essential feature in the analysis of interim data. This type of assessment may be leveraged to provide meaningful context accompanying the textual recommendation of a DMC to enable more informed decision making, while upholding the sponsor's commitment to remain blinded to particular study results.

Andreas Sashegyi, PhD
Research Advisor, Decision Sciences Lilly Corporate Center, Eli Lilly and Company, Indianapolis, Indiana

Key Words
Data monitoring committee; Multiattribute model; Benefit-risk; Blinding

Correspondence Address
Andreas Sashegyi, PhD, Drop Code 1512, Lilly Corporate Center, Eli Lilly and Company, Indianapolis, IN 46285 (email: aisasheg@lilly.com).

INTRODUCTION

Over the past two decades the use of independent data monitoring committees (DMCs) by pharmaceutical companies to monitor the accruing data, especially of late-stage clinical trials with mortality or serious morbidity outcomes, has become a norm. Ellenberg et al. (1) provide a broad overview of the necessity, structure, and operation of DMCs, and both the Committee for Medicinal Products for Human Use in Europe and the US Food and Drug Administration have issued DMC guidances (2,3). The intent of interim data monitoring is primarily to ensure patient safety but may also involve evaluation of whether early stopping due to futility or overwhelming evidence of efficacy is warranted. Importantly, transferring the critical obligation of interim monitoring to an independent party helps ensure trial integrity and minimize bias, to a large extent by allowing the sponsor to remain blinded to interim treatment results. However, this also sets up a fundamental dilemma with respect to the roles and jurisdictions of the DMC and the trial sponsor, which remains unresolved. Following an interim analysis of unblinded data, the DMC makes a recommendation to the sponsor regarding the continuation of the ongoing trial, and whether or not any modifications are warranted. The sponsor is called upon to make a decision based on this recommendation. But the unblinded results, which alone are the ideal basis for sound decision making, are seen only by the DMC and must by design be withheld from the sponsor. In certain circumstances, perhaps where a promising efficacy trend is offset by a safety profile that is more serious than anticipated or marked by an unexpected finding, arriving at a satisfactory textual recommendation may be very difficult for a DMC, and choosing an appropriate course of action based on just this next to impossible for the sponsor. DeMets et al. (4) have documented the challenges faced by a DMC seeing increasingly worrisome negative trends in accumulating trial data, a situation in which the possibility of some meaningful communication with the sponsor regarding the balance of benefits and risks would be desirable.

Submitted for publication: February 14, 2011
Accepted for publication: May 3, 2011

Sashegyi (5) has highlighted the issues in DMC-sponsor interactions in the context of monitoring blinded adaptive trial designs, in which the sponsor's decision-making role is even more accentuated.

As a recommended solution to this dilemma, I propose using a multiattribute benefit-risk model to facilitate communication between a DMC and the sponsor and to enable more informed decision making. The decision-analytic framework described by Mussen et al. (6) as well as Felli et al. (7), and leveraged in the benefit-risk assessment model (BRAM) introduced by Felli et al. (7), is ideally suited to this problem. The BRAM was recently published and this article illustrates how the model may be leveraged by a DMC to effectively communicate the balance of emerging benefits and risks of active therapy relative to control, without revealing specific study outcomes. The structure and implementation of the model will be reviewed only at a high level in this article, leaving much of the technical details to the published report on the BRAM; I focus instead on aspects of interpretation and practical implementation in addressing the problem at hand. My aim is to provide sufficient guidance here that in conjunction with the more detailed methodological report on the BRAM, the reader will be in a position to implement this approach as part of a DMC-sponsor communication plan.

METHODS

The BRAM addresses three fundamental requirements of a solution to the dilemma incurred when a DMC makes a recommendation based on unblinded data, but the sponsor must ultimately decide on a course of action in a blinded manner:

1. It allows the sponsor to be informed of certain features of the emerging trial data, beyond a terse DMC recommendation, to enable a good decision concerning the continuing conduct of the study.
2. The information conveyed does not reveal any specifics about trial endpoints or other study measurements, thus effectively protecting study blinding.
3. The information conveyed allows an assessment of whether emerging risks for patients are tolerable in view of accruing benefits.

The BRAM graphically presents contextual beliefs about benefits and risks in a framework conducive to focused discussion. Felli et al. (7) underscore the importance of selecting an analytic frame at the outset, fixing the disease state, population of interest, and perspective for assessments and trade-offs. In this application, the disease state is the one under investigation in the trial for which the DMC is engaged. The population of interest is specifically the set of trial participants; if the experimental therapy is already available to a broader population (say, as a marketed therapy for another indication), a benefit-risk evaluation for that population would likely require a different frame and is not within the purview of the DMC. The perspective for assessments and trade-offs is that of the DMC, with a primary focus on patient safety.

The structure of the model is defined by identifying relevant measures or subattributes that inform key attributes which collectively characterize the benefit and risk profiles of a particular intervention. In a randomized, controlled trial the interventions under investigation might be an experimental therapy and a control (active or placebo) arm. Figure 1 depicts the model structure graphically as advanced by Felli et al. (7). The "value" of a treatment alternative is a function of an overall assessment of risks and benefits, where risks are broadly categorized into safety and tolerability attributes, and benefits are similarly categorized either into specific efficacy attributes or characteristics having to do with life effects (indirect consequences of treatment that are not safety related) or the convenience of treatment administration. Each attribute is informed by a number of measures that are collectively used to evaluate that attribute. These measures are dependent on the specific application, with the exception perhaps of those that inform the convenience attribute, for which the measures listed in Figure 1 may be considered a more or less standard set. In using such a model for the

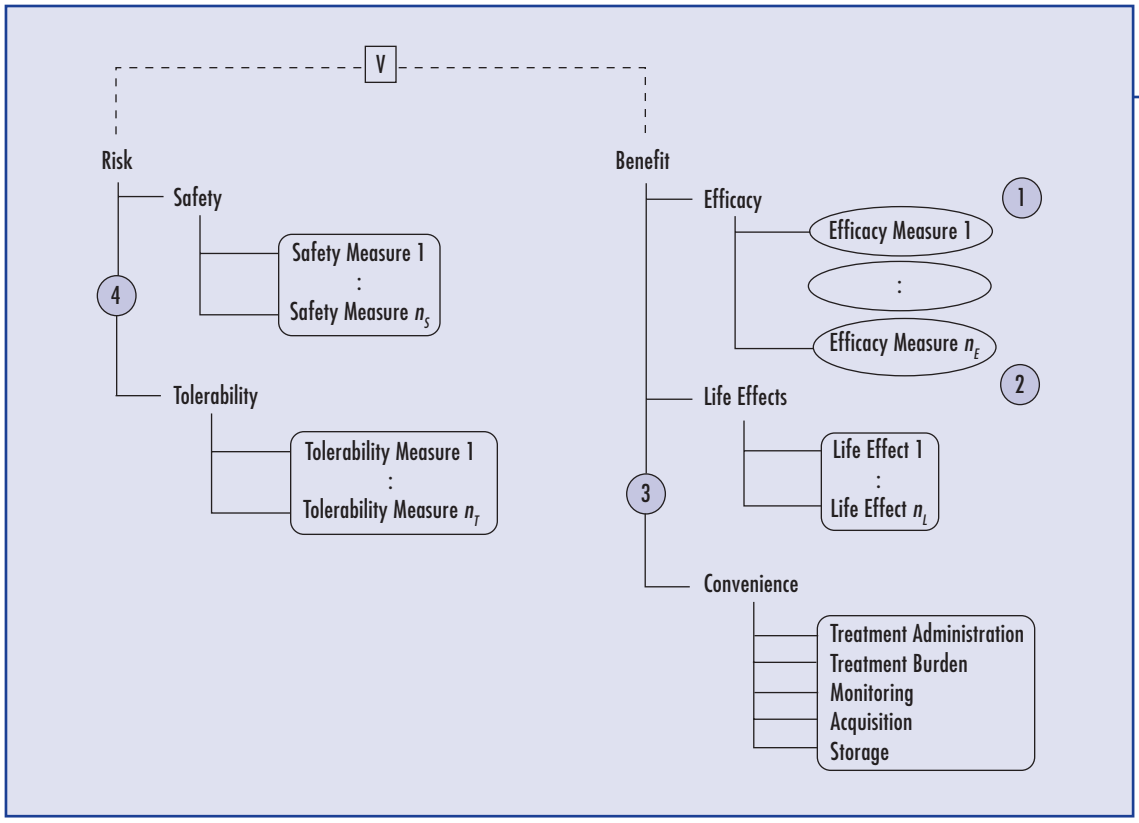

FIGURE 1

General structure of the benefit-risk assessment model. A DMC may leverage this model in evaluating interim clinical trial data by assessing (1) the primary endpoint, along with (2) other efficacy measures, as well as (3) other benefit attributes, in the context of (4) key safety and other risk measures.

evaluation of treatments in a clinical trial, it makes sense that the first (presumably most important) measure under efficacy be an assessment of the primary endpoint of the trial. But other measures should reflect additional features impacting the overall evaluation of efficacy, such as the secondary endpoints specified in the protocol. All these measures collectively then contribute to the efficacy assessment, which is combined with similar assessments of life effects and convenience to provide a summary of benefits. This benefit evaluation must then be balanced against the safety and tolerability profile that leads in the same manner to a summary of risks.

In terms of implementation, each measure in a particular set is weighted according to its relative importance to the evaluation of the attribute in question. Likewise, each attribute itself is weighted according to its relative importance to the evaluation of the benefit or risk dimension it informs. Clearly the specific weights selected for an evaluation through this model depend on the perspective for assessments and trade-offs; that is, they should be chosen by the DMC. Finally, in an effort to standardize the output of this model and render results interpretable, a transformation function is associated with each measure, which takes the observed data for that measure and maps it to a value between 0 and 1. By convention, the extremes of 0 and 1 correspond, respectively, to the least and most desirable scores under the benefit dimension, and to the least and most detrimental scores under the risk dimension, such that more desirable benefit and more detrimental risk ratings correspond to higher values. Thus, within a given attribute, the weighted average of the transformed measures yields an evaluation of the attribute (a value between 0 and 1), and similarly the weighted average of the safety and tolerability attributes yields an overall evaluation of risks, whereas the weighted average of efficacy, life effects, and convenience yield an overall evaluation of efficacy, again both in terms of values between 0 and 1.

For a given intervention (treatment or control), a particular set of data points reflecting

the observed measures of the BRAM thus results in a unit-independent (x, y) pair of coordinates depicting a position on a benefit (y) – risk (x) plane $P = [0,1]x[0,1]$. Clearly, the most desirable treatments will have benefit scores close to 1 and risk scores close to 0. Of interest to both a DMC and the sponsor of the trial in question is the position on P of the experimental therapy relative to the control.

In practice, the sponsor should discuss specifics of the BRAM with members of the DMC at their organizational meeting. This is ideally the time to agree on the structure of the model, that is, the subattributes to be measured and the specific variables or other data sources that will be used as the basis for the evaluation, as well as the transformation functions that will be leveraged to transform observed data of various types and scales to the unit interval. The basic structure of the final model and its intended use should be captured in the DMC charter and should not be altered during the conduct of the study. Furthermore, the sponsor should share with the DMC the weights the study team would place on each of the subattributes as well as the main attributes of the model. In the subsequent evaluations undertaken by the DMC, I recommend that each DMC member specify his or her own weights, thus arriving at a unique benefit-risk assessment that can be examined in concert with the assessment of the other DMC members. But knowing the sponsor's weights up front provides an important reference point to the DMC that may appropriately impact their deliberations. In the same vein, the study team should discuss sample scenarios with the DMC and provide an indication of the kinds of benefit-risk balances that the sponsor would find acceptable, and ones that would be unacceptable. This is accomplished by simply thinking about particular hypothetical outcomes for each BRAM measure and using the proposed weights to aggregate these outcomes into a benefit-risk assessment as discussed above, for the active therapy as well as the control. In considering numerous such scenarios, by appreciating each time the difference in position between active therapy and control (along both the benefit and risk axes) and relating this back to the outcomes that led to these positions, one is gradually able to develop a sense of the practical implications represented by certain differences in benefit and risk scores, without having to tie these back to any one particular outcome. A priori discussion of such sample scenarios between the sponsor and the DMC should be enlightening for both parties. It is designed to reveal important information about key stakeholders' beliefs of what is acceptable and what is not, thus enabling potential DMC members to decide prospectively whether they can accept a particular sponsor's risk tolerance, and vice versa. Currently, sponsors typically have no real insight into the varying attitudes among DMC members toward benefit-risk trade-offs. Yet this insight alone could have significant impact on the final composition of the DMC, and on the communication and interpretation of recommendations.

RESULTS

How, specifically, should a DMC use the BRAM for benefit-risk assessments based on interim data, and how should model results be interpreted? An ostensibly key differentiating feature of this versus other applications of the BRAM is the fact that incorporating the inherent uncertainty in the data that inform the BRAM is not only desirable but essential, since these data are by definition only interim results and thus furnish incomplete information at best. Thus, in evaluating the various BRAM subattributes based on the observed data, one should think in each case about what would constitute an appropriate range of results rather than a single point estimate. In some cases, this range could be defined through standard statistical approaches, such as using a confidence interval; in other cases, defining a reasonable range may require different criteria.

For example, suppose a DMC is monitoring a trial comparing a new diabetes compound against standard of care. The primary efficacy outcome is HbA1c lowering, and this has consequently been assigned as the first and most

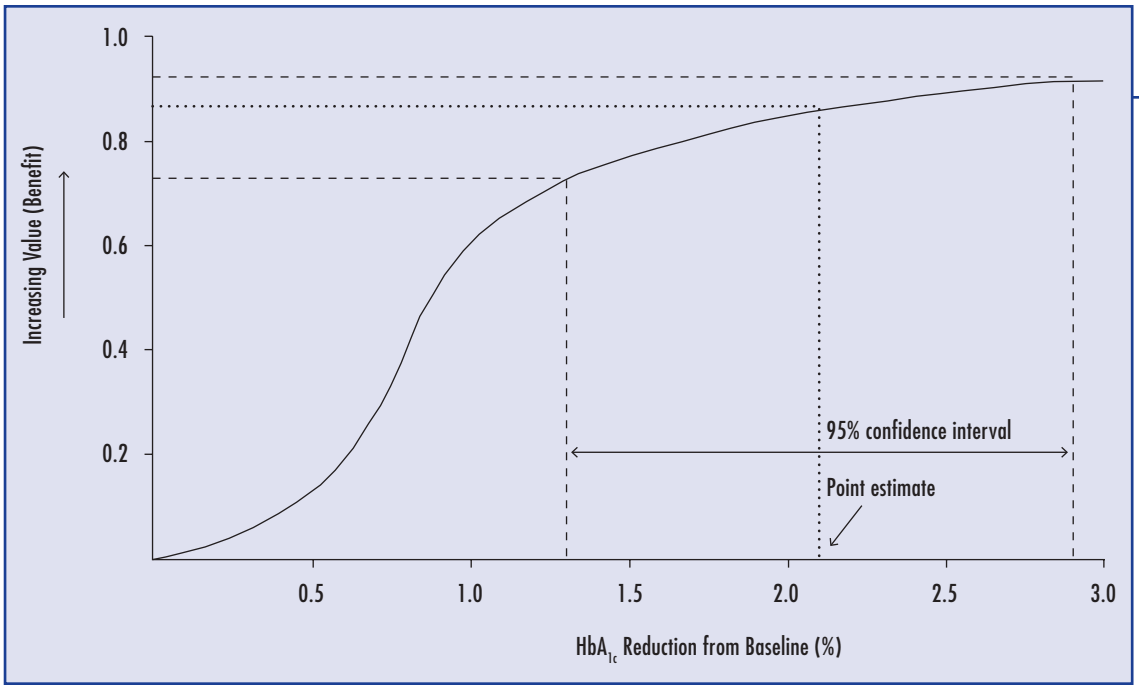

FIGURE 2

Transformation function for reduction in HbA1c. On the basis of this function, a 95% CI (1.3, 2.9) around the estimated mean reduction from baseline based on interim data translates to an interval of (0.73, 0.92) on the transformed value scale.

important efficacy subattribute of the BRAM. (Other efficacy subattributes include fasting glucose, insulin sensitivity, beta cell function, lipid profile [based on a composite of variables], and weight loss.) At its first interim analysis, the DMC observes a 2.1% mean decrease from baseline in HbA1c among patients on the experimental therapy, with a 95% CI (1.3%, 2.9%). Using the agreed-upon transformation function for HbA1c (see Figure 2), the point estimate and plausible range of outcomes for the HbA1c measure of the BRAM at this interim time point is 0.87 (0.73, 0.92) on the transformed [0,1] value scale. Point and interval estimates for the other efficacy measures are obtained similarly. Figure 3 shows a hypothetical set of outcomes for the measures of the efficacy attribute. The weights (in italics) assigned to the various measures can be used in conjunction with the point estimates to obtain an estimated overall efficacy score, which can be likewise combined with estimates of the life effects and convenience scores (hypothetical values also shown in Figure 3) to obtain a point estimate for the benefit dimension. The same is done for the risk dimension. Then a straightforward simulation exercise using repeated random sampling from all the various interval estimates of the benefit and risk measures can be used to generate a benefit-risk "cloud," or plausible region around the point estimate pair. Figure 4 graphically depicts the output generated in this fashion, for the experimental therapy as well as the standard of care control. This naturally invites the question of whether the increase in risk implied by active therapy over control is justified in view of the anticipated benefit. If one were to repeat such an analysis at several interim time points over the course of a trial, one would expect the benefit-risk coordinates to move about within their plausible ranges, and the uncertainty clouds around them to get smaller and smaller as the data used in the evaluation become more and more mature.

A DMC may elect to generate a single benefit-risk evaluation, using main and subattribute weights agreed to as a group; alternatively and indeed preferably, each DMC member may generate an individual assessment using his or her individually assigned weights. For an n-member DMC, the latter approach would lead to a graph with a set of n benefit-risk coordinates for the experimental arm as well as the control arm; appropriate regions of uncertainty can again be generated by simulation for each set of points, in this case incorporating also the observed

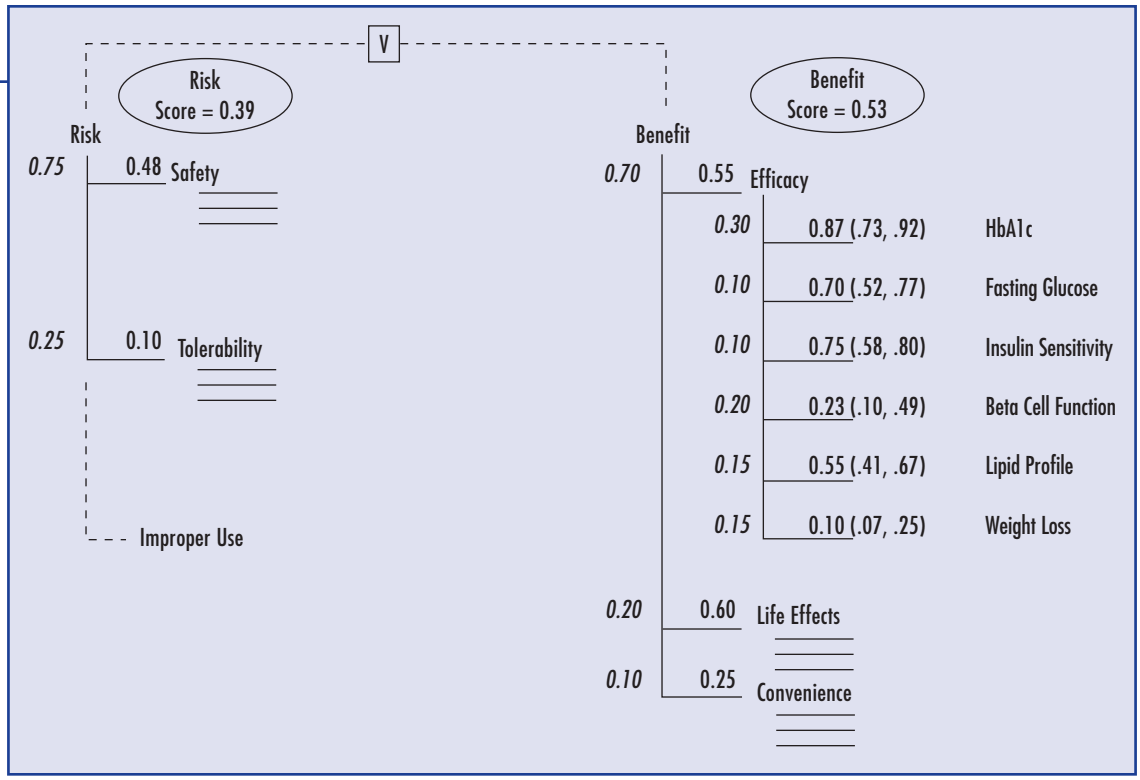

FIGURE 3

Hypothetical set of weights (in italics) and outcomes for the measures of the efficacy attribute in the BRAM for a diabetes application. The weighted average of the point estimates for the various measures yields the summary score for efficacy, and similar calculations with the hypothetical scores for the other attributes yields the overall benefit and risk scores. Simulation using repeated random sampling from all the various interval estimates of the benefit and risk measures can be used to generate a benefit-risk cloud, or plausible region around the point estimate pair.

range of weights employed by the various DMC members. Figure 5 gives an illustration. This kind of output provides substantial and readily interpretable insight into the emerging benefit-risk profile as interpreted by the DMC: the variability in assessments across DMC members within each treatment arm may be compared to the separation of the benefit-risk clouds of the two treatments overall, to help inform whether apparent treatment differences on either dimension are likely to be real or not. The extent of agreement among DMC members in their as-

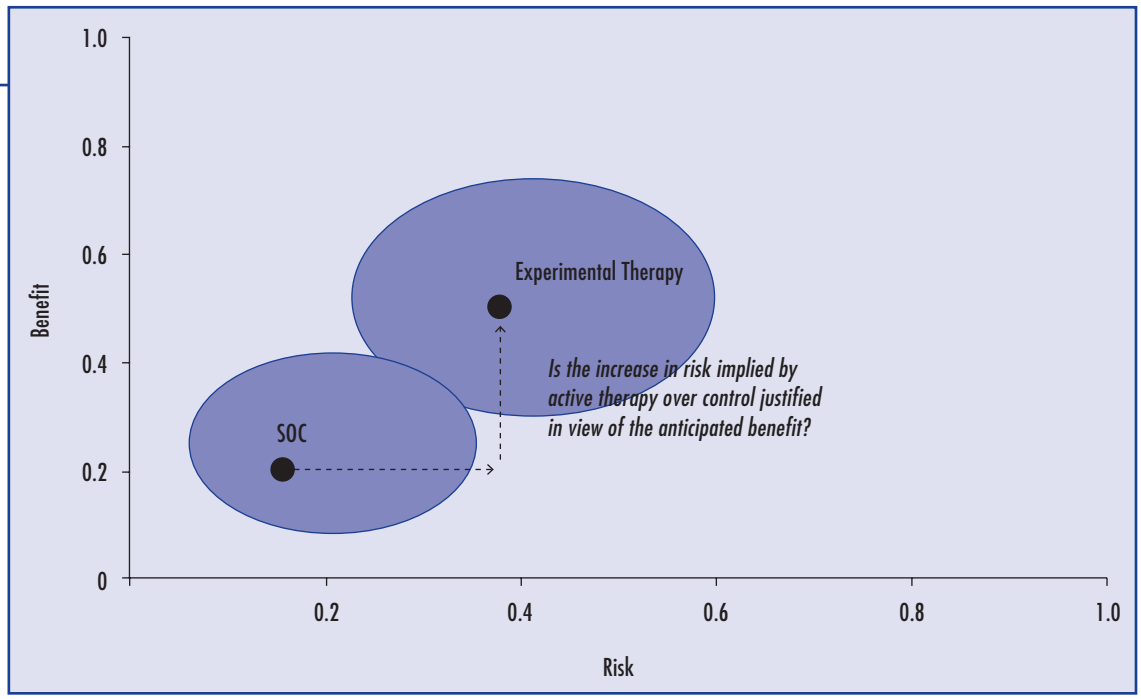

FIGURE 4

Graphical depiction of the benefit-risk assessment of a hypothetical experimental diabetes therapy versus standard of care (SOC). Shaded areas represent uncertainty regions.

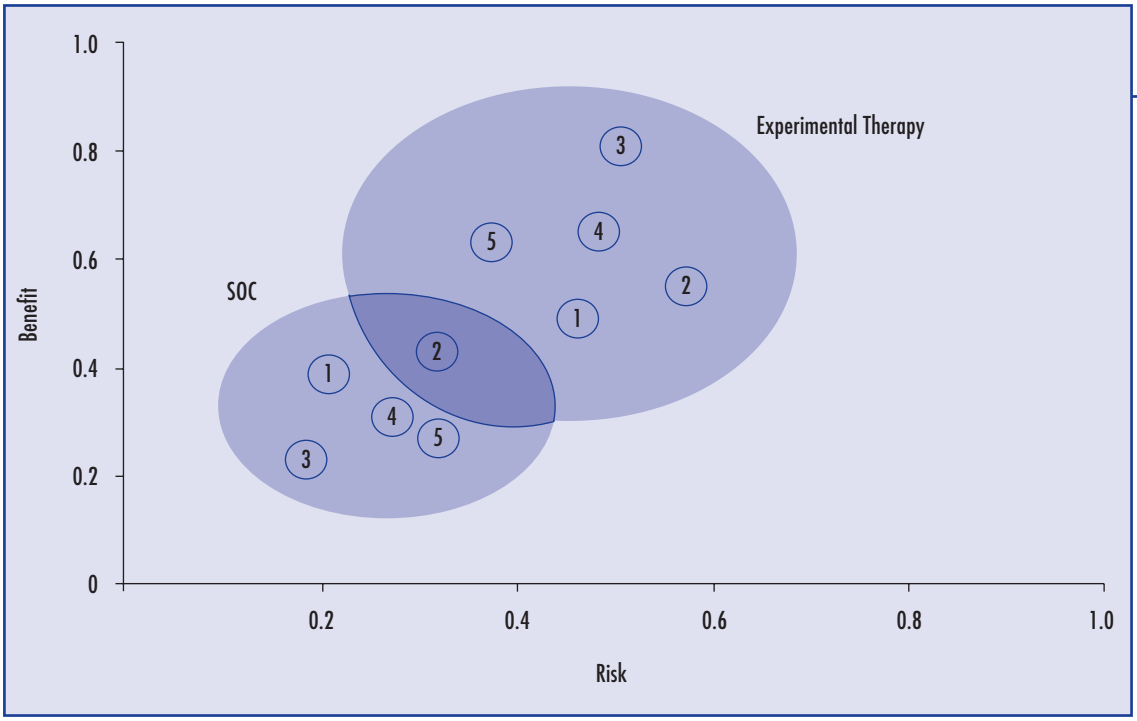

FIGURE 5

Possible benefit-risk assessment of experimental therapy versus control from a five-member DMC, with members labeled 1 through 5.

sessment of benefits and risk can thus be made apparent, which is potentially important information for the sponsor but very difficult to communicate succinctly as part of a textual recommendation. Even more light is shed on the consistency, or lack thereof, among DMC members' evaluations if the links between the individual members' assessments of control and active therapy are explicitly revealed. Thus from Figure 5 we conclude that on the whole the DMC sees a distinct benefit of the experimental therapy over control, but at the expense of an equally marked liability in terms of increased risk. The assessments of members 1 and 2 are consistent with each other in their assessment of active therapy relative to control; even though member 2 arrived at a higher risk score for both treatment arms as compared to member 1, both agree the incremental benefit of active therapy is minimal, and thus would not likely advocate continuing the study. Member 3 also concludes that there is a marked increase in risk with active therapy, but based on how he has weighted the attributes of the BRAM, he notes a significant benefit as well. This individual may struggle with his vote on whether to recommend continuing or terminating the trial. For member 4, the benefits of the experimental therapy outweigh the risks somewhat, and she may be comfortable recommending that the study continue. Finally, member 5 is the most optimistic among the DMC, noting clear benefits with minimal increase in risk. This example illustrates a particularly difficult but nonetheless plausible situation a DMC may be confronted with: a trial in which a promising benefit of the experimental therapy is offset by greater-than-anticipated risk, coupled with significantly varying assessments of the data on the part of the individual members. Formulating a recommendation to the sponsor under these conditions seems virtually impossible without an agreed-upon framework for communicating the balance of benefit and risk. The information content in a graphic such as Figure 5, were it to be shared directly with the sponsor, would ease the burden of effective communication for the DMC and serve to facilitate the ultimate decision by the sponsor to continue the trial because the benefit-risk balance of doing so appears at least minimally acceptable, or to terminate due to futility or safety concerns. We emphasize again that despite the richness inherent in this form of communicating an

emerging benefit-risk profile, it is not possible in this framework to deduce any specific trial outcome on either treatment arm. As an additional safeguard against potential back-calculation, a DMC may elect not only to utilize measure and attribute weights that are distinct from the sponsor's, but also different transformation functions that reflect the committee's own assessment of the value accorded to each observed data element, as opposed to that dictated by the sponsor's transformation functions. DMCs and sponsors should thus feel comfortable including BRAM output in their communications, knowing that study blinding is not compromised.

DISCUSSION

Although the example captured in Figure 5 may seem contrived, it illustrates two realistic and related challenges that DMCs must be prepared to face:

1. The optimal recommendation to the sponsor on whether to continue or stop an ongoing trial based on interim results may be very difficult to formulate, especially when emerging benefits appear to be offset by proportional risks.
2. Given the intentionally varied backgrounds of the members composing a DMC, it may not be realistic to achieve a consensus position among these members; yet the committee is responsible for articulating a single recommendation to the sponsor.

Regarding these points, proposing the continuation of a study with specific modifications to minimize a certain risk, say, may be helpful in arriving at a recommendation, but this is ultimately an unsatisfactory tactic for communicating a DMC's concerns regarding the emerging benefit-risk balance. In particular, it does not call out the potential legitimate disagreement among DMC members. If such lack of consensus is not reflected in some manner in the DMC's recommendation, this may leave certain members feeling very uncomfortable. Communicating to the sponsor the output of a BRAM in which the links between the individual members' assessments of control and active therapy are revealed would address this concern directly and uniquely help to shape the sponsor's interpretation of the DMC's conclusions. We suggest that the ability of the BRAM to reflect individual members' assessments, not just an overall benefit-risk evaluation from the DMC as a group, is a key advantage of this model.

Despite the fact that the output of the BRAM as described here effectively guards against study unblinding, in that no determination of specific study results is possible based on the proposed graphical representation of benefit and risk, concerns may nonetheless remain that leveraging the BRAM in DMC-sponsor communication gives the appearance of revealing too much information. Conceivably regulators as well as some DMC members themselves may retain a level of discomfort with use of the BRAM, especially in pivotal registration trials. Discussing its proposed use in a given study with key stakeholders in advance, including regulators, is therefore essential. Restrictions on when BRAM output may be shared with the sponsor should be captured in the DMC charter and may help alleviate these concerns. For instance, rather than providing a benefit-risk assessment to the sponsor as a matter of course with every DMC recommendation, sharing the BRAM may be intentionally reserved just for cases in which achieving consensus or recommending a path forward is especially difficult. That said, a DMC may find that routinely leveraging the BRAM just for its own internal use may be a welcome help in formulating recommendations—especially if the committee and the sponsor have taken sufficient time to discuss the structure of the model, as well as actions based on various scenarios that might arise, in advance. Agreement should also be reached on how to incorporate new findings. For instance, the BRAM must be able to take into account an unexpected side effect that appears for the first time in the trial in question. This can be done by leaving one of the measures of the safety attribute as an unspecified placeholder, intended to capture such unexpected findings. Depending on the seriousness of the finding, DMC members may then accord more or less weight to this measure. This constitutes a legitimate excep-

tion to the general mandate to settle on a set of weights in advance, and not change these based on observed results.

Use of the BRAM as a tool for effective communication between a DMC and the sponsor of a study has certain practical limitations that warrant discussion. Implementation of this approach requires significant cooperation and effort from members of the study team, the DMC, and the statistical analysis group supporting the DMC. Agreeing on the structure of the model may be a nontrivial exercise (yet insightful in terms of revealing the varying perspectives of the DMC and sponsor). Getting the DMC and the (potentially independent) statistician supporting the committee comfortable with working with the model will also require an investment of time spent in education. The former will need to be at ease as model users, whereas the supporting statistician will have the added responsibility of actually running the model based on members' assessments, including the simulations required to generate the uncertainty clouds. On this latter point, an accurate reflection of the uncertainty in benefit-risk assessments is crucial in helping to determine whether apparent differences in benefit and risk are likely to be real or not. The extent of this uncertainty is determined on the one hand by the differences across DMC members in weights assigned to the BRAM attributes, but to a large extent it is driven by the variability in the observed data informing the BRAM. As indicated above, in some cases uncertainty in observed BRAM measures may be reflected directly by using confidence intervals of the appropriate outcomes. In other cases, establishing a range of uncertainty may be less straightforward and subject to debate. The study team statistician should help the DMC think through appropriate definitions of ranges on outcomes when an obvious solution is not available.

The feasibility of this approach itself thus poses the question of whether the effort required to implement it is warranted in view of the expected gains. Large, long-term trials for which numerous interim analyses are planned appear to hold the greatest potential for realizing the benefits of having the possibility of leveraging a BRAM in DMC-sponsor communication. The promise of facilitating a DMC discharging its duties by offering a vehicle for communicating important, complex facets of the committee's deliberations, and facilitating the sponsor's ultimate decision making based on clearer, more insightful recommendations suggests that it may be worth piloting this model.

REFERENCES

1. Ellenberg S, Fleming T, DeMets D. *Data Monitoring Committees in Clinical Trials: A Practical Perspective*. West Sussex, UK: Wiley; 2002.
2. CHMP. Guideline on data monitoring committees. 2005. http://www.emea.europa.eu/pdfs/human/ewp/587203en.pdf (accessed February 3, 2011).
3. Food and Drug Administration. Guidance for clinical trial sponsors: establishment and operation of clinical trial data monitoring committees. 2002. http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm127073.pdf (accessed February 3, 2011).
4. DeMets DL, Pocock SJ, Julian DG. The agonizing negative trend in monitoring of clinical trials. *Lancet*. 1999;354:1983–1988.
5. Sashegyi AI. Role of the data monitoring committee and sponsor in monitoring adaptive designs. *Pharm Outsourcing*. 2010;11:28–31.
6. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines—part 1: the development of a new model using multi-criteria decision analysis. *Pharmacoepidemiol Drug Saf*. 2007;16:S2–S15.
7. Felli JC, Noel RA, Cavazzoni PA. A multiattribute model for evaluating the benefit-risk profiles of treatment alternatives. *Med Decis Making*. 2009; 29:105–115.

The author has disclosed that he has received grants/research support as well as other financial material support from Eli Lilly and Company, where he is also a stock shareholder.