VALUE IN HEALTH 18 (2015) 250–259

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/jval

A Unified Framework for Classification of Methods for Benefit-Risk Assessment

Mehdi Najafzadeh, PhD[1,], Sebastian Schneeweiss, MD, ScD[1], Niteesh Choudhry, MD, PhD[1], Katsiaryna Bykov, PharmD, MS[1], Kristijan H. Kahler, RPh, PhD[2], Diane P. Martin, PhD[2], Joshua J. Gagne, PharmD, ScD[1]*

[1]*Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA;* [2]*Novartis Pharmaceutical Corporation, East Hanover, NJ, USA*

ABSTRACT

Background: Patients, physicians, and other decision makers make implicit but inevitable trade-offs among risks and benefits of treatments. Many methods have been proposed to promote transparent and rigorous benefit-risk analysis (BRA). **Objective:** To propose a framework for classifying BRA methods on the basis of key factors that matter most for patients by using a common mathematical notation and compare their results using a hypothetical example. **Methods:** We classified the available BRA methods into three categories: 1) unweighted metrics, which use only probabilities of benefits and risks; 2) metrics that incorporate preference weights and that account for the impact and duration of benefits and risks; and 3) metrics that incorporate weights based on decision makers' opinions. We used two hypothetical antiplatelet drugs (*a* and *b*) to compare the BRA methods within our proposed framework. **Results:** Unweighted metrics include the number needed to treat and the number needed to harm. Metrics that incorporate preference weights include those that use maximum acceptable risk, those that use relative-value–adjusted life-years, and those that use quality-adjusted life-years. Metrics that use decision makers' weights include the multicriteria decision analysis, the benefit-less-risk analysis, Boers' 3 by 3 table, the Gail/NCI method, and the transparent uniform risk benefit overview. Most BRA methods can be derived as a special case of a generalized formula in which some are mathematically identical. Numerical comparison of methods highlights potential differences in BRA results and their interpretation. **Conclusions:** The proposed framework provides a unified, patient-centered approach to BRA methods classification based on the types of weights that are used across existing methods, a key differentiating feature.

Keywords: benefit-risk analysis, harm-benefit analysis.

Introduction

Benefit-risk analysis (BRA) is a formal approach for evaluating the balance between benefits and risks of drugs or other health care interventions. A BRA involves synthesizing the existing evidence about the benefits and safety of treatments from different data sources. Data generated from rigorously designed pre- and post-marketing studies and formal statistical analysis of those data within the frequentist paradigm usually constitute the evidence for risks and benefits of medical treatments. The actual decisions made by regulators, physicians, or patients about the overall risk-benefit balance, however, rely on subjective judgments and require making implicit but inevitable trade-offs among different risks and benefits of treatments. These judgments often have a strong influence on the decisions and yet they are prone to considerable ambiguity and potential bias. There is a consensus that more transparent and rigorous methods for BRA are needed [1–4].

Several recent reviews have discussed the advantages and disadvantages of more than a dozen BRA methods [2–4]. Choosing an appropriate BRA method for a given clinical question is not straightforward. In this article, we evaluate and compare different BRA methods on the basis of key factors of BRA that matter most for patients [5]. This patient-centered perspective yields a simple, unified framework for classifying BRA methods. We formulate existing BRA methods within this framework using common mathematical notation and use a simple hypothetical example to illustrate their use with this notation.

Methods

Framework

Previous efforts for comparisons of BRA methods have generally focused on technical aspects including type of data used, analytical methodology, ease of implementation, and ability to communicate or present the results to compare different BRA

* *Address correspondence to:* Mehdi Najafzadeh, Department of Medicine, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA.

E-mail: mnajafzadeh@partners.org.

1098-3015$36.00 – see front matter

http://dx.doi.org/10.1016/j.jval.2014.11.001

methods [3,4,6]. Although these criteria are important, they result in classes of methods that share more core commonality than previously acknowledged.

From the patient's perspective, a BRA should be able to measure the impact of treatment on expected quality of life. This is possible only if the method can account for 1) probabilities of different treatment outcomes and 2) the magnitude and duration of impact of different treatment outcomes. The impact of treatment on patient quality of life is not limited to clinical outcomes (i.e., efficacy or adverse effects); factors related to the treatment process (e.g., oral vs. injection route of administration) should be ideally incorporated into the BRA. By considering a patient-centered perspective, BRA methods can be classified and compared according to the extent to which they incorporate each of these aspects. Because all BRA methods fulfill the first criterion by incorporating outcome probabilities, we propose classifying methods in the following way according to whether and how they account for the impact of these outcomes on the patient.

Unweighted metrics

These methods only use probabilities of outcomes (both risks and benefits) or a functional form of those probabilities in the analysis. Neither the extent nor the duration of impact is considered.

Metrics that incorporate patient-derived preference weights

In addition to probabilities of outcomes, methods that incorporate patient-derived preference weights can account for the impact of different outcomes on quality of life. Weights can reflect patients' preferences for different treatment-related outcomes and processes. In addition, some preference-weighted BRA methods are able to incorporate the duration and timing of different events and therefore can provide a more accurate estimate of overall treatment impact on patients' quality of life.

Metrics that incorporate weights from other perspectives

These methods use either a decision maker's or other expert's opinion to derive a scheme to weigh different treatment outcomes and processes. These methods can implicitly account for the perceived impacts of different outcomes (i.e., risks and benefits) on patients' quality of life.

Mathematical Notation

We use the following notation to formulate different BRA methods within the proposed framework. This common notation facilitates comparisons among different BRA methods and illustrates the inherent similarity of methods within and across each of the three proposed classes:

- p_{it} is the probability of experiencing a beneficial outcome i with treatment t. This is expressed as the probability of having the beneficial complement of a binary outcome. For example, if the benefit is survival, p_{it} is the probability of surviving with treatment t. If the benefit is reduction in myocardial infarction (MI), p_{it} is the probability of *not* having an MI with treatment t.
- q_{jt} is the probability of experiencing a harmful outcome j with treatment t. This is expressed as the probability of having the harmful complement of a binary outcome. For example, if the adverse event is mortality, q_{jt} is the probability of dying with treatment t. If the harm is increase in major bleeding, then q_{jt} is the probability of experiencing a major bleeding event with treatment t.
- Δp_i is the difference in probabilities (or risk difference) of the beneficial outcome i with treatment a versus treatment b; that is, $\Delta p_i = p_{ia} - p_{ib}$.
- Δq_j is the difference in probabilities (or risk difference) of the harmful outcome j with treatment a versus treatment b; that is, $\Delta q_j = q_{ia} - q_{ib}$.
- μ is a coefficient that represents the decision maker's threshold for the relative trade-off between benefits and risks. The value $\mu = 1$ means that benefits and risks are equally important to the decision maker, whereas $\mu > 1$ indicates that harms are more important. For simplicity, we assume $\mu = 1$ in the hypothetical example.

Using this notation, we propose the following general formula to unify a large family of existing benefit-risk methods:

$$\mathrm{INB} = \sum_{i=1}^{I} v_i T_i \Delta p_i - \mu \sum_{j=1}^{J} v_j T_j \Delta q_j \quad (1)$$

where v_i and T_i are the weights and duration of impact of outcome i, respectively, on patients' quality of life.

Hypothetical Example

We use antiplatelet therapy for acute coronary syndrome to illustrate the implementation of different BRA methods within the proposed framework and using the above notation. The benefits of antiplatelet therapy in reducing the risk of MI and thromboembolism have been shown in several large clinical trials [7–11]. Antiplatelet therapy, however, is also associated with increased risk of bleeding. Different antiplatelet agents can have different effects on ischemic and bleeding events. This exemplifies a situation in which using a BRA method can help patients and physicians optimize their treatment decision. We consider four relevant clinical outcomes: MI, stroke, minor bleeding, and major bleeding. Table 1 presents the probabilities of each outcome for two hypothetical antiplatelet drugs. In Table 2, we present the preference weights for each health state, measured using several different approaches. These preference weights can reflect utilities measured using standard gamble (SG), time trade-off (TTO), or indirect (e.g., EuroQol five-dimensional questionnaire, six-dimensional health state short form [derived from short-form 36 health survey], and health utilities index 3) methods. We have assigned EuroQol five-dimensional questionnaire utilities to each health state on the basis of published studies [12–15] and present preference weights elicited using a discrete choice experiment (DCE) [16]. We also provide assumptions about the duration of each health state.

Results

Review and Comparison of BRA Methods

In this section, we provide general mathematical formulations for different BRA methods within each of the three proposed categories and based on the unified equation (Equation 1). These equations can be extended to incorporate more details (e.g., more granular health states and timing of events) when more data are available. We also use the hypothetical example to demonstrate calculations using each BRA method. We provide further details about the BRA methods in Table 3.

Unweighted Metrics

Number needed to treat and number needed to harm

Number needed to treat (NNT) represents the number of patients that need to be treated in order for one of those patients to benefit from treatment. Number needed to harm (NNH) represents the number of patients that need to be treated in order for

Table 1 – Assumptions about probabilities of different events in the hypothetical antiplatelet example.

Event	Probability (per year)*		p_{ia}†	p_{ib}	Δp_i	q_{ja}‡	q_{jb}	Δq_i	NNT_i	NNH_i
	Treatment *a*	Treatment *b*								
Myocardial infarction	5.0%	7.0%	95.0%	93.0%	2.0%	–	–	–	50	–
Stroke	1.2%	1.4%	98.8%	98.6%	0.2%	–	–	–	500	–
Minor bleeding	5.0%	2.5%	–	–	–	5%	2.5%	2.5%	–	40
Major bleeding	3.5%	2.5%	–	–	–	3.5%	2.5%	1%	–	100

NNH, number needed to harm; NNT, number needed to treat.

* The rates are based on the outcomes of several large clinical trials [7–11].

† p_{it} is the probability of experiencing a beneficial outcome i with treatment t. Δp_i is the difference in probabilities (or risk difference) of the beneficial outcome i with treatment a vs. treatment b; i.e., $\Delta p_i = p_{ia} - p_{ib}$.

‡ q_{jt} is the probability of experiencing a harmful outcome j with treatment t. Δq_j is the difference in probabilities (or risk difference) of the harmful outcome j with treatment a vs. treatment b; i.e., $\Delta q_j = q_{ja} - q_{jb}$.

one of them to be harmed from adverse effects of the treatment. NNT and NNH can be calculated as follows:

$$NNT_i = \frac{1}{\Delta p_i}; \quad NNH_j = \frac{1}{\Delta q_j} \tag{2}$$

Because both NNT and NNH are counts, they can be directly contrasted and the criterion for determining the net favorability of treatment t assuming a benefit i and risk j can be defined as follows:

$$NNT|NNH = \sum_{i=1}^{I} \frac{1}{NNT_i} - \mu \sum_{j=1}^{J} \frac{1}{NNH_j} \tag{3}$$

Note that NNT and NNH are inverse risk differences. Therefore, this criterion can be alternatively formulated as follows:

$$NNT|NNH = \sum_{i=1}^{I} \Delta p_i - \mu \sum_{j=1}^{J} \Delta q_j \tag{4}$$

This criterion is a special case of the general formula (Equation 1) in which the weights, v_i, and duration of impact, T_i, of different outcomes are assumed to be 1.

Minimum clinical efficacy

Minimum clinical efficacy is mathematically equivalent to the NNT\NNH method. We prove this by deriving the standard minimum clinical efficacy equation from Equation 4 in Supplemental Materials found at http://dx.doi.org/10.1016/j.jval.2014.11.001.

Metrics That Incorporate Preference Weights

Relative-value–adjusted NNT\NNH

Relative-value–adjusted NNT\NNH uses the NNT\NNH framework but incorporates patients' valuation about intensity of risk and benefit [17].

For a given benefit, this relative value is expressed in the form of RV_j:

$$RV_j = \frac{1 - u_j}{1 - u_{base}} \tag{5}$$

where u_j is the preference weight for a health state associated with a harmful outcome j and u_{base} is the preference weight for the health state in the absence of the beneficial outcome. That is, if the beneficial outcome is MI reduction, u_{base} represents the health state associated with having an MI.

Table 2 – Assumptions about preference weights associated with four health states derived from different approaches in the hypothetical antiplatelet example.

Health state	Utility weights* (HRQOL)	Preference weights† (DCE)	Preference weights (DCE)	Weights using the MCDA method (%)‡	Scores for the MCDA method‡	Duration of event impact§
	u	v	$r = v/\Delta p_i$ or Δq_j	w	$\Delta S = S_a - S_b$	T
Myocardial infarction	0.84 (12)	0.18	0.09	30	5	Lifetime
Stroke	0.6 (13)	0.014	0.07	40	2	Lifetime
Minor bleeding	0.8 (15)*	0.025	0.01	5	1	1 mo
Major bleeding	0.46 (14)*	0.03	0.03	25	3	1 mo

DCE, discrete choice experiment; EQ-5D, EuroQol five-dimensional questionnaire; HRQOL, health-related quality of life; MCDA, multicriteria decision analysis.

* The utility weights were multiplied by the duration of event impact to calculate the overall effect in terms of quality-adjusted life-years. As such, the overall impact of MI and stroke is larger than that of minor and major bleeding, which has short duration. All utility weights are EQ-5D utilities measured in independent studies [12–15].

† These preference weights are derived from a DCE in patients with cardiovascular diseases [16].

‡ Weights and scores for MCDA are based on assumptions.

§ Duration of events is based on our assumptions.

Table 3 – Classification and comparison of existing benefit-risk assessment (BRA) methods.

BRA methods	Able to incorporate patients' preferences?	Independent BRA method?*	Can statistical uncertainty be incorporated?	Comments	Result based on our example
Unweighted metrics					
Number needed to treat and number needed to harm (NNT|NNH)	No	Yes	Yes	Equivalent to using inverse of risk differences	−1.3% net probability of benefit if treated with *a* vs. *b*
Minimum clinical efficacy (MCE)	No	Yes	Yes	Mathematically equivalent to NNT|NNH and risk difference methods	−1.3% net probability of benefit if treated with *a* vs. *b*
Metrics that incorporate patient-derived preference weights					
Relative value–adjusted number needed to treat (RV-NNT)	Yes	Yes	Yes	Multiple harms or multiple benefits can be included, but not simultaneously	−4.5% net probability of benefit if treated with *a* vs. *b*
Maximum acceptable risk (MAR)	Yes	Yes	Yes	Requires estimation using stated preference methods (e.g., discrete choice experiment)	4.64% net probability of benefit if treated with *a* vs. *b*
Incremental net health benefit (INB) based on relative-value–adjusted life-years (RVALYs)	Yes	Yes	Yes	Requires estimation using stated preference methods (e.g., discrete choice experiment)	3.83 RVALYs gained per 100 patients treated with *a* vs. *b*
INB based on quality-adjusted life-years (QALYs)	Yes	Yes	Yes	Requires estimates of preference weights of different health states using generic utility measures (e.g., standard gamble, time trade-off, EQ-5D, SF-6D, and HUI)	5.43 QALYs gained per 100 patients treated with *a* vs. *b*
Metrics that incorporate weights from other perspectives					
Multicriteria decision analysis (MCDA)	Potentially yes, if patients determine criteria, scores, and weights	Yes	Potentially yes, but most MCDAs are deterministic	Requires decision makers' agreement on criteria, scores, and weights	1.5 is a dimensionless score, only suggesting that *a* is preferred to *b*
Benefit-less-risk analysis (BLRA)	Potentially yes	Yes	No	Conceptually similar to the MCDA method	–
Gail-NCI method	Potentially yes	Yes	No	Requires assigning weights to the three categories of outcomes	–
Transparent uniform risk benefit overview (TURBO)	Potentially yes	Yes	No	Focuses on documenting qualitative discussions and judgments that take place in a regulatory decision-making process. Limited to the inclusion of two risks and two benefits	–
Boers' 3 by 3 table	Potentially yes	Yes	No	Results are in the form of a table rather than a single number	–

continued on next page

Table 3 – *continued*

BRA methods	Able to incorporate patients' preferences?	Independent BRA method?*	Can statistical uncertainty be incorporated?	Comments	Result based on our example
Other approaches cited in the literature					
Quantitative framework for risk and benefit assessment (QFRBA)	NA	No	NA	Suggests using standard epidemiological measures (e.g., risk ratios, odds ratios, and risk differences) for BRA. Therefore, NNT|NNH and MCE are examples of QFRBA	–
(Quality-adjusted) time without symptoms and toxicity (TWiST or Q-TWiST)	Yes	Yes	No	Same as INB using QALYs as the outcome metric	–
Net clinical benefit	NA	No	NA	Can be classified in any of the three families of metrics depending on the metric that is used to quantify risks and benefits	–
Probabilistic simulation model (PSM)	NA	No	NA	An approach to incorporate uncertainty in any method that is selected for BRA. For example, one can use INB using risk difference, QALYs, or RVALYs, and then PSM to quantify the degree of uncertainty in estimated INB	-
Risk-benefit contour (RBC)	Yes	Yes	Potentially yes	A method to elicit individuals' risk preferences and cannot be directly used for BRA analysis	–
Risk-benefit plane (RBP) and risk-benefit acceptability threshold (RBAT)	NA	No	NA	An approach for presenting BRA results. One can conduct a BRA using NNT|NNH, QALYs, or RVALYs and then present the incremental risk and incremental benefit in an RBP	–

EQ-5D, EuroQol five-dimensional questionnaire; HUI, health utilities index; NCI, National Cancer Institute; SF-6D, six-dimensional health state short form (derived from short-form 36 health survey).

* Whether this is an explicit and distinct approach to quantify and compare benefits and risks and therefore can be considered as an independent BRA method *per se*. For example, Monte Carlo simulation is a method to quantify uncertainty in general terms rather than being a standalone and independent BRA method.

Thus, assuming that patients attain or remain in their state of perfect health as a result of the treatment, the magnitude of treatment effect on quality of life is $1 - u_{base}$. The RV_j ratio indicates how patients compare the reduction in health associated with a harmful outcome (i.e., $1 - u_j$) to the gain in health associated with experiencing the beneficial outcome $1 - u_{base}$. These preference weights can be determined using standard health-related quality-of-life measures (e.g., SG and TTO). Once we have the RV_j for a particular harm j and benefit pair, we can define the relative-value–adjusted NNH (NNT|RV_NNH) as follows:

$$\text{NNT|RV\_NNH} = \frac{1}{\sum_{j=1}^{J}\left(\text{RV}_j \times \Delta q_j\right)} \tag{6}$$

Note that we can include several harmful outcomes in the formulation of RV_NNH. This approach, however, accommodates only a single beneficial outcome. We can then compare the RV_NNH to the NNT to determine net favorability:

$$\text{NNT|RV\_NNH} = \frac{1}{\text{NNT}} - \mu\frac{1}{\text{RV\_NNH}} \tag{7}$$

After replacing NNT and RV_NNT definitions, we have:

$$\text{NNT|RV\_NNH} = \Delta p - \mu\sum_{j=1}^{J}\left(\text{RV}_j \times \Delta q_j\right) \tag{8}$$

Note that this is a special case of the general formula (Equation 1)

in that a single beneficial outcome is compared with multiple harmful outcomes and using the RV_j as weights, $v_j \times T_j = RV_j$

Alternatively, we could use an RV_NNT and compare it to the unweight NNH. This would allow the incorporation of multiple beneficial outcomes but would be limited to only a single harmful outcome. This approach is not amenable to the simultaneous inclusion of multiple benefits and harms.

Incremental net benefit using maximum acceptable risk

Unlike the methods described above, the incremental net benefit (INB) approach provides a flexible framework that simultaneously accommodates probabilities and weights for multiple harms and benefits. Formulated in a unified way, the weights can be derived using any method (e.g., SG and DCE). Maximum acceptable risk (MAR) indicates patients' willingness to trade off harmful and beneficial outcomes of a treatment [18] and requires eliciting patients' preference weights using stated preference methods such as DCE. Assume that a DCE has been conducted in a representative patient population and v_i is the estimated preference weight for a given change in probabilities of having a beneficial outcome i (denoted by Δp_i), and v_j is the estimated preference weight for a given change in probabilities of having a harmful outcome j (denoted by Δq_i). MAR_{ij}, which is the maximum probability of a harmful event j that individuals are willing to accept in exchange for a 1% increase in the probability of achieving a beneficial outcome i, can be defined as follows:

$$MAR_{ij} = \frac{v_i/\Delta p_i}{v_j/\Delta q_j} \times \frac{1}{100} \tag{9}$$

A separate MAR is calculated for each outcome in relation to one of the outcomes (either harmful or beneficial) that is selected as the referent. These MARs can then be used to generate a single measure of INB:

$$\text{INB} = \sum_{i=1}^{I} MAR_i \Delta p_i - \mu \sum_{j=1}^{J} MAR_j \Delta q_j \tag{10}$$

This is a special case of the general formula (Equation 1) using MAR as weights ($v_i = MAR_i$) and assuming that the durations of impact are all equal to 1.

INB using relative-value–adjusted life-years

Individuals' relative preferences for different treatment outcomes obtained from a DCE can be used to calculate INB in terms of relative-value–adjusted life-years (RVALYs) [19]. Assume that r_j is the estimated preference weight for a 1% change in probability of having a harmful outcome j; r_i is the estimated preference weight for a 1% change in probability of having a beneficial outcome i; T_i is the duration of effect of the beneficial outcome i; and T_j is the duration of effect of the harmful outcome j. The INB can then be defined as follows:

$$\text{INB}_{\text{RVALYs}} = \sum_{i=1}^{I} r_i T_i \Delta p_i - \mu \sum_{j=1}^{J} r_j T_j \Delta q_j \tag{11}$$

where $r_j T_j$ and $r_i T_i$ are RVALYs gained or lost when experiencing beneficial or harmful events, respectively. This is a special case of the general formula (Equation 1) assuming r_i as weights ($v_i = r_i$).

Preference estimates using choice experiments can evaluate aspects of health care interventions that otherwise cannot be captured using health-related quality-of-life measures [2]. For instance, the use of oral medications instead of injections often positively impacts patients' quality of life. Yet this cannot be captured adequately using standard health-related quality-of-life measures. Including treatment modality as an attribute in a DCE can quantify the impact of this aspect on patients' preferences and their willingness to trade off modality against efficacy or safety outcomes can be directly estimated.

INB using quality-adjusted life-years

Assume that u_i is the utility of a given health state after treatment measured (directly or indirectly) using SG or TTO. Also, assume that u_{base} is the baseline utility of patients before starting the treatment; T_i is the duration of effect of the beneficial outcome i; and T_j is the duration of effect of the harmful outcome j. The INB in terms of QALYs can then be defined as follows:

$$\text{INB}_{\text{QALY}} = \sum_{i=1}^{I} (u_{base} - u_i) T_i \Delta p_i - \mu \sum_{j=1}^{J} (u_{base} - u_j) T_j \Delta q_j \tag{12}$$

where $(u_{base} - u_i)T_i$ are the QALYs gained when experiencing a beneficial event and $(u_{base} - u_j)T_j$ are the QALYs lost when experiencing a harmful event.

This is a special case of the general formula (Equation 1) assuming $(u_{base} - u_i)$ as weights ($v_i = u_{base} - u_i$).

Health-related quality-of-life measures have been validated and widely used for different health conditions, particularly in cost-effectiveness studies. All techniques that are available for the estimation and presentation of cost-effectiveness studies can be directly adapted for BRA. A shortcoming of this approach, however, is that unlike relative values obtained using DCE, it does not consider the trade-offs that patients are willing to make among different outcomes and it cannot capture non–health-related outcomes (e.g., invasiveness of treatment).

Metrics That Incorporate Weights from Other Perspectives

The previous set of methods incorporate weights on the basis of patient preferences. In particular, the INB formula is a unified approach that enables the integration of various kinds of preference weights. In the following section, we present a diverse set of methods that incorporate weights that are often subjectively determined from other perspectives.

Multicriteria decision analysis

Multicriteria decision analysis (MCDA) is a powerful tool that can be used to facilitate joint decision making during group discussions. MCDA can be used to organize different elements of a complex decision, provide a systematic and transparent assessment of the problem, and document the thought process leading to a decision in a tractable format [20]. The first step of an MCDA is to establish the decision context that involves identifying key decision makers, defining the treatment options of interest, and discussing the aims and context of the problem. The group then defines objectives and criteria pertinent to the problem and formulates the problem as a decision tree by organizing the objectives and criteria in a hierarchy. Knowledge from clinical trials and safety studies can be used to identify important benefits and risks that need to be included in the decision tree. For example, a decision tree for BRA can consist of two branches assigned to benefits and risks at the first level, sources of data or pivotal clinical trials at the second level, and types of beneficial or harmful events at the third level of hierarchy. Decision makers then discuss the scores that should be given to each treatment option by considering how they perform on the basis of those certain criteria. For example, if a criterion is incidence of MI, the probability of MI given a treatment is implicitly considered in the scoring of that treatment. The group then agrees on weights for each criterion that reflect their perceptions about its relative importance compared with the other criteria. Finally, by calculating weighted scores for each criterion in the decision tree, an overall weighted score can be estimated that summarizes participants' judgments about the net benefit of the treatment options. The results of MCDA are in numerical format, and sensitivity analysis can be performed to explore the effect of different scores and weights on the final evaluation. Nevertheless, it should be noted that scores and weights are determined on the basis of decision makers' judgments and

perceptions about different criteria and the underlying trade-offs among different benefits and risks. MCDA can incorporate multiple benefits and risks in the BRA and can account for qualitative judgment and uncertainty of the decision.

During the conduct of an MCDA, the group agrees on scores (S_{it}) and weights (w_i) for each treatment option criterion. The calculation of the final weighted score can be interpreted as an INB of treatment a versus treatment b and using the formula

$$\mathrm{INB}_{\mathrm{MCDA}} = \sum_{i=1}^{I} w_i \Delta S_i - \mu \sum_{j=1}^{J} w_j \Delta S_j \tag{13}$$

where $\Delta S_i = S_{ia} - S_{ib}$, $\Delta S_j = S_{ja} - S_{jb}$, and $\sum w = 100\%$.

Although the criteria might represent any number of features of the treatments, they can be used to describe the benefits and risks. Although this approach has some similarities with the general formula ((Equation 1), it cannot be directly derived from the general formula. For example, ΔS_i reflect a group's judgment about the magnitude of risk differences (Δp_i) and their relevant importance (w_i).

Benefit-Less-Risk Analysis

In benefit-less-risk analysis (BLRA), important harmful outcomes associated with the treatment are identified [21]. Each patient is then assigned a score between 0 and 3, where 0 indicates experiencing the minimum and 3 indicates the maximum level of harm (e.g., severity) for each outcome. Weights are assigned to each harmful outcome on the basis of the decision maker's perceived weight for each of those outcomes. Outcome frequencies, scores, and weights are then combined to calculate a weighted score that reflects the overall risk for each treatment option. The original BLRA method was developed to include only one fixed beneficial outcome that is compared against the calculated overall weighted risk. A parallel approach, however, can be used to summarize multiple benefits. The strength of BLRA is that it uses patient-level data to build an aggregate measure of INB. As with MCDA, the calculated quantities are based on decision makers' selection of scores and weights, rather than on patient preferences.

Boers' 3 by 3 Table

In this method, the investigator creates a table with three rows for different severities of risk and three columns for different levels of benefits [22]. Similar to BLRA, this method requires patient-level data. Each patient is categorized into one of the nine cells on the basis of the severity of the observed treatment outcomes. Counts in the table cells can then be transformed to percentages. Outcome severity can be based on physician judgment or patient opinion. Using physician judgment to assign patients into different levels of risks and benefits requires making implicit and often subjective assumptions when comparing and weighting multiple risks and benefits. Finally, the results are in the form of a table rather than a single number.

Gail/NCI Method

Gail's method emphasizes the estimation of events within strata defined by patient characteristics (e.g., age, sex, and race) [23]. All harmful and beneficial outcomes are divided into three categories on the basis of severity of their expected impact: life-threatening, severe, and other (i.e., neither life-threatening nor severe). The probabilities of different events within each category (both beneficial and harmful events) are then added to calculate the net risk of life-threatening, severe, and other outcomes. A weight is assigned to each of these three categories on the basis of the perceived impact on the patients and a summary score is calculated. Gail's method was originally developed to estimate risk-benefit balance on a patient level by taking into account patients' personal characteristics and relative weights for each of the three categories of outcomes. It can be extended, however, to estimate benefit-risk balance on the population level as well. Gails' method can be formulated as follows:

$$I(w_1,w_2,w_3) = w_1\left(\sum_{\text{life-threatening}} \Delta p_i - \sum_{\text{life-threatening}} \Delta q_j\right) + w_2\left(\sum_{\text{severe}} \Delta p_i - \sum_{\text{severe}} \Delta q_j\right) + w_3\left(\sum_{\text{other}} \Delta p_i - \sum_{\text{other}} \Delta q_j\right) \tag{14}$$

where w_1, w_2, and w_3 are the weights that are assigned to life-threatening, severe, and other outcomes, respectively.

Transparent Uniform Risk Benefit Overview

Transparent uniform risk benefit overview (TURBO) considers only two harmful outcomes and assigns a score between 1 and 5 to the more important risk and a score of 1 or 2 to the other, comprising an R factor (i.e., "Risk factor") of between 1 and 7 [24]. A B factor (i.e., "Benefit factor") is similarly calculated by considering the most and the second most important benefits. Finally, a T score that ranges from 1 to 7 is assigned to each patient by subtracting his or her R factor from his or her B factor. The T scores are based on the decision maker's perception about the frequency and intensity of risks or benefits. For example, $T = 1$ indicates that treatment a, when compared with treatment b, is associated with a large risk and a small benefit, whereas $T = 7$ indicates a relatively small risk and large benefit. TURBO is an approach to document the qualitative discussions and judgments that take place in a regulatory decision-making process.

Comparing BRA Results Using Hypothetical Example

Unweighted Metrics

NNT and NNH

In Table 1, we present NNT (for MI and stroke reduction) and NNH (for major and minor bleeding) calculated using Equation 2 for the antiplatelet example. Using Equation 3 with $\mu = 1$, we get:

$$\mathrm{NNT|NNH} = \sum_{i=1}^{I} \frac{1}{\mathrm{NNT}_i} - \mu \sum_{j=1}^{J} \frac{1}{\mathrm{NNH}_j} = \left(\frac{1}{50} + \frac{1}{500}\right) - \left(\frac{1}{40} + \frac{1}{100}\right) = 2.2\% - 3.5\% = -1.3\%$$

Using risk, the difference formulation provides identical results:

$$\mathrm{NNT|NNH} = \sum_{i=1}^{I} \Delta p_i - \mu \sum_{j=1}^{J} \Delta q_j = 2.2\% - 3.5\% = -1.3\%$$

The result, −1.3%, represents the net probability of benefit if treated with treatment a compared with treatment b. A negative value indicates that the sum of probabilities of benefits is smaller than the sum of probabilities of risks. This method therefore suggests an unfavorable risk- benefit balance of treatment a versus treatment b. Note that all unweighted methods implicitly assume equal weights for all outcomes. As such, probabilities of different outcomes are directly added or subtracted without consideration of their relative impact on quality of life. For example, probabilities of minor and major bleeding have been directly summed up to represent the total probability of risks.

Metrics That Incorporate Preference Weights

Relative-Value–Adjusted NNT|NNH

In the hypothetical example, if we consider reduction in MI ($u_{\mathrm{base}} = 0.84$) as the benefit of antiplatelet therapy, then using Equation 5 we have:

$$RV_j = \frac{1-0.8}{1-0.84} = 1.25 \text{(for minor bleeding)}$$

and

$$RV_j = \frac{1-0.46}{1-0.84} = 3.375 \text{(for major bleeding)}$$

This implies that patients value not having a minor bleed 25% more than they value not having an MI and they value not having a major bleed 3.375 times more than they value not having an MI. Note that this may be counterintuitive, but it is because the impact of minor bleeding is assessed over only 1 month whereas the impact of MI is assessed over a lifetime and this approach does not reconcile differences in the duration of event impacts. Using Equation 6 for RV_{NNH} and Δq_j for minor and major bleeding from Table 1, we have:

$$RV\_NNH = \frac{1}{\sum_{j=1}^{J}\left(\Delta q_j \times RV_j\right)} = \frac{1}{(2.5\% \times 1.25) + (1\% \times 3.375)} = 15.38$$

This suggests that approximately 15 patients need to be treated with treatment *a* instead of treatment *b* in order to have one additional patient harmed from adverse effects of treatment *a*. We can then use the RV_{NNH} in place of the crude NNH in Equation 3 to calculate the preference-weighted net probability of benefit. Because only one benefit can be included in this method, we have chosen MI risk reduction. Given that the NNT for MI is 50, and assuming $\mu = 1$, we have:

$$NNT|RV\_NNH = \frac{1}{NNT_{MI}} - \mu\frac{1}{RV\_NNH} = \frac{1}{50} - \frac{1}{15.38} = -4.5\%$$

This represents a −4.5% net probability of benefit if treated with treatment *a* compared with treatment *b*. The negative sign indicates that the sum of probabilities of risks (i.e., minor and major bleeding) exceeds the probability of benefit (reducing MI), after accounting for their relative preference weights, suggesting an unfavorable risk-benefit balance for treatment *a* versus *b* (when MI reduction is the only benefit considered).

INB Using MAR

Based on the DCE preference weights presented in Table 2 and using major bleeding as the referent outcome, MAR_{ij} can be calculated for MI, stroke, and minor bleeding as follows:

MAR for MI versus major bleeding:

$$MAR_{ij} = \frac{v_i/\Delta p_i}{v_j/\Delta q_j} = \frac{0.18/2\%}{0.03/1\%} = 3$$

MAR for stroke versus major bleeding:

$$MAR_{ij} = -\frac{v_i/\Delta p_i}{v_j/\Delta q_j} = \frac{0.014/0.2\%}{0.03/1\%} = 2.33$$

MAR for minor bleeding versus major bleeding:

$$MAR_{ij} = -\frac{v_i/\Delta p_i}{v_j/\Delta q_j} = \frac{0.025/2.5\%}{0.03/1\%} = 0.33$$

Note that MAR for major bleeding versus itself is 1 by definition. Also, note that MAR_{ij} can be defined for any two attributes or outcomes, and indicates the marginal rate of substitution between the risks of those attributes.

A 3% MAR for MI indicates that individuals in our sample, on average, were willing to accept 3% additional risk of major bleeding to reduce the risk of MI by 1%.

Assuming $\mu = 1$, INB can be calculated using MAR_{ij} as follows:

$$INB_{MAR} = \sum_{i=1}^{I} MAR_i \Delta p_i - \mu \sum_{j=1}^{J} MAR_j \Delta q_j$$

$$INB_{MAR} = (3 \times 2\% + 2.33 \times 0.2\%) - (0.33 \times 2.5\% + 1 \times 1\%) = 4.64\%$$

After simultaneously including all outcomes and accounting for preferences of each, the results suggest that in a cohort of patients using treatment *a* instead of treatment *b*, 4.64% will experience a positive INB.

Although we have not explicitly included the duration of events in this analysis, the scenarios and attributes of a DCE can be designed to describe the assumptions about the impact and duration of various outcomes. In that case, estimated preference weights from DCE will reflect individuals' preferences for different outcomes by considering probability, impact, and duration of treatment on quality of life.

INB Using RVALYs :

INB using RVALYs also uses preference weights estimated using DCEs but explicitly incorporates the duration of events in the calculation of the INB. Based on the numbers in Tables 1 and 2, and assuming an average life expectancy of 20 years in our study population, the INB in terms of RVALYs can be estimated using Equation 11 as (assuming $\mu = 1$) follows:

$$INB_{RVALYs} = \sum_{i=1}^{I} r_i T_i \Delta p_i - \mu \sum_{j=1}^{J} r_j T_j \Delta q_j$$

$$INB_{RVALYs} = (0.09 \times 20 \times 2\% + 0.07 \times 20 \times 0.2\%) - \left(0.01 \times \frac{1}{12} \times 2.5\% + 0.03 \times \frac{1}{12} \times 1\%\right)$$

$$INB_{RVALYs} = 3.6 + 0.28 - 0.0021 - 0.0025 = 3.83\%$$

Compared with treatment *b*, treatment *a* results in a gain of 3.83 RVALYs per 100 patient-years. Note that including the time dimension affects the contribution of short-term risks in the overall calculation of INB because avoiding MI or stroke is more important than avoiding bleeding events given that MI and stroke have long-lasting effects on patients' quality of life.

INB Using QALYs

INB using QALYs also explicitly incorporates weights and the duration of events in the analysis.

Applying this method to the example (assuming $\mu = 1$ yields the following:

$$INB_{QALY} = \sum_{i=1}^{I} (u_0 - u_i) T_i \Delta p_i - \mu \sum_{j=1}^{J} (u_0 - u_j) T_j \Delta q_j$$

$$INB_{QALY} = [(0.95 - 0.84) \times 20 \times 2\% + (0.95 - 0.6) \times 20 \times 0.2\%] - \left[(0.95 - 0.8) \times \frac{1}{12} \times 2.5\% + (0.95 - 0.46) \times \frac{1}{12} * 1\%\right]$$

$$INB_{QALY} = (4.4 + 1.4) - (0.031 + 0.041) = 5.43\%$$

Assuming an average life expectancy of 20 years in the patient population, these results suggest that using treatment *a* instead of treatment *b* results in 5.43 additional QALYs per 100 patient-years treated. Note that health-related quality-of-life weights often are numerically different, with preference weights derived using DCEs. This can lead to different INB estimates using RVALYs and QALYs methods, as is the case in our example.

Metrics That Incorporate Weights from Other Perspectives

The unifying formula also enables INB analyses with MCDA. In the hypothetical example, the relevant criteria for comparing the two treatment options are risk of MI and stroke (indicating benefits) and the risk of minor and major bleeding (indicating risks). Assuming the scores (S_{it}) and weights (w_i) in Table 2, the final weighted scores can be calculated as (assuming $\mu = 1$ follows:

$$\sum_{i=1}^{I} w_i \Delta S_i - \mu \sum_{j=1}^{J} w_j \Delta S_j = (30\% \times 5 + 40\% \times 2) - [5\% \times (1) + 25\% \times (3)] = 1.5$$

This implies that treatment *a* is net favorable to treatment *b* on the basis of assumed weights and scores.

Discussion

We have proposed a unifying framework for classifying BRA methods into three families on the basis of the way in which the methods measure and balance benefits and risks. This classification emphasizes the aspects of BRA that are important to patients. Metrics that incorporate patient-derived preference weights account for the three dimensions of harms and benefits that directly affect patients' quality of life: probability, impact, and duration of impact of the outcomes. These methods make explicit and transparent the trade-offs among outcomes. Unweighted metrics, such as NNT\NNH, are commonly used for BRA but do not capture the impact and duration of benefits or risks or patient preferences for those outcomes. Metrics that use other weights are generally easy to implement but may not reflect patient preferences.

The numerical differences in results from different BRA methods demonstrate the impact that weights can have on the assessments of overall balance between effectiveness and safety of drugs. In our hypothetical example, unweighted methods resulted in an unfavorable benefit- risk balance of treatment *a* versus treatment *b* while methods based on MAR, RVALY, and QALY arrived at an opposite conclusion. The results of different BRA methods can diverge in particular in situations in which beneficial and harmful outcomes differ in terms of their clinical significance and duration. Our hypothetical example presented a situation in which the harmful outcomes (bleeding) had smaller impacts on patients' quality of life as compared with the beneficial outcomes (reduction in stroke and MI). BRA methods, however, can yield different results under other plausible scenarios. For example, if treatment *a* conferred a small incremental benefit and caused a rare but serious adverse event (e.g., death), using the weighted metrics could indicate an unfavorable benefit-risk balance that might not be evident with unweighted BRA methods.

Metrics that incorporate patient-derived preferences are expected to produce BRA estimates that are relevant to patients. When unweighted metrics are used, the assumption of equal weights and duration of events should be explicitly stated. Although metrics that use decision makers' weights are easy to implement, it is not clear that they will represent patient preferences. Some studies have shown significant differences between patients' and physicians' preferences for different treatment attributes [25,26]. Furthermore, BRA methods that use decision makers' preferences are highly dependent on the perceptions and assumptions of a small group of decision makers who participate in the study process [27]. Despite these limitations, alternative BRA methods, such as MCDA, might help to organize and clarify the thought process and underlying assumptions made by decision makers in BRA, particularly when direct, rigorous measurement of patients' preferences is not possible or feasible.

Metrics that incorporate patient-derived preference weights require more sophisticated approaches to elicit weights than do those in the other families. For instance, preference weights derived from stated preference methods (e.g., DCE) are scenario specific. BRA using stated preference methods often requires an initial step to elicit preferences for the attributes of interest in the relevant population. This limits the practicability of stated preference methods (i.e., MAR and RVALYs) compared with preference-based methods that use generic health-related quality-of-life weights (i.e., QALY). BRA methods based on stated preference methods, such as DCE, however, can capture trade-offs that patients are willing to make among different beneficial and harmful outcomes. Such trade-off information is generally not captured by the use of health-related quality-of-life weights (e.g., the EuroQol five-dimensional questionnaire) because they reflect preferences for individual outcomes measured in distinct studies and populations. Additional work is needed to examine correlations between stated preferences and utility weights and whether relative utility weights might well approximate state preferences.

Other methods have been proposed for use in BRA [3,4]. These, however, cannot be considered as independent BRA methods *per se*. For example, Monte Carlo simulation, which has been included as a method in other BRA frameworks, is a methodology to quantify uncertainty rather than being an explicit approach to compare benefits and risks. As such, Monte Carlo simulation can be used in conjunction with various BRA methods (e.g., NNT\NNH, INB using RVALYs, and INB using QALYs) to analyze uncertainty around results. Furthermore, we found substantial overlap in BRA methods that are often considered distinct methods in the literature [3,4]. For example, quality-adjusted time without symptoms and toxicity and INB using QALYs can be hardly differentiated but are generally described as separate methods. Also, NNT and NNH are identical to using risk difference or minimum clinical efficacy methods.

Although our unifying framework is necessarily simplistic, we believe that it will help investigators and decision makers understand the major commonality among existing BRA techniques. Others have discussed more nuanced considerations of the desired properties of BRA methods, and these should be considered within our framework, as well. For example, Lynd et al. [6] have proposed 10 criteria for evaluating and comparing different BRA techniques. These criteria emphasize the ability of a BRA technique to include individuals' valuation of risks and benefits, uncertainty, multiple risks and benefits, and duration and intensity of risks and benefits. They also suggest that BRA results should be easy to interpret and the techniques should be easily applied across different interventions with the ability to quantify both objective and subjective risks and benefits. Puhan et al. [4] have proposed additional criteria for comparing different BRA methods. Their criteria include type of data needed for BRA (e.g., individual vs. aggregate level), type of analysis (e.g., simulation vs. empirical), type of metric (e.g., QALY, count, or probability), ability to account for the joint distribution of harms and benefits, and type of presentation of results (e.g., differences, ratios, or graphic methods). These considerations may be useful when selecting among methods within a given family of metrics.

Our study has limitations. We have not conducted a formal, systematic review of the literature on BRA methods. Rather, we have relied on previous reviews to identify existing BRA methods for our analysis. We supplemented these reviews with searches for emerging BRA methods using terms such as "benefit risk," "harm benefit," and "incremental net benefit" using PubMed and Google Scholar. Our hypothetical example presents only one possible scenario that may arise in the context of BRA. Other clinical examples need to be examined to compare the performance of different methods. We have used hypothetical weights and scores for the MCDA approach to illustrate the implementation of BRA in our example. Therefore, the numerical results for MCDA should be interpreted with caution.

Approaches to facilitate comparison and selection of BRA techniques are increasingly important because stakeholders around the world are focusing more attention on BRAs. For example, the Center for Drug Evaluation and Research at the US

Food and Drug Administration is working to improve communication of risks and benefits [28], consider patients' perception about risks and benefits [5], and develop a qualitative framework for BRA [29]. Others, including the European Medicines Agency, Canada, Australia, Switzerland, and Singapore, and the Pharmaceutical Research and Manufacturers of America, are also developing frameworks for qualitative and quantitative methodologies for BRA. Examining the methods within the scope of our proposed unified framework and, mathematically, using standardized notation make it clear that existing methods share substantial core commonality. It also facilitates comparisons of the types of weights that are used across existing methods, a key differentiating feature.

Source of financial support: This project was funded by an unrestricted research grant from Novartis Pharmaceuticals Corporation to the Brigham and Women's Hospital. Dr. Najafzadeh received Fellowship support from Canadian Institutes of Health Research (CIHR) during the course of this study.

Supplemental Materials

Supplemental material accompanying this article can be found in the online version as a hyperlink at http://dx.doi.org/10.1016/j.jval.2014.11.001 or, if a hard copy of article, at www.valueinhealthjournal.com/issues (select volume, issue, and article).

REFERENCES

[1] The future of drug safety—promoting and protecting the health of the public: FDA's response to the Institute of Medicine's 2006 report. 2007. Available from: http://www.fda.gov/downloads/drugs/drugsafety/postmarketdrugsafetyinformationforpatientsandproviders/ucm171627.pdf. [Accessed October 7, 2014].

[2] Garrison LP Jr, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood) 2007;26:684–95.

[3] Guo JJ, Pandey S, Doyle J, et al. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy—report of the ISPOR Risk-Benefit Management Working Group. Value Health 2010;13:657–66.

[4] Puhan MA, Singh S, Weiss CO, et al. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. BMC Med Res Methodol 2012;12:173.

[5] Food and Drug Administration. FDA working with patients to explore benefit/risk: opportunities & challenges. Available from: http://www.fda.gov/ForConsumers/ByAudience/ForPatientAdvocates/ucm298136.htm. 2012. [Accessed October 7, 2014].

[6] Lynd LD, Coombes ME, Rashidi AA, et al. A systematic review and critical evaluation of harm-benefit analysis methods for therapeutic interventions. Presented at the 27th Annual Meeting of the Society for Medical Decision Making (SMDM). October 21-25, 2005, San Francisco, CA.

[7] The effect of low-dose warfarin on the risk of stroke in patients with nonrheumatic atrial fibrillation. The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators. N Engl J Med 1990;323:1505–11.

[8] Risk factors for stroke and efficacy of antithrombotic therapy in atrial fibrillation: analysis of pooled data from five randomized controlled trials. Arch Intern Med 1994;154:1449–57.

[9] Connolly SJ, Ezekowitz MD, Yusuf S, et al. Dabigatran versus warfarin in patients with atrial fibrillation. N Engl J Med 2009;361:1139–51.

[10] Patel MR, Mahaffey KW, Garg J, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med 2011;365:883–91.

[11] Granger CB, Alexander JH, McMurray JJ, et al. Apixaban versus warfarin in patients with atrial fibrillation. N Engl J Med 2011;365:981–92.

[12] Levin LA, Wallentin L, Bernfort L, et al. Health-related quality of life of ticagrelor versus clopidogrel in patients with acute coronary syndromes—results from the PLATO trial. Value Health 2013;16:574–80.

[13] Pickard AS, Johnson JA, Feeny DH, et al. Agreement between patient and proxy assessments of health-related quality of life after stroke using the EQ-5D and Health Utilities Index. Stroke 2004;35:607–12.

[14] Spiegel BM, Chiou CF, Ofman JJ. Minimizing complications from nonsteroidal antiinflammatory drugs: cost-effectiveness of competing strategies in varying risk groups. Arthritis Rheum 2005;53:185–97.

[15] Shah SV, Gage BF. Cost-effectiveness of dabigatran for stroke prophylaxis in atrial fibrillation. Circulation 2011;123:2562–70.

[16] Najafzadeh M, Gagne JJ, Choudhry NK, et al. Patients' preferences in anticoagulant therapy: a discrete choice experiment. Circ Cardiovasc Qual Outcomes 2014;7:912–9.

[17] Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: a proposal using quantitative methods. Pharmacoepidemiol Drug Saf 2003;12:611–6.

[18] Johnson FR, Ozdemir S, Mansfield C, et al. Are adult patients more tolerant of treatment risks than parents of juvenile patients? Risk Anal 2009;29:121–36.

[19] Lynd LD, Najafzadeh M, Colley L, et al. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making—a case study of alosetron in irritable bowel syndrome. Value Health 2010;13:411–7.

[20] 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(Suppl. 1):S2–15.

[21] Chuang-Stein C. A new proposal for benefit-less-risk analysis in clinical trials. Control Clin Trials 1994;15:30–43.

[22] Boers M, Brooks P, Fries JF, et al. A first step to assess harm and benefit in clinical trials in one scale. J Clin Epidemiol 2010;63:627–32.

[23] Gail MH, Costantino JP, Bryant J, et al. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst 1999;91:1829–46.

[24] European Medicine Agency. Report of the Committee for Medicinal Products WorkingGroup on Benefit-Risk Assessment Models and Methods (Contract No.: Doc. Ref. EMEA/CHMP/15404/2007). 2007. Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf. [Accessed February 16, 2012].

[25] Devereaux PJ, Anderson DR, Gardner MJ, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 2001;323:1218–22.

[26] Muhlbacher AC, Juhnke C. Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? Appl Health Econ Health Policy 2013;11:163–80.

[27] Marsh K, Lanitis T, Neasham D, et al. Assessing the value of healthcare interventions using multi-criteria decision analysis: a review of the literature. Pharmacoeconomics 2014;32:345–65.

[28] Fischhoff B., Brewer N., Downs J.; Food and Drug Administration. Communicating risks and benefits: an evidence-based user's guide. 2011. Available from: http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM268069.pdf. [Accessed October 07, 2014].

[29] Frey P. Director, Office of Planning and Analysis, Center for Drug Evaluation and Research. Benefit risk considerations in CDER: development of a qualitative framework. Available from: http://www.fda.gov/downloads/AboutFDA/CentersOffices/OfficeofMedicalProductsandTobacco/CDER/UCM317788.pdf. 2012. [Accessed October 7, 2014].