

Quantitative Risk/Benefit Assessment: Where Are We?

Christy Chuang-Stein

Abstract Pharmaceutical sponsors use a variety of approaches to make important benefit/risk decisions about their products internally. Benefit/risk assessment is equally important when regulators evaluate a product for marketing approval and payers evaluate it for reimbursement decision. Once a product receives marketing authorization, it is critical to communicate pertinent benefit and risk information to patients and health-care providers. All of the above can be made easier by the use of a common framework. In this paper, we review where we are in benefit/risk assessment. This includes endeavors by academic institutions, regulators, and the pharmaceutical industry. Despite concerns about quantitative benefit/risk assessment expressed by some, we argue that without a way to quantitatively incorporate the relative importance of factors impacting benefit/risk assessment, it will be hard to bring transparent decisions to questions such as “does the benefit of this product outweigh the risk.”

1 Introduction

I saw a cartoon in the summer of 2010. The cartoon is a metaphor for a pharmaceutical developer who, like a hurdler, needs to cross a set of hurdles. The first hurdle is labeled quality, the second safety and the third efficacy. The fourth one is unlabeled, but looks menacing. The hurdle takes the form of a solid wall. There are sharp spikes coming out of the wall, facing the hurdler. The caption reads—There was general agreement that the fourth hurdle was the one to look out for.

I asked myself—Is the cartoon an exaggeration of the environment a product developer is in? If there is some truth to the cartoon, then what is this fourth hurdle?

C. Chuang-Stein (✉)

Statistical Research and Consulting Center, Pfizer Inc., Kalamazoo, MI 49007, USA

e-mail: christy.j.chuang-stein@pfizer.com

Is it benefit/risk? Is it relative effectiveness, a term frequently used in Europe? Is it comparative effectiveness, a term gaining momentum in the USA? Is it Health Technology Assessment, a cost-effectiveness evaluation pharmaceutical developers have to go through to have their products reimbursed in socialized health-care systems in Europe, Canada, and Australia? Or, is it all of the above plus other emerging value propositions?

I was told by a friend who was familiar with the origin of the cartoon that the cartoonist used the fourth hurdle as a symbol for health technology assessment. In an editorial in *Clin Pharmacology and Therapeutics*, Honig [1] described comparative effectiveness as the fourth hurdle in drug development. I cannot but feel that the cartoon could easily apply to benefit/risk evaluation because of the lack of a common approach for articulating the trade-off between benefit and risk to reach a transparent decision.

When preparing my presentation for the fourth Seattle Symposium, I checked the FDA Advisory Committee (AC) meetings from September to early November 2010 to see how the benefit/risk question was presented to the AC members. There were six AC meetings altogether. Among the six meetings, three were to decide whether approved products should remain on the market. One was to review a supplemental application of an approved product. Two were to review new molecular entities, one of which was lorcaserin hydrochloride (with diet and exercise) for weight management for obese patients on September 16th. The public meeting on lorcaserin attracted a lot of attention. The interest level for weight management is generally high. In this case, the interest was elevated by the diet drug subutramine that was reviewed the day before for possible regulatory actions including product withdrawal. On September 16th, the Endocrinologic and Metabolic Drugs AC was asked to vote on only one question. The question was whether available data demonstrated that the potential benefits of lorcaserin outweighed the potential risks to allow marketing approval.

There was a general agreement that lorcaserin's efficacy data met FDA's requirement, albeit marginally. Concerns were raised about lorcaserin's safety. These included the fact that lorcaserin is chemically similar to two weight-loss drugs that were withdrawn in 1997 due to their links to the valvular heart disease. In addition, two-year studies in rats reported an excess number of malignant mammary tumors in female rats. The cancer concerns were not confirmed in clinical trials. Nevertheless, it was felt that the duration of the trials might be too short and the study populations not diverse enough to allow a potential cancer risk to be detected.

When it was time to vote, the votes were five (36%) for approval and nine (64%) against approval. On October 22, 2010, FDA rejected lorcaserin for the proposed indication, signaling that, in the eyes of the agency, the safety concerns outweighed what the agency called lorcaserin's marginal effectiveness. Had the efficacy of lorcaserin been much better than what had been observed, would FDA approve lorcaserin for the indication sought? On August 2 2011, the Bloomberg news reported that lorcaserin's manufacturers announced that a newly completed study showed the concentrations of lorcaserin to be lower in human brains than in

rat models. This finding helped ease concerns that lorcaserin may be linked to brain tumors. On July 22 2012 FDA granted marketing authorization to lorcaserin.

In general, how do we make decisions with opposing needs? A sensible approach is to adopt a framework where all relevant factors are first collected. This first step is then followed by articulating the relative importance of these factors, identifying a sensible way to combine factors with weights that reflect the relative importance of the factors, checking out the properties of the combination algorithms, settling on a decision rule and identifying conditions where the rule would lead to clear and unequivocal decision. Finally, make a decision and communicate the decision to individuals who have an interest in the outcome.

The rest of the paper is organized as follows. In Sect. 2, we discuss selected approaches to examine benefit and risk simultaneously. Some of these approaches combine benefit and risk into one measure for easy interpretation, while others consider benefit and risk jointly. Section 3 describes a benefit/risk framework developed by the Benefit Risk Action Team (BRAT) of the Pharmaceutical Research and Manufacturers of America (PhRMA). In Sect. 4, we discuss communicating benefit/risk to the public. Section 5 describes a report on current tools and processes for regulatory benefit–risk assessment issued by the Benefit–Risk Methodology Project of the European Medicines Agency. Section 6 describes briefly a recent FDA draft guidance on factors to consider when making benefit–risk determinations in medical device premarket review. We end this paper by acknowledging challenges of benefit/risk assessment of pharmaceutical products in general and offer some additional comments in Sect. 7.

2 Measures or Approaches to Assess Benefit and Risk Simultaneously

2.1 Quality-Adjusted Life Without Toxicity Q-TWiST

One of the earlier attempts to discount benefit by risk of cancer drugs was to calculate time without symptoms of disease and toxic effects (TWiST) [2]. This concept was further developed to form quality-adjusted TWiST (Q-TWiST) [3]. Q-TWiST was obtained by discounting survival by a utility weighting that reflected quality of life in different physical conditions. For example, the utility weighting could be different for days with toxic effects and days after disease progression. More recently, Hughes et al. [4] used quality-adjusted life-years within a decision-analytical framework.

While discounting survival by treatment-related toxicity and/or poor quality of life was intuitive, the discounting process could be subjective. As such, Irish et al. [5] suggested conducting a threshold utility analysis as a form of sensitivity analysis by comparing treatments across all combinations of the utility weightings for days with toxic effects and days after disease progression. Such sensitivity analysis allows

researchers to observe how the comparison varies with different utility weightings. Variations of the threshold utility analysis are possible. For example, if data are available, one can incorporate patients' experiences or functions on days after disease progression instead of relying on utility weighting solely on such days.

The idea of discounting benefit by risk was also adopted by Chuang-Stein [6] who proposed a benefit-less-risk analysis. Under this analysis, benefit was discounted linearly by a risk measure via the use of a conversion factor. The conversion factor serves to convert benefit and risk to a similar scale to allow their integration into a single measure.

2.2 *Clinical Utility Index*

Pharmaceutical manufacturers regularly assess the benefit/risk profiles of their products. This applies to selecting a dose and making go/no go decisions. Regarding dose selection, some sponsors try to maximize benefit while keeping the risk at an (pre-specified) acceptable level. Alternatively, a pharmaceutical manufacturer can apply the concept of a clinical utility index (CUI) to facilitate dose selection. A clinical utility index is a composite measure that combines several measures (some of which may be desirable while others are not) into one to facilitate decision making.

Ouellet et al. [7] constructed a CUI when investigating the potential value of a new treatment for insomnia. Five efficacy endpoints had been used previously to evaluate benefit from an insomnia drug. They were latency to persistent sleep, wake after sleep onset, quality of sleep, and sleep architecture measured by the percentages of stage 1 and stages 3–4 sleep. An undesirable consequence of using an insomnia drug is the residual drug effect, which could make a user feel lethargic on the morning after taking the medication. Residual drug effect could be assessed by two measures from a commonly used Leeds questionnaire for insomnia research. If a withdrawal effect is a potential concern for an insomnia drug, it should be appropriately measured and included as a risk endpoint.

Faced with these seven endpoints recorded on different scales, Ouellet et al. first normalized the scales so that the endpoints were combinable. For example, a change of 25 min for wake after sleep onset was considered to be approximately equivalent to 15 min change in time to persistent sleep (see Table 1 in Ouellet et al. [7]). Next, they surveyed 581 physicians engaged in insomnia research and developed a weighting scheme to combine the seven endpoints into a CUI.

Using the CUI in a dose–response study, Ouellet et al. concluded that it would not be worthwhile to continue the development of the new compound for an insomnia indication.

2.3 Incremental Benefit/Risk Ratio

Assume that both benefit and risk could be described by a binary endpoint. We use pe_N and pe_C to denote the probability of experiencing the benefit in the new treatment and the control groups, respectively. We use pr_N and pr_C to denote the corresponding probabilities of experiencing risk. We will assume that the new treatment delivers more benefit, at the expense of more risk. If the new treatment delivers more benefit and less risk, then there is no need to discuss the benefit/risk tradeoff between these two treatments.

A measure frequently used in cost-effectiveness analysis is the incremental cost-effectiveness ratio, which looks at the increase in cost relative to every unit of increase in effectiveness. This concept could be used to form the incremental benefit/risk ratio (IBRR) in (1). As with any ratio-based measure, it is important to interpret (1) in conjunction with the magnitude of the numerator and the denominator that form the ratio. The construction of the ratio in (1) does not imply that the benefit and the risk are of equal clinical relevance.

$$IBRR = \frac{pe_N - pe_C}{pr_N - pr_C} \quad (1)$$

The IBRR in (1) could be re-expressed as in (2).

$$IBRR = \frac{\left(\frac{1}{pr_N - pr_C}\right)}{\left(\frac{1}{pe_N - pe_C}\right)} \quad (2)$$

The numerator in (2) is often interpreted as the number needed to treat to harm (NNTH) and the denominator is the number needed to treat to benefit (NNTB). So, if $NNTH = 10$ and $NNTB = 5$, then $IBRR = 2$. This ratio has the interpretation that, on the average, for each additional individual experiencing the adverse event under the new treatment, two more individuals will benefit from the new treatment compared to the control. Obviously, large IBRR values will make the new treatment more attractive. The question is—for the target patient population—is there an IBRR threshold beyond which the new treatment will be considered to have a more favorable benefit/risk profile (compared to the control)? If this threshold is not achieved in the entire target population, is there a clinically meaningful subgroup in which the new treatment is likely to have a more favorable IBRR?

The concept of IBRR was used in an FDA Cardiovascular and Renal Drugs Advisory Committee meeting on February 3, 2009. The committee was to evaluate prasugrel (an antiplatelet agent) for reducing cardiovascular events in patients with acute coronary syndrome (ACS) undergoing primary or delayed percutaneous coronary intervention. Data came from a single large trial TRITOM of 13,608 patients [23]. The primary endpoint was a composite endpoint of cardiovascular death, nonfatal myocardial infarction, and nonfatal stroke. Results from the study

showed prasugrel to be more efficacious than the comparator clopidogrel, but at the expense of more bleeding. While some concerns were raised about a possible increase in malignancy risk associated with prasugrel, we will focus on bleeding here since bleeding is a common (and major) side effect of antiplatelet (and anticoagulant) agents.

By TIMI (Thrombolysis in Myocardial Infarction Trial) convention, bleeding was broadly classified as major, minor, or minimal. Figure 1 was extracted from an FDA presentation by Dr. Ellis Unger at the Advisory Committee meeting. In his presentation, Dr. Ellis Unger applied the ratio concept and showed the number of composite events prevented for each additional bleeding event with prasugrel. This ratio is represented on the y-axis. The x-axis notes number of days after the intervention. The bleeding events plotted in Fig. 1 correspond to serious events (the blue curve at the bottom), TIMI major bleeding events (the red curve on the top) or TIMI major and minor bleeding events combined (the black curve in the middle). The ratio was done in a cumulative fashion in that once an individual experienced an endpoint (a composite efficacy endpoint or a bleeding event), the individual was said to have experienced that endpoint at all subsequent time points. In other words, at a given number of days x after the intervention, pe_N and pe_C in (1) represent the probabilities that patients on prasugrel and control have not yet experienced the composite efficacy endpoint up to days x . As for pr_N and pr_C , they correspond to the probabilities that patients in these two groups have experienced a bleeding event by days x .

Figure 1 was used in an exploratory manner to help interpret the results from TRITON at the AC meeting. It is quite likely that the February 2009 meeting was the first time many AC members (as well as others in the audience) saw the use of IBRR. We will focus on the curve corresponding to major bleeding (the red curve on the top) in Fig. 1. The curve was high at the beginning. It gradually came down over time and eventually settled around a value of 3. Was 3 a good IBRR value in this case? There was no discussion of a minimum IBRR for prasugrel in order to receive marketing approval.

Prasugrel was approved on July 10, 2009 with a black box warning on bleeding risk. The black box warning also includes patient subpopulations for which prasugrel is contraindicated.

2.4 Graphic Display

Chuang-Stein et al. [8] proposed to use a multinomial random variable to capture efficacy (benefit) and safety (risk) outcomes simultaneously. The multinomial random variable Chuang-Stein et al. proposed has five outcome categories. They are *benefit and no serious adverse events*, *benefit and serious adverse events*, *no benefit and no serious adverse events*, *no benefit and serious adverse events*, and *side effects leading to withdrawal*. Here the term “serious adverse events” should be interpreted in the context of the patient population. They may not necessarily

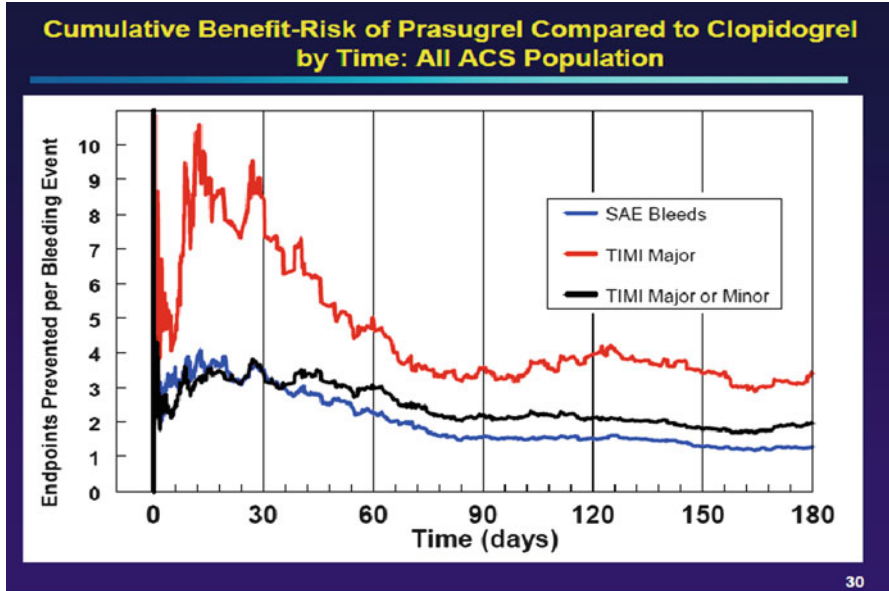


Fig. 1 Incremental benefit/risk ratio over time for comparing prasugrel with clopidogrel. The graph was taken from an FDA presentation at the February 3, 2009 Cardiovascular and Renal Drugs Advisory Committee Meeting

mean serious adverse events by the official regulatory definition. Chuang-Stein et al. proposed to use observed proportions of these categories, along with weights reflecting the desirability of these outcomes, to construct linear or ratio scores to compare treatments. Since weights reflect the clinical relevance of the categories, there is no need to assume that the categories are of equal relevance clinically. This idea was later extended by Entsuah and Gorman [9], Entsuah and Gao [10] and Pritchett and Tamura [11] to include more outcome categories in real-case applications. Entsuah et al. also discussed a simple case of sensitivity analysis to see how the comparison between two treatment groups could vary as a function of the chosen weights.

Norton [12, 13] used graphics to display the distribution of the five categories over time. Labeling the five outcomes as “Benefit Only,” “Benefit + AE,” “Neither,” “AE only,” and “Withdraw”, he plotted each individual’s outcome category at each of the six post-randomization assessment points in a 12-week trial (Fig. 2). Individuals in Fig. 2 were arranged in such a way that dropouts were grouped together for easy visualization of the dropout pattern. Except for withdrawal, an individual could stay in or move to another response category from one assessment period to the next. In this sense, the graphic displays a snapshot in time on the response (i.e., not cumulative experience up to that point). Displaying response in this manner requires one to have access to observations on all patients at all assessment periods or until patients dropout of the study. Consequently, one needs to prespecify a

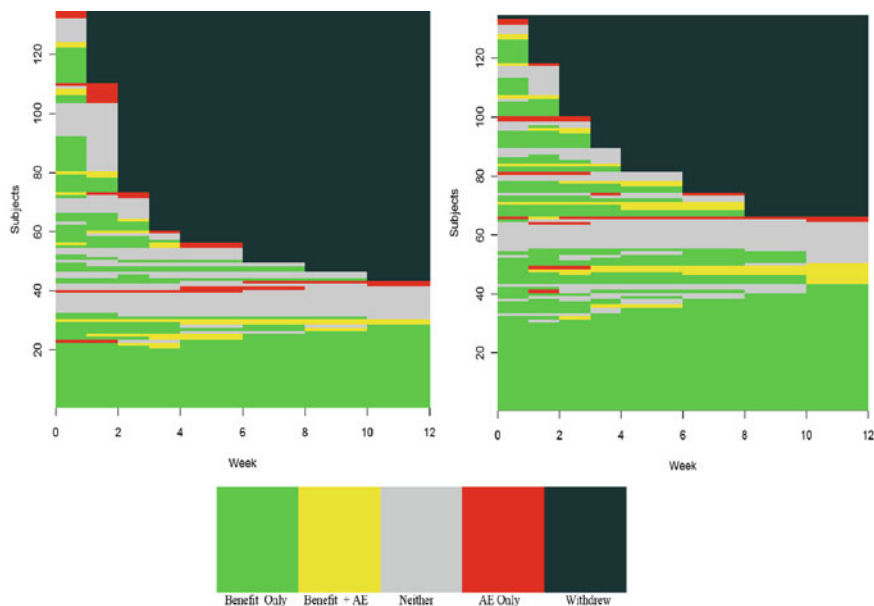


Fig. 2 Display of 5 (benefit, risk) outcomes over time. The *left panel* pertains to the control group while the *right panel* pertains to the investigational treatment group (Display is a courtesy from Norton [12].)

method to handle the situation where a patient missed a clinical visit between two completed ones. Possible approaches include carrying the last available response category forward or creating an extra category to represent the missing intermediate response.

The left panel in Fig. 2 corresponds to the control arm while the right panel corresponds to the investigational treatment group. There were no missing intermediate data in this example. Visual inspections of the figure reveal that compared to the investigational arm, the control group exhibited a higher and early dropout pattern. The graph shows clearly how the distribution of these five outcome categories changes over time and how the pattern differs between the two groups. One could compute the % of different colors (e.g., green) over a period of interest (e.g., weeks 8–12) to make a simple qualitative comparison between the groups. For both groups, Fig. 2 shows that some individuals derived benefit early and continued to do so without experiencing the adverse events.

The display in Fig. 2 could be extended to include more categories or multiple outcomes. For example, if positive outcomes on two equally important efficacy endpoints are twice as good as a single positive outcome on only one endpoint and that experiencing two distinct types of adverse events is twice as bad as experiencing only one, then one can form a net benefit/risk outcome by calculating $(\# \text{ of beneficial outcomes} - \# \text{ of untoward adverse events})$ for each individual. There are possible five values for the net outcome (i.e., 2, 1, 0, -1, and -2). One could plot the

distribution of these five net outcomes over time. Implicit in this calculation is the assumption that one good outcome can offset one bad outcome, an assumption that may not be valid in many situations.

3 Benefit Risk Framework Developed by PhRMA

The Pharmaceutical Research and Manufactures of America (PhRMA) have long recognized the importance of and the need for a transparent benefit/risk assessment process. In 2006, PhRMA sponsored a Benefit Risk Action Team (BRAT). The objectives were to formulate a framework for the ideal benefit–risk approach and to provide greater structure to assist sponsor-regulator discussions. BRAT partnered with epidemiologists at the Research Triangle Institute Health Solutions on the task.

Before developing the framework, members of BRAT agreed that the framework should be considered as a set of processes and tools to guide decision makers. In addition, the framework should be flexible to handle different contexts. The proposed framework was required to go through three rounds of development and testing, using mock products in the statin, tumor necrosis factor, and triptan classes. Early experience with the BRAT framework has been published in Coplan et al. [14] and Levitan et al. [15]. In early 2011, the testing was completed and BRAT developed a software application to assist graphic displays of the framework as well as its output. BRAT offered the software tool to PhRMA member companies for internal pilots, hoping to receive additional comments on the framework and gain support for broad implementation.

The framework could be described as a series of six steps [14, 15]. They are: define decision context (step 1), identify outcomes (step 2), identify and extract source data (step 3), customize framework (step 4), assess outcome importance (step 5) and display and interpret key B-R metrics (step 6). These steps can be slightly modified to better fit a particular situation. Ideally, the six steps should be completed before a new drug application (or biologics license application) and that the first four steps should be completed before conducting the pivotal trials. In theory, the framework can be applied at any stage during product development or post-approval. One difficulty in establishing the framework after the outcome data are known is that the process could be influenced by the outcome, thus creating potential bias. This is usually not a problem for a mature field with well-articulated efficacy endpoints and classes of products with well-characterized side effects. For a product of novel mechanism, it might actually be necessary to rely on safety data from late-stage trials to help characterize product risk.

Steps 3 and 4 above involve identifying data sources (randomized clinical trials or observational studies) and assessing the relevance of the information. Step 6 discusses the display of benefit and risk summary. BRAT strongly encouraged displaying the summary graphically such as in a forest plot. The latter is illustrated in Coplan et al. [14] and Levitan et al. [15]. Fig. 3 shows a forest plot presented by the manufacturer of rivaroxaban at an FDA Cardiovascular and Renal Drugs

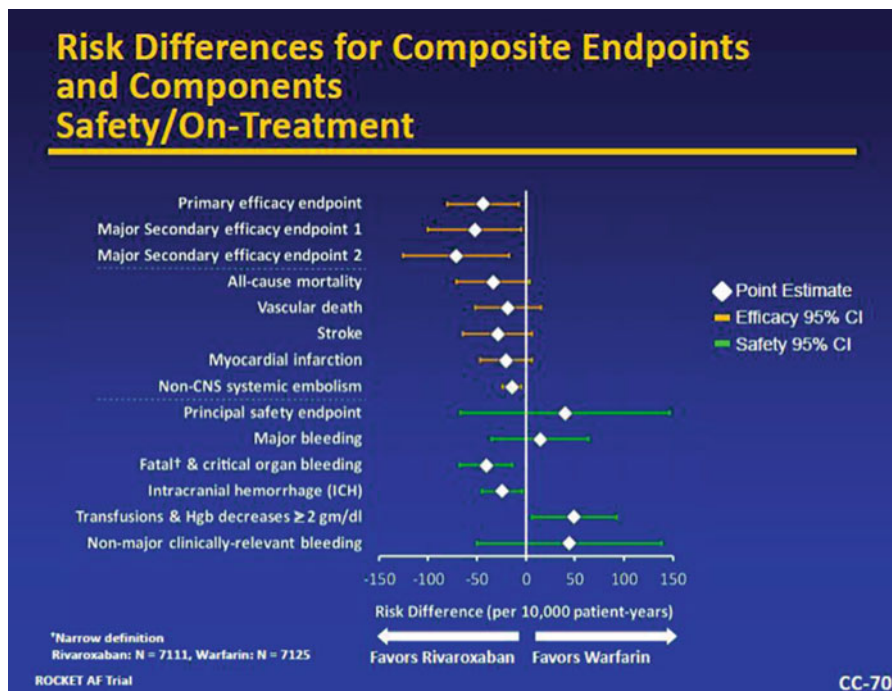


Fig. 3 Forest plot of eight benefit endpoints along with six risk endpoints for comparing rivaroxaban with warfarin, presented at the September 8 2011 Cardiovascular and Renal Drugs Advisory Committee meeting by the manufacturer of rivaroxaban

Advisory Committee meeting on September 8 2011. The meeting was to assess the benefit and risk of rivaroxaban against warfarin in preventing stroke and systemic embolism in patients with non-vascular arterial fibrillation. Details for the AC meeting could be found at the FDA website.

Fig. 3 displays results related to eight benefit and six risk endpoints. Some of the endpoints (e.g., all cause mortality, vascular death, stroke, MI) are components of a composite endpoint. For each endpoint, Fig. 3 shows the difference in the observed number of individuals/10,000 patient years who reported experiencing that endpoint. The difference is noted by a diamond in the plot with a companion 95% confidence interval. For all endpoints, the difference was calculated by subtracting the response in the warfarin group from that in the rivaroxaban group. Since all endpoints are undesirable, a difference that is greater than 0 will signal a better outcome for the warfarin group.

In general, point estimates and confidence intervals should be based on meta-analyses of relevant data sources. It is important that the data sources be systematically searched and critically appraised for inclusion so that the statistics in Fig. 3 are defensible.

In its initial work, PhRMA BRAT did not recommend any particular benefit/risk measure and did not promote the use of weights either. BRAT pointed out that data behind the forest plot in Fig. 3 could be used to support any chosen benefit/risk measure with weights chosen before seeing the data. In addition, getting into debates on weight selection early on could distract the team from focusing on the development of the framework. The decision probably also reflected a common concern that any effort to reduce multiple endpoints into a signal measure could result in information loss. In my opinion, without a mechanism to appropriately debate the roles the various benefit and risk endpoints play, without a way to quantitatively reflect the relative importance of these endpoints and how conclusion vary with the relative importance, answers to question such as “does the benefit of this product outweigh its risk” will continue to lack the transparency and structured deliberations we desire.

PhRMA transferred all work related to the BRAT framework to the Center for Innovation in Regulatory Science Ltd. for further development in January 2012.

4 Communication on Benefit/Risk to the Public

One major challenge in modern medicine is our ability to effectively communicate essential information about medicines to users. Surveys have repeatedly shown that many Americans don't understand the drugs they are taking. Lack of understanding often leads to noncompliance, contributing to medication errors, ineffective disease management, and considerable risks. The US FDA has a Risk Communication Advisory Committee. According to the information posted at the FDA web site, this Committee is to advise the Commissioner of the FDA or designee on methods to effectively communicate risk associated with products regulated by the FDA. The Committee reviews and evaluates strategies and programs designed to communicate with the public about the risks and benefits of FDA-regulated products. It also reviews and evaluates research relevant to such communication.

During a Committee meeting on February 26–27 in 2009, Steven Woloshin and Lisa Schwarz proposed to use a drug facts box to communicate the benefits and side effects of prescription drugs. The idea of the drug facts box was based on the successful implementation of a standardized nutrition facts box that is required of all packaged food sold publicly in the USA. Under Woloshin and Schwarz's proposal, the top panel of a drug facts box (Fig. 4) contains critical information on benefit while the bottom panel contains critical information on risk. The drug facts box can be viewed as a simplified tabular version of the forest plot display in Fig. 3 where data should come from all relevant sources and summarized in meta analyses.

Interestingly enough, the Patient Protection and Affordable Care Act H.R. 3590 (also known as the Health Care Bill) mentions a drug facts box. Specifically, it states that “*The Secretary of Health and Human Services . . . shall determine whether the addition of quantitative summaries of the benefits and risks of prescription drugs in a standardized format (such as a table or drug facts box) to the promotional labeling*

Prescription Drug Facts: AMCID (amoditine)

What is this drug for?	To relieve heartburn
Who might consider taking it?	Men and women bothered by heartburn or acid reflux disease
Who should NOT take it?	Women who are pregnant or breastfeeding
Recommended testing	None
Other things to consider doing	Eat frequent small meals; avoid fatty foods (and others which trigger your symptoms); excessive alcohol and eating close to bedtime; don't smoke; look into other medications.

AMCID STUDY FINDINGS BOX

500 people with bothersome heartburn were given AMCID or a sugar pill for 2 weeks. Here's what happened:

What difference did AMCID make?	People given a sugar pill	People given AMCID (20 mg a day)
Did AMCID help? Fewer people taking AMCID had heartburn (17% fewer)	81% 810 in 1000	64% 640 in 1000
Did AMCID have side effects? <i>Life threatening side effects</i> No difference between AMCID and a sugar pill	None observed	
<i>Symptom side effects</i> No difference in headache	About 5% in both groups 50 in 1000	
No difference in diarrhea	About 2% in both groups 20 in 1000	
No difference in dizziness	About 1% in both groups 10 in 1000	

How long has the drug been in use?
Amcid was approved by FDA in 1991 - Studies show that most serious side effects or recalls of new drugs happen during their first 5 years of approval.

Fig. 4 An example of a drug facts box shown by Woloshin and Schwartz at the Feb 26–27 2009 FDA Risk Communication Advisory Committee meeting

or print advertising of such drugs would improve health care decision making by clinicians and patients and consumers..”

One cannot underestimate the power of a standardized format to assist public understanding of the facts about a drug. Because drug facts are typically more

complicated than food, a drug facts box, if it becomes standardized, should have a drill-down option to offer additional summaries for individuals who desire more detailed information.

5 Report from EMA Benefit–Risk Methodology Project

On August 31, 2010, the Benefit–Risk Methodology Project sponsored by the European Medicines Agency (EMA) issued a report on current tools and processes for regulatory benefit–risk assessment (EMA/549682/2010). The report describes qualitative and quantitative approaches. It mentions ongoing work by PhRMA, the Center for Medical Research and the FDA. The latter are all classified as qualitative approaches by the report.

The report also describes 18 quantitative approaches and the view of the authors on each approach. Among the 18 quantitative approaches, the report comments that only three (Bayesian statistics, decision trees & influence/relevance diagrams, and multi-criteria decision analysis) incorporate the value or utilities of benefit and risk, along with probabilities representing the uncertainties of those effects, to numerically represent the benefit–risk balance.

The Bayesian statistics approach uses Bayes' Theorem to update the degree of prior belief as new information becomes available for incorporation. Prior belief may come from information related to similar products in the same class or opinions of key opinion leaders. In the latter case, it is important to ensure that the solicited prior belief is free of bias arising from the involvement of individuals with real or potential conflict of interest. The Bayesian approach calculates decision-relevant posterior probabilities. The decision trees & influence/relevance diagrams approach is derived from decision theory. It is based on three basic assumptions: (1) probabilities exist; (2) utilities exist; and (3) the action associated with the highest expected utility will be the most preferred. The multi-criteria decision analysis (MDCA) was developed by Mussen et al. [16, 17]. It was the prototype for the PhRMA BRAT framework. MDCA goes beyond the qualitative framework. It includes scoring and weighting. MDCA defines scoring as the process of measuring the value of options and uses weighting to ensure that the units of value on all the criteria are comparable to be combinable. The development of the clinical utility index (CUI) discussed in Sect. 2.1 is a simple case of applying these concepts.

The report acknowledges that any quantitative method requires a qualitative framework. On many occasions, utilizing qualitative and quantitative approaches in tandem can be useful. Since efforts to improve benefit/risk assessment are continuing at regulatory agencies, academic institutions and within the pharmaceutical industry, we can expect additional summary reports on this topic in the future.

6 FDA Draft Guidance on Factors to Consider When Making Benefit–Risk Determinations in Medical Device Premarket Review

On August 15, 2011, FDA issued a draft guidance, for public comment, on factors to consider when making a benefit–risk determination in medical device premarketing review. The agency hopes that a guidance on this topic can provide greater clarity for both the reviewers at the agency and the device industry. The document discusses three hypothetical examples in detail with likely assessment outcomes. The document also briefly describes six real cases and how the decision was made in each case.

The draft document contains a worksheet in the appendix. The worksheet offers users a systematic way to articulate factors that should be considered when making benefit–risk assessment. Major factor headings for the worksheet are: type of benefit, magnitude of the benefit, probability of the patient experiencing a benefit, duration of effect, severity and types of harmful events, probability and duration of a harmful event, risk of false-positive or false-negative for diagnostics, uncertainty, patient tolerance for risk, availability of alternative treatments or diagnostics, risk mitigation and novelty of technology.

The above factors reflect the core principles for benefit/risk assessment. They are (1) the risk of a product should be evaluated with respect to its potential benefit; (2) benefit/risk assessment should be conducted with respect to the target population and in view of available alternative therapies; (3) the strength of the supporting data (benefit and risk) is crucial to the evaluation.

In the draft guidance, FDA is hesitant to consider quantitatively weighting the factors because the appropriate weighting could change over time. We feel that this concern alone does not justify not using a quantitative approach if it can help make a better decision based on the available evidence at a particular point in time. Circumstances can indeed change and our priorities may shift over time. Temple [18] explained FDA’s decision to remove terfenadine (a non-sedating antihistamine approved in the USA in 1985) from the market in 1998 when terfenadine’s active metabolite fexofenadine became available. Terfenadine was linked to fatal ventricular arrhythmia *torsade de pointes* while fexofenadine was not. FDA’s decision underscored a continuous benefit/risk assessment process that began in 1992 when *torsade de pointes* was first reported in patients taking terfenadine. When a new drug with a pharmacologically identical effect but without a major serious adverse reaction becomes available, the benefit/risk profile of the older product may no longer be considered favorable.

7 Discussion

It is well accepted that benefit/risk assessment is necessary for a new treatment. Over the past ten years, several workshops have been dedicated to this topic including the one sponsored by the Institute of Medicine entitled “*Understanding the Benefits*

and Risks of Pharmaceuticals” on May 30, 2006. Our coverage on benefit/risk assessment in this paper was purposely kept simple. We did so to help readers gain some overall perspective on the benefit/risk assessment movement and not to get bogged down by technical details. In addition, we feel that any systematic approach for benefit/risk assessment needs to be intuitive, logical, and easy to understand to have a chance for broad uptake.

Many have pointed out the challenges of benefit/risk assessment [19, 20]. For one thing, while benefit might be realized shortly after initiating a treatment, serious risk might take a long time to surface. Randomized trials conducted during the premarketing phase are usually too short to observe long-term risk. As such, benefit/risk assessment needs to occur regularly after a product is available to the public. A product’s benefit/risk profile could change over time with emerging post-marketing data and the availability of newer products. In the USA, the Drug Safety and Risk Management Advisory Committee plays an important advisory role to the FDA on this question.

Equally important is the need to quantify uncertainty in our estimates of benefit and of risk and therefore the chosen benefit/risk measures. Uncertainty affects the strength of the data. Figure 1 does not include any information concerning the variability around the reported incremental benefit risk ratio. One could apply the bootstrap methodology to construct confidence intervals and include them in the figure.

How can statisticians help? At a recent Joint Statistical Meetings, Hoerl and Snee [21] discussed the concept of “statistical engineering.” They defined statistical engineering as the study of how best to utilize statistical concepts, methods, and tools and integrate them with other relevant sciences to generate improved results. The idea behind statistical engineering is that we have a list of statistical science parts. As statisticians, we need to assemble these statistical science parts to improve our systems and our processes. Hoerl and Snee claimed that statistical engineering could be applied to improve anything. I believe statistical engineering could help us develop a better process to conduct risk/benefit assessment also. To do this, we need to collaborate with other disciplines and bring alive the relevant statistical science parts on our list.

So, where are we now in terms of quantitative benefit/risk assessment? The good news is that we have started to see some concerted attention dedicated to this topic, both by the regulatory agencies and by the pharmaceutical industry. Several pharmaceutical companies piloted the BRAT framework within their organizations in 2011. Several references included in this paper describe the applications of benefit/risk assessment approaches to actual cases. Despite these, advancements are being made in small steps. There is no doubt that benefit/risk assessment is complicated and situation dependent. Many questions remain, some of which will not have easy solutions. For example, how do we address conflicting findings from different meta-analyses (or between observational studies and randomized trials) on the safety of a marketed product? How do we resolve conflicting recommendations from different branches within the same regulatory agency?

The last question in the above paragraph came into full focus during a joint meeting among three FDA Advisory Committees (Pediatric, Pulmonary-Allergy Drugs, Drug Safety and Risk Management) on December 11 2008. The three ACs along with FDA's Office of Surveillance and Epidemiology (OSE) and Division of Pulmonary and Allergy Products (DPAP) were to weigh the public health implications of real and serious but relatively infrequent occurrences of severe asthma exacerbations and asthma-related death against the asymptomatic benefits of bronchodilation and asthma control of long-acting beta-agonists (LABAs). During the FDA presentation, the Director of DPAP acknowledged differing views on how to manage LABA risk within the agency. While OSE preferred withdrawing asthma indication for all LABAs for the pediatric patients and removing asthma indication and contraindicating the use of single ingredient LABAs for all ages, DPAP preferred continued marketing of products containing LABAs and managing safety risk through labeling. Readers interested in this debate are referred to Kramer [22].

While we should continue to strive for the best approach possible, we should be mindful not to let the perfect become the enemy of the good. In my opinion, we should start experimenting with quantitative benefit/risk assessment using a common framework, sharing experience from these efforts collectively and deciding the next step among partnerships with academia, regulatory agencies, and the pharmaceutical industry.

Acknowledgements The author wants to thank Jon Norton for the use of Fig. 2. The author also wants to thank PhRMA BRAT, especially Bennett Levitan, Paul Coplan, Rebecca Noel, Marilyn Metcalf, and Diana Hughes, for BRAT-generated materials. In addition, the author wants to thank Leila Zelnick and Tom Fleming for their comments which have helped improve the quality of the paper.

References

1. Honig PK (2011) Comparative effectiveness: the fourth hurdle in drug development and a role for clinical pharmacology. *Clin Pharmacol Ther* 89(2):151–156
2. Gelber RD, Gelman RS, Goldhirsch A (1989) A quality of-life oriented endpoint for comparing treatments. *Biometrics* 45:781–795
3. Glasziou PP, Simes RJ, Gelber RD (1990) Quality adjusted survival analysis. *Stat Med* 9: 1259–1276
4. Hughes DA, Bayoumi AM, Pirmohamed M (2007) Current assessment of risk-benefit by regulators: is it time to introduce decision analyses? *Clin Pharmacol Ther* 82:123–127
5. Irish W, Sherrill B, Cole B, Gard C, Gledennig GA, Mouridsen H (2005) Quality-adjusted survival in a crossover trial of letrozole versus tamoxifen in postmenopausal women with advanced breast cancer. *Ann Oncol* 10:1458–1462
6. Chuang-Stein C (1994) A new proposal for benefit-less-risk analysis in clinical trials. *Control Clin Trials* 15:30–43
7. Ouellet D, Werth J, Parekh N, Feltner D, McCarthy B, Lalonde RL (2009) The use of a clinical utility index to compare insomnia compounds: a quantitative basis for benefit-risk assessment. *Clin Pharmacol Ther* 85:277–282

8. Chuang-Stein C, Mohberg NR, Sinkula M (1991) Three measures for simultaneously evaluating benefits and risks using a categorical data from clinical trials. *Stat Med* 10:1349–1359
9. Entsua R, Gorman JM (2002) Global benefit-risk assessment of antidepressants: venlafaxine XR and fluoxetine. *J Psychiatr Res* 36:111–118
10. Entsua R, Gao B (2002) Global benefit-risk evaluation of antidepressant action: comparison of pooled data for venlafaxine, SSRIs and placebo. *CNS Spectr* 7:882–888
11. Pritchett Y, Tamura R (2008) The application of global benefit-risk assessment in clinical trial design and some statistical considerations. *Pharm Stat* 7:170–178
12. Norton J (2010) A longitudinal model for medical benefit-risk analysis, with case study. Presented at the 19th Annual International Chinese Statistical Association Applied Statistics Symposium, Indianapolis, IN, June 20–23
13. Norton J (2011) A longitudinal model and graphic for benefit-risk analysis, with case study. *Drug Inf J* 45:741–747
14. Coplan PM, Noel RA, Levitan BS, Ferguson J, Mussen F (2011) Development of a framework for enhancing the transparency, reproducibility and communication of the benefit-risk balance of medicines. *Clin Pharmacol Ther* 89:312–315
15. Levitan BS, Andrews EB, Gilseman A, Ferguson J, Noel RA, Coplan PM, Mussen F (2011) Application of the BRAT framework to case studies: observations and insights. *Clin Pharmacol Ther* 89:217–224
16. Mussen F, Salek S, Walker S (2007) 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* 16(Suppl 1):S2–S15
17. Mussen F, Salek S, Walker S (2007) A quantitative approach to benefit-risk assessment of medicines – part 2: the practical application of a new model. *Pharmacoepidemiol Drug Saf* 16(Suppl 1):S16–S41
18. Temple R (2007) Quantitative decision analysis: a work in progress. *Clin Pharmacol Ther* 82:127–130
19. Chuang-Stein C, Entsua R, Pritchett Y (2008) Measures for conducting comparative benefit-risk assessment. *Drug Inf J* 42:223–233
20. O’Neill R (2008) A perspective on characterizing benefits and risks derived from clinical trials: can we do more? *Drug Inf J* 42:235–245
21. Hoerl RW, Sneer RD (2010) Statistical engineering: an idea whose time has come? A discussion in honor of Gerry Hahn’s 80th birthday. Presented at the Joint Statistical Meetings, Vancouver British Columbia, August 1–5
22. Kramer JM (2009) Balancing the benefits and risks of inhaled long-acting beta-agonists – the influence of values. *N Engl J Med* 360:1592–1595
23. Wiviott SD, Braunwald E, McCabe CH, Montalescot G, Ruzyllo W, Gottlieb S, Neumann FJ, Ardissino D, De Servi S, Murphy SA, Riesmeyer J, Weerakkody G, Gibson CM, Antman EM (2007) TRITON-TIMI 38 Investigators. Prasugrel versus clopidogrel in patients with acute coronary syndromes. *N Engl J Med* 357:2001–2015