

Quantifying Benefit–Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature

A. Brett Hauber · Angelyn O. Fairchild ·
F. Reed Johnson

Published online: 1 May 2013
© Springer International Publishing Switzerland 2013

Abstract Decisions regarding the development, regulation, sale, and utilization of pharmaceutical and medical interventions require an evaluation of the balance between benefits and risks. Such evaluations are subject to two fundamental challenges—measuring the clinical effectiveness and harms associated with the treatment, and determining the relative importance of these different types of outcomes. In some ways, determining the willingness to accept treatment-related risks in exchange for treatment benefits is the greater challenge because it involves the individual subjective judgments of many decision makers, and these decision makers may draw different conclusions about the optimal balance between benefits and risks. In response to increasing demand for benefit–risk evaluations, researchers have applied a variety of existing welfare-theoretic preference methods for quantifying the tradeoffs decision makers are willing to accept among expected clinical benefits and risks. The methods used to elicit benefit–risk preferences have evolved from different theoretical backgrounds. To provide some structure to the literature that accommodates the range of approaches, we begin by describing a welfare-theoretic conceptual framework underlying the measurement of benefit–risk preferences in pharmaceutical

and medical treatment decisions. We then review the major benefit–risk preference-elicitation methods in the empirical literature and provide a brief overview of the studies using each of these methods. The benefit–risk preference methods described in this overview fall into two broad categories: direct-elicitation methods and conjoint analysis. Rating scales (6 studies), threshold techniques (9 studies), and standard gamble (2 studies) are examples of direct elicitation methods. Conjoint analysis studies are categorized by the question format used in the study, including ranking (1 study), graded pairs (1 study), and discrete choice (21 studies). The number of studies reviewed here demonstrates that this body of research already is substantial, and it appears that the number of benefit–risk preference studies in the literature will continue to increase. In addition, benefit–risk preference-elicitation methods have been applied to a variety of healthcare decisions and medical interventions, including pharmaceuticals, medical devices, surgical and medical procedures, and diagnostics, as well as resource-allocation decisions such as facility placement. While preference-elicitation approaches may differ across studies, all of the studies described in this review can be used to provide quantitative measures of the tradeoffs patients and other decision makers are willing to make between benefits and risks of medical interventions. Eliciting and quantifying the preferences of decision makers allows for a formal, evidence-based consideration of decision-makers' values that currently is lacking in regulatory decision making. Future research in this area should focus on two primary issues—developing best-practice standards for preference-elicitation studies and developing methods for combining stated preferences and clinical data in a manner that is both understandable and useful to regulatory agencies.

A. Brett Hauber (✉) · A. O. Fairchild · F. Reed Johnson
RTI-Health Solutions, 200 Park Office Drive,
Research Triangle Park, NC 27709, USA
e-mail: abhauber@rti.org

Key Points for Decision Makers

- Regulatory approval decisions for medical interventions require both clinical evidence and subjective judgments.
- Eliciting and quantifying the preferences of decision makers is necessary for a formal, evidence-based consideration of decision-makers' values that is currently lacking in regulatory decision making.
- A growing body of literature provides quantitative estimates of benefit–risk preferences that can be used to evaluate the relative importance of clinical benefits and risks in drug-approval decisions.
- Future research in this area should focus on two primary issues—developing best-practice standards for preference-elicitation studies and developing methods for combining stated preferences and clinical data in a manner that is both understandable and useful to regulatory agencies.

1 Background

Decisions regarding the development, regulation, sale, and utilization of pharmaceutical and medical interventions require an evaluation of the balance between benefits and risks. Such evaluations are subject to two fundamental challenges: measuring the clinical effectiveness and harms associated with the intervention, and determining the relative importance of these different types of outcomes. Clinical measurement challenges include accounting for multiple potential efficacy and safety endpoints, large uncertainty about the likelihood, causality, reversibility, and latency of low-frequency adverse events, and heterogeneity of effects among patient subgroups. Because the health and financial consequences of decisions based on these evaluations can be enormous, substantial resources are devoted to improving the measurement of treatment-related effectiveness and harms. However, even if clinical effectiveness and harms could be measured perfectly, the fundamental problem of determining whether decision makers are willing to accept the risks of an intervention in order to achieve the expected benefits of that intervention remains.

In some ways, determining the willingness to accept treatment-related risks in exchange for treatment benefits is the greater challenge because it involves the individual subjective judgments of many decision makers, and these decision makers may draw different conclusions about the optimal balance between benefits and risks. In drug-approval decisions, the fundamental question facing

decision makers is what likelihood of adverse events is acceptable for a treatment that offers substantial improvements in efficacy relative to the current standard of care. For some decision makers, a relatively low rate of effectiveness in a sufficiently large population could offset a given adverse-event risk, while for others, any risk of a serious treatment-related adverse event would make a treatment unacceptable.

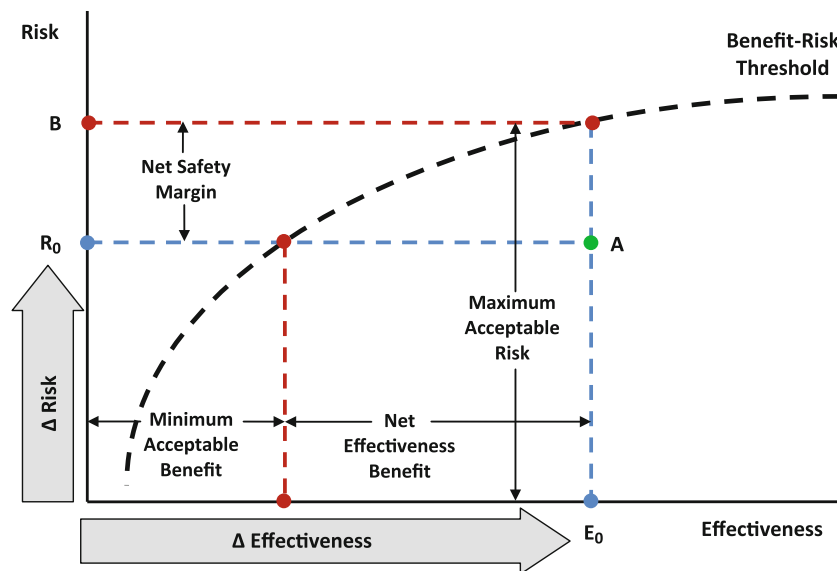
In response to increasing demand for benefit–risk evaluations, researchers have applied a variety of existing welfare-theoretic preference methods to quantify the tradeoffs decision makers are willing to accept among expected clinical benefits and risks. The primary objective of this review is to provide an overview of the growing empirical literature quantifying benefit–risk tradeoff preferences for pharmaceutical and medical interventions. The methods used to elicit benefit–risk preferences have evolved from different theoretical backgrounds in health economics, the decision sciences, and shared decision making. To provide some structure to the literature that accommodates the range of approaches, we begin by describing a welfare-theoretic conceptual framework underlying the measurement of benefit–risk preferences in pharmaceutical and medical treatment decisions. We then review the major benefit–risk preference-elicitation methods in the empirical literature and provide a brief overview of the studies using each of these methods.

2 Conceptual Framework

Neoclassical welfare theory posits that an individual makes choices among competing goods and services to maximize his or her own wellbeing or utility. Integral to this theory is that people have well-defined preferences for different goods and services or states of the world, and that these preferences are revealed by the choices they make. Further, goods and services are differentiated, meaning that they have varying characteristics over which people have preferences and among which people are willing to trade. In many markets, people reveal their preferences by making actual choices. In other markets, real-world choices are limited by regulation, institutional arrangements, product availability, or other market-limiting factors. However, this does not mean that individuals do not have preferences for products or product characteristics that are unavailable in the traditional market sense. Such is the case with health and healthcare.

In 1998, the Council for International Organizations of Medical Sciences (CIOMS) Working Group IV presented a simple graph to represent the relationships among the benefits and risks of pharmaceutical products [1]. In this graph, benefits are plotted on the horizontal axis, risks are

Fig. 1 Quantitative measures of benefit–risk tradeoffs—a framework for conceptualizing and contextualizing quantitative measures of tradeoffs between treatment-related benefits and risks. A is the observed level of benefit and risk for treatment A ; E_0 is the effectiveness of treatment A ; R_0 is the risk of treatment A ; B is the maximum acceptable risk of treatment A



plotted on the vertical axis, and a threshold is plotted above which risks outweigh benefits and below which benefits outweigh risks. The Working Group described the threshold separating these two conclusions in terms of judgments and tradeoffs. Neoclassical welfare provides a method for quantifying those judgments and tradeoffs and, therefore, a method for estimating the threshold in any given situation.

Lynd and O'Brien [2] proposed a similar approach to comparing benefits and risks in an evaluation of prophylaxis of deep vein thrombosis (DVT). Specifically, these authors presented a benefit–risk plane in which a benefit (the rate of DVT) is measured on the horizontal axis and a risk (the rate of major bleeding) is measured on the vertical axis. The authors demonstrated a simulation method for using data from randomized controlled trials to plot a joint density function of benefits and risks. In addition, Lynd and O'Brien describe how to evaluate these results relative to different hypothetical thresholds. The authors alluded to using preferences of patients or the general public as a source of data to estimate such as threshold; however, they did not estimate a specific threshold in this application.

The CIOMS IV report took a regulatory risk-management perspective, implying that the threshold should reflect preferences or judgments of regulators on behalf of the public. Lynd and O'Brien suggested that preferences of patients or the general public could be used to define the threshold. Incorporating such preferences requires a principled framework for defining and quantifying values for the outcomes of interest. Neoclassical welfare theory is the most widely accepted conceptual framework for analyzing decisions involving the allocation of scarce resources. That theory requires that the threshold reflect the preferences of individual consumers of the intervention; that is, patients. We use a single general welfare-theoretic framework based

on patient preferences that can be applied to multiple decision contexts. Therefore, in the remainder of this section, we refer generally to decision makers, whether they are patients themselves, healthcare providers, or regulators.

Figure 1 presents a benefit–risk threshold that describes the benefit–risk tradeoffs decision makers are willing to accept. Similar to both the CIOMS IV framework and that presented by Lynd and O'Brien, treatment effectiveness is measured on the horizontal axis and treatment-related harms are measured on the vertical axis. For the purpose of this review, we assume that treatment benefits and risks correspond to defined beneficial and harmful outcomes. From the patients' perspective prior to initiating therapy, both beneficial and harmful outcomes are probabilistic. Clinical data can indicate mean or most likely outcomes, but decision makers must evaluate options with the understanding that a particular patient's experience could be quite different from the average or typical outcome. For the purpose of describing this framework, we define the benefit–risk threshold as the maximum risk of harm decision makers would accept for a realized improvement in a beneficial outcome. Thus, the horizontal axis is an index of that improvement and the vertical axis is the associated risk or probability of harm.

The benefit–risk threshold is thus the locus of combinations of effectiveness and the probability of harm, which are equally as acceptable as standard of care or another reference condition [3]. Points below the threshold indicate combinations that are better than the reference condition, while points above the threshold indicate combinations that are worse than the reference condition. The threshold slopes upward because decision makers are averse to bearing risk, and thus greater risk must be sufficiently compensated with better efficacy. The threshold is likely to

be nonlinear because of the law of diminishing marginal utility; that is, we expect the benefit–risk threshold to increase at a decreasing rate with increases in treatment effectiveness.

The benefit–risk threshold can be used to construct several quantitative measures of benefit–risk preferences. Consider treatment A, with effectiveness E_0 and risk R_0 . Point B on the vertical axis corresponds to the maximum acceptable risk (MAR) the decision maker is willing to accept in exchange for the treatment benefit represented by E_0 . The difference between MAR and the observed or expected risk is therefore a measure of the net safety margin; that is, the risk in excess of the observed or expected risk that a decision maker would be willing to accept in exchange for the effectiveness of treatment A.

Alternatively, benefits can be described relative to a fixed level of risk. For the observed or expected level of risk (R_0), there is a corresponding minimum level of benefit required for a decision maker to accept this level of risk. This minimum acceptable benefit (MAB) is measured as the horizontal distance between the vertical axis and the level of effectiveness corresponding to the point on the benefit–risk threshold at R_0 . Because treatment A is to the right of the benefit–risk threshold, this treatment yields a net effectiveness benefit equal to the difference between the observed or expected effectiveness and MAB; that is, the benefit of treatment A in excess of the minimum benefit required for a decision maker to accept a risk of R_0 .

3 Benefit–Risk Preference-Elicitation Methods and Examples from the Literature

We conducted a search of the existing empirical literature quantifying risk–benefit preferences for medical interventions to understand the range of methods that have been employed and to assess the size of the empirical literature on this topic. Using PubMed, we searched for studies published before May 2012 with abstracts containing combinations of the keywords “maximum acceptable,” “risk,” “benefit,” “health,” “tradeoff,” “preferences,” “valuation,” and “willing to accept.” We screened the >3,000 studies that resulted from this search and retained 131 studies that appeared to report one or more of the quantitative measures of benefit–risk preferences described in Fig. 1. Although these measures have been given different names in different studies, the terminology used in the description of the conceptual framework is consistent with that used by the majority of benefit–risk preference studies. When different terminology was used in the literature to describe the same measures, these studies were included. A review of these abstracts yielded 40 studies that met these criteria. While our coverage of the literature is thorough, it is not intended to be

exhaustive. There are many more studies that estimate relative preferences for benefits and risks of treatment but that do not quantify benefit–risk tradeoffs explicitly. This review is limited to those studies that quantify one of the measures described in Fig. 1.

A variety of preference-elicitation methods have been used to quantify benefit–risk preferences. These methods fall into two broad categories: direct-elicitation methods that require decision makers to explicitly identify their MAR or MAB at a single point on the benefit–risk threshold, and conjoint-analysis methods with which benefit–risk preferences are inferred based on patterns of responses to a series of questions. Rating scales, threshold techniques, and standard gamble are examples of direct elicitation methods. Conjoint analysis studies are categorized by the question format used in the study, including ranking, graded pairs, and discrete choice. Figure 2 provides a list of empirical studies quantifying benefit–risk preferences grouped by preference-elicitation methodology.

3.1 Direct-Elicitation Methods

Direct-elicitation methods present respondents with a hypothetical medical intervention. Respondents then are asked to indicate the amount of risk they would be willing to accept to achieve the benefits of the intervention (MAR) or the amount of benefit they would require to accept a medical intervention with known risks (MAB). Each direct-elicitation task yields a single point on the benefit–risk threshold, because each direct-elicitation task involves eliciting either MAR or MAB for one medical intervention. To recover multiple points on the benefit–risk threshold, the benefit or the risk associated with the medical intervention is varied and the direct-elicitation task is repeated for each new scenario. Although there are statistical techniques that can be used to analyze these data, nonparametric measures typically are used to summarize the results of direct-elicitation questions.

Three direct-elicitation methods have been used to elicit benefit–risk tradeoffs associated with medical interventions. These include rating-scale, threshold-technique, and standard-gamble methods. While all of these direct-elicitation methods ask each respondent to consider a single target medical intervention relative to a reference condition or standard of care and state explicitly his or her MAR or MAB, they differ in the way that the direct-elicitation task is constructed. We consider each of these direct-elicitation methods below.

3.1.1 Rating Scales

The rating-scale method is the most direct method of eliciting benefit–risk preferences. Respondents are

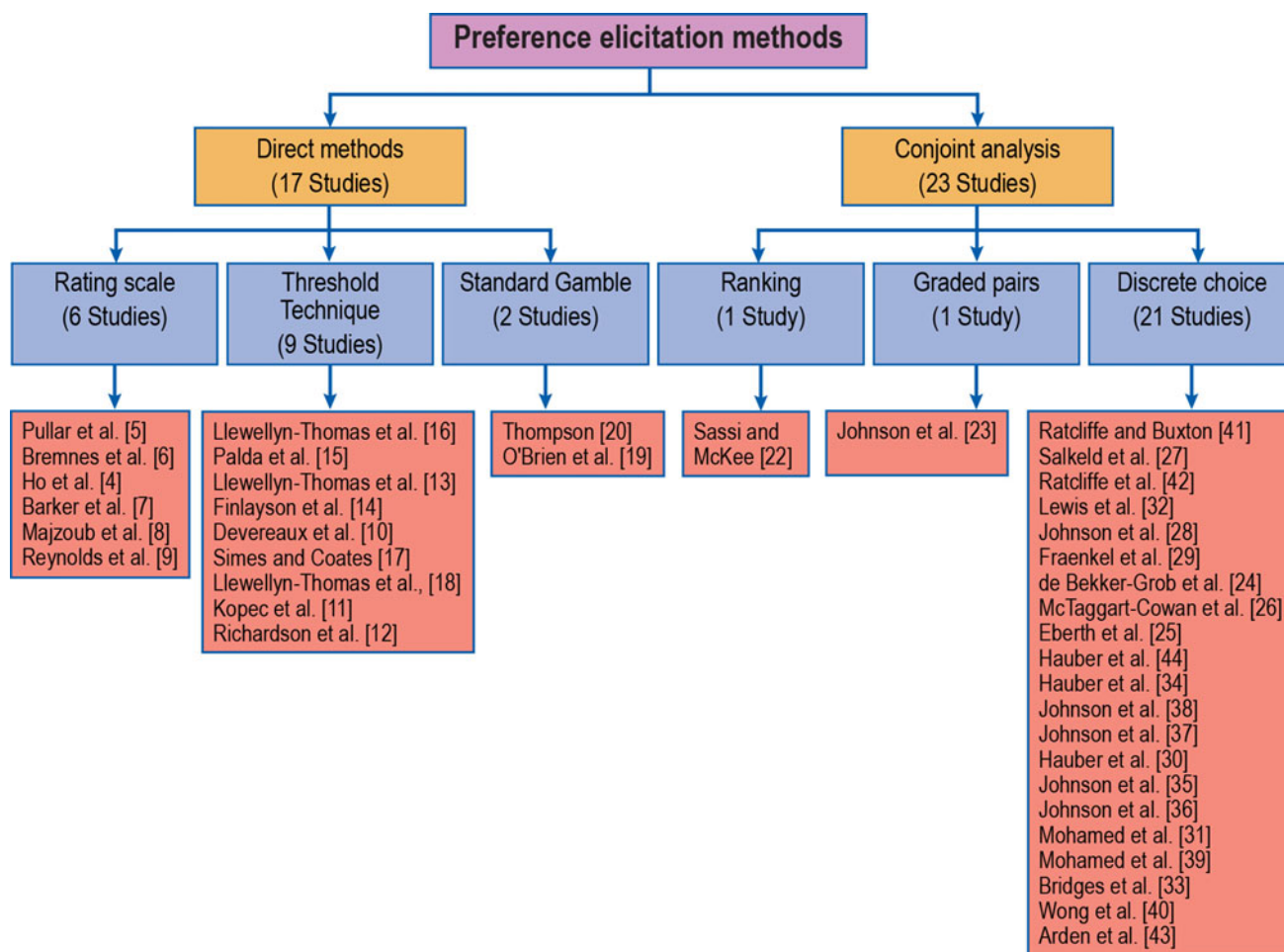


Fig. 2 List of empirical studies—empirical literature quantifying benefit–risk preferences grouped by preference-elicitation method. Within each method, studies are presented in chronological order

presented with a description of a medical intervention. The description defines the salient characteristics of the benefits or risks of the intervention; however, when the benefits are defined, the probability of one adverse event is left undefined. Likewise, when the risks are defined, the likelihood of achieving a given benefit is left undefined. Respondents then use a rating scale to indicate the highest probability of experiencing the given adverse event they would accept to achieve the defined benefit or the lowest probability of the given benefit they would require to accept the risks of the intervention.

We identified six empirical studies that used a rating-scale technique to elicit benefit–risk preferences. Ho et al. asked patients with rheumatoid arthritis to indicate the maximum risk of death they would accept in exchange for four different levels of treatment benefit [4]. Pullar et al. [5] used a rating scale to elicit the maximum acceptable risk of three different adverse events that rheumatologists and rheumatology patients would be willing to accept to achieve each of five levels of benefit, defined by the

likelihood of a cure, the likelihood of symptom improvement, and the likelihood of no further deterioration. Bremnes et al. [6] estimated the minimum acceptable probability of cure or the minimum acceptable level of symptom relief required for cancer patients, doctors, nurses, and healthy subjects to accept a toxic chemotherapy regimen.

The authors of three additional studies described their methods as standard gamble; however, each of these studies used rating scales to elicit benefit–risk preferences. These three studies quantified the maximum acceptable risk of decreased longevity due to immunosuppressant drugs used to prevent rejection after different types of transplants and the maximum risk of transplant rejection that patients would accept from a transplant procedure [7–9]. All three studies used questions from the Louisville Instrument for Transplantation, which asked respondents to circle the probability corresponding to the maximum risk of rejection a patient would tolerate and still choose to have a transplant.

3.1.2 Threshold technique

The threshold technique (also referred to as the probability tradeoff technique and the probability threshold technique) presents respondents with a pair of medical interventions, each of which is defined by its salient characteristics. One intervention is the target intervention or intervention of interest. The other is referred to as the reference intervention. Respondents then are asked to indicate which medical intervention they prefer. Depending on the objectives of the study, one characteristic is then varied such that the preferred alternative is made unambiguously less attractive or the alternative that was not chosen is made unambiguously more attractive and the question is repeated. The probability of benefit or risk is varied systematically until a respondent changes his or her choice. The probability of benefit or risk that induces the respondent to switch provides a point estimate of the minimum acceptable benefit (often referred to as “minimal required benefit” in threshold-technique studies) or maximum acceptable risk (often referred to as “maximal acceptable risk” in threshold-technique studies) of the target intervention, respectively.

We identified nine empirical studies that used the threshold technique to elicit benefit–risk preferences. Devereaux et al. [10] calculated the minimum decrease in the probability of a stroke and the maximum increase in the probability of bleeding that patients would accept for prescription and non-prescription treatments for atrial fibrillation. Kopec et al. [11] and Richardson et al. [12] estimated the maximum acceptable risk of stomach bleed, heart attack or stroke that patients and physicians would be willing to accept in exchange for two different reductions in osteoarthritis pain. Llewellyn-Thomas et al. [13] also estimated risk tolerance for improvements in arthritis symptoms. Other studies used the threshold technique to calculate patients’ maximum acceptable risk of death associated with variations in the location of the treatment facility during pancreatic cancer resections [14], the maximum acceptable incremental risk of breast cancer recurrence that patients would accept in order to avoid adverse events [15], the minimum acceptable efficacy for different treatments for benign prostatic hyperplasia [16], and the minimum required increase in 5-year survival necessary for patients to accept adjuvant chemotherapy for early-stage breast cancer [17]. Finally, Llewellyn-Thomas et al. [18] estimated the required reduction in the risk of cardiovascular events resulting from cholesterol-lowering and antihypertensive medications with known potential risks or side effects.

3.1.3 Standard Gamble

In standard gamble, respondents are presented with a choice between a certain health state and a gamble with

two possible outcomes—one better (often perfect health) and one worse (often death) than the certain health state. The probability of the gamble resulting in the worse health state is increased systematically until respondents are indifferent between the certain health state and the gamble. The standard gamble technique typically is used to estimate health-state utilities, and 1 minus the probability at which the respondent is indifferent between the certain health state and the gamble is the utility of the particular health state. O’Brien et al. [19] applied this framework to benefit–risk preference analysis by defining the probability in the gamble to be the risk of a fatal side effect of an intervention that would improve a person’s health from their current health state to a better health state. Using this approach, O’Brien et al. estimated the level of fatal treatment-related risk that patients with rheumatoid arthritis would be willing to accept for a treatment that would yield one of four possible health-state improvements. Thompson [20] used a similar approach to estimate the maximum acceptable risk of a fatal adverse event that patients would accept for a treatment that cured their rheumatoid arthritis.

3.2 Conjoint-Analysis Methods

Conjoint-analysis methods present respondents with two or more hypothetical profiles. Each profile represents a medical intervention and is defined by multiple attributes, each of which represents a salient feature of the medical intervention. Each profile in the set is defined by the same attributes, but the levels of these attributes vary across the alternatives. Respondents are asked to indicate their preferences from among the alternatives presented by rating or ranking the alternatives or by choosing the most preferred alternative. The task is then repeated with different profiles according to an experimental design. By analyzing the pattern of responses to these questions, the rate at which respondents are willing to trade off among the attributes and attribute levels can be quantified. When benefits and risks are included among the attributes, the tradeoff results can be used to calculate MAR or MAB. Because the levels of each benefit and risk attribute vary over a range, the results of a conjoint-analysis study can be used to recover the benefit–risk threshold over that range.

Conjoint analysis data typically are analyzed using limited dependent variable analysis, and these analyses can be quite sophisticated. A detailed review of the statistical methods required to analyze conjoint-analysis data is beyond this scope of this paper; however, the type of analysis required is driven largely by the question format. We identified three primary conjoint question formats that have been used to elicit benefit–risk preferences. These included ranking, graded pairs, and discrete-choice experiments. While all of these methods ask each respondent

to indicate their preferences among sets of alternatives, they differ in the way that the task is presented. We consider each of these conjoint-analysis methods below.

Our search of the literature returned 23 studies that use conjoint analysis to estimate explicitly the tradeoffs decision makers are willing to make among benefits and risks of medical interventions. Other conjoint studies estimate the relative importance of benefits and risks of medical interventions, and the resulting data likely could be used to estimate MAR or MAB; however, these studies do not present the rate at which respondents are willing to tradeoff among benefits and risks, nor do they present any of the benefit–risk measures presented in Fig. 1. For example, Fraenkel et al. [21] used conjoint analysis to elicit women’s preference weights for characteristics of cytotoxic medications for lupus, including probabilistic benefits and risks; however, these authors did not explicitly calculate the rates at which patients would be willing to accept tradeoffs among benefits and risks, and thus did not estimate any of the measures presented in the conceptual framework in Fig. 1.

3.2.1 Ranking

In a ranking experiment, respondents are shown several profiles and asked to rank them from most preferred to least preferred. Our search returned only one study which used a ranking experiment to elicit benefit–risk preferences. Sassi et al. [22] presented vascular surgeons and cardiologists with 10 hypothetical diagnostic strategies that featured positive predictive value, negative predictive value, test-related morbidity, and novelty of technology. Respondents were asked to order these diagnostic strategies from best to worst for a hypothetical patient profile. The authors estimated the minimum acceptable negative and positive predictive values that respondents would accept in exchange for a 0.1 % increase in test-related morbidity.

3.2.2 Graded Pairs

Graded pairs is a question format in which respondents are presented with two profiles and asked to indicate their strength of preference for one of the two alternatives. Typically, this is done by presenting the respondent with a rating scale in which the ends of the scale can be used to indicate a strong preference for an alternative, the middle rating represents indifference between the alternatives, and the intermediate points can be used to indicate somewhat weaker preferences for the alternatives. Only one study in our review, Johnson et al. [23], utilized this method; the authors find that menopausal women would be willing to accept increased risks of serious adverse events, including heart attack and breast cancer, in exchange for better symptom relief when taking hormone-replacement therapy.

In addition, these authors estimate different risk tolerances for risks that are presented as absolute risks and those that are presented as relative risks.

3.2.3 Discrete-Choice Experiments

The majority of conjoint-analysis studies employ a discrete-choice format. In a discrete-choice experiment, respondents are asked to choose the most-preferred alternative from a set of profiles, assuming that these are the only alternatives available. Some discrete-choice experiments allow respondents to opt out; that is, to indicate that they prefer no medical intervention to the treatment alternatives presented in the choice task. Other discrete-choice experiments require that respondents choose from among a set of medical-intervention profiles, even if they would prefer no medical intervention to the options presented in the choice task.

Our search returned four studies that allowed respondents to select among two profiles or an opt-out alternative. De Bekker-Grob [24] used this approach to estimate women’s preferences for osteoporosis drugs and found that women were willing to accept treatment-related nausea if the treatment resulted in a 40 % reduction in the risk of hip fracture. Eberth et al. [25] estimated the rate at which men with benign prostatic hyperplasia were willing to accept treatment-related side effects and risks to reduce prostate size. McTaggart-Cowan et al. [26] found that asthma patients were willing to accept fewer symptom-free days to reduce treatment-related side effects, and Salkeld et al. [27] estimated the maximum acceptable risk of a false-positive screening result that older adults would be willing to accept to prevent one death from colorectal cancer.

The most common question format used in discrete choice experiments presents respondents with a forced choice in which respondents are asked to choose from among a set of treatment alternatives. We identified 17 studies that used this question format. Johnson et al. [28] estimated the maximum levels of three fatal risks that patients with Crohn’s disease would be willing to accept for different levels of treatment benefit. Fraenkel et al. [29] estimated the minimum acceptable reduction in fracture risk that women with osteoporosis would expect from a treatment that requires daily injections. Four other studies used a forced-choice discrete-choice experiment to estimate MAB. Hauber et al. [30] estimated MAB among patients for treatments for idiopathic thrombocytopenia purpura, Mohamed et al. [31] estimated patients’ and physicians’ MAB for hepatitis B treatments, Lewis et al. [32] estimated physicians’ and midwives’ MAB for prenatal tests for Down’s syndrome, and Bridges et al. [33] estimated the minimum additional years of progression-free survival required for patients with non-small-cell lung cancer to accept various treatment-related adverse events.

In 11 additional studies, researchers estimated MAR in various therapeutic areas, including Alzheimer's disease [34], irritable bowel syndrome [35], Crohn's disease [36, 37], multiple sclerosis [38], renal cell carcinoma [39, 40], liver transplantation [41], osteoarthritis [42, 43], and HIV and AIDS [44].

4 Discussion

In the past two decades, legislators and regulators in the US and European Union have expressed increasing interest in formal benefit–risk evaluation for pharmaceutical and medical interventions. In recent years, there also has been increasing interest in incorporating the perspectives of patients into regulatory decisions [45, 46]. At the same time, researchers from different disciplines have used different methods to quantify the benefit–risk tradeoff preferences of patients, physicians, and other healthcare decision makers. Benefit–risk preference-elicitation methods have been applied to different types of medical interventions, including pharmaceuticals, medical devices, surgical and medical procedures, and diagnostics, as well as resource-allocation decisions such as facility placement. While preference-elicitation approaches may differ across studies, all of the studies described in this overview of the literature can be used to provide quantitative measures of the tradeoffs patients and other decision makers are willing to make between the benefits and risks of medical interventions.

As noted above, this overview of the literature was not intended to identify every study used to estimate benefit–risk preferences. In addition, the studies described in this overview are limited to those that provide only a subset of possible quantitative benefit–risk measures. Therefore, the collection of studies described in this overview may not include some studies that contribute important information about the benefit–risk tradeoffs decisions makers are willing to make when considering a medical intervention. However, even with these caveats, this overview of the literature demonstrates that this body of research already is substantial, and it appears that the number of benefit–risk preference studies in the literature will continue to increase.

All of the studies described here can be viewed based on a common conceptual framework consistent with the one presented in Fig. 1. The framework is based on the fundamental axioms of economic theory and is flexible enough to accommodate a broad range of risks and benefits associated with a medical intervention. As demonstrated by the range of studies presented here, it also can be applied to multiple decision makers, including both patients and healthcare providers. However, while the framework includes both a benefit–risk threshold and the observed or expected benefits and risks associated with an intervention,

most of the studies described here estimated only one or more points on the benefit–risk tradeoff curve. Few of these studies compared risk tolerance estimates to actual clinical data [5, 28, 29], a necessary condition if the results of these studies are to be used to inform regulatory benefit–risk decisions.

The simplest method of combining benefit–risk preference information and clinical results in benefit–risk decisions is to plot the clinical data corresponding to point A in Fig. 1 relative to the benefit–risk threshold. As described above, Lynd and O'Brien [2] suggested one method for plotting this point. Holden [47] proposed an alternative approach to evaluating the benefit–risk balance of a medical intervention by combining preference data with clinical outcomes data. Specifically, Holden suggested applying a relative preference weight to a clinical outcome to estimate a weighted measure of number needed to treat (NNT) or number needed to harm (NNH).

Interestingly, the framework presented in Fig. 1 can be used to estimate weighted NNT and weighted NNH. Specifically, for any given level of efficacy, MAR can be interpreted as the ratio of the maximum acceptable number of adverse events (N_{AE}) to the size of the exposed population (N). Number needed to harm (NNH) is a measure of the number of people that need to be exposed to a treatment to yield one case of the adverse event assuming a constant therapeutic benefit. Thus, the inverse of MAR ($1/MAR$) is a measure of the minimum acceptable number needed to harm that can be compared directly with the clinical measure of NNH. For example, if clinical data reveal that the NNH for a treatment-related cancer is 1,000, that means that after 1,000 exposures to the treatment, researchers would expect to see one case of treatment-related cancer. If, given the expected benefits of treatment, decision makers would accept a maximum level of risk of only 0.067 % (or less than 1 in 1,000) then the level of treatment-related cancer risk would be too high. Conversely, MAB represents the minimum level of efficacy that patients require to accept a given level of risk. Thus, for any given level of adverse-event risk, MAB is the ratio of the minimum number of patients achieving a therapeutic benefit (N_B) to the total size of the treated population (N). Thus, the inverse of MAB ($1/MAB$) is a preference-based measure of the maximum acceptable number needed to treat that can be compared directly with the clinical measure NNT. For example, if clinical data reveal that the NNT for a treatment is 2, meaning that two exposures are required to see one beneficial outcome (a response rate of 50 %), and the MAB for that treatment given all its associated risks is 40 %, then the benefits of this treatment would outweigh the risks.

The approach implicit in the conceptual framework presented here and the NNH/NNT approach are applicable

when the critical benefit–risk decision involves comparing a single benefit outcome and a single adverse event. However, when a medical intervention or a set of medical interventions yield multiple beneficial outcomes or are associated with more than one adverse event, the evaluation becomes more complicated. Lynd et al. [48] combine clinical outcomes data and relative benefit–risk preference weights to estimate the net clinical benefits of a treatment for irritable bowel syndrome. The clinical data were estimated using a discrete-event simulation model, and preference weights from a conjoint-analysis study were applied to each outcome in the study. Because it can be used to estimate relative benefit–risk preference weights for multiple outcomes simultaneously, conjoint analysis probably is more appropriate than direct elicitation methods when estimating net clinical benefits for medical interventions associated with multiple benefits, multiple risks, or both.

One shortcoming of the methods described here is that most benefit–risk assessors are unlikely to be familiar with stated-preference methods, and may be unconvinced of their validity and reliability. The most serious limitation of these methods is that they employ judgments among hypothetical alternatives. Hypothetical choices do not have the same clinical, financial, and emotional consequences of actual choices. Successful elicitation of patient-preference data requires that the sample shares a common, clear understanding of the features they are evaluating. In the case of eliciting preferences including probabilistic outcomes, researchers also must cope with the generally low level of numeracy in the general population. Some well-established findings include [49, 50]: (1) people’s difficulty evaluating small probabilities, (2) the sensitivity of risk evaluations to characteristics of the risk itself, such as the timing, source, and familiarity of the risk, whether risk exposure is voluntary, and whether outcome is catastrophic, (3) framing effects, and (4) heuristics that individuals use to translate population risks into personal risks. Obtaining valid and reliable measures of patient preferences is challenging, and researchers will need to demonstrate that it is possible to quantify patient preferences and to subject the results to the same rigorous standards as those applied to clinical, epidemiological, and patient-reported outcomes data.

Benefit–risk evaluations inevitably require judgments about the relative importance of endpoints measured in clinical data. Currently, these judgments typically are made by clinical scientists who have no special expertise in making societal value judgments. Physician input into benefit–risk evaluations often is limited to small, unrepresentative samples. Physician experts often are members of advisory panels; however, these experts are few in number and may not be well suited to assessing the values of the population of physicians who recommend or conduct the medical interventions under consideration.

While there is growing agreement that benefit–risk evaluations should take the patient perspective into consideration, current evaluation practices do not require the values of patients to be quantified in drug approval decisions. However, because patients are both the potential beneficiaries of new treatments and also those who bear the risks associated with those treatments, their preferences arguably warrant consideration as well. In addition, patient representatives often are asked to provide anecdotes or opinions as part of the review process, but are unlikely to be representative of an entire patient population. Public outreach efforts may give voice to patient concerns about regulatory decisions; however, it is not obvious how such outreach efforts can and should influence decision making. Employing stated-preference methods to quantify benefit–risk preferences is consistent with the growing interest in greater patient involvement in health-care decision making.

5 Conclusions

Eliciting and quantifying the preferences of decision makers allows for a formal, evidence-based consideration of decision-makers’ values that currently is lacking in regulatory decision making. There is a substantial and growing literature describing stated preference methods which offer one option for gathering systematic, theoretically sound, and valid evidence of the values of patients and physicians, and should be considered in benefit–risk evaluations. Future research in this area should focus on two primary issues—developing best-practice standards for preference-elicitation studies, and developing methods for combining stated preferences and clinical data in a manner that is both understandable and useful to regulatory agencies.

Acknowledgments The authors received no direct funding for this manuscript, and declare no financial conflicts of interest.

Author Contributions A. Brett Hauber led all aspects of this study, including defining the objectives of the research, defining the search strategy, developing criteria for study inclusion, reviewing studies, and writing the manuscript. A. Brett Hauber also contributed to developing the conceptual model. Angelyn O. Fairchild conducted the literature search and contributed to defining the search strategy, developing criteria for study inclusion, reviewing studies, and writing the manuscript. F. Reed Johnson contributed to defining the objectives of the research and developing the conceptual model and writing the manuscript. A. Brett Hauber had access to all data and take full responsibility for the content of this manuscript.

References

1. Cioms, IV Working Group. Benefit–risk balance for marketed drugs: evaluating safety signals. Report of the CIOMS Working Group IV. Geneva: Council for International Organizations of Medical Sciences; 1998.

2. Lynd L, O'Brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. *J Clin Epidemiol.* 2004;57(8):795–803.
3. Van Houtven G, Johnson FR, Kilambi V, Hauber AB. Eliciting benefit–risk preferences and probability-weighted utility using choice-format conjoint analysis. *Med Decis Making.* 2011;31(3):469–80.
4. Ho M, Lavery B, Pullar T. The risk of treatment. A study of rheumatoid arthritis patients' attitudes. *Br J Rheumatol.* 1998;37(4):459–60.
5. Pullar T, Wright V, Feely M. What do patients and rheumatologists regard as an 'acceptable' risk in the treatment of rheumatic disease? *Br J Rheumatol.* 1990;29(3):215–8.
6. Bremnes RM, Andersen K, Wist EA. Cancer patients, doctors and nurses vary in their willingness to undertake cancer chemotherapy. *Eur J Cancer.* 1995;31A(12):1955–9.
7. Barker JH, Furr A, Cunningham M, Grossi F, Vasilic D, Storey B, et al. Investigation of risk acceptance in facial transplantation. *Plast Reconstr Surg.* 2006;118(3):663–70.
8. Majzoub RK, Cunningham M, Grossi F, Maldonado C, Banis JC, Barker JH. Investigation of risk acceptance in hand transplantation. *J Hand Surg Am.* 2006;31(2):295–302.
9. Reynolds CC, Martinez SA, Furr A, Cunningham M, Bumpous JM, Lentsch EJ, et al. Risk acceptance in laryngeal transplantation. *Laryngoscope.* 2006;116(10):1770–5.
10. Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. *BMJ.* 2001;323(7323):1218–22.
11. Kopec JA, Richardson CG, Llewellyn-Thomas H, Klinkhoff A, Carswell A, Chalmers A. Probabilistic threshold technique showed that patients' preferences for specific trade-offs between pain relief and each side effect of treatment in osteoarthritis varied. *J Clin Epidemiol.* 2007;60(9):929–38.
12. Richardson CG, Chalmers A, Llewellyn-Thomas HA, Klinkhoff A, Carswell A, Kopec JA. Pain relief in osteoarthritis: patients' willingness to risk medication-induced gastrointestinal, cardiovascular, and cerebrovascular complications. *J Rheumatol.* 2007;34(7):1569–75.
13. Llewellyn-Thomas HA, Arshinoff R, Bell M, Williams JI, Naylor CD. In the queue for total joint replacement: patients' perspectives on waiting times. Ontario Hip and Knee Replacement Project Team. *J Eval Clin Pract.* 1998;4(1):63–74.
14. Finlayson SR, Birkmeyer JD, Tosteson AN, Nease RF Jr. Patient preferences for location of care: implications for regionalization. *Med Care.* 1999;37(2):204–9.
15. Palda VA, Llewellyn-Thomas HA, Mackenzie RG, Pritchard KI, Naylor CD. Breast cancer patients' attitudes about rationing postlumpectomy radiation therapy: applicability of trade-off methods to policy-making. *J Clin Oncol.* 1997;15(10):3192–200.
16. Llewellyn-Thomas HA, Williams JI, Levy L, Naylor CD. Using a trade-off technique to assess patients' treatment preferences for benign prostatic hyperplasia. *Med Decis Making.* 1996;16(3):262–82.
17. Simes RJ, Coates AS. Patient preferences for adjuvant chemotherapy of early breast cancer: how much benefit is needed? *J Natl Cancer Inst Monogr.* 2001;30:146–52.
18. Llewellyn-Thomas HA, Paterson JM, Carter JA, Basinski A, Myers MG, Hardacre GD, Dunn EV, D'Agostino RB, Naylor CD. Primary prevention drug therapy: can it meet patients' requirements for reduced risk? *Med Decision Making.* 2002;22:326–39.
19. O'Brien BJ, Elswood J, Calin A. Willingness to accept risk in the treatment of rheumatic disease. *J Epidemiol Community Health.* 1990;44(3):249–52.
20. Thompson MS. Willingness to pay and accept risks to cure chronic disease. *Am J Public Health.* 1986;76(4):392–6.
21. Fraenkel L, Bodardus S, Wittnik DR. Understanding patient preferences for the treatment of lupus nephritis with adaptive conjoint analysis. *Med Care.* 2001;39(11):1203–16.
22. Sassi F, McKee M. Do clinicians always maximize patient outcomes? A conjoint analysis of preferences for carotid artery testing. *J Health Serv Res Policy.* 2008;13(2):61–6.
23. Johnson FR, Ozdemir S, Hauber B, Kauf TL. Women's willingness to accept perceived risks for vasomotor symptom relief. *J Womens Health.* 2007;16(7):1028–40.
24. de Bekker-Grob EW, Essink-Bot ML, Meering WJ, Pols HA, Koes BW, Steyerberg EW. Patients' preferences for osteoporosis drug treatment: a discrete choice experiment. *Osteoporos Int.* 2008;19(7):1029–37.
25. Eberth B, Watson V, Ryan M, Hughes J, Barnett G. Does one size fit all? Investigating heterogeneity in men's preferences for benign prostatic hyperplasia treatment using mixed logit analysis. *Med Decis Making.* 2009;29(6):707–15.
26. McTaggart-Cowan HM, Shi P, Fitzgerald JM, Anis AH, Kopec JA, Bai TR, et al. An evaluation of patients' willingness to trade symptom-free days for asthma-related treatment risks: a discrete choice experiment. *J Asthma.* 2008;45(8):630–8.
27. Salkeld G, Solomon M, Short L, Ryan M, Ward JE. Evidence-based consumer choice: a case study in colorectal cancer screening. *Aust N Z J Public Health.* 2003;27(4):449–55.
28. Johnson FR, Ozdemir S, Mansfield C, Hass S, Miller DW, Siegel CA, et al. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. *Gastroenterology.* 2007;133(3):769–79.
29. Fraenkel L, Gulanski B, Wittink D. Patient willingness to take teriparatide. *Patient Educ Couns.* 2007;65(2):237–44.
30. Hauber AB, Johnson FR, Grotzinger KM, Ozdemir S. Patients' benefit–risk preferences for chronic idiopathic thrombocytopenic purpura therapies. *Ann Pharmacother.* 2010;44(3):479–88.
31. Mohamed AF, Johnson FR, Hauber AB, Lescrauwaet B, Masterson A. Physicians' stated trade-off preferences for chronic hepatitis B treatment outcomes in Germany, France, Spain, Turkey, and Italy. *Eur J Gastroenterol Hepatol.* 2012;24(4):419–26.
32. Lewis SM, Cullinane FN, Bishop AJ, Chitty LS, Marteau TM, Halliday JL. A comparison of Australian and UK obstetricians' and midwives' preferences for screening tests for down syndrome. *Prenat Diagn.* 2006;26(1):60–6.
33. Bridges JF, Mohamed AF, Finfern HW, Woehl A, Hauber AB. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. *Lung Cancer.* 2012;77(1):224–31.
34. Hauber AB, Johnson FR, Fillit H, Mohamed AF, Leibman C, Arrighi HM, Grundman M, Townsend RJ. Older Americans' risk-benefit preferences for modifying the course of Alzheimer disease. *Alzheimer Dis Assoc Disord.* 2009;23(1):23–32.
35. Johnson FR, Hauber AB, Ozdemir S, Lynd L. Quantifying women's stated benefit–risk trade-off preferences for IBS treatment outcomes. *Value Health.* 2010;13(4):418–23.
36. Johnson FR, Hauber B, Ozdemir S, Siegel CA, Hass S, Sands BE. Are gastroenterologists less tolerant of treatment risks than patients? Benefit–risk preferences in Crohn's disease management. *J Manag Care Pharm.* 2010;16(8):616–28.
37. Johnson FR, Ozdemir S, Mansfield C, Hass S, Siegel CA, Sands BE. Are adult patients more tolerant of treatment risks than parents of juvenile patients? *Risk Anal.* 2009;29(1):121–36.
38. Johnson FR, Van Houtven G, Ozdemir S, Hass S, White J, Francis G, et al. Multiple sclerosis patients' benefit–risk preferences: serious adverse event risks versus treatment efficacy. *J Neurol.* 2009;256(4):554–62.

39. Mohamed AF, Hauber AB, Neary MP. Patient benefit–risk preferences for targeted agents in the treatment of renal cell carcinoma. *Pharmacoeconomics*. 2011;29(11):977–88.
40. Wong MK, Mohamed AF, Hauber AB, Yang JC, Liu Z, Rogerio J, Garay CA. Patients rank toxicity against progression free survival in second-line treatment of advanced renal cell carcinoma. *J Med Econ*. 2012;15(6):1139–48.
41. Ratcliffe J, Buxton M. Patients’ preferences regarding the process and outcomes of life-saving technology. An application of conjoint analysis to liver transplantation. *Int J Technol Assess Health Care*. 1999;15(2):340–51.
42. Ratcliffe J, Buxton M, McGarry T, Sheldon R, Chancellor J. Patients’ preferences for characteristics associated with treatments for osteoarthritis. *Rheumatology (Oxford)*. 2004;43(3):337–45.
43. Arden NK, Hauber AB, Mohamed AF, Johnson FR, Peloso PM, Watson DJ, et al. How do physicians weigh benefits and risks associated with treatments in patients with osteoarthritis in the United Kingdom? *J Rheumatol*. 2012;39(5):1056–63.
44. Hauber AB, Mohamed AF, Watson ME, Johnson FR, Hernandez JE. Benefits, risk, and uncertainty: preferences of antiretroviral-naive African Americans for HIV treatments. *AIDS Patient Care STDS*. 2009;23(1):29–34.
45. European Medicines Agency (2010) Road map to 2015. http://www.ema.europa.eu/docs/en_GB/document_library/Report/2011/01/WC500101373.pdf. Accessed 15 April 2012.
46. Maxmen A. Law spurs regulator to heed patients’ priorities. *Nature*. 2012;487(7406):154. doi:10.1038/487154a.
47. Holden WL. Benefit–risk analysis: a brief review and proposed quantitative approaches. *Drug Saf*. 2003;26(12):853–62.
48. Lynd LD, Naiafzadeh M, Colley L, Byrne MF, Willan AR, Sculpher MJ, Johnson FR, Hauber AB. 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(4):411–7.
49. Glickman TS, Gough M. Readings in risk. Washington, DC: Resources for the Future; 1990.
50. Kahneman D, Tversky A. Choices, values, and frames. Oxford: Cambridge University Press; 2000.