Use of Patient Preferences Data Regarding Multiple Risks to Inform Regulatory Decisions

J. Felipe Montano-Campos, Juan Marcos Gonzalez⬤, Timothy Rickert, Angelyn O. Fairchild⬤, Bennett Levitan, and Shelby D. Reed⬤

Abstract

Background and Objectives. Risk-tolerance measures from patient-preference studies typically focus on individual adverse events. We recently introduced an approach that extends maximum acceptable risk (MAR) calculations to simultaneous maximum acceptable risk thresholds (SMART) for multiple treatment-related risks. We extend these methods to include the computation and display of confidence intervals and apply the approach to 3 published discrete-choice experiments to evaluate its utility to inform regulatory decisions. **Methods.** We generate MAR estimates and SMART curves and compare them with trial-based benefit-risk profiles of select treatments for depression, psoriasis, and thyroid cancer. **Results.** In the depression study, SMART curves with 70% to 95% confidence intervals portray which combinations of 2 adverse events would be considered acceptable. In the psoriasis example, the asymmetric confidence intervals for the SMART curve indicate that relying on independent MARs versus SMART curves when there are nonlinear preferences can lead to decisions that could expose patients to greater risks than they would accept. The thyroid cancer application shows an example in which the clinical incidence of each of 3 adverse events is lower than the single-event MARs for the expected treatment benefit, yet the collective risk profile surpasses acceptable levels when considered jointly. **Limitations.** Nonrandom sample of studies. **Conclusions.** When evaluating conventional MARs in which the observed incidences are near the estimated MARs or where preferences demonstrate diminishing marginal disutility of risk, conventional MAR estimates will overstate risk acceptance, which could lead to misinformed decisions, potentially placing patients at greater risk of adverse events than they would accept. **Implications.** The SMART method, herein extended to include confidence intervals, provides a reproducible, transparent evidence-based approach to enable decision makers to use data from discrete-choice experiments to account for multiple adverse events.

Highlights

- Estimates of maximum acceptable risk (MAR) for a defined treatment benefit can be useful to inform regulatory decisions; however, the conventional metric considers one adverse event at a time.
- This article applies a new approach known as SMART (simultaneous maximum acceptable risk thresholds) that accounts for multiple adverse events to 3 published discrete-choice experiments.
- Findings reveal that conventional MARs could lead decision makers to accept a treatment based on individual risks that would not be acceptable if multiple risks are considered simultaneously.

Corresponding Author
Shelby D. Reed, Duke Clinical Research Institute, Duke University School of Medicine, 300 W Morgan Street, Durham, NC 27701, USA; (shelby.reed@duke.edu).

Introduction

Assessing whether treatment benefits outweigh risks of potential adverse events is central to regulatory and clinical decisions in health care. When considering whether to approve new treatments, regulators are tasked with identifying treatments that are both safe and effective.[1] They must judge whether multiple treatment-related risks are acceptable given a treatment's benefit. Similarly, physicians and patients engage in treatment decisions that seek to provide therapeutic benefits while minimizing risks, and patients make ongoing choices about continuing treatments based on their daily experience with multiple symptoms and side effects. In all these contexts, treatments are accepted or rejected as a bundle of benefits and risks.

Discrete-choice experiments (DCEs) and related methods are often used to quantify patients' and other stakeholders' health care preferences and particularly their willingness to accept risks, because these methods are uniquely capable of quantifying compensatory tradeoffs that are inherent in multifaceted decisions.[2,3] This versatility can be exploited to estimate relative preference weights for features of medical interventions based on stated choices that mirror real-world decisions. In DCE

Duke Clinical Research Institute, Duke University School of Medicine, Durham, NC, USA (JFMC, JMG, TR, AOF, SDR); Department of Population Health Sciences, Duke University School of Medicine, Durham, NC, USA (JMG, SDR); Janssen Research & Development, Titusville, NJ, USA (BL).

surveys, patients are often asked to evaluate tradeoffs among features of medical interventions, and these features often include 2 or more adverse event risks expressed as probabilities of negative outcomes. In the conventional approach, researchers quantify these tradeoffs by calculating the maximum acceptable risk (MAR) for each adverse event by equating the utility of the treatment benefit and the expected disutility of bearing the adverse event risk.[4]

Fairchild et al.[5] argued that the standard method for calculating risk acceptance does not make full use of robust experimental data from stated-preference surveys. The conventional approach estimates the maximum increase of a single adverse event risk that would be acceptable given a specified level of benefit, assuming that all other risks are unchanged.[6] An MAR estimate is produced for each individual adverse event risk included in the study. These individual MAR estimates usually are the only information available to compare the incremental incidence of the adverse outcome as observed in the clinical data. Any consideration of multiple simultaneous risks must be done in an ad hoc manner and with limited information on the interplay of the multiple risks and the correlation of MAR estimates. This could be particularly impactful in the regulatory context, where even small differences in MARs could affect the preference signals considered by agencies for approval decisions.

However, nearly all treatments are associated with multiple potential adverse events. Expanding the concept of individual MARs to account for multiple simultaneous risks could provide product developers and regulatory decision makers a transparent framework to consider more complex tradeoffs. Such a framework could affect both strategic product development questions pertaining to what tradeoffs are acceptable for target product profiles as well as regulatory questions pertaining to what profiles of risks and benefits in submitted treatments are acceptable.

Fairchild et al.[5] recently proposed a method for calculating a simultaneous maximum acceptable risk threshold (SMART) that makes better use of the multidimensionality of discrete-choice data by estimating the various combinations of probabilistic risks of adverse events that

would be acceptable given a specified benefit. In the present study, we apply this method to generate continuums of MAR vectors or SMART curves using data from published DCE preference studies. We also extend the approach to include confidence intervals along SMART curves to represent joint levels of uncertainty associated with the acceptance of 2 risks. We compare the resulting estimates against 1) the originally published, individual MARs and 2) the relevant clinical evidence on observed benefits and risks of treatment. The goal is to demonstrate situations in which MAR and SMART methods might result in qualitatively different interpretations of risk acceptance.

In the next section, we briefly outline the SMART calculation and then consider 3 published preference studies. For each study, we review the clinical context and the structure of the discrete-choice study, including an assessment of any specific risks and benefits that are observed clinically and included in the DCE. We then use preference weights to replicate the reported mean MAR estimates for each adverse event risk individually and to calculate the SMART, comparing both estimates against benefits and risks reported from clinical trials. We conclude with a discussion of the interpretation of these findings and their relevance for decision makers in regulatory and clinical contexts.

Methods

We identified 3 studies in the literature to demonstrate how the SMART approach can be applied. We selected these 3 studies because each demonstrates a unique aspect of the SMART approach as applied in clinical decision making. Also, in each case, clinical data on the incidence of relevant benefits and risks are available in the published literature and analogous to the risk and benefit attributes in the preference study. In general, it is possible to calculate the mean SMART threshold using published log-odds preference weights, which are typically reported in a table or figure in the main body of an article or in the supplemental materials. The confidence intervals can also be calculated based on the variance-covariance matrix associated with these preference weights. Although the authors of this study were involved in designing and publishing the preference studies used as examples here, the core insights gleaned by applying the SMART method are based on published data and do not require proprietary, respondent-level data.

For each study, we extract the raw (i.e., log odds) preference weights. We then used these preference weights to replicate the reported MAR estimates for individual adverse events. Following this confirmatory replication exercise, we identify treatments in the relevant therapeutic area that can be characterized by the attributes and levels included in the preference study. We identify the relevant levels of each attribute that apply to the candidate treatment. We use these clinical data, combined with the published preference weights, to quantify survey respondents' perceived valuation of a treatment profile using relevant attributes as the "utility of benefit."

We then follow the procedures described above and in Fairchild et al.[5] to identify all possible combinations of relevant adverse event risks that would exactly offset the expected utility of benefit of treatment. These combinations define the continuous threshold of acceptable combinations of risks (i.e., the SMART curve) corresponding to the main DCE results.

We calculate confidence regions around the SMART curves to account for the uncertainty of the DCE preference weights using the mean preference weights and the associated variance-covariance matrix. The confidence regions are calculated through a generalization of the Krinsky and Robb method outlined in Cooper.[7] This method is similar to bootstrapping in that repeated draws are made from specified distributions. We parametrize a multivariate normal distribution using each study's reported preference weight means and the associated variance-covariance matrix, from which we draw 1,000 sets of preference weights to generate 1,000 SMART curves.

For each draw of preference weights, we estimate the full SMART curve representing all of the combinations of probabilistic risks that would be acceptable given the level of benefit, as described in Fairchild et al.[5] Specifically, in cases with exactly 2 adverse event risks, 1 and 2, we calculate the probability of adverse event 2 that would exactly offset the expected utility of treatment benefits, evaluated at each potential level of adverse event 1. We can generalize this to n adverse events, for which we identify all combinations of probabilities for the $n - 1$ adverse events, and calculate the probability of adverse event n that would exactly offset the expected utility of treatment benefits, at each of the combinations of the other $n - 1$ adverse events.

The result of this exercise is 1,000 simulated SMART curves, each defined by a set of jointly acceptable risk-risk combinations. For each level of adverse event 1, we identify the 2.5% and 97.5% percentiles of adverse event 2 to define the 95% confidence region. Larger or smaller
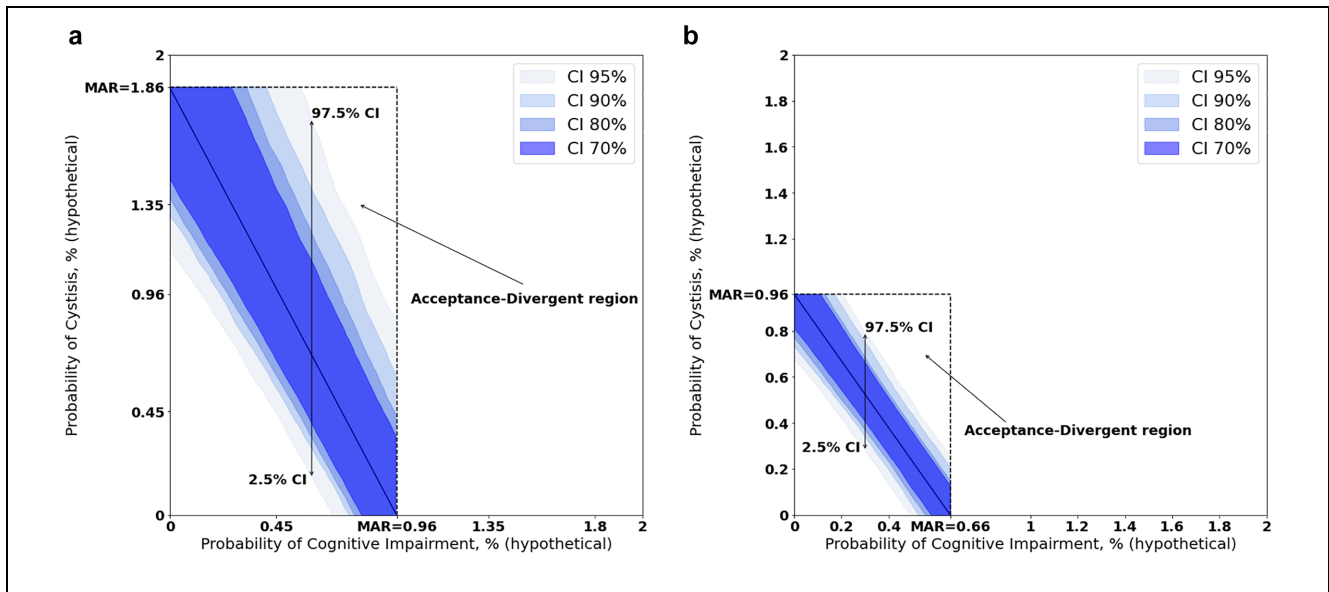
Figure 1 Simultaneous maximum acceptable risk thresholds (SMART) for esketamine-induced Montgomery-Åsberg Depression Rating Scale improvement of 4.1 in the (a) clinical trial sample and (b) online panel sample of 2 hypothetical risks based on reported events with ketamine abuse. CI, confidence interval; MAR, maximum acceptable risk.

confidence regions can use corresponding percentiles. Note that the mean SMART curve was estimated based on the mean utility estimates derived from the main model, not the mean of the simulations. Thus, the confidence intervals represent the uncertainty surrounding the SMART estimate derived from uncertainty in the underlying preferences estimates.

Results

Study 1: Treatment-Resistant Major Depressive Disorder

Fairchild et al.[8] report the results of a patient preference study eliciting benefit-risk tradeoff preferences among adults with treatment-resistant major depressive disorder. Participants were recruited from both an online sample of respondents who self-report to the condition (n = 297) and among English-speaking participants in 2 clinical trials of a novel ketamine-based therapy, esketamine (n = 159). The published results include the MAR of either cognitive impairment or ulcerative cystitis, along with a series of SMART estimates for specific benefit levels, and it is the first publication to present both types of MAR analysis. The results of this study were also included in the regulatory submission for esketamine.[9] While esketamine was not observed to result in either cognitive impairment or ulcerative cystitis side

effects, this preference study was conducted due to these side effects occurring in some individuals who abused ketamine.

We expand on the originally published SMART analysis by comparing the published risk acceptance results against an approximation of esketamine's clinical profile. Kim et al.[10] document a 4.1-point improvement in the Montgomery-Åsberg Depression Rating Scale (MADRS) mood score for 56 mg of esketamine combined with a newly initiated antidepressant relative to the control group, which received placebo plus a newly initiated antidepressant. By interpolating that treatment benefit using the preference weights for MADRS score improvement, we compute the utility benefit for both the clinical trial (0.639) and the online panel sample (0.319) participating in the initial DCE survey.[5] We then compute the SMART and corresponding confidence regions to represent measurement uncertainty using the generalized Krinsky-Robb method outlined above (Figure 1a and b).

In both figures, the point estimate of the single-event MARs for each potential adverse event is found at the *x* or *y* intercept, where the other adverse event risk is equal to zero. The preference weights for each adverse event were modeled linearly[5]; thus, the SMART appears as a linear line connecting the 2 independent MAR point estimates. If the point representing the probabilities for both adverse events on the *x-y* plane falls above the SMART
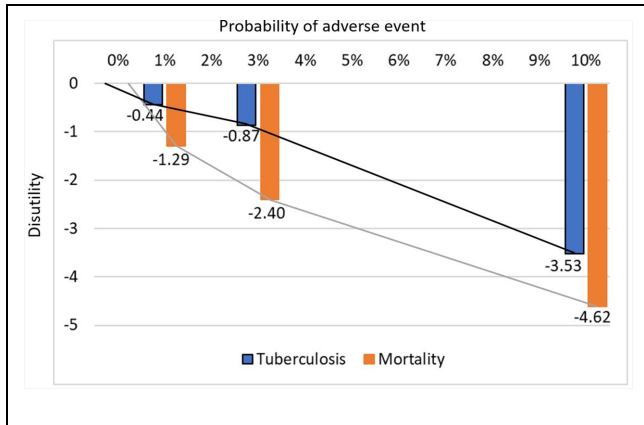
Figure 2 Preference weights for treatment-related risk levels of tuberculosis and mortality for plaque psoriasis on the hands from Fairchild et al. (2017).[11] Lines are included to display nonlinearity of preference weights.
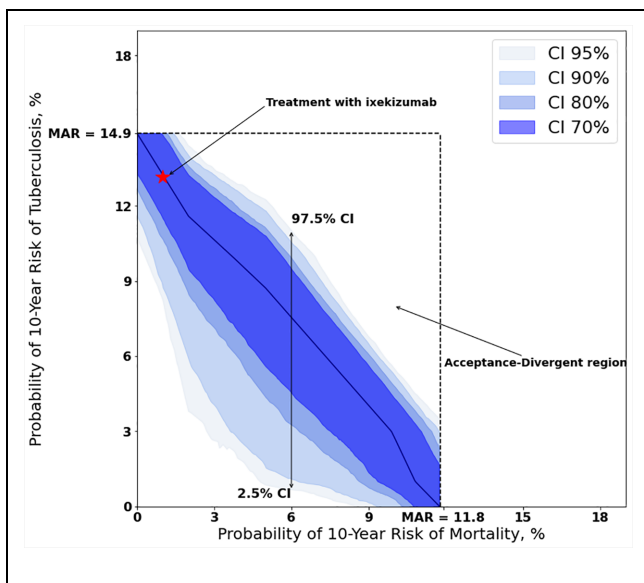

Figure 3 Simultaneous maximum acceptable risk threshold (SMART) for a reduction from 25% to 0% area of hands with moderate plaque psoriasis. The red dot indicates the probabilities for the adverse events of mortality and tuberculosis in Gordon et al.[12] extrapolated across 10 y. Tuberculosis served as a proxy for severe infection.

curve, one can conclude that the benefits of a treatment do not outweigh the combined disutility associated with both risks. The zone representing situations in which decision making may differ when considering individual MARs versus considering SMARTs is the region above

the SMART line but within the dashed lines that represent point estimates of the single-risk MARs. We refer to this area as the acceptance-divergent region. Without SMART, a decision maker assessing information on adverse events has to adjust the individual MAR information to determine their joint acceptability, generally a process done by intuition without mathematical support. This could lead to an erroneous conclusion that treatment benefits outweighed its risks if the actual probability of cystitis and memory loss associated with treatment fell in the acceptance-divergent region. Such an error could occur if the assessment considers the MAR estimates for each adverse event individually or if the adjustment used to account for multiple types of adverse events is not consistent with respondents' preferences. In the case of esketamine, neither adverse event was observed in the clinical trials; however, other treatments for treatment-resistant depression could potentially have such outcomes.

Study 2: Biologics for Psoriasis

In Fairchild et al.,[11] the researchers quantified benefit-risk preferences among 927 participants with psoriasis pertaining to the alleviation of varying degrees of plaque psoriasis on the face, body, and hands when different treatment regimens result in different risks for severe tuberculosis infection and mortality. An applicable treatment of interest to this study is ixekizumab. In randomized clinical trials, ixekizumab was documented to provide near-total alleviation of plaque psoriasis on all body parts among patients with a median area coverage of 27%, with annual risks of both severe infection and mortality at 1.4% and 0.1%, respectively.[12]

This study included multiple risk levels of the adverse event attributes and modeled risk preferences as a piecewise linear function. In other words, the model allows for the possibility that the preference weights could fluctuate nonlinearly; that is, the preference weight for 10% risk may be more or less than 10 times the preference weight for 1% risk. The researchers observed such nonlinear preferences in their model for psoriasis reduction in the patient's hands (Figure 2). These preference weights indicate a decreasing marginal disutility of mortality risk and a fairly linear functional form for tuberculosis risk. We built the SMART curve based on the utility gain associated with total alleviation of moderate plaque psoriasis that has an initial 25% coverage on the patient's hands (Figure 3).

As above, the point estimate single-event MARs are located at the intercepts where the risk of the other

Table 1 Progression-Free Survival and Adverse Events Incidence of Study Treatment versus Placebo in 2 Clinical Trials for Iodine Refractory Thyroid Cancer from Fleeman et al. (2019)[14]

	SELECT Trial		DECISION Trial	
	Placebo	Lenvatinib	Placebo	Sorafenib
Progression-free survival (months, median)[a]	3.6	18.3	5.8	10.8
Incidence of severe hand-foot skin reaction (grade 3 +), %[b]	0.0	3.4	0.0	20.3
Incidence of severe proteinuria (grade 3 +), %[b]	0.0	10.0	0.0	0.0
Incidence of severe hypertension (grade 3 +), %[b]	3.8	42.9	2.4	9.7

[a]Data extracted from table 2 in Fleeman et al. (2019).[14]
[b]Data extracted from table 5 in Fleeman et al. (2019)[14] reproduced according to terms of the Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/.

adverse event is zero and the mean SMART curve is bounded by different levels of confidence regions. The key difference between Figures 1a, b and 3 is that the mean SMART curve is slightly nonlinear and its confidence intervals are asymmetric. As expected given the simulated results in Fairchild et al.,[5] jointly modeling nonlinear preference weights can affect the area of the acceptance-divergent region bounded by the SMART curve and the single-event MARs. When the confidence regions bow toward the axes, as is clearly the case for the 90% and 95% CI contours, the acceptance-divergent region is larger than with a linear model. This occurs because the preference study revealed less incremental risk tolerance at lower levels of mortality risks.

Furthermore, we plot the corresponding risks reported in clinical results of ixekizumab, using tuberculosis as a proxy for severe infection (red marker in Figure 3). To equate the annual risks reported in the clinical trial to the 10-y risks from the DCE, we assume uniform hazard of the annual clinical risk, yielding a 10-y risk of mortality of 0.99% and 10-y risk of serious infection (which the DCE proxies with tuberculosis) of 13.15%. When we extrapolate the reported clinical risks in this way, the risk profile of ixekizumab appears much closer to the point-estimate SMART curve than each MAR threshold individually, indicating that the benefit-risk balance is just on the borderline of acceptable based on these criteria. However, the lower limits of the confidence regions for the SMART curve indicate a much lower risk acceptance for both adverse events. With a 10-y mortality risk of 1%, the conservative lower limits of the 90% and 95% SMART confidence intervals indicate that acceptance of 10-y risk of serious infection falls from about 13% (SMART curve) to about 10% and 8%, respectively, assuming no change in the risk of mortality.

Study 3: Radioactive Iodine for Refractory Differentiated Thyroid Cancer

The SMART approach can be applied when there are more than 2 adverse events of interest. In a preference study by Mohamed et al.,[13] the authors present the results of a DCE survey among 134 patients with a self-reported diagnosis of differentiated thyroid cancer who had received at least 1 round of treatment with radioactive iodine. The study focused on key treatment features that differentiate 2 treatments indicated for this cancer, sorafenib and lenvatinib, which are approved by the United States Food and Drug Administration (FDA) for radioactive iodine refractory differentiated thyroid cancer. The differentiating features included progression-free survival (PFS) and 3 probabilistic adverse-event attributes: hand-foot skin reaction, hypertension, and proteinuria. To be able to calculate confidence intervals, we obtained the raw preference weights and corresponding variance-covariance matrix from the study's corresponding author. The original preference study presented individual MAR estimates associated with switching from a treatment with a median PFS of about 10 mo (similar to sorafenib) to a treatment with a median PFS of 16 or 18 mo (similar to lenvatinib), which we successfully replicated prior to proceeding with our SMART analysis.

Table 1 presents a summary of the clinical benefits and adverse-event risks of the 2 approved treatments as they relate to the attributes included in this preference study, based on a comparison of the clinical trials conducted for each drug in this indication.[14] Using the original specification with linear and quadratic terms to model treatment attributes, we computed the approximate preference values associated with the differences in PFS between placebo and treatment groups. We then use
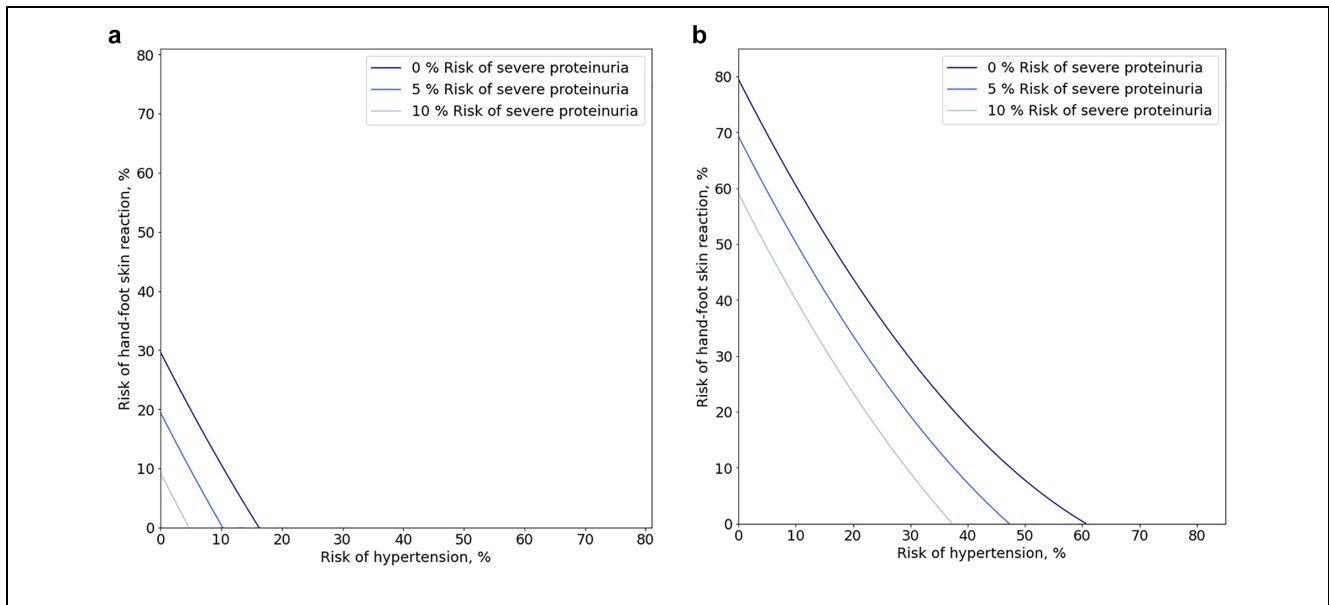
Figure 4 Simultaneous maximum acceptable risk thresholds (SMART) curves for 3 probabilistic adverse-events for (a) 5 additional months of progression-free survival and (b) 14.7 additional months of progression-free survival. In both figures, 3 SMART curves are shown for hypertension and hand-foot skin reaction for 3 different fixed probabilities of severe proteinuria.

these values to compute traditional MARs and SMART curves associated with medicine-specific improvements in PFS. Figure 4a depicts SMART curves for risks of hand-foot skin reaction and hypertension corresponding to PFS gains of 5 mo for sorafenib versus placebo. Figure 4b depicts SMART curves for risks of hand-foot syndrome and hypertension corresponding to PFS gains of 14.7 mo for lenvatinib versus placebo. The figures show 3 curves, corresponding to risk levels of 0%, 5%, and 10% for severe proteinuria. Because confidence intervals for each of the lines obscure the tradeoffs across the 3 risks displayed in the figure, confidence regions are not depicted.

Single-risk MARs for hypertension and hand-foot syndrome correspond to the point on the figure where the curves intersect with the *x* and *y* axes when the risk of severe proteinuria is 0%. For sorafenib, the single-event MARs are 16.4% for hypertension, 29.6% for hand-foot skin reaction, and 14.6% for severe proteinuria. Similarly for lenvatinib, the single-event MARs are 60.8% for hypertension, 79.5% for hand-foot skin reaction, and 39.0% for severe proteinuria The different locations of the corresponding curves in Figure 4a and b suggest that there will be a broad range of adverse event probabilities that are acceptable for lenvatinib that are not acceptable for sorafenib, because lenvatinib offers substantially higher treatment benefit.

Figure 5a and b show the SMART curve for risk of hypertension and hand-foot skin reaction that is associated with the observed clinical benefit of each treatment, holding this risk of severe proteinuria at the level observed in placebo-controlled clinical trials of sorafenib and lenvatinib, respectively.[14] Stars in each figure mark the combined risk profile associated with each treatment. For both treatments, the 3 individual MAR estimates are each greater than the observed incremental risks for each of the adverse events. However, when all 3 risks are considered jointly, the combined risk profiles lie to the northeast of the SMART threshold, indicating that the combination of the 3 adverse-event risks exceeds patients' average willingness to accept the risks presented in the study. In this case, the individual and joint MAR approaches yield qualitatively different implications for decision makers, who might conclude that the medicine is acceptable based on the individual MAR estimates but unacceptable based on the more comprehensive SMART curves.

Discussion

Regulatory benefit-risk assessments where there are multiple, distinct adverse events to consider represent complex, multifaceted challenges. While it is intuitive that increasing the risk of one adverse event lessens the
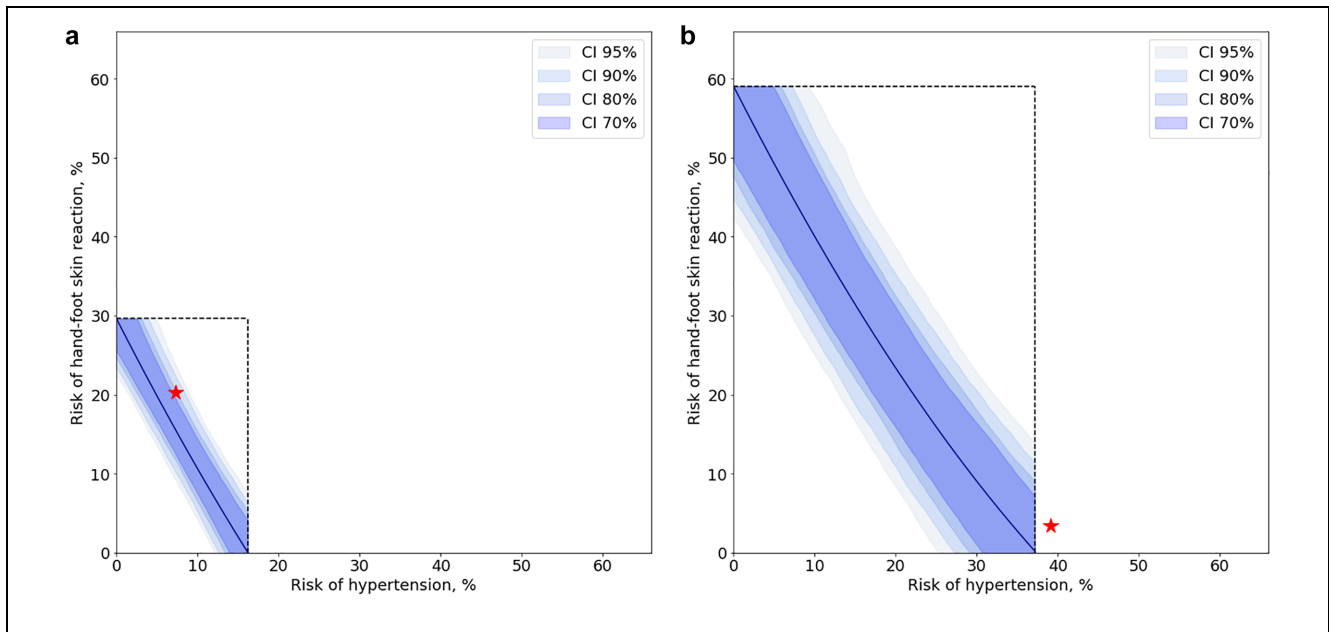
Figure 5 Risk profiles for: (a) risk profile for sorafenib versus placebo: progression-free survival (PFS) gain = 5 mo with risk of severe proteinuria of 0% and (b) risk profile for lenvatinib versus placebo: PFS gain = 14.7 mo with risk of severe proteinuria of 10%.

acceptability of other adverse events in the benefit-risk balance, conventional reporting approaches rely on decision makers to mentally make these adjustments. In this article, we show that the SMART (i.e., simultaneous MAR threshold) approach provides a reproducible and transparent representation of benefit-risk assessments with multiple risks using empirical stated-preference data from discrete choice studies. We also introduce methods to generate confidence intervals along SMART curves. With varying confidence limits, results can be displayed as confidence contours. These contours convey important information to decision makers about the degree of certainty that a treatment's benefits outweigh risks.

The finding that SMART analysis and conventional MAR analysis can yield qualitatively different conclusions regarding the acceptability of treatments highlights the importance of applying the SMART approach in cases in which patient preference data are used to inform decision making. In our first application to the treatment-resistant depression study, we graphically demonstrated the set of hypothetical risk profiles for which the individual MAR point estimates would be deemed acceptable but the SMART analysis would not, which we called the acceptance-divergent region. In this case, the adverse events chosen for the DCE represented hypothetical adverse events that were not detected in

esketamine clinical trials. Despite the adverse events not playing a major role in the regulatory decision, findings from the patient preference study were submitted to the FDA and were found by members of the advisory committee to be informative.[15,16] In our second application examining preferences for biologic agents used to treat psoriasis, asymmetric SMART confidence intervals indicate less risk tolerance when considering multiple adverse events, thereby demonstrating that mental adjustments such as halving the individual MARs (a heuristic that the authors have found reasonable when there are 2 risks with linear SMARTs) to account for 2 adverse-event risks may be unreliable when the SMARTs are nonlinear. In our third application to treatments for refractory thyroid cancer, we portray 2 treatments, both of which satisfy individual maximum acceptable risk criteria for adverse events but do not satisfy them when considering the joint disutility of these risks in the SMART framework. These findings may have important implications for how decision makers apply clinical data to findings from stated-preference benefit-risk studies.

The degree to which differences between the conventional single adverse-event MAR and the SMART approaches matter depends on the modeling strategy and on marginal preferences for avoiding risk. Cases with decreasing marginal disutility across risk levels, as

exemplified in the psoriasis example, are identified as most problematic.[5] Previous studies, including those examining rank-dependent expected utility and probability weighting functions,[17] have documented that people generally overweight small probabilities and underweight larger probabilities, consistent with observations of diminishing marginal disutility for increasing probability or frequency of adverse events, suggesting that the acceptance-divergent region may often be an important consideration.[18,19] Cases with nonlinear marginal patterns are hard to predict and therefore important to evaluate. Finally, cases in which traditional MARs are orders of magnitude higher than the observed incidence of adverse events are not likely to result in different regulatory conclusions.

In this article, we have expanded the presentation of SMART curves to include confidence intervals between 2 or 3 adverse-event risk levels up to the point-estimate MAR for each individual risk. It is important to note that these confidence regions represent measurement uncertainty associated with the SMART continuum rather than preference heterogeneity across patients. Thus, it is not appropriate to interpret the 70% confidence region as though 85% (i.e., $1 - (0.5 \times [1 - 0.7])$) of patients would find acceptable all risk profiles to the southeast of this region. However, since simulations are used to generate confidence regions, reporting findings in a probabilistic manner could be helpful. For instance, if a treatment's actual risk profile ranked at the 75th percentile of simulated SMART values, one could loosely infer a 25% probability that its risk profile is acceptable. In the ixekizumab example, its adverse-event profile corresponded to the 49th percentile of increasing SMART values, thus a 51% probability that the joint risk profile with ixekizumab is considered acceptable given its specified benefit.

Although we showed that adverse-event risks associated with sorafenib and lenvatinib may outweigh their expected benefits in terms of gains in PFS, these drugs are routinely used in patients with radioactive iodine refractory differentiated thyroid cancer. This does not necessarily indicate that patient-level benefit-risk tradeoffs with these treatments are unacceptable. Several factors should be considered to determine acceptability. First, the preference results represent average preferences and, therefore, average risk acceptance. It is possible that these average risk acceptance results mask preference heterogeneity across the sample; individual patients or groups of patients who opt for these treatments may have greater than average risk tolerance. They may also have diagnostic or demographic characteristics that modify their personal chances of benefit or adverse events. These patients may have discussed treatment options with their providers and decided to accept the benefit-risk profile upon initiating therapy. Second, the data presented in Fleeman et al.[14] also notes that both lenvatinib and sorafenib are poorly tolerated, with approximately 14% of trial participants receiving lenvatinib and 19% of trial participants receiving sorafenib discontinuing therapy and nearly two-thirds with dose reductions (68% with lenvatinib; 64% with sorafenib) (Schlumenberger et al.)[20] and (Brose et al.)[21]. This suggests that patients who are aware of the risks might accept experiencing adverse events but anticipate changing their dose or discontinuing therapy if they experience toxicities. Furthermore, patients with differing access to clinical consultation may have a different understanding of the risks and benefits of treatment compared with those participating in the survey, which featured choices between medicines defined by only a subset of potentially relevant features. Finally, it is also possible that patients in real clinical situations placed more weight on potentially extending their lives than similar patients who participated in the survey.

In addition to the evident usefulness of the SMART approach to support regulatory decision making, it may also be useful to medical product developers throughout a product's life cycle. In the earlier phases of development, the approach can be used to generate a continuum of target product profiles in which varying levels of acceptable benefit-risk combinations are taken into account. The SMART approach can be used to monitor the benefit-risk profile for a given product as its body of evidence grows during development and postapproval. SMART curves may also be able to identify cases in which more precision about the incidence of an adverse event is needed to increase the likelihood that the product is considered acceptable from a benefit-risk perspective.

To facilitate the adoption of this new approach for use in benefit-risk decision making, we recommend that authors of DCEs evaluating benefit-risk tradeoffs include the variance-covariance matrix and raw (log-odds) preference weights among a manuscript's supplemental materials. Development of an online interactive tool or sharable programming code could also assist less technical researchers to generate SMART curves and confidence contours.

Conclusions

We have identified certain cases in which the generation of SMART curves would lead to qualitatively different regulatory conclusions, potentially affecting the interpretation of risk-tolerance estimates in regulatory applications. Therefore, we recommend that any time stated-preference

data are considered for regulatory decision making, the SMART approach should be used to yield a more comprehensive representation of benefit-risk tradeoffs than the conventional MAR approach.

Authors' Note

Data supporting the findings presented herein are available from the corresponding author upon reasonable request. Table 1 reproduces data reported in table 5 in Fleeman et al. 2019[14] according to terms of the Creative Commons Attribution 4.0 International License: http://creativecommons.org/licenses/by/4.0/.

ORCID iDs

Juan Marcos Gonzalez https://orcid.org/0000-0002-5386-0907
Angelyn O. Fairchild https://orcid.org/0000-0002-5548-5987
Shelby D. Reed https://orcid.org/0000-0002-7654-4464

References

1. US Food and Drug Administration. Factors to consider when making benefit-risk determinations in medical device premarket approval and De Novo classifications: guidance for industry and food and drug administration staff. 2019. Available from: https://www.fda.gov/media/99769/download. [Accesssed 16 November, 2022.].
2. Hauber AB, Fairchild AO, Johnson F. Quantifying benefit-risk preferences for medical interventions: an overview of a growing empirical literature. *Appl Health Econ Health Policy*. 2013;11:319–29.
3. Ho MP, Gonzalez JM, Lerner HP, et al. Incorporating patient-preference evidence into regulatory decision making. *Surg Endosc*. 2015;29(10):2984–93.
4. 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.
5. Fairchild AO, Reed SD, Gonzalez JM. Methods for calculating the simultaneous maximum acceptable risk threshold (SMART). *Med Decis Making*. Epub ahead of print November 2022. DOI: 10.1177/0272989X221132266
6. Gonzalez JM, Boeri M. The impact of the risk functional form assumptions on maximum acceptable risk measures. *Patient*. 2021;14(6):827–36. DOI: 10.1007/s40271-021-00518-y
7. Cooper J. A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous choice contingent valuation surveys. *Land Econ*. 1994;70:111–22.
8. Fairchild AO, Katz EG, Reed SD, et al. Patient preferences for ketamine-based antidepressant treatments in treatment-resistant depression: results from a clinical trial and panel. *Neurol Psychiatry Brain Res*. 2020;37:67–78.
9. Janssen Research & Development, LLC. Esketamine nasal spray for patients with treatment-resistant depression. Advisory committee briefing document. Available from: https://www.fda.gov/media/121377/download. [Accessed 16 November, 2022].
10. Kim J, Farchione T, Potter A, Chen Q, Temple R. Esketamine for treatment-resistant depression—first FDA-approved antidepressant in a new class. *N Engl J Med*. 2019;381(1):1–4.
11. Fairchild AO, Reed SD, Johnson FR, Anglin G, Wolka AM, Noel RA. What is clearance worth? Patients' stated risk tolerance for psoriasis treatments. *J Dermatolog Treat*. 2017;28(8):709–15.
12. Gordon KB, Blauvelt A, Papp KA, et al; UNCOVER-1 Study Group; UNCOVER-2 Study Group; UNCOVER-3 Study Group. Phase 3 trials of ixekizumab in moderate-to-severe plaque psoriasis. *N Engl J Med*. 2016;375(4):345–56.
13. Mohamed AF, González JM, Fairchild A. Patient benefit-risk tradeoffs for radioactive iodine-refractory differentiated thyroid cancer treatments. *J Thyroid Res*. 2015;2015:1–8.
14. Fleeman N, Houten R, Chaplin M, et al. A systematic review of lenvatinib and sorafenib for treating progressive, locally advanced or metastatic, differentiated thyroid cancer after treatment with radioactive iodine. *BMC Cancer*. 2019;19(1):1209.
15. Katz EG, McNulty P, Levitan B, Treichler P, Martynowicz J, Jamieson C. U.S. Food and Drug Administration's patient-focused drug development initiative: experience with integration of patient-experience data in a new drug application for esketamine nasal spray plus a newly initiated oral antidepressant for treatment-resistant depression. *Ther Innov Regul Sci*. 2022;56(1):38–46.
16. Reardon S. Antidepressant based on party drug gets backing from FDA advisory group. *Nature*. Available from: https://www.nature.com/articles/d41586-019-00559-2. [Accessed 16 June, 2022].
17. Tversky A, Kahneman D. Advances in prospect theory: cumulative representation of uncertainty. *J Risk Uncertain*. 1992;5(4):297–323.
18. Johnson FR, Ozdemir S, Mansfield C, et al. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. *Gastroenterology*. 2007;133(3):769–79.
19. Lauridsen JT, Lønborg J, Gundgaard J, Jensen HH. Diminishing marginal disutility of hypoglycaemic events: results from a time trade-off survey in five countries. *Qual Life Res*. 2014;23(9):2645–50.
20. Schlumberger M, Tahara M, Wirth LJ, et al. Lenvatinib versus placebo in radioiodine-refractory thyroid cancer. *N Engl J Med*. 2015;372(7):621–30.
21. Brose MS, Nutting CM, Jarzab B, et al.; DECISION investigators. Sorafenib in radioactive iodine-refractory, locally advanced or metastatic differentiated thyroid cancer: a randomised, double-blind, phase 3 trial. *Lancet*. 2014;384(9940):319–28.