



# Patient-centered clinical trials

Shomesh E. Chaudhuri<sup>1,2</sup>, Martin P. Ho<sup>3</sup>, Telba Irony<sup>4</sup>, Murray Sheldon<sup>3</sup>  
and Andrew W. Lo<sup>1,2,5</sup>

<sup>1</sup> Laboratory for Financial Engineering, MIT Sloan School of Management, USA

<sup>2</sup> Department of Electrical Engineering and Computer Science, MIT, USA

<sup>3</sup> Center for Devices and Radiological Health, Food and Drug Administration, USA

<sup>4</sup> Center for Biologics Evaluation and Research, Food and Drug Administration, USA

<sup>5</sup> Computer Science and Artificial Intelligence Laboratory, MIT, USA

**We apply Bayesian decision analysis (BDA) to incorporate patient preferences in the regulatory approval process for new therapies. By assigning weights to type I and type II errors based on patient preferences, the significance level ( $\alpha$ ) and power ( $1 - \beta$ ) of a randomized clinical trial (RCT) for a new therapy can be optimized to maximize the value to current and future patients and, consequently, to public health. We find that for weight-loss devices, potentially effective low-risk treatments have optimal  $\alpha$ s larger than the traditional one-sided significance level of 5%, whereas potentially less effective and riskier treatments have optimal  $\alpha$ s below 5%. Moreover, the optimal RCT design, including trial size, varies with the risk aversion and time-to-access preferences and the medical need of the target population.**

## Introduction

Determining the acceptable level of uncertainty associated with clinical evidence has been an important and challenging decision when regulators conduct benefit–risk assessments of novel technologies, especially for unmet medical needs. Traditional clinical trial designs typically set the one-sided significance level [i.e., the maximum allowed value for the rate of type I error (approving a therapy for which there is not a reasonable assurance of safety and effectiveness)] at 5% regardless of the context in which the decision is made or the public health implications of the consequences. However, the context could matter for making rational and sensible decisions with significant public health impact. In some circumstances, the consequences of making a type I error can be less important than those of a type II error (not approving a therapy for which there is a reasonable assurance of safety and effectiveness), particularly when the therapy can treat a life-threatening or irreversibly debilitating disease or condition for which there are no other available treatments. Moreover, the standard

value of 5% for type I error is, itself, arbitrary and not tied to any context-specific considerations.

To address this important regulatory science challenge, the Center for Devices and Radiological Health (CDRH) at the U.S. Food and Drug Administration (FDA) has used a stepwise strategy. First, the CDRH has conveyed its approach to making benefit–risk assessments more robust and systematic through the release of a guidance document in 2012 (and subsequently updated in 2016) on benefit–risk determinations for premarket approval and *de novo* classification decisions [1]. The guidance document is intended to explain the FDA's thinking on the factors to be taken into account when making benefit–risk determinations for premarket approval of medical devices and has explicitly listed patient perspectives as one of the important factors for the CDRH staff to consider. Next, the CDRH has made a commitment to make its regulatory decision-making more patient-centered by engaging patient stakeholders and exploring the use of quantitative methods to elicit and incorporate patient preferences in a valid scientific manner. In 2013, the CDRH held a public Patient Preference Initiative Workshop to engage stakeholders.

Corresponding author: Lo, A.W. (alo-admin@mit.edu)

To explore ways to include patient perspectives in its regulatory decision-making, the CDRH sponsored a proof-of-concept pilot study in 2012 to elicit quantitative preferences on benefit–risk tradeoffs for weight-loss devices [2]. Weight-loss devices were chosen for the following reasons. First, obesity has a profound impact on public health owing to its prevalence and associated serious comorbidities. Second, patients with obesity are facing difficult benefit–risk tradeoffs when choosing treatments that are of low risk and low effectiveness versus treatments of higher risk with higher effectiveness. Third, the effectiveness of treatments has a significant behavioral component. Finally, preliminary communications with several medical device innovators enabled the CDRH to anticipate several investigative weight-loss devices with diverse benefit–risk profiles being developed. A weight-loss device study that used a scientific method to elicit quantitative patient preferences with a large sample of subjects would provide an unprecedented opportunity to inform the CDRH’s decision-making.

The CDRH recognized the importance of heterogeneity across the spectrum of patient preferences and chose a method called discrete choice experiments, which could capture not only the average but also, more importantly, the distribution of patient preferences, including its variability among patients of various genders, ages, body mass indexes and previous experience with weight-loss surgeries. Moreover, the device attributes and levels considered in the study, including benefits (amount of weight loss, weight-loss duration, improvement in comorbidities), risks (side-effect duration, chance of hospitalization, chance of dying from getting the device) and other device characteristics (dietary restrictions, type of operation), were selected by CDRH regulators based on a portfolio of devices that were in the development pipeline but not yet on the market. Each question in the study involved a choice between two hypothetical weight-loss devices with different attribute profiles and each attribute had varying levels. Each subject in the CDRH study answered eight choice questions, and the choices made revealed the relative importance of these attributes and their levels to that subject.

Since the study results became available, they have been informing reviewers at the CDRH when making their approval decisions across a wide range of weight-loss devices with differing benefit–risk profiles. After a gap of eight years, since the last class of weight-loss devices (gastric bands) was approved, the CDRH approved an electrical stimulation system and two gastric balloon systems in 2015, and a gastric emptying system in 2016. However, with no objective, explicit and transparent method to directly relate the specific patient preference evidence developed in the study to the acceptable level of uncertainty associated with the submitted clinical evidence, CDRH review staff have had to subjectively consider the evidence.

In recent years, the CDRH has achieved significant milestones to facilitate the design and conduct of patient-preference studies by sponsors and patient groups. In 2015 the Medical Device Innovation Consortium released a patient preference framework report sponsored by the FDA. It discusses how patient preference information can be used at various stages of the total product lifecycle and includes a catalog of existing methods for eliciting patient preferences compiled by a panel of experts. In 2016, the FDA Patient Preference Information guidance document was released,

**BOX 1****FDA Guidance: Patient preference information (p. 6)**

This guidance focuses on the specific type of patient input referred to as patient preference information (PPI), which, for the purposes of this guidance, is defined as: qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions . . . The specific role of quantitative PPI is to provide estimates of how much different outcomes, endpoints or other attributes are valued by patients, and the tradeoffs that patients state or demonstrate they are willing to make among them. Such outcomes or other attributes of a device include demonstrated or posited measures of effectiveness, safety and other device characteristics that can impact benefit–risk considerations, including (but not limited to) means of implantation, duration of effect, duration and frequency of use, and utility of the device.

containing the mandate quoted in Box 1 [3]. However, there is a missing quantitative link between specific patient preferences and acceptable levels of uncertainty. Furthermore, methodologies that objectively, transparently and reproducibly determine these quantitative links are elusive. The CDRH is working with scientists to create a solution to this regulatory science question.

In this paper, we use a Bayesian method proposed by Lo *et al.* [4,5] and quantitative patient preference data from the CDRH patient preference obesity study to calculate acceptable levels of uncertainty (significance level and power) when designing pivotal clinical trials for clinical evidence required by regulatory decision-making. The Bayesian approach has long been applied to clinical trial design and analysis [6–12]. Bayesian decision analysis (BDA), the particular method presented in this article, aims to optimize the balance between type I and type II error rates and the severity of the consequences of making these errors based on patient preferences. The appropriate patient preference scores are scientifically elicited and estimated across safety and effectiveness, and used to construct hypothetically optimal balanced two-arm fixed-sample randomized clinical trials (RCTs) to maximize the expected value for patients.

We take into account the fact that lengthy clinical trials provide more power but can negatively impact public health because they delay access of effective treatments to patients. In addition, we weigh the consequences of approving an ineffective treatment versus rejecting an effective intervention. If we set the significance level to be smaller and consequently more stringent, we not only reduce the chance of approving an ineffective treatment but also increase the chance of rejecting an effective treatment.

We find that the BDA optimal design is often substantially different in significance level, power and sample size from the conventional approach using a fixed one-sided significance level of 5%. Of course, the BDA optimal design depends on several key assumptions. Although this framework provides a systematic and quantitative method of incorporating multifaceted tradeoffs into RCT design, the usefulness of its recommendations relies on the appropriateness of these assumptions and on accurately calibrated model parameters. Although we apply BDA to a specific medical device in this study, the framework applies more broadly to other therapeutic devices and drugs; and this application is meant to serve as a proof-of-concept for a more general and systematic

approach to incorporating clinical context, including patient preferences, into the medical device and drug approval processes.

## BDA optimal clinical trial design

Here, we propose a quantitative framework to explicitly take into account patient preferences across multiple device attributes when determining the optimal sample size and critical value of a balanced two-arm fixed-sample RCT. We first define a patient-centered value model associated with given medical device attributes. We then assign prior probabilities to each possible combination of these attributes, and formulate the expected value of the trial. The optimal trial size ( $2n$ , where there are  $n$  patients in each arm of the study) and the one-sided significance level ( $\alpha$  or critical value  $\lambda_\alpha$ ) are then jointly determined to maximize the expected value of the trial. Note that maximizing the value of the trial means providing access to a safe and effective treatment to patients as soon as possible, or concluding that the treatment has not demonstrated a reasonable assurance of safety and effectiveness as soon as possible. It is equivalent to minimizing its losses, which include the consequences of incorrect decisions for all current and future patients, and the inefficiency of delaying access of a potentially safe and effective treatment to patients. Although the value model we now introduce is based on preference data for a specific device, our methods are similar to a previous oncology analysis [4,5] and are applicable to other therapeutics for which patient-preference data are available.

### Patient-centered value model

The CDRH weight-loss device study elicited and quantified the importance of safety, effectiveness and other attributes of weight-loss devices into patient preference scores [2]. eTable 1 (see Supplementary material online) shows the preference scores estimated from the survey data for total body weight loss as a percentage of initial weight (%TBWL) and mortality risk. The scores are on a scale from  $-10$  (least preferred) to  $+10$  (most preferred), where  $-10$  is the estimated value of a 5% mortality risk to patients. %TBWL and mortality risk were modeled as continuous variables, and their preference weights were linearly interpolated between observations.

Patient preference scores for each attribute were then mapped directly to relative values. For example, the change in value of an increase in mortality risk from 0 to 1% can be quantified by the preference score difference between these two levels. The change in value of this increase from the patient's perspective is therefore  $-3.5$ . Similarly, the change in value of an increase in %TBWL from 0 to 30% is  $+4.3$ . Given both these changes, and holding other attributes constant, the net change in value would then be  $+0.8$ , implying the additional weight loss would more than compensate for the increased mortality risk according to the patient preference information. The relative loss of value per patient,  $L$ , of using one intervention with a lower overall preference score over another is then defined in terms of this net change in value. The number of patients affected can be used to scale  $L$  to estimate a collective loss of value.

The value associated with a clinical trial for a superiority claim can be categorized into in-trial and post-trial value. In-trial value depends on the number of subjects in each arm of the trial, and is independent of the outcome of the trial if both arms have the same

number of subjects. In contrast, post-trial value is completely dependent on the outcome of the trial and affects patients beyond the scope of the trial. In particular, we assume there is no post-trial loss in value with a correct decision [i.e., rejecting (approving) a device that is less (more) preferred relative to the control] except for the wait time during the regulatory review process. We further define the relative loss in value per person of using the investigational device under the null hypothesis ( $H = 0$ ) as  $L_0$ , and the relative loss in value per person of forgoing the use of the investigational device under the alternative hypothesis ( $H = 1$ ) as  $L_1$ . If the size of the target population is  $N$ , then the aggregate loss in value of a type I or II error will be  $DF_t \cdot N \cdot L_0$  and  $N \cdot L_1$ , respectively, where  $DF_t$  is a discount factor that decreases from 1 to 0 and accounts for the wait time,  $t$ , caused by the regulatory review process. In other words, patients place a lower value on an effective treatment if it is not accessible immediately. Therefore, the aggregate loss in value caused by the length of the regulatory review process under the alternative hypothesis is  $[1 - DF_t] \cdot N \cdot L_1$ .

Finally, if the investigational device is less preferred to the control, then the  $n$  subjects in the investigational arm collectively experience a loss of value of  $n \cdot L_0$ . However, if the investigational device is preferred to the control treatment, then the  $n$  subjects in the control arm forgo a better treatment and experience a collective loss of value of  $n \cdot L_1$ . The potential losses in value associated with a fixed sample trial are shown in eTable 2 (see Supplementary material online). Note that there is no loss in value (i.e., there is maximum value) in the hypothetically optimal scenario where the correct approval decision is made immediately and without running a trial.

We additionally assume time-consistent (i.e., exponential) discounting, and suggest that time-horizon preferences, which measure the capacity of patients to tolerate waiting, be elicited directly in future patient survey experiments. If the annual discount rate is  $r$ , then the discount factor is given by  $DF_t = e^{-rt}$ , where  $t$  is length of the regulatory-approval process. This proposed discount factor ensures that patient preferences do not change over time in such a way that they become inconsistent with one another. The duration of the regulatory review process is assumed to be determined by the size of the study ( $2n$ ), the patient accrual rate for the study ( $\eta$ ), the time required to setup the study ( $s$ ), the follow-up time of the final patient to complete the study ( $f$ ) and the FDA review time ( $\tau$ ) such that  $t = s + 2n/\eta + f + \tau$ .

### Bayesian decision analysis

A quantitative primary endpoint based on %TBWL is assumed for the trial. We further assume that subjects in the treatment arm receive the investigational device and each subject's response is independent of all other responses. Diet and exercise are assumed to be administered to patients in the control arm. The response variables in the treatment arm, denoted by  $\{T_1, \dots, T_n\}$ , are assumed to be independent and identically distributed, where  $T_i \sim \mathcal{N}(\mu_t, \sigma_t^2)$ . Similarly, the control arm responses, represented by  $\{P_1, \dots, P_n\}$ , are assumed to be independently and identically distributed as  $P_i \sim \mathcal{N}(\mu_p, \sigma_p^2)$ . We further confine ourselves to superiority trials where the device is likely to have either a positive effect ( $\mu_t > \mu_p$ ) or no effect ( $\mu_t = \mu_p$ ). In such cases, the treatment effect of the device,  $\delta$ , is defined as the difference between the response means in the two arms (i.e.,  $\delta \equiv \mu_t - \mu_p$ ). In a fixed-

sample trial with  $n$  subjects in each arm, we collect observations from the treatment and control arms, and form the following  $t$ -statistic (Eq. (1)).

$$T = \frac{\hat{\mu}_t - \hat{\mu}_p}{\sqrt{\frac{\hat{\sigma}_t^2}{n} + \frac{\hat{\sigma}_p^2}{n}}} \quad (1)$$

where  $\hat{\mu}$  and  $\hat{\sigma}$  represent the sample mean and standard deviation, respectively, and  $T$  has a noncentral  $t$ -distribution with noncentrality parameter  $\delta \cdot \sqrt{\frac{n}{\sigma_t^2 + \sigma_p^2}}$ . Under the assumption that the two variances are equal, this distribution has  $2(n - 1)$  degrees of freedom. The  $t$ -statistic,  $T$ , is then compared to the critical value,  $\lambda_\alpha$ . Finding that  $T > \lambda_\alpha$  supports rejection of the null hypothesis (i.e., that the device has no effect). The probability of failing to reject the null hypothesis, for a device that provides a treatment effect  $\delta$  with response variances  $\sigma_t^2$  and  $\sigma_p^2$ , is therefore  $P(T \leq \lambda_\alpha)$ .

Assuming prior probabilities  $p_0$  and  $p_1$  (where  $p_0 + p_1 = 1$ ) for the cases where the investigational device is equally effective ( $H = 0$ ) and more effective ( $H = 1$ ) to the control treatment, respectively, and letting  $V_0$  and  $V_1$  be the value created in the hypothetically optimal scenarios where the correct approval decision is made immediately and without running a trial, it is straightforward to calculate the expected value associated with an RCT design with

parameters  $\{n, \lambda_\alpha\}$  (Eqs. (2)–(4)).

$$E[\text{Value}; n, \lambda_\alpha] = p_0(V_0 - E[\text{Loss} | H_0]) + p_1(V_1 - E[\text{Loss} | H_1]) \quad (2)$$

where

$$E[\text{Loss} | H_0] = L_0 \cdot [\alpha \cdot DF_t \cdot N + n], \quad (3)$$

$$E[\text{Loss} | H_1] = L_1 \cdot [N - (1 - \beta) \cdot DF_t \cdot N + n], \quad (4)$$

$\alpha$  is the significance level and  $1 - \beta$  is the power of the trial. The optimal sample size ( $n^*$ ) and critical value ( $\lambda_\alpha^*$ ) are jointly determined such that the expected value of the trial is maximized subject to an upper bound,  $\text{Power}_{\max}$ , for the power level, which we set to 80% in our simulations. This power constraint is intended to represent the practical considerations of the medical device industry. In solving the constrained optimization problem, we observe that the expected value of the trial is maximized when the expected loss,  $E[\text{Loss}; n, \lambda_\alpha] = p_0 E[\text{Loss} | H_0] + p_1 E[\text{Loss} | H_1]$ , is minimized.

### Weight-loss device case study

Using BDA and the estimated preference scores, we can now formulate the BDA optimal fixed-sample test for weight-loss devices. We assume an annual discount rate of 10%, a %TBWL

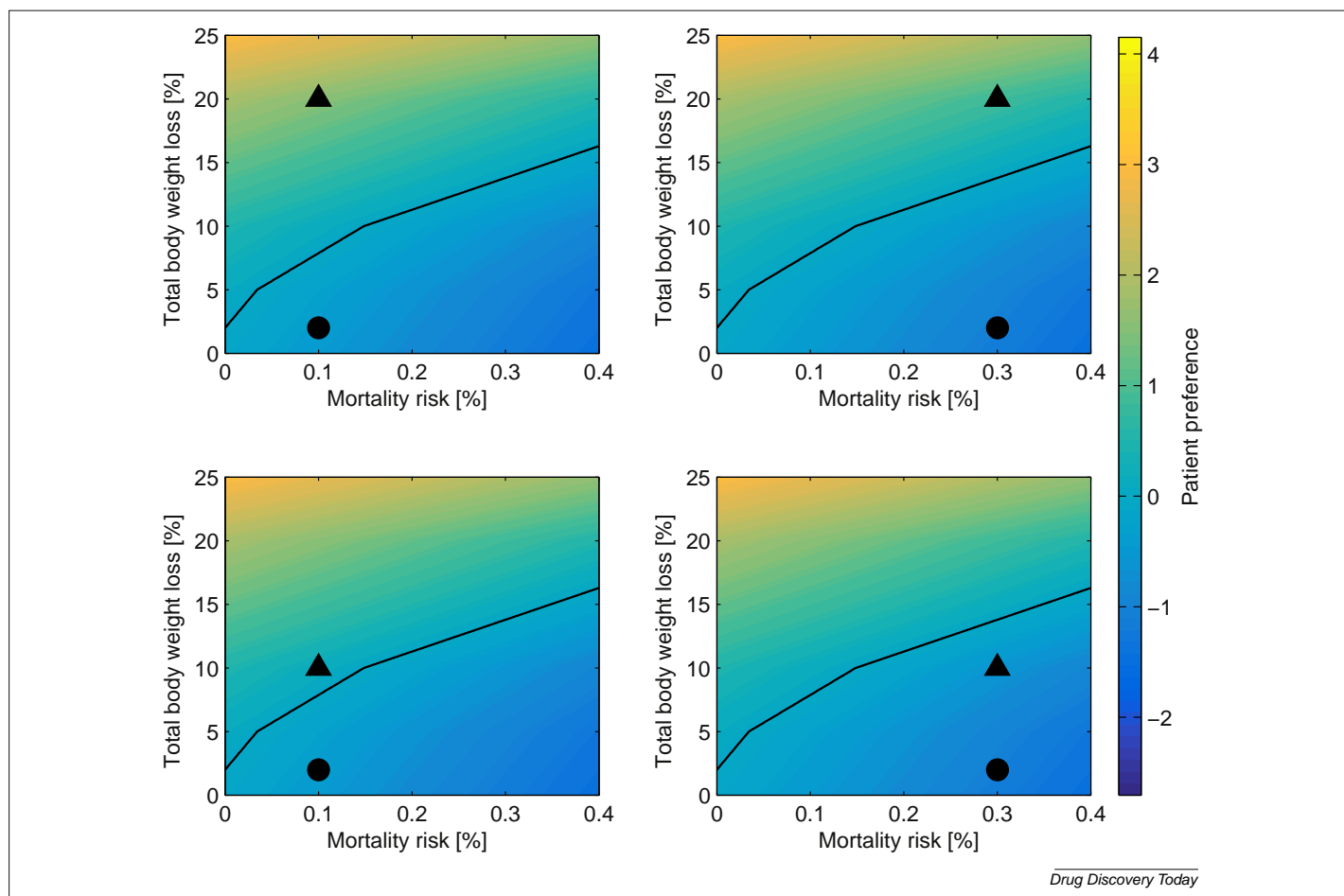


FIGURE 1

Possible device characteristics: low-risk with high weight loss (top left), low-risk with low weight loss (bottom left), high-risk with high weight loss (top right) and high-risk with low weight loss (bottom right). Circles and triangles represent the investigational device characteristics under the null hypothesis ( $H = 0$ ) and alternative hypothesis ( $H = 1$ ), respectively.



TABLE 1  
BDA-optimal RCTs for weight loss devices.

Device characteristics under $H = 1$	Trial size ( $2n$ )	Critical Value ( $\lambda_\alpha$ )	Significance ( $\alpha$ )	Power ( $1 - \beta$ )
Low risk, high %TBWL	44	1.54	6.5%	80%
Low risk, low %TBWL	278	1.83	3.5%	80%
High risk, high %TBWL	78	2.33	1.1%	80%
High risk, low %TBWL	0	–	–	–

standard deviation of 25% for both arms of the study, a patient accrual rate of 100 patients per year (and a fixed study startup time of 6 months, final observation period of 1 year and FDA review time of 9 months) and a target population of 100,000 patients. We also consider two separate categories of interventions: low risk and high risk, which represent devices that require noninvasive and invasive surgeries, and have mortality risks of 0.1% and 0.3%, respectively. In each case, we assume the device is either ineffective ( $\mu_t = \mu_p$ ) or effective ( $\mu_t > \mu_p$ ), with equal prior probability. This is consistent with the equipoise principle of two-arm clinical trials, which states that it is only ethical to assign the same number of patients to both arms if there is no prior information in the medical profession that favors one arm over the other [13]. We further subcategorize effective weight loss into low effectiveness ( $\mu_t = 10\%$ ) and high effectiveness ( $\mu_t = 20\%$ ). Figure 1 summarizes the multiple categories of investigational devices considered. Finally, the control treatment (e.g., diet and exercise) is assumed to provide moderate weight loss ( $\mu_p = 2\%$ ) and have no additional mortality risk. The authors with domain-specific expertise have calibrated the parameters in these hypothetical examples based on multiple years of first-hand knowledge and participation in regulatory review. In the Supplementary material we show sensitivity analyses to investigate the robustness of our analysis to perturbations in our model’s key assumed parameter values (see eTables 3–5). Table 1 lists the optimal RCTs for the devices described above. As can be seen, the device that is assumed to be low risk but is potentially highly effective has a relatively large BDA optimal significance level ( $\alpha$ ) of 6.5%. This value is greater than the traditional 5% level, reflecting the fact that patients are willing to bear increased uncertainty of receiving an ineffective device because (i) the device is thought to be safer and/or (ii) they value the weight loss benefit and do not want to miss the opportunity of receiving an effective weight-loss device and/or (iii) they want to access a potentially effective weight-loss device sooner. The pref-

erence and ability to shorten the regulatory approval process is especially apparent because the trial size is set to 44 patients (22 in each arm), which is approximately six-times smaller than the trial size recommended for the low-risk, low potential weight-loss device.

By contrast, the optimal significance level for the device that is assumed to have a high mortality risk is 1.1%. The additional mortality risk causes the preference score of this device to be relatively low, and hence patients require greater evidence of clinical effectiveness. To achieve this goal, the trial size is set to 78 patients, almost twice the size of the trial recommended for the low-risk device with an equivalent potential effectiveness. However, the preference to keep the trial size small remains evident, especially when compared with the 278-patient trial size recommended for the low-risk, low potential weight-loss device. Finally, note that, because the high-risk, low potential weight-loss device is not preferred to the standard treatment under the alternative hypothesis (Fig. 1), the BDA framework recommends this device be rejected without conducting a trial.

Although the previous analysis studied the benefit–risk preferences of the average patient, we can also calculate BDA optimal trial designs for more-risk-tolerant early adopters. In this analysis, we derive counterpart BDA optimal RCTs for a subset of patients who have a fraction ( $\gamma$ ) of the risk aversion of the average survey respondent. For example, if  $\gamma = 2/3$  a patient’s risk aversion is one-third smaller in magnitude when compared to the average patient’s risk preferences seen in eTable 1 (see Supplementary material online). We use this coefficient for illustrative purposes only and recommend that preference scores from population subsets, including categorization by disease severity and other demographics, be elicited directly from survey responses. Table 2 compares the BDA optimal designs for patients who are one-third ( $\gamma = 2/3$ ) and two-thirds ( $\gamma = 1/3$ ) less risk averse than the average patient.

TABLE 2  
BDA-optimal RCTs for weight loss devices for three risk-tolerance groups.

Device characteristics under $H = 1$	Risk aversion ( $\gamma$ )	Trial size ( $2n$ )	Critical value ( $\lambda_\alpha$ )	Significance ( $\alpha$ )	Power ( $1 - \beta$ )
Low risk, high %TBWL	1	44	1.54	6.5%	80%
	2/3	34	1.26	10.9%	80%
	1/3	20	0.77	22.5%	80%
Low risk, low %TBWL	1	278	1.83	3.5%	80%
	2/3	190	1.36	8.7%	80%
	1/3	94	0.71	24.0%	80%
High risk, high %TBWL	1	78	2.33	1.1%	80%
	2/3	62	1.99	2.6%	80%
	1/3	44	1.54	6.5%	80%
High risk, low %TBWL	1	0	–	–	–
	2/3	0	–	–	–
	1/3	278	1.83	3.5%	80%

The BDA optimal RCT designs vary substantially across risk-tolerance groups. With a one-third decrease in risk aversion, the sample size decreases by a factor of 21%–32% and the significance level increases in magnitude by 1.5%–5.2%, such that the recommended  $\alpha$ s for low- and high-risk devices are on the order of 10% and 2.5%, respectively. The small number of patients and large  $\alpha$ s for the risk-tolerant subgroups relative to the average population are observed because the BDA optimal RCT, by being less conservative, aims to grant faster access to these ‘early adopters’. The decrease in the length of the regulatory approval process is a consideration to offset the excess risk from the extra permissiveness in the trial, and the overall penalty—the expected harm to current and future risk-tolerant patients—can be minimized. Finally, note that, if effective, the high-risk, low potential weight-loss device is preferred to the standard treatment by the most-risk-tolerant subgroup (in fact they view the device similarly to how the average population views the low-risk, low potential weight-loss device), and the BDA framework recommends conducting a conservative trial.

### Concluding remarks

We present a quantitative framework in which patient preferences are the center of RCT design. We quantify the loss in value to public health associated with different actions in any fixed-sample RCT, use a BDA framework to aggregate the value of the trial and then determine an optimal RCT in which the expected value is maximized. We tailor this framework to weight-loss devices, using quantitative preference evidence elicited from patients through conjoint analysis, and assumptions for RCT statistics to design BDA optimal RCTs for average and more-risk-tolerant patient populations.

Our results demonstrate that the traditional RCT design with a fixed statistical significance level does not necessarily maximize overall value (or equivalently minimize harm) to current and future patients of an investigational treatment. For low-risk devices and risk-tolerant populations, the inefficiency is mainly caused by lengthy RCTs that are too conservative and overprotective of the type I error rate (i.e., too focused on rejecting ineffective treatments and on avoiding the harm caused by false-positives). Missed treatment opportunities do indeed harm patients, and should be considered along with the risk of approving ineffective or risky treatments.

Conversely, for some high-risk devices, such as those that require open surgery, traditional one-sided significance levels of 5% are more permissive than the BDA optimal thresholds. These RCTs enable a larger chance of approving ineffective or riskier treatments, such that the expected benefits are not justified by the risk to patient health. We believe that the more nuanced consideration of significance level and power described here is instructive to the design of future clinical trials. Although we made strong assumptions here for illustrative purposes, these assumptions can be readily relaxed in future work. For example, although we have used point prevalence for simplicity in this article, period prevalence, incidence rates and other epidemiological measures can be used to estimate the total population affected by the outcome of the trial. Moreover, other factors (including the time until the adverse effects of a type I error are discovered after a device is inadvertently approved, measures of disease burden and the

expected time until a new treatment is discovered that is at least as safe and effective as the investigational device) can easily be incorporated into the model [5].

Our findings must therefore be qualified in several respects. First, many clinical trials are noninferiority trials instead of the superiority trials we have considered. Second, we have considered fixed sample clinical trials, when in reality clinical trials for regulatory purposes could be adaptive and might include interim analyses for early signals of effectiveness, futility or lack of safety. Any of these possible adaptations in any given trial could alter the optimal significance level and power and appropriate modifications to our calculations are required to determine the optimal designs in these situations. Third, the trials considered here use the percent of total bodyweight loss as the primary endpoint, and mortality risk as the only safety concern. For weight-loss devices, these attributes are clear and of unambiguous importance. Moreover, total bodyweight loss is a surrogate endpoint for morbidity and mortality; hence, this trial resembles those of many other devices and drugs. Other attributes can also be included, such as weight-loss duration, co-morbidities, side-effect duration, among others, which are more difficult to gauge. Study-specific definitions of type I error and type II error loss would require more nuanced treatment in these trials, but can easily be included in our BDA framework. Fourth, we acknowledge that hypothetical patient choices such as the ones obtained in a discrete choice experiment do not have the same clinical and emotional consequences as actual choices. However, advances in patient preference elicitation methods and best practices have helped increase the reliability of such results by ensuring that respondents are well informed, and that hypothetical biases are minimized. Despite these limitations, the estimated preference scores allow us to develop quantitative models to compare benefit–risk tradeoffs across device attributes. This information is required for making patient-centered, evidence-based regulatory decisions. Finally, we have constrained our attention to patient medical outcomes without considering the financial cost to patients and their families, industry or society. New therapies often come at a very high financial cost, which, when taken into account, can raise the bar of success for new agents, thus lowering the acceptable significance level. By contrast, the larger the target population, the more robust the results will be to the accuracy of  $N$  and the more palatable higher financial costs might be for developing beneficial therapies. Although other decision makers (insurance companies, etc.) might integrate financial costs into consideration, these concerns would not be included in regulatory decisions. The increased significance levels that we have proposed could lower the cost of clinical trials, which has grown to an average of US\$36 500 per patient as of 2013 [14], and reduce the risk to sponsors, which might encourage device development, lower device costs and further accelerate clinical research.

To incorporate perspectives from the entire biomedical ecosystem and realize the value of patient input to the device development process, the CDRH has developed a patient engagement advisory committee consisting of key stakeholder groups: patient advocates, caregivers, physicians, medical device and biopharma executives, regulators and policymakers. It is possible this committee would be an appropriate forum to consider in formulating explicit cost estimates for type I and type II errors. These estimates

can then be incorporated into the FDA decision-making process as additional inputs to their quantitative and qualitative deliberations. This ability of the BDA framework to systematically weigh multifaceted tradeoffs that reflect a variety of perspectives combined with its flexibility and practicality make it a potentially valuable tool for optimal RCT design. Although the framework is robust, we emphasize that careful consideration must be applied to the assumptions underlying the specific models to produce useful recommendations. If correctly implemented, a Bayesian perspective has the potential to benefit all stakeholders.

## Acknowledgments

Research support from the MIT Laboratory for Financial Engineering is gratefully acknowledged. We thank Thomas Gross

and Ram Tiwari from the FDA/Center for Devices and Radiologic Health for helpful discussion and edits to the manuscript. We also thank Don Berry, Leah Isakov and Vahid Montazerhodjat for helpful comments and discussion, and Jayna Cummings for editorial assistance. The views and opinions expressed in this article are those of the authors only and do not necessarily represent the views and opinions of any institution or government agency, any of their affiliates or employees, or any of the individuals acknowledged above.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.drudis.2017.09.016>.

## References

- 1 FDA (2016) *Factors to Consider When Making Benefit–risk Determinations in Medical Device Premarket Approval and De Novo Classifications*. Available at: <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm506679.pdf>
- 2 Ho, M. *et al.* (2014) Incorporating patient-preference evidence into regulatory decision making. *Surg. Endosc.* 29, 2984–2993
- 3 FDA (2016) *Patient Preference Information — Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling*. Available at: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM446680.pdf>
- 4 Isakov, L. *et al.* (2015) *Is the FDA Too Conservative or Too Aggressive? A Bayesian Decision Analysis of Clinical Trial Design*. SSRN <http://dx.doi.org/10.2139/ssrn.2641547>
- 5 (2017) Use of Bayesian Decision Analysis to Minimize Harm in Patient-Centered Randomized Clinical Trials in Oncology. *JAMA Oncol.* 3 (9), e170123 <http://dx.doi.org/10.1001/jamaoncol.2017.0123>
- 6 Berry, D. (2004) Bayesian statistics and the efficiency and ethics of clinical trials. *Stat. Sci.* 19, 175–187
- 7 Berry, D. (2006) Bayesian clinical trials. *Nat. Rev. Drug Discov.* 5, 27–36
- 8 Irony, T. and Simon, R. (2006) Application of Bayesian methods to medical device trials. In *Clinical Evaluation of Medical Devices, Principles and Case Studies* (2nd edn) (Becker, K.M. and Whyte, J.J., eds), pp. 99–116, Humana Press
- 9 Irony, T. *et al.* (2007) Evolving methods: evaluating medical device interventions in a rapid state of flux. In *The Learning Healthcare System: Workshop Summary* (Olsen, L., ed.), pp. 93–99, The National Academies Press
- 10 FDA (2010) *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. Available at: <https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf>
- 11 Bonangelino, P. *et al.* (2011) Bayesian approaches in medical device clinical trials: a discussion with examples in the regulatory setting. *J. Biopharm. Stat.* 21, 938–953
- 12 Irony, T. *et al.* (2012) Statistical methods in clinical trials. In *Statistical Methods in Healthcare* (Faltin, F., ed.), pp. 22–54, Wiley
- 13 Freedman, B. (1987) Equipoise and the ethics of clinical research. *N. Engl. J. Med.* 317, 141–145
- 14 Battelle Technology Partnership Practice (2015) *Biopharmaceutical Industry-Sponsored Clinical Trials: Impact on State Economies*. Battelle Memorial Institute, Columbus, OH