



Bayesian Statistics for Medical Devices: Progress Since 2010

Gregory Campbell^{1,2} · Telba Irony² · Gene Pennello³ · Laura Thompson⁴

Received: 8 September 2022 / Accepted: 24 December 2022 / Published online: 3 March 2023
© This is a U.S. government work and its text is not subject to copyright protection in the United States 2023

Abstract

The use of Bayesian statistics to support regulatory evaluation of medical devices began in the late 1990s. We review the literature, focusing on recent developments of Bayesian methods, including hierarchical modeling of studies and subgroups, borrowing strength from prior data, effective sample size, Bayesian adaptive designs, pediatric extrapolation, benefit-risk decision analysis, use of real-world evidence, and diagnostic device evaluation. We illustrate how these developments were utilized in recent medical device evaluations. In Supplementary Material, we provide a list of medical devices for which Bayesian statistics were used to support approval by the US Food and Drug Administration (FDA), including those since 2010, the year the FDA published their guidance on Bayesian statistics for medical devices. We conclude with a discussion of current and future challenges and opportunities for Bayesian statistics, including artificial intelligence/machine learning (AI/ML) Bayesian modeling, uncertainty quantification, Bayesian approaches using propensity scores, and computational challenges for high dimensional data and models.

Keywords Prior Information · Hierarchical Bayesian modeling · Bayesian adaptive designs · Benefit-risk decision analysis · Real-world evidence · Diagnostic test accuracy

Introduction

This paper is aimed at providing for non-statisticians an accounting on the use of Bayesian statistics in medical device clinical studies conducted for regulatory purposes. Our primary emphasis will be on the period after 2010, the year the Food and Drug Administration (FDA) issued their guidance document “Use of Bayesian Statistics in Medical Device Clinical Trials” [1]. Our paper updates the 2011 report [2] After describing a brief early history, we discuss Bayesian developments in borrowing from prior information,

exchangeability, effective sample size, dynamic borrowing, pediatric extrapolation, benefit-risk assessment, real-world evidence, diagnostic applications, and illustrating concepts with recent medical device examples. In the Supplementary Material, we provide a recent list of FDA-approved medical devices for which Bayesian studies were submitted in support of approval.

For a parameter such as a treatment effect or a safety endpoint, the Bayesian approach combines prior information (*prior distribution*) with information from the newly observed data through the *likelihood function* to obtain the *posterior distribution* for the parameter, where the *likelihood function* is a mathematical model that provides information about the parameter from the new data alone. The Bayesian approach provides an alternative to the frequentist (non-Bayesian) approach, which does not incorporate any prior information. By incorporating prior information, posterior estimates of parameters are often (but not always) more precise than frequentist estimates.

✉ Gregory Campbell
GCStat@verizon.net

¹ GCStat Consulting LLC, 14605 Sandy Ridge Road, Silver Spring, MD 20905, USA

² Quantitative Sciences Consulting, Statistics and Decision Sciences, The Janssen Pharmaceutical Companies of Johnson & Johnson, 1125 Trenton-Harbourton Road, Titusville, NJ 08560, USA

³ Center for Devices and Radiological Health, U.S. Food and Drug Administration, Silver Spring, MD 20993, USA

⁴ Center for Biologics Evaluation and Research, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD 20993, USA

Brief Early History

In the late 1990s, the FDA Center for Devices and Radiological Health (CDRH) started to consider the use of Bayesian statistics in medical device clinical trials. Early success of the Bayesian initiative has been described in several publications [3–6]. The four of us, then in the Division of Biostatistics in CDRH, were the primary writing team for drafting the FDA guidance, which was finalized in 2010 by CDRH and the Center for Biologics Evaluation and Research. The guidance distinguishes two basic approaches: (1) borrowing information (data) from past studies using hierarchical Bayes or other methodologies and (2) designing a Bayesian adaptive study usually with no prior information but rather relying on accumulating data within the clinical study to potentially make preplanned changes to the design. For making statistical inferences, the guidance outlines using posterior probabilities for trial success (e.g., superiority or non-inferiority) rather than frequentist p -values and Bayesian credible intervals as opposed to the frequentist confidence intervals.

Two key messages in the guidance are that Bayesian approaches require preplanning, and when using prior studies, the appropriateness of prior studies needs to be evaluated and agreed upon by the FDA. An entirely subjective prior, that is, a prior based entirely on expert opinion, is discouraged because expert opinion is more susceptible to generating disagreement among stakeholders than data from prior studies. The guidance explicitly states “device approval could be delayed or jeopardized if FDA advisory panel members or other clinical evaluators do not agree with the opinions used to generate the prior.”

Borrowing Prior Information

Prior information may play an important role in evaluating medical devices. Medical devices often advance rapidly as in coronary devices [7]. Prior information from a previous study may be leveraged.

Bayesian methods are particularly well suited to combining information from different data sources [8]. The guidance focused on borrowing from prior information based on hierarchical modeling, but more recently developed Bayesian methods may also be used. We review hierarchical models and describe additional borrowing methods.

Bayesian Hierarchical Modeling and Exchangeability

A typical Bayesian hierarchical model across studies assumes that participants within a study are exchangeable and that, at a higher hierarchical level, the studies are

exchangeable with each other. Units, whether they are studies, participants within a study, or subgroups (see [Bayesian Subgroup Analysis](#)), are exchangeable if the probability of observing any particular set of outcomes on those units is invariant to any re-ordering of the units. Consider multiple studies of the performance of the same or similar devices. If all participants across the studies were exchangeable, then the data from these separate studies could be completely pooled together as if all participants were coming from the same population. However, this is inadvisable since there are always differences among studies even if the protocols are quite similar. For an example of exchangeability of trials for an adverse event, all trials could be considered exchangeable if the adverse event rate for any trial is no more likely to be larger or smaller than that of the other trials. Consequently, exchangeability of studies is not reasonable if, for example, a new device under study is expected to perform better than previously studied devices. It is important to highlight that the assessment of exchangeability requires both clinical and statistical expertise. In addition, hierarchical modeling usually assumes that each study has an underlying performance parameter, e.g., a mean, coming from a super-population of means that itself has an overall mean and standard deviation. A hierarchical model “shrinks” the estimated mean from a primary study toward the means from prior studies by borrowing strength from these studies, which increases precision of estimation. Borrowing increases as the standard deviation of the means decreases (assuming all other aspects of the shrinkage estimate are equal). However, when only a few prior studies are available, the estimate of the standard deviation of the means can be unstable.

The prior and current studies may not be exchangeable if their baseline covariates differ. However, the studies might still be exchangeable if conditioned on the covariates. See Pennello and Thompson [4] for an example using a hypothetical device for coronary plaque reduction.

Several approved devices listed in the Supplementary Material used Bayesian hierarchical models. For example, to assess non-inferiority of a drug-eluting coronary stent compared to another coronary stent with respect to 12-month target lesion failure (TLF) rate, a Bayesian hierarchical model was used to borrow data from two prior trials [9]. The model included a bias parameter between the TLF rates in the two prior trials and the TLF rate of the current trial, in each treatment group, to reflect the potential for different primary endpoint results from the current compared to prior trials. The model also adjusted for age, diabetes, history of PCI, ischemic status, and LVEF to mitigate the effect of possible differences among the 3 studies. The posterior mean TLF event rate at 12 months was 6.32% in the treatment group compared to 8.90% in the control group (with 95% CI on the difference: -5.47% to 0.13%). The posterior probability that the difference in 12-month TLF rates was less than

the pre-specified margin of 3.85% was 100.0%, which was greater than 97.5% pre-specified criteria for success.

Power Prior

In the Bayesian power prior approach, the likelihood function for the prior data is raised to a power “alpha” between 0 and 1 representing how much the prior data will be discounted, where 1 represents no discounting (complete borrowing) and 0 represents complete discounting (no borrowing) [10]. Alpha can be fixed before the observing the new data (static borrowing), or dynamic where it adapts to the similarity between the new and prior datasets.

As an example, Ye et al. [11] reanalyzed the US study of 686 subjects from the approved placental immunoassay diagnostic device for spontaneous preterm delivery by incorporating 511 prior subjects from a supplementary European study reported in the Summary of Safety and Effectiveness (SSED) [12] using a power prior approach. The posterior mean of alpha was 0.216, corresponding to information borrowed from approximately $511 * 0.216 = 111$ prior subjects. The posterior mean of sensitivity and specificity were 49.9% and 98.0%, respectively, which were similar to the estimates without borrowing, but more precise.

Commensurate Priors

Commensurate prior methods [13, 14] use hierarchical models with a commensurability parameter to control borrowing from prior sources. The prior on the mean (or other parameter of interest) of the current data is centered around the mean of the prior data to represent a bias between the historical and current populations. A “spike and slab” prior [15] on the commensurability parameter is used to favor the current information when there is evidence of heterogeneity by specifying only a spike of probability that the prior and current data sources are homogeneous.

The power and commensurate prior methods may be applicable when there is only one prior study, as it may be easier to provide a prior distribution on a bias or power parameter versus a between-study variance as for the hierarchical model.

Effective Sample Size

In Bayesian statistics, the *effective sample size (ESS)* for a parameter (e.g., device effect) is the size of the data that the Bayesian analysis effectively used to estimate the parameter based on its posterior distribution. Malec [16] defined *ESS* as the sample size N times the ratio of the posterior variances of the parameter when its prior information is ignored vs. when it is utilized. This definition is natural because under common data sampling designs the variance of an estimator is approximately inversely proportional to sample size. *ESS* was calculated for the rate of major adverse cardiovascular events of a coronary stent to quantify how much data the Bayesian model might borrow from the prior information on previous generation stents [4, 17, 18].

To illustrate, consider data on the successful placement of three versions of the intracranial stent [19] as prior information for a hypothetical fourth generation stent to be studied in $N = 60$ patients (Table 1). A hierarchical model was implemented in the freely available software OpenBUGS (<https://www.mrc-bsu.cam.ac.uk/software/bugs/openbugs/>). The OpenBUGS code for the intracranial stent data is provided (Supplementary Material). Based on the variance of a binomial proportion, the *ESS* for the proportion of successfully placed 4th generation stents can be defined as $(\text{posterior mean}) * (1 - \text{posterior mean}) / (\text{posterior variance})$. For a study of size $N = 60$, when the sample proportion for the 4th generation stent is 40/60, the posterior mean and standard deviation are 0.6712 and 0.05258, yielding $ESS = 0.6712(0.3288) / 0.05258^2 = 79.8$, which rounds to 80. Based on this *ESS*, the sample size that is effectively borrowed from the prior, known as the *prior effective sample size*, is $PESS = ESS - N = 80 - 60 = 20$, the difference

Table 1 Effective sample size (ESS) and prior effective sample size (PESS) for successful placement rate of 4th generation intracranial stent under hierarchical model borrowing strength from data on three previous generation stents

Stent	Data			Posterior			
	N	X	Pct.	Mean	Std	ESS	PESS
4th	60	NA	NA	65.8	16.9	NA	8
4th	60	40	66.7	73.5	4.9	81	20
4th	60	30	50.0	52.9	6.2	66	6
3rd	56	42	75.0				
2nd	59	37	62.7				
1st	8	6	75.0				

N = study size; X = number of successful placements; Pct = $100 * X / N$; ESS = $\text{Mean} * (100 - \text{Mean}) / \text{Std}^2$; PESS = $ESS - N$; Std = standard deviation

between ESS and the actual sample size N (Table 1, 2nd row). If instead the sample proportion is 30/60, then the posterior mean and standard deviation are 0.5291 and 0.06162, yielding $ESS = 0.5291(0.4709)/0.06162^2 = 65.6$, which rounds to 66, and $PESS = 66 - 60 = 6$ (Table 1, 3rd row). The expected value of $PESS$ at the outset, denoted $E(PESS)$, may be important to calculate at the planning stage of a Bayesian study. For the 4th generation stent, $E(PESS)$ is obtained from the posterior distribution of the success proportion by designating the number of successfully placed stents as missing (“NA”) in the OPENBugs code. The posterior distribution is then really the *prior* distribution (sometimes called the *prior predictive* distribution) because it is based on the prior data alone from the earlier generation stents. In this case, the posterior mean and standard deviation are 0.6577 and 0.1697, yielding $E(PESS) = 0.6577(0.3423)/0.1697^2 = 7.8176$, which rounds to 8. (Table 1, 1st row). $E(PESS)$ is the amount of borrowing expected from the prior distribution at the outset, but will likely differ from the actual $PESS$ after the data are observed. Definitions of $PESS$ at the outset have been proposed based on Fisher information and have shown promise [20, 21].

Dynamic Borrowing

Bayesian models that permit dynamic borrowing from prior information allow a trialist to adapt the amount of borrowing based on similarity between the prior historical data and data from accrued subjects. This contrasts with static borrowing which pre-specifies the amount of borrowing before beginning the current study, such as with a fixed alpha parameter in the power prior approach [22]. If the current study results end up being different from the historical study results, dynamic borrowing allows an adjustment to the amount borrowed to avoid potentially biased study conclusions.

The methods discussed in [Bayesian Hierarchical Modeling and Exchangeability](#), [Power Prior](#) and [Commensurate Priors](#) can be considered as dynamic borrowing approaches or extended to incorporate dynamic borrowing. Hierarchical models, in general, can be viewed as dynamic borrowing because the amount borrowed is influenced by parameters that are updated based on similarity of current study data with prior study data. For power priors, if a non-degenerate prior distribution is placed on the alpha parameter, then discounting (i.e., lack of borrowing) is based on its posterior distribution given current and prior data [10]. However, most dynamic approaches to power priors have empirically estimated the alpha parameter based on similarity of current and prior data [22]. Similarity could be determined at an interim look or at the end of the study, and can incorporate constraints or additional information based on clinical input or regulatory consideration [23–25]. Commensurate priors can also be extended to incorporate dynamic borrowing for

purpose of optimizing the number of subjects randomized to a current control group by assessing the similarity of current control with the historical controls [26].

Baseline covariates can be included with all these methods. Kotalik et al. [27] illustrate how to do dynamic borrowing in the presence of treatment effect heterogeneity where covariates that modify a device effect are differentially distributed across studies.

Pediatric Extrapolation

Medical device trials in pediatric populations may be difficult to conduct due to recruitment difficulties, ethical considerations, and variable age range for the defined pediatric population. While device performance may be similar in adult and pediatric populations, outcome differences may exist. If outcome data are available in both pediatric and adult subjects, then a Bayesian approach can be used to borrow strength from the adult data to estimate device performance in the pediatric population, while recognizing possible differences among the populations. This borrowing is commonly referred to as “[pediatric extrapolation](#)” because for some variables (e.g., age, weight, height) the adult distribution will overlap little or not at all with the pediatric distribution.

The FDA Guidance “[Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices](#)” [28] includes advice for extrapolating from adult to pediatric populations, with an appendix that describes the use of Bayesian hierarchical modeling. Bayesian methods for borrowing from adult data should consider differences in age-related variables. If the device treatment effect on a pediatric population is of lower (or higher) magnitude than on adults due to the pediatric population having different levels of one or more covariates that influences the treatment effect, treatment-by-covariate interactions may be modeled. A proportional interactions model [29] could be used where some of the interactions have the same proportionality across the covariates to simplify model fitting.

An example where pediatric extrapolation was used to support device approval is a vagus nerve stimulator that was approved for children aged 4 to 11 years with partial onset epilepsy. The company had an indication for patients 12 and older. The main prospective study was limited to 30 subjects aged 4–11 years. Evaluation of effectiveness used a simple two-level hierarchical model that borrowed strength from four prior studies of the device on patients ranging in age from 4 to adult, but these prior studies contained limited data on patients who were 4 to 11 [30]. The primary effectiveness endpoint was the proportion of patients with at least a 50% reduction in frequency of seizures after 12 months of treatment. The model attenuated the effect in the main study from the observed responder rate of 47% down to 39%, with a tighter credible

interval, illustrating the borrowing strength of the hierarchical model. The outlying 47% rate was shrunk toward the overall rate, making it appear closer to the other estimated rates.

Bayesian Subgroup Analysis

The sample estimate of a treatment effect will vary with each sample. Thus sample estimates of treatment effects within subgroups will tend to have more variation than the actual treatment effects [31]. In frequentist analysis, this extra variation can lead to an abundance of false declarations of significant differences between subgroups in the sample effect when the actual treatment effect differences are zero. In Bayesian hierarchical subgroup analysis [32, 33] this unwanted extra variation is reduced, reducing the proportion of falsely significant differences. For subgroups in a one-way array (i.e., defined by a single factor), the subgroup treatment effects are assumed exchangeable, which enables borrowing of strength, leading to posterior means that shrink the sample estimates of the subgroup effects toward the overall effect. In a decision analysis framework, if the losses are 0, 1, and A ($0 < A < 1$), for making a correct decision, an incorrect decision, and no decision on the sign of the difference in treatment effect between two subgroups, then the resulting Bayes procedure controls at A the proportion of incorrect sign declarations, known as the *directional false discovery rate* [34]. For subgroups defined by multiple factors, main factor effects and factor interaction effects are separately modeled as exchangeable, resulting in posterior means that shrink the sample estimates according to the evidence of variation in each set of effects [35].

Bayesian hierarchical subgroup analysis can be used to adjust an exploratory subgroup finding for multiplicity before utilizing that finding as prior information in a confirmatory study of that subgroup [5]. This approach was taken in the Acute Myocardial Infarction with HyperOxyemic Therapy (AMIHOT) II trial to confirm the subgroup with anterior acute myocardial infarction (AMI) identified in AMIHOT I [36]. The subgroup identified in AMIHOT I was considered exploratory. AMIHOT II was then conducted in that subgroup only to confirm the therapeutic effect. In AMIHOT II, the prior distribution on the therapeutic effect in the subgroup was taken to be the posterior distribution in AMIHOT I based on a hierarchical model across all subgroups considered in that trial.

Bayesian Adaptive Design and Predictive Modeling

The second major approach discussed in the FDA guidance is the Bayesian adaptive design to make preplanned changes based on accumulating data. Chow and Chang [37] review frequentist and Bayesian adaptive designs for clinical trials.

Hobbs and Carlin [38] review Bayesian adaptive designs for drug and device trials. In order for an adaptive design to work well, subjects should not be recruited too quickly because there may be no time to make adaptations.

A Bayesian adaptive design, where the adaptation for sample size reassessment is to increase the sample size, provides an opportunity to get the sample size right for the targeted power. Whereas most fixed sample size trials rarely achieve the desired sample size (either too small or too large), a Bayesian adaptive design can adopt a Goldilocks approach [39]. In a Bayesian adaptive approach to sample size re-estimation, interim results are used to modify the sample size based on observed results. Sample size re-estimation can be based on the predictive distribution not only for subjects already in the trial who have not yet reached the requisite follow-up time for the outcome, but also for subjects yet to be recruited. While the advantage of adaptive designs can be clear-cut, a potential disadvantage of sizing the trial to exactly its optimal sample size is that it leaves no room for error, missing data or for adjudication differences after the trial has been stopped; thus, it is advisable to intentionally slightly overpower the study.

An advantage of a Bayesian adaptive design is that it is sometimes possible to build a model with accumulating data on intermediate and final endpoints that have been observed in the study to predict (via predictive distribution) unobserved final outcomes with a predictive model using intermediate outcomes. Piecewise exponentials are often used to model the primary outcome using time-dependent intermediate ones [40].

A contrast of the similarities and differences between Bayesian designs and general adaptive designs [41] provided the impetus for the development of the FDA guidance on adaptive designs for medical devices [42].

For an electrosurgical ablation system [43], a Bayesian adaptive design was used in the pivotal study for sample size re-estimation at interim looks. The study was fully Bayesian in that statistical evaluation of primary outcomes was performed using Bayesian methodology for all analyses (interim, for sample size determination, and final). At each planned interim look (with maximum sample size of 100), the predictive probabilities of meeting the primary safety and effectiveness endpoints at the end of the trial were used to decide whether to stop patient accrual, stop the trial for futility, or continue enrolling subjects. After the first interim look, there were 55 subjects for safety and 50 for the effectiveness analysis. The predictive probability of meeting the 30-day safety endpoint was 100%, and of meeting the 6-month effectiveness endpoint when all 50 subjects were followed to 6 months was 98.8%. Thus, accrual was stopped. The final analyses were based on the posterior probability of meeting the endpoints, using all enrolled subjects' information.

A Bayesian adaptive design was also utilized for a novel carotid sinus stimulator that delivers resynchronization therapy to reduce symptoms of heart failure in patients⁴⁴. The trial Baroreflex Activation Therapy (BAT) in Patients with Heart Failure and Reduced Ejection Fraction Ineligible for Resynchronization Therapy (BeAT-HF) consisted of two phases: expedited and extended. The expedited phase supported approval of the device based on BAT meeting the safety performance goal of major adverse neurological & cardiac (MANCE) event free rate > 85%, BAT plus medical management (MM) being superior to MM alone in the three intermediate, symptom endpoints of N-terminal pro-B-type natriuretic peptide (NT-proBNP), Minnesota Living with Heart Failure Questionnaire (MLWHF), and quality of life (QOL) at six months follow-up, and a 55% predictive probability of superiority in a 2-year composite of heart failure morbidity and cardiovascular mortality (HFM&CVM). The extended phase of the trial, designed to show superiority of BAT plus MM to MM alone in a 2-year composite endpoint, incorporated a Bayesian adaptive sample size algorithm based on the predictive probability of passing the HFM&CVM hypothesis test [45].

Diagnostic Devices and Bayesian Statistics

Binary diagnostic devices (tests) render (without loss of generality) either a positive or negative test result, indicating likely presence or absence of a medical condition, e.g., a disease. The pre-test (prior) probability of disease is called the *disease prevalence*. The post-test (posterior) probability of disease is called the *predictive value* of the test. *Sensitivity* is the probability that the test is positive in diseased subjects. *Specificity* is the probability that the test is negative in non-diseased subjects. Bayes Theorem determines the positive and negative predictive values (*PPV*, *NPV*) of the test from sensitivity, specificity, and prevalence. More generally, Bayes Theorem determines the predictive value of any type of test output for any type of disease state (e.g., binary, polychotomous, continuous).

In diagnostic test accuracy studies, the test result may be missing or disease status may be unverified for some subjects [46]. When such data are missing, *data augmentation* can be used to simplify Bayesian calculation of the posterior distribution via Gibbs sampling [47, 48], assuming the data are missing at random. Pennello [49] developed Bayesian models and Gibbs sampling algorithms for evaluating diagnostic test accuracy when disease status is *missing not at random*.

Diagnostic tests can be intended to predict the future as exemplified by the diagnostic test to predict spontaneous preterm delivery mentioned earlier [12].

Bayesian Benefit-Risk Assessments

According to the final guidance released by the FDA in 2019 [50], benefit-risk regulatory assessments for market approval of medical devices involve several factors including magnitude of treatment effects, probabilities of adverse events, uncertainty, and patient perspectives. In Bayesian decision analysis, the consequences of factors are quantified by utilities and the choice to use or not use a medical device for a population is the one that provides the maximum expected utility, where the expectation is taken over the posterior distribution. Fu et al. [51] propose how to quantify a medical product's benefit-risk within a Bayesian decision analysis framework in which prior uncertainties are incorporated.

In the design of a clinical trial in emergency medicine, Lewis and Berry [52] used a Bayesian decision-theoretic approach with group sequential methods to balance benefit versus risk. In a review of Bayesian methods in regulatory science, Rosner [53] argues for the use of Bayesian decision theory in clinical research, including early considerations of its use.

The values that patients place on benefits and risks of treatments are *patient preference* information. The FDA guidance on voluntary submission of patient preference information for inclusion in decision summaries and device labeling [54] cites the ISPOR Task Force Report [55] on statistical methods for discrete choice experiments, in which a hierarchical Bayes approach is described for estimating individual patient preferences when available preference information is insufficient.

For post-market surveillance of medical device safety, Hatfield et al. [56] proposed a Bayesian decision-theoretic framework with hierarchical modeling of safety data to formalize the process of trading off real-world costs and benefits of regulatory actions that may be taken in response to device safety problems.

The clinical utility of a diagnostic device (medical test) depends on the clinical consequences of correct and incorrect test classifications of disease status (presence or absence) and possibly of testing itself (e.g., from an invasive test procedure or imaging induced radiation exposure). *Decision curve analysis* [58–60], *relative utility curve* [61], and *net benefit* [62] are measures of clinical utility that have been developed based on considering a rule-in threshold for the risk of disease above which treatment would be recommended according to clinical guidelines [57]. The rule-in risk threshold is the risk at which the expected benefit and expected cost (non-monetary harm) of referring untreated

subjects for treatment are thought to be equal based on the clinical consequences (which may otherwise be intangible to quantify). Conversely, a rule-out risk threshold may also be considered and defined as the risk at which the expected benefit and expected cost of withholding treatment from those scheduled to receive it are thought to be equal. Rule-in and rule-out risk thresholds are likely to not apply to everyone, but utility measures based on risk thresholding may nonetheless be useful for approximating the net benefit of a test in a given clinical application.

For a binary test based on applying a threshold to an underlying continuous or ordinal score, the receiver operating characteristic (ROC) curve is a plot of the test's sensitivity vs. 1-specificity as the threshold is varied. The decision of where to place the threshold on the ROC curve to maximize expected utility is fundamentally a Bayesian decision-theoretic exercise depending on the benefits and costs of accurate and inaccurate classifications, respectively [63].

Bayesian FDA Submission Activity (2011–2021)

A list through 2010 of publicly available FDA documents concerning FDA Bayesian submissions is available in Campbell [2]. In the Supplementary Material, we provide an updated list of FDA Summaries of Safety and Effectiveness (SSEDs) for original PreMarket Applications (PMA) and PMA Supplements (PMASs) that mentioned Bayesian methodology. For some Bayesian studies submitted to the FDA, the Bayesian methods are not described in the SSED but the use of Bayesian methods for their FDA submission are described in publications or other publicly available information as demonstrated in Campbell [2]. Based on our list, FDA approved 16 medical devices from 1998 to 2010 based on PMAs and PMASs for which Bayesian studies were submitted [2]. Since 2010, we identified an additional 31 medical devices approved by FDA based on PMAs and PMASs that relied on Bayesian methodology. For publicly available original PMAs, 14 occurred before 2011, and 23 after 2010. Whereas there were only two publicly identified PMASs for the period from 1999 to 2010, there have been 8 PMASs since. This trend shows more Bayesian PMAs and significantly more PMASs since the 2011 report [2].

There were only 2 hierarchical Bayes and no Bayesian adaptive PMAs pre-2011, but there were at least 8 Bayesian adaptive designs and 6 hierarchical Bayes PMAs after 2010. So again, the trend shows an increase for both designs but more so for Bayesian adaptive. Some of the power prior activity was for choosing a fixed (static) discount factor for the prior information rather than having it updated using current data.

Concerning adaptive Bayesian designs, Yang et al. [64] reported 75 Bayesian submissions out of a total of 251 adaptive designs for the period from January 2007 to May 2013. Here submissions could be multiple ones for an Investigational Device Exemption (IDE), PMA, PMA Supplement, 510(k) (Premarket Notification) or Humanitarian Device Exemption (HDE)) or a pre-submission request for one of these. Of 251, 32 were completed studies (PMAs, PMASs, 510(k)s), of which 14 were Bayesian. There were 225 Original PMA submissions (adaptive and not) during this time period, of which 17 were adaptive and 8 adaptive Bayesian. Of these 75 Bayesian designs, most (76%) used non-informative priors.

Of 15 PMAs and PMASs that have been approved as of April 15, 2022 through the FDA's Breakthrough Designation Program [65], which began in 2017 and is the successor to the Expedited Access pathway of 2015, six were Bayesian [44, 66–70]

Label Expansion by Using Real-World Evidence

Regarding the FDA Guidance on using real-world evidence in regulatory decisions for medical devices [71] there have been many submissions that have used real-world evidence [72]. In particular, Bayesian hierarchical models have been used in at least two examples for label expansion: (1) indications for a drug-eluting coronary stent to patients with diabetes mellitus using data from four clinical trial databases as the source of prior information [73], and (2) expansion of an indication for an implanted autonomic nerve simulator for epilepsy [30] where effectiveness was based on a hierarchical model in which a Japanese regulatory-mandated post-approval study served as the source of observed current data (30 patients), and data from five previous trials as the sources of prior information.

If a medical device that is considered for an indication is already on the market for a different indication, safety data from real-world use of the device for the approved indication may be leveraged to build a prior distribution for safety for the new indication.

The Present and the Future

With regard to medical device trials for regulatory purposes, there has been continued use of Bayesian statistics in general and an increase in the use of Bayesian adaptive trial designs with non-informative priors for sample size reassessment. Bayesian statistics in concert with propensity score methodology may play an important role in the use of real-world evidence (RWE). Bayesian methods in regulatory science

are reviewed by Campbell [74] and Rosner [53]. An increasing emphasis is being placed on teaching medical students Bayesian reasoning for evidence-based medicine [75, 76].

There has been NIH funding for Bayesian medical device trials that started with the Adaptive Designs Accelerating Promising Trials Into Treatments (ADAPT-IT) initiative [77].

While propensity score methodology has proven quite useful to augment clinical trials with other data or to use real-world evidence [78] to build a control group for a single-arm study, Bayesian methods have been developed to borrow strength differentially for stratification. The propensity scores-integrated power prior approach uses both propensity score stratification and power priors to integrate real-world data into a single-arm clinical study or to augment a two-arm trial. When prior and current populations are different at baseline, the propensity score approach utilizing Dirichlet priors for propensity quantiles allows for borrowing more data from the prior study when there is more overlap between the two data sources within a stratum [79, 80].

There is now the tantalizing possibility of using virtual patients (stochastic engineering) generated from *in-silico* models of lead failures [81] and using phantoms to provide prior information for use in clinical trials [82–84], both supported by The Medical Device Innovation Consortium (MDIC), a public–private partnership among industry, non-profit organizations, and federal agencies including the FDA [85]. In 2015, the MDIC virtual patient working group collaborated with FDA to demonstrate how to implement the virtual patient framework in a mock trial design and IDE submission to assess fatigue fracture in a hypothetical new ICD lead. Prior information was generated using an *in-silico* model of lead fracture, which was constructed using past information from similar ICD leads. Details about the mock submission can be found at <https://mdic.org/project/virtual-patient-vp-model/>.

With regard to the Breakthrough Devices Program [65] and Humanitarian Device Exemptions [86] for rare diseases using medical devices, Bayesian methods may be utilized to confidently predict longer-term effectiveness and safety for the former [45] and “probable benefit” for the latter.

More generally, since 2010, the use of Bayesian methods in drug and biologic trials has increased; one example in biologics is the recent BNT162b2 COVID-19 Vaccine Phase 3 Trial [87]. In their Complex Innovative Design Initiative for drug and biologics, FDA has chosen for the Pilot Program several Bayesian proposals [88].

The use of Bayesian methods may also be an aid for Data Monitoring Committees (DMCs or DSMBs) for monitoring ongoing frequentist as well as Bayesian trials [89].

Diagnostics are increasingly being designed for polychotomous classification (more than 2 categories, possibly ordered). For example, in patients with chronic hepatitis C, real-time

shear wave elastography (SWE) and transient elastography (TE) were considered for classification of METAVIR liver fibrosis stages F0–1, F2, F3, or F4 [90]. Bayesian models and computation algorithms for analyzing nominal or ordinal polychotomous data have been developed [91, 92].

Bayesian credible intervals for parameters may be based on highest posterior density, or central posterior probability. Rice and Ye [93] (2022) proposed a unified view of Bayesian credible intervals for single and multiple parameters that may be adopted in future.

While the vast interest in artificial intelligence/machine learning (AI/ML) has led to an explosion in regulatory submissions of diagnostic devices based on AI/ML, especially imaging diagnostics [94], many non-Bayesian ML models lack uncertainty quantification of model output [95]. An advantage of Bayesian ML models is that they automatically provide uncertainty quantification of model output through the posterior distribution. Unfortunately, since current computational algorithms for fitting fully Bayesian ML models have not been satisfactory, strong priors may be necessary [95]. Nonetheless, Bayesian belief networks (BNs) may provide probability interpretations of associations and conditional independence among variables, which are not provided by non-Bayesian ML models. BNs have been developed to predict LVAD mortality using the INTERMACS device database [96] to predict successful wound healing following combat-related trauma based on predictive biomarker data from serum and wound effluent [97] and to predict health outcomes in war wounded for clinical decision support [98].

Bayesian statistical software has become more available and includes STAN, FACTS, SAS PROCs, R packages, and Python libraries. In addition, user-friendly Bayesian interfaces have been created to make Bayesian analysis accessible, such as BEANZ for subgroup analysis [99].

The National Science Foundation, in stating that the field of statistics is at a crossroads, cited nonparametric Bayesian methods as a great development and Bayesian computation as useful in a wide range of applications because it accommodates complex modeling [100].

The future for Bayesian statistics in clinical trials generally and in medical device trials in particular has never been brighter.

Acknowledgements

The authors acknowledge the assistance of Dr. Xuefeng Li of CDRH in helping to provide publicly available information about Bayesian PMA submissions.

Author contributions

All authors contributed to the conception and design of the manuscript. All authors contributed sections and provided comment, input, and edits for the manuscript.

Funding

No source of funding was provided for this manuscript.

Data availability

All data generated or analysed during this study are included in this published article (and its supplementary information files).

Declarations

Conflict of interest

Gregory Campbell is an independent statistical and regulatory consultant for medical devices and pharmaceutical products. Telba Irony is currently an employee of Janssen Pharmaceutical Companies of Johnson & Johnson, receiving salary and equities. Gene Pennello and Laura Thompson declare no conflict of interest.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s43441-022-00495-w>.

References

- U.S. Food and Drug Administration. *The Use of Bayesian Statistics in Medical Device Clinical Trials: Guidance for Industry and Food and Drug Administration Staff*, 2010. <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm>. Accessed 2 Aug 2022.
- Campbell G. Bayesian statistics in medical devices: Innovation sparked by FDA. *J Biopharm Stat*. 2011;21:871–87.
- Campbell G. The experience in the center for devices and radiological health with Bayesian strategies. *Clin Trials J*. 2005;2:359–63.
- Irony T, Simon R. Application of Bayesian methods to medical device trials. In: Becker KM, Whyte JJ, editors. *Clinical evaluation of medical devices, principles and case studies*. 2nd ed. New York: Humana Press; 2006. p. 99–116.
- Pennello GA, Thompson L. Experience with reviewing Bayesian medical device trials. *J Biopharm Stat*. 2008;18(1):81–115.
- Bonangelino P, Irony T, Liang S, Li X, Mukhi V, Ruan S, Xu Y, Yang X, Wang C. Bayesian approaches in medical device clinical trials: a discussion with examples in the regulatory setting. *J Biopharm Stat*. 2011;21(5):938–53.
- O'Malley AJ, Normand S-LT. Statistics: keeping pace with the medical technology revolution. *Chance*. 2003;16(4):41–4. <https://doi.org/10.1080/09332480.2003.10554874>.
- National Research Council. *Combining information: statistical issues and opportunities for research*. Washington, DC: The National Academies Press; 1992. <https://doi.org/10.17226/20865>.
- U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P170030 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170030B.pdf. Accessed 2 Aug 2022.
- Ibrahim JG, Chen M-H. Power prior distributions for regression models. *Stat Sci*. 2000;15(1):46–60.
- Ye K, Han Z, Duan Y, Bai T. Normalized power prior Bayesian analysis. 2022; [arXiv:2204.05615](https://arxiv.org/abs/2204.05615) [stat.ME]. Accessed at [arXiv:2204.05615](https://arxiv.org/abs/2204.05615).
- U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P160052 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf16/P160052B.pdf. Accessed 2 Aug 2022.
- Hobbs BP, Carlin BP, Mandrekar SJ, et al. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*. 2011;67(3):1047–56.
- Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear model. *Bayesian Anal*. 2012;7(3):639–74.
- Mitchell TJ, Beauchamp JJ. Bayesian variable selection in linear regression. *J Amer Statist Assoc*. 1988;83(404):1023–32.
- Malec D. A closer look at combining data among a small number of binomial experiments. *Stat Med*. 2001;20:1811–24.
- O'Malley AJ, Normand SL, Kuntz RE. Sample size calculation for a historically controlled clinical trial with adjustment for covariates. *J Biopharm Stat*. 2002;12(2):227–47.
- O'Malley AJ, Normand SL, Kuntz RE. Application of models for multivariate mixed outcomes to medical device trials: coronary artery stenting. *Stat Med*. 2003;22(2):313–36.
- Kadhodayan Y, Somogyi CT, Cross DT, et al. Technical, angiographic and clinical outcomes of neuroform 1, 2, 2 treo and 3 devices in stent-assisted coiling of intracranial aneurysms. *J Neurointerv Surg*. 2012;4:368–74.
- Morita S, Thall PF, Müller P. Determining the effective sample size of a parametric prior. *Biometrics*. 2008;64:595–602.
- Neuenschwander B, Weber S, Schmidli H, O'Hagan A. Predictively consistent prior effective sample sizes. *Biometrics*. 2020;76(2):578–87.
- Viele K, Berry S, Neuenschwander B, et al. Use of historical control data for assessing treatment effects in clinical trials. *Pharm Stat*. 2014;13(1):41–54.
- Thompson L, Chu J, Xu J, et al. Dynamic borrowing from a single prior data source using the conditional power prior. *J Biopharm Stat*. 2021;31(4):403–24.
- Jiang L, Nie L, Yuan Y. Elastic priors to dynamically borrow information from historical data in clinical trials. *Biometrics*. 2021. <https://doi.org/10.1111/biom.13551>.
- Psioda M, Ibrahim J. Bayesian clinical trial design using historical data that inform the treatment effect. *Biostatistics*. 2019;20(3):400–15.
- Hobbs BP, Carlin BP, Sargent DJ. Adaptive adjustment of the randomization ratio using historical control data. *Clin Trials J*. 2013;10:430–40.
- Kotalik A, Vock D, Donny E, et al. Dynamic borrowing in the presence of treatment effect heterogeneity. *Biostatistics*. 2021;22(4):789–804.
- U.S. Food and Drug Administration. *Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices: Guidance for Industry and Food and Drug Administration Staff*, 2016. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/leveraging-existing-clinical-data-extrapolation-pediatric-uses-medical-devices>. Accessed 23 Aug 2022.
- Kovalchik SA, Varadhan R, Weiss CO. Assessing heterogeneity of treatment effect in a clinical trial with the proportional interactions model. *Stat Med*. 2013;32(28):4906–23.
- U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P970003/S207 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf/p970003s207b.pdf. Accessed 3 Aug 2022.

31. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. *J Res Educ Eff*. 2012;5(2):189–211.
32. Alosch M, Fritsch K, Huque M, et al. Statistical considerations on subgroup analysis in clinical trials. *Stat Biopharm Res*. 2015;7(4):286–303.
33. Henderson NC, Louis TA, Wang C, Varadhan R. Bayesian analysis of heterogeneous treatment effects for patient-centered outcomes research. *Health Serv Outcomes Res Methodol*. 2016;16(4):213–33.
34. Lewis C, Thayer DT. A loss function related to the FDR for random effects multiple comparisons. *J Stat Plan Inference*. 2004;125:49–58.
35. Pennello G, Rothmann M. Bayesian subgroup analysis with hierarchical models. In: Menon S, Peace KE, Chen D-G, editors. *Biopharmaceutical applied statistics symposium*. Springer: Biostatistical Analysis of Clinical Trials; 2019.
36. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P170027 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170027B.pdf. Accessed 3 Aug 2022.
37. Chow SC, Chang M. Adaptive design methods in clinical trials—a review. *Orphanet J Rare Dis*. 2008. <https://doi.org/10.1186/1750-1172-3-11>.
38. Hobbs BP, Carlin BP. Practical Bayesian design and analysis for drug and device clinical trials. *J Biopharm Stat*. 2008;18(1):54–80.
39. Broglio KR, Connor JT, Berry SM. Not too big, not too small: a goldilocks approach to sample size selection. *J Biopharm Stat*. 2014;24(3):685–705.
40. Berry SM, Carlin BP, Lee JJ, et al. *Bayesian adaptive methods for clinical trials*. Boca Raton, FL: CRC Press; 2011.
41. Campbell G. Similarities and differences of Bayesian designs and adaptive designs for medical devices: a regulatory view. *Stat Biopharm Res*. 2013;5:356–68.
42. U.S. Food and Drug Administration. 2016. *Adaptive Designs for Medical Device Clinical Studies: Guidance for Industry and Food and Drug Administration Staff*. Available at <https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm446729.pdf>. Accessed 2 Aug 2022.
43. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P100045 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf10/P100046B.pdf. Accessed 2 Aug 2022.
44. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P180050 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf18/P180050b.pdf. Accessed 2 Aug 2022.
45. Zile MR, Abraham WT, Lindenfeld J, Weaver FA, Zannad F, Graves T, Rogers T, Galle EG. First granted example of novel FDA trial design under expedited access pathway for premarket approval: BeAT-HF. *Am Heart J*. 2018;204:139–50.
46. Campbell G, Pennello G, Yue L. Missing data in the regulation of medical devices. *J Biopharm Stat*. 2011;21(2):180–95.
47. Tanner M. *Tools for statistical inference: methods for the exploration of posterior distributions and likelihood functions*. 3rd ed. Springer; 1996.
48. Little R, Rubin D. *Statistical analysis with missing data*. 3rd ed. Wiley; 2019.
49. Pennello GA. Bayesian analysis of diagnostic test accuracy when disease state is unverified for some subjects. *J Biopharm Stat*. 2011;21:954–70.
50. U.S. 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 at <https://www.fda.gov/media/99769/download>. Accessed 2 Aug 2022.
51. Fu B, Li X, Scott J, He W. A new framework to address challenges in quantitative benefit-risk assessment for medical products. *Contemp Clin Trials*. 2020;95:106073. <https://doi.org/10.1016/j.cct.2020.106073>.
52. Lewis RJ, Berry DA. Group-sequential clinical trials: a classical evaluation of Bayesian decision-theoretic designs. *J Amer Statist Assoc*. 1994;89:1528–34.
53. Rosner GL. Bayesian methods in regulatory science. *Stat Biopharm Res*. 2020;12(2):130–6. <https://doi.org/10.1080/19466315.2019.1668843>.
54. U.S. Food and Drug Administration. *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: Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders*, 2016. Available at <https://www.fda.gov/media/92593/download>. Accessed 2 Aug 2022.
55. Hauber AB, González JM, Groothuis-Oudshoorn CG, Prior T, Marshall DA, Cunningham C, IJzerman MJ, Bridges JF. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health*. 2016;19(4):300–15. <https://doi.org/10.1016/j.jval.2016.04.004>.
56. Hatfield LA, Baugh CM, Azzone V, et al. Regulator loss functions and hierarchical modeling for safety decision making. *Med Decis Mak*. 2017;37(5):512–22.
57. Pepe MS, Janes H, Li CI, Bossuyt PM, Feng Z, Hilden J. Early-phase studies of biomarkers: what target sensitivity and specificity values might confer clinical utility? *Clin Chem*. 2016;62(5):737–42.
58. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26(6):565–74.
59. Kerr KF, Brown MD, Zhu K, Janes H. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34(21):2534–40.
60. Kerr KF, Marsh TL, Janes H. The importance of uncertainty and opt-in v. opt-out: best practices for decision curve analysis. *Med Decis Mak*. 2019;39(5):491–2.
61. Baker SG. Putting risk prediction in perspective: relative utility curves. *J Natl Cancer Inst*. 2009;101(22):1538–1542. Erratum in: *J Natl Cancer Inst*. 2014;106(11):dju337.
62. Marsh TL, Janes H, Pepe MS. Statistical inference for net benefit measures in biomarker validation studies. *Biometrics*. 2020;76(3):843–52.
63. Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem*. 1993;39(4):561–577. Erratum in: *Clin Chem*. 1993;39(8):1589.
64. Yang X, Thompson L, Chu J, et al. Adaptive design practice at the Center for Devices and Radiological Health (CDRH), January 2007 to May 2013. *Therap Innov Reg Sci*. 2016;50(6):710–7.
65. U.S. Food and Drug Administration. *Breakthrough Devices Program: Guidance for Industry and Food and Drug Administration Staff*, December 2018. Available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/breakthrough-devices-program>. Accessed 24 Aug 2022.
66. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P180007 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf18/P180007b.pdf. Accessed 2 Aug 2022.

67. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P180036 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf18/P180036b.pdf. Accessed 2 Aug 2022.
68. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P190016 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf19/P190016b.pdf. Accessed 2 Aug 2022.
69. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P210034 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf21/P210034b.pdf. Accessed 2 Aug 2022.
70. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P170019 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170019b.pdf. Accessed 2 Aug 2022.
71. U.S. Food and Drug Administration (2017). The Use of Real World Evidence to Support Regulatory Decision-Making: Guidance for Industry and Food and Drug Administration Staff. Available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-real-world-evidence-support-regulatory-decision-making-medical-devices>. Accessed 2 Aug 2022.
72. U.S. Food and Drug Administration (2021). Center for Devices and Radiological Health. Examples of Real-World Evidence (RWE) Used in Medical Device Regulatory Decisions. Available at <https://www.fda.gov/media/146258/download>. Accessed 2 Aug 2022.
73. U.S. Food and Drug Administration, Center for Devices and Radiological Health, "PMA P070015/S128 and P110019/S075 FDA Summary of Safety and Effectiveness Data," Available at https://www.accessdata.fda.gov/cdrh_docs/pdf11/P110019S075B.pdf. Accessed 2 Aug 2022.
74. Campbell G. Regulatory acceptance of Bayesian statistics. In: Lesaffre E, Baio G, Boulanger B, editors. Bayesian methods in pharmaceutical research. Boca Raton: CRC Press; 2020. p. 41–51.
75. Kurzenhäuser S, Hoffrage U. Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. *Med Teach*. 2002;24(5):516–21.
76. Sedlmeier P, Gigerenzer G. Teaching Bayesian reasoning in less than two hours. *J Exp Psych*. 2001;130:380–400.
77. Meurer WJ, Lewis RJ, Tagle D, et al. An overview of the adaptive designs accelerating promising trials into treatments (ADAPT-IT) project. *Ann Emerg Med*. 2012;60(4):451–7.
78. Yue L. Regulatory considerations in the design of comparative observational studies using propensity scores. *J Biopharm Stat*. 2012;22:1272–9.
79. Wang C, Li H, Chen WC, et al. Propensity score-integrated power prior approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat*. 2019;29(5):731–48.
80. Li H, Chen WC, Wang C, Lu N, Song C, Tiwari R, Xu Y, Yue LQ. Augmenting both arms of a randomized controlled trial using external data: an application of the propensity score-integrated approaches. *Stat Biosci*. 2022;14(1):79–89.
81. Haddad T, Himes A, Thompson L, et al. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. *J Biopharm Stat*. 2017;27(6):1089–103.
82. Badano A. In silico imaging clinical trials: cheaper, faster, better, safer, and more scalable. *Trials*. 2021. <https://doi.org/10.1186/s13063-020-05002-w>.
83. Jang KJ, Pant YV, Zhang B, et al. Robustness evaluation of computer-aided clinical trials for medical devices. In: Proceedings of 10th ACM/IEEE International Conference on CyberPhysical Systems, April 16–18, 2019, Montreal, QC, Canada. ACM, New York, NY 2019;163–173. Accessed at dl.acm.org/doi/https://doi.org/10.1145/3302509.3311058
84. Badano A, Graff CG, Badal A, et al. Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial. *JAMA Netw Open*. 2018;1(7):e185474.
85. Medical Device Innovation Consortium. Incorporation of stochastic engineering models as prior information in Bayesian medical device trials. 2018 Dec. Available at <https://mdic.org/news/incorporation-of-stochastic-engineering-models-as-prior-information-in-bayesian-medical-device-trials/>. Accessed 23 Aug 2022.
86. U.S. Food and Drug Administration. Humanitarian Device Exemption. Available at <https://www.fda.gov/medical-devices/premarket-submissions-selecting-and-preparing-correct-submission/humanitarian-device-exemption>. Accessed 23 Aug 2022.
87. Polack F, Thomas S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. *N Engl J Med*. 2020;383:2603–15.
88. U.S. Food and Drug Administration. Complex Innovative Design Pilot Program Trial Design Case Studies. Available at <https://www.fda.gov/drugs/development-resources/complex-innovative-trial-design-meeting-program>. Accessed 2 Aug 2022.
89. Fayers PM, Ashby D, Parmar MK. Tutorial in biostatistics Bayesian data monitoring in clinical trials. *Stat Med*. 1997;16(12):1413–30. [https://doi.org/10.1002/\(sici\)1097-0258\(19970630\)16](https://doi.org/10.1002/(sici)1097-0258(19970630)16).
90. Ferraioli G, Tinelli C, Zicchetti M, et al. Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. *Eur J Radiol*. 2012;81(11):3102–6.
91. Albert JH, Chib S. Bayesian analysis of binary and polychotomous response data. *J Amer Statist Assoc*. 1993;88:669–79.
92. Johnson VE, Albert JH. Ordinal data modeling. Springer; 1999.
93. Rice K, Ye L. Expressing regret: a unified view of credible intervals. *Am Stat*. 2022. <https://doi.org/10.1080/00031305.2022.2039764>.
94. Benjamins S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *Npj Digit Med*. 2020;3(1):18. <https://doi.org/10.1038/s41746-020-00324-0>.
95. Dunson DB. Statistics in the big data era: Failures of the machine. *Stat Probab Lett*. 2018;136:4–9.
96. Loghmanpour NA, Kanwar MK, Druzdzel MJ, et al. A new Bayesian network-based risk stratification model for prediction of short-term and long-term LVAD mortality. *ASAIO J*. 2015;61(3):313–23.
97. Forsberg JA, Potter BK, Wagner MB, et al. Lessons of war: turning data into decisions. *EBioMedicine*. 2015;2(9):1235–42. Accessed at <https://walterreed.tricare.mil/Health-Services/Specialty-Care/Murtha-Cancer-Center/Orthopaedic-Oncology/Jonathan-Agner-Forsberg-MD>. Accessed 23 Aug 2022.
98. Stojadinovic A, Eberhardt J, Brown TS, et al. Development of a Bayesian model to estimate health care outcomes in the severely wounded. *J Multidiscip Healthc*. 2010;3:125–35.
99. Wang C, Louis T, Weiss C, et al. Beanz: an R package for Bayesian analysis of heterogeneous treatment effect with graphical user interface. *J Stat Softw*. 2018;85(7):1–31.
100. He X, Madigan C, Wellner J, et al. (2019). Statistics at a crossroads: Who is for the challenge? NSF Workshop report. National Science Foundation. https://www.nsf.gov/mps/dms/documents/Statistics_at_a_Crossroads_Workshop_Report_2019.pdf. Accessed 23 Aug 2022.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.