External Evidence Methods (EEM) Framework

Statistical Methods for Leveraging External Data in Regulatory Decision-Making

A Report of the EEM Working Group
of the Medical Device Innovation Consortium (MDIC)

January 21, 2021

TABLE OF CONTENTS

EXTERNAL EVIDENCE METHODS (EEM) WORKING GROUP MEMBERS

MDIC has assembled a Working Group comprising member organizations and other subject matter experts to guide work on this project.

- Vandana Bhatia, PhD, FDA|CDRH*
- Heng Li, PhD, FDA|CDRH*
- Nelson Lu, PhD, FDA|CDRH*
- Theodore Lystig, PhD, Medtronic
- Changhong Song, PhD, FDA|CDRH*
- Ram Tiwari, PhD, FDA|CDRH*
- Yun-Ling Xu, PhD, FDA|CDRH*
- Lilly Q. Yue, PhD, FDA|CDRH*

**This publication reflects the views of the contributor and should not be construed to represent FDA's views or policies.*

Additional Contributors:

- Jithesh Veetil, PhD, Program Director, Data Science and Technology, MDIC

Acknowledgments:

We appreciate the input from Timothy E. Hanson, PhD and Bradley P. Carlin, PhD. We acknowledge many industry participants who provided responses to the MDIC EEM survey.

DISCLAIMER

The views and opinions expressed in this publication are those of the authors and do not necessarily reflect the views and policies of their respective employers; its management, subsidiaries, affiliates, professionals; or any other agency, organization, or company. The views and opinions in this publication are subject to change and revision.

The general recommendations in this document:
Do not imply FDA concurrence for specific applications
Do not represent the opinion or policy of the FDA or of the companies represented
Do not necessarily reflect the official policy or position of MDIC

ABOUT THE MEDICAL DEVICE INNOVATION CONSORTIUM

The Medical Device Innovation Consortium (MDIC) is the first public-private partnership created to advance the medical device regulatory process for patient benefit.

MDIC was formed in 2012 to bring the FDA and industry together to share vital knowledge that can help bring safe, affordable, and effective devices to patients and providers more quickly. MDIC membership and participation is open to nonprofit, industry, and government organizations that are substantially involved in medical device research, development, treatment, or education; or in the promotion of public health; or that have expertise or interest in regulatory science.

MDIC has been designed to pursue several strategies that support its mission:

- Create a forum for collaboration and dialogue
- Make strategic investments in regulatory science, utilizing working groups to identify and prioritize key issues, and to request, evaluate, and implement project proposals
- Provide and enable implementation of tools from these projects that drive cost-effective innovation

The activities and outputs from MDIC are intended to:

- Ensure that innovative technology is readily available to U.S. patients
- Provide industry and government with methods and tools that may be used to expedite medical device development and the regulatory process
- Reduce the risk and expense of clinical research
- Reduce time and cost of medical device development

MDIC members provide guidance and leadership through collaboration to develop solutions for regulatory, scientific, health, and economic challenges within the medical device and diagnostic industry.

MDIC External Evidence Methods (EEM) Program

MDIC's External Evidence Methods (EEM) program aims to establish a more predictable pathway for use of external evidence methods, such as new, innovative (Frequentist/Bayesian) methods and the cataloging of existing methods for evidence fusion from data external to a clinical trial. External trial data include but are not limited to real-world data (RWD), real-world evidence (RWE), engineering modeling and simulation, and similar device clinical trial data to support regulatory medical device decisions and other stakeholder decisions.

EXECUTIVE SUMMARY

When planning a prospective medical device clinical study, it is sometimes the case that the potential incorporation of data external to the planned study warrants consideration. The increased utilization of such external data, often falling into the category of real-world data (RWD), is partly due to the passage of the 21st Century Cures Act,[1] which requires an expanded role of real-world evidence (RWE) in the approval process of medical products. The purpose of incorporating external data is often to reduce the length of the study, thereby bringing new safe and effective technologies to market sooner to help patients in need. External data may also provide insights into the clinical performance of the device being studied.

The Medical Device Innovation Consortium External Evidence Methods (MDIC EEM) Framework is a document intended to help stakeholders navigate their way through the nuts and bolts of leveraging external data. It catalogs different sources of external data and some of the traditional and novel statistical methods (Frequentist and Bayesian) applicable to the design and analysis of a clinical study in which external data play a role. It also provides references to actual past studies leveraging external data in which some of these statistical methods were successfully applied to support the approval/clearance of medical devices, or the modification of their indications.

There is a wide variety of study designs that involve the leveraging of external data. These external data can be leveraged to augment a single-arm study, or construct or augment either the control arm or the treatment arm (or both) of a comparative clinical study, and they may come from a single source or from multiple sources. The statistical methods cataloged in this Framework can be used for evidence generation in all the scenarios.

Prior to leveraging external data, it is essential to confirm that they are fit for purpose, that is, relevant and reliable. Relevance is the extent to which the data apply to the regulatory question at hand, are amenable to sound clinical and statistical analysis, and are interpretable using informed clinical/scientific judgment. Reliability concerns how the data were collected (data accrual), and whether the people and processes in place during data collection and analysis provide assurance that errors are kept at a minimum and that data quality and integrity are adequate (data assurance).

In many cases, the external data being leveraged are collected in the past so that patients' clinical outcomes have already been recorded. To maintain study objectivity and integrity, careful study design is needed to guard against data dredging (finding patterns in data that can be presented as statistically significant). A study protocol and analysis plan should be created prior to accessing, retrieving, and analyzing external data, regardless of whether the external data are already collected or if they are to be collected in the future. The clinical protocol and analysis plan should address the same elements that a traditional clinical study protocol and statistical analysis plan would cover. In addition, it should include elements such as identification of the external data source(s), a discussion of the relevance and reliability of the data, and a description of the statistical methodology that will be used to leverage external data to make statistical inferences. It is essential that the outcome-free principle[2] be upheld, that is, outcome data should be kept out of sight during study design. Propensity score-based study designs, which are featured in this Framework, do not need any outcome data, and hence can readily conform to the outcome-free principle.

This Framework is in alignment with MDIC's initiatives, including the National Evaluation System for health Technology Coordinating Center's (NESTcc's) mission of accelerating the timely, reliable, and cost-effective development of RWE to enhance regulatory and clinical decision-making. It is also in alignment with other initiatives of MDIC including the Real-World Clinical Evidence Generation: Advancing Regulatory Science and Patient Access for

In Vitro Diagnostics (IVDs) Framework,[3] and data quality and methodological framework documents developed by the NESTcc's data quality and methods subcommittees on RWD and RWE.[4,5]

Purpose of the EEM Framework

This framework is intended to help stakeholders navigate their way through leveraging external data by:

1. Cataloging different sources of external data
2. Cataloging statistical methods that can be considered to leverage external data
3. Considering uses of external data, when appropriate, for premarket regulatory decision-making for medical devices
4. Providing examples to illustrate the application of various statistical methods where external data have been leveraged

1 INTRODUCTION AND SCOPE

The statutory standard for the level of evidence for approval or clearance of devices is one of reasonable assurance of safety and effectiveness based on valid scientific evidence.[6]

The Code of Federal Regulations 21 CFR 860.7(c)(2)[6] defines valid scientific evidence as evidence from "well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use."

The 21st Century Cures Act[1] directed the U.S. Food and Drug Administration (FDA) to expand the role of RWE in the approval process of medical products. In 2017, the FDA (Centers for Devices and Radiological Health/Center for Biologics Evaluation and Research [CDRH/CBER]) then issued a guidance document, *Use of Real-World Evidence To Support Regulatory Decision-Making For Medical Devices*.[7]

The guidance defines RWD as:

> *"Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources. Examples of RWD include data derived from electronic health records (EHRs), claims and billing data, data from product and disease registries, patient-generated data including in home-use settings, and data gathered from other sources that can inform on health status, such as mobile devices. RWD sources (e.g., registries, collections of EHRs, and administrative and healthcare claims databases) can be used as data collection and analysis infrastructure to support many types of trial designs, including, but not limited to randomized trials such as large simple trials, pragmatic clinical trials, and observational studies (prospective and/or retrospective)."*[7]

The guidance defines RWE as the clinical evidence regarding the usage, and potential benefits or risks, of a medical product derived from analysis of RWD. The guidance document states that "RWD and associated RWE may constitute valid scientific evidence depending on the characteristics of the data."[7]

This document provides a framework for leveraging data from sources outside of a prospectively designed study, referred to as external data. There is a large variety of data sources from which external data can be extracted, which includes RWD sources, but also includes data from historical clinical studies and simulated data. This document categorizes these different sources of external data and catalogs some of the statistical methods that can be applied when leveraging external data in a prospectively designed clinical study to make statistical inference. Clinical evidence generated by appropriate analysis of external data can provide insights into the clinical outcomes associated with medical device use. This information can potentially be used in premarket regulatory decision-making throughout the total product life cycle (TPLC) and help establish a streamlined pathway for the use of external data in regulatory submissions.

Whether external evidence derived from an external data source can support a specific type of regulatory decision-making is determined by assessing whether:

(1) Leveraging external data is "fit for purpose" for the specific objectives of the current study
(2) The external data are relevant to the clinical question being asked (relevance)
(3) The external data have adequate quality and integrity essential for regulatory decision-making (reliability)

Such assessment requires considerations from clinical, regulatory, and statistical perspectives. Under the right conditions, evidence derived from leveraging external data can potentially be used to support a wide range of regulatory decision-making for premarket medical devices such as approval/clearance of a new device or an expansion of the indications for use of an already approved device. This framework presents previously approved regulatory case examples of applications where external data were utilized.

Scope of the Framework

This framework identifies:

1. Various sources of external data
2. Traditional and novel statistical methods for generating external evidence
3. Potential uses of external data for premarket medical device regulatory decision-making
4. Examples to illustrate application of statistical methodology where external data have been leveraged

2 EXTERNAL DATA AND EXTERNAL EVIDENCE

Regulatory decision-making regarding medical devices is based on valid scientific evidence. In many cases, such evidence needs to be derived from clinical data. Traditionally, to collect clinical data, an investigational clinical study needs to be planned prospectively. In some circumstances there is an interest in obtaining (supplementary) clinical data from alternative sources to save time, reduce cost, or otherwise alleviate the burden of data collection.

The phrase "external data" is an overarching term for data external to the planned clinical study (from these other sources). There is a large variety of data sources from which external data can be extracted, including what is generally referred to as RWD. The purpose of this document is to catalog statistical methods, so we need a catch-all phrase to capture all the situations in practice where these methods can be applied. The phrase external data serves our purpose well. External evidence is the clinical evidence generated by appropriate analysis of external data. When the external data come from an RWD source, the term external evidence is synonymous with RWE.

Below are some examples of data sources from which external data can be obtained.

Historical clinical studies: Traditional prospective clinical studies conducted in the past. Usually, patient-level data are available. Often, study protocol and study results have been reviewed by regulatory authorities.

Medical device registry: An organized system that continuously and consistently collects relevant data in conjunction with routine clinical care, evaluates meaningful outcomes, and comprehensively covers the population defined by exposure to particular device(s) at a reasonably generalizable scale (e.g., international, national, regional, and health system) with a primary aim to improve the quality of patient care[7]

Lab test databases: Results of tests performed on clinical specimens in order to get information concerning diagnosis, treatment, and prevention of disease

Electronic health record (EHR): An electronic record of health-related information on an individual that conforms to nationally recognized/utilized interoperability standards and that can be created, managed, and consulted by authorized clinicians and staff across more than one healthcare organization[7]

Medical administrative claims data: Claims data that arise from a person's use of the healthcare system and the reimbursement of healthcare providers for that care[7]

Simulated data: Data generated by computer simulation on a virtual patient population using mathematical models that incorporate information on physiological systems

Medical literature: Publications reporting results of clinical studies. They usually contain summary statistics only, but sometimes patient-level data can be obtained upon request.

3 GENERATION OF EXTERNAL EVIDENCE

3.1 Relevance and Reliability of the External Data

Whether external evidence derived from an external data source can support a specific type of regulatory decision-making is determined mainly by the relevance and reliability of the external data source.

Relevance

Relevance is the extent to which the data apply to the regulatory question at hand, are amenable to sound clinical and statistical analysis, and are interpretable using informed clinical/scientific judgment.[7] For an external data source to be relevant to the regulatory question, generalizability to the intended target patient population needs to be established; the data source needs to contain enough information for the assessment of the outcomes of interest; and depending on the application, specific device identification information (e.g., unique device identifier) needs to be captured. In order for data to be relevant for the application of statistical methods covered in this framework, key baseline covariates must be available to adjust for confounding factors that may impact the exposure or outcomes of interest.

Reliability

Reliability concerns how the data were collected (data accrual), and whether the people and processes in place during data collection and analysis provide assurance that errors are kept at a minimum and that data quality and integrity are adequate (data assurance).[7] To ensure reliability for data generated and collected by healthcare providers, the individual sites must implement defined processes, employ qualified personnel, and conduct appropriate training. Furthermore, reliability is enhanced if all the sites share a common data capture form, adopt a common definitional convention, and adhere to a common temporal framework for collection of key data points.

Quality Control

Quality control is an important component of reliability. Historical clinical studies conducted for regulatory purposes typically have good quality control due to regulatory oversight. For registries, quality control can generally be realized by following published recommendations such as those by the International Standards Organization,[8] the Agency for Health Care Quality,[9] the Patient-Centered Outcomes Research Institute (PCORI),[10] the National Medical Device Registry Task Force (*reference to come*), and the International Medical Device Regulators Forum (IMDRF) Registry.[11]

Certain external data sources, such as some administrative and healthcare claims databases or EHRs, may not have established data quality control processes and may not be capable of fully implementing or following the above recommendations. This limitation should be kept in mind when considering the regulatory use of these data sources. Regardless of the original purpose for collecting the external data, procedures of quality assurance should be put into place during the data source design and development stages (when applicable) to optimize the reliability. The quality assurance procedures should be clearly defined and described in a detailed data management standard operating procedures (SOP) manual.

Additional Considerations for Assessing Relevance and Reliability

The assessment of relevance and reliability includes a wide range of factors, which are not exhaustively covered in the above short summary. A comprehensive list of these factors can be found in the FDA RWE guidance.[7] The National Evaluation System for health Technology Coordinating Center's (NESTcc's) Data Quality Framework[4] also contains some discussion of relevance and reliability, applicable specifically for EHR.

3.2 Statistical Methods for Generating External Evidence

This section catalogs some of the Frequentist and Bayesian statistical methods that can be applied when leveraging external data.

Propensity score-based methods require availability of patient-level external data for both clinical outcomes and baseline covariates of interest. In these methods, the study design is based on propensity score methodology, which is utilized to balance baseline covariates between the prospectively enrolled patients and those from external data sources or between the investigational device arm and control arm from different data sources. A key procedural element that enforces objectivity when using these methods involves ensuring those who carry out the prospective study design activities have no access to outcome data.

The Bayesian modeling methods generally only require availability of patient-level external data for clinical outcomes. For some modeling techniques it is possible to incorporate patient-level baseline covariate information when "borrowing" strength from external sources. An important underlying assumption is that patients are exchangeable within each data source and data sources are exchangeable.

(a) Propensity Score Method

This is a statistical method that can be applied when one is to leverage external data to construct a control arm for a non-randomized comparative clinical study with the investigational device group constituted by prospectively enrolled patients.

The propensity score methodology involves two phases: (1) study design and (2) outcome analysis. The study design phase refers to estimating propensity scores and forming groups of participants with similar propensity scores. Techniques such as matching and stratification are often used. Among participants with the same propensity score, the observed baseline covariates are balanced between the two treatment groups to reduce bias. A key advantage of the propensity score methodology is that outcome data are not used in the study design phase and therefore can be concealed.

The two-stage study design proposed by Yue et al.[2] and described in Li et al.[12] provides a paradigm for conducting a comparative observational, non-randomized study within the premarket regulatory setting.

Stage 1 occurs when the clinical protocol is being developed. The key aspects include selection of external data, estimation of sample size, and specification of baseline covariates. To ensure outcome-free study design, an independent statistician is identified who remains blinded to any outcome data.

Stage 2 starts ideally as soon as information on all baseline covariates is available. In this stage, the independent statistician identified in the first design stage estimates propensity scores, implements a propensity score strategy (e.g., stratification or matching), assesses balance in covariate distribution, and finalizes the statistical analysis plan

for future outcome analysis. This iterative process ends when adequate covariate balance is reached. After clinical outcome data have been collected from all patients, statistical inference for the parameter(s) of interest is made.

(b) Propensity Score-Integrated Power Prior Approach for Incorporating RWE in Single-Arm Clinical Studies

This is a statistical procedure that can be applied when one is to leverage external data to augment a single-arm clinical study. The information in the external data is incorporated into the clinical study via the Bayesian method of the power prior.[13] However, rather than directly constructing the power prior based on external data, propensity score methodology is used to pre-select a subset of external data excluding those patients that are dissimilar to those in the current study in terms of observed covariates, and to stratify the selected external patients together with those in the clinical study into more homogeneous strata. Within those strata, the covariates are more balanced between external patients and those enrolled into the clinical study. The power prior method is then implemented in each stratum to obtain stratum-specific posterior distributions, which are then combined to complete the Bayesian inference for the parameters of interest.

(c) Propensity Score-Integrated Composite Likelihood Approach for Incorporating RWE in Single-Arm Clinical Studies

This is a statistical procedure that can be applied when one is to leverage external data to augment a single-arm clinical study. The information in the external data is incorporated into the clinical study by constructing a composite likelihood in which the external data are discounted.[14] However, rather than constructing one composite likelihood based on the combined external data and clinical study, propensity score methodology is used to pre-select a subset of external data excluding those patients who are dissimilar to those in the current study in terms of observed covariates, and to stratify the selected external patients together with those in the clinical study into more homogeneous strata. Within those strata, the covariates are more balanced between external patients and those enrolled into the clinical study. In each stratum, a composite likelihood is constructed for the parameter of interest with the external data being discounted, and the maximum likelihood estimate is derived accordingly. The parameter of interest is then estimated as a weighted average of parameter estimates across all strata.

(d) Propensity Score-Integrated Composite Likelihood Approach for Augmenting the Control Arm of a Randomized Controlled Trial (RCT) by Incorporating RWD

This is a statistical procedure that can be applied when one is to leverage external data to augment the control arm of an RCT.[15] The information in the external data is incorporated into the clinical study by constructing a composite likelihood in which the external data are discounted.

Propensity score methodology is used to design the study. First, the propensity score, defined as the probability of the patient data being from the RCT rather than the external data, is estimated for every patient. Second, external participants are selected if their propensity scores are similar to those in the current study.

Different strata are formed based on all RCT data and selected external data. Within these strata, the covariates are more balanced between external patients and those enrolled in the clinical study. Composite likelihoods are then constructed to analyze the data. In each stratum, a likelihood function is constructed based on participants using investigational intervention, and a composite likelihood function is constructed based on control participants, with the external participants being discounted. The parameter of interest within each stratum is derived based on maximum likelihood estimates. The parameter of interest is then estimated as a weighted average of these parameter estimates across all strata.

(e) Bayesian Hierarchical Modeling

Bayesian hierarchical modeling is a methodology to combine results from multiple studies (or data sources) in which the information is available on several different levels, utilizing Bayesian methods.

In EEM applications, in which data external to a clinical trial are used, a model typically can be written based on the patient level and study (or data source) level so that observations and parameters are structured in a hierarchical manner. Patient outcomes within a study are generated from a population with distribution specified by some study-specific parameters, and the study-specific parameters are generated exchangeably from a common (super) population that is governed by some hyperparameters. A hyperprior can then be specified for these hyperparameters. With such a structure, the posterior distribution of the parameter of interest can be derived using an analytical approach or, more commonly, some numerical methods such as Markov chain Monte Carlo (MCMC).

In regulatory applications, the frequentist properties of a Bayesian design need to be evaluated.[16] A design's operating characteristics, such as type I error rate and power, which are usually derived via simulations, need to be demonstrated as acceptable.

An important assumption in utilizing this methodology is the exchangeability: patients within each study are assumed to be exchangeable, and studies are exchangeable. This assumption enables one to "borrow strength" from the other data sources while acknowledging that the studies are not identical in all respects.

The exchangeability among studies usually depends on the justifications from clinical and engineering perspectives. Sometimes the exchangeability assumption may not hold; however, it is possible that studies are conditionally exchangeable when accounting for some variables. In such a case, these variables need to be incorporated into the model.

Such a situation often arises when leveraging external studies on adult use of a device for extrapolation to the pediatric use. When clinical outcomes differ between adults and pediatric patients, the adult studies and the pediatric study are not exchangeable. However, if clinical outcomes depend on an age-related covariate, the studies may be exchangeable within each given level of the covariate. More details can be found in the FDA guidance, *Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices.*[17]

(f) Bayesian Power Priors

Power priors offer a convenient and statistically principled way to downweight (or discount) the external data. In the first step of this two-step approach, Bayes' rule is applied to generate a posterior on a parameter of interest using only the external data in which the likelihood of external data is raised to a power between 0 and 1. When this "power parameter" is set to zero, no external data are used in the analysis, whereas when the power parameter is set to one, no downweighting occurs. In the second step, Bayes' rule is again applied once the current data are obtained. The posterior of the parameter is derived by combining the likelihood of the current data and the posterior obtained in the first step.

In the case of a Gaussian (normal) external likelihood with a certain number of observations, it can be shown that the effective external sample size is the product of the external number and the power parameter.[18,19] As a result, in practice, the power parameter is often thought of as the "proportion of the external data" being used by the procedure.

While the power parameter can be treated as fixed, it may also be considered random. A hierarchical power prior structure may be used to introduce a distribution (prior) on the power parameter, so there is a joint power prior consisting of the parameter of interest and the power parameter. Adding a hyperprior on the power parameter will allow for greater uncertainty in the analysis compared with a fixed power parameter. However, several authors (*references to come*) caution against the use of this joint power prior as it violates the likelihood principle. Instead, many authors (*references to come*) modify the joint power prior to the product of a normalized conditional power prior and an independent proper prior for the power parameter.

(g) Commensurate Priors

This method was proposed by Hobbs et al.[20-22] It is assumed that the current data likelihood is governed by the parameter of interest, and the external data likelihood is governed by a corresponding parameter that is possibly different from the main parameter of interest. A hierarchical model is built to incorporate a commensurate prior as the primary mechanism for weighting the influence of external information relative to its consistency with data from the current study. The commensurate prior for the parameter of interest is typically specified to be centered around the corresponding parameter for the external data and conditional on a precision parameter, which parameterizes the prior precision. The amount of external data borrowing can thus be modified by tuning this precision parameter. A larger variance would imply that there is less faith in the similarity of the current and external data, and therefore allow the current estimate to stray further from the external data. The posterior kernel of all parameters is the product of likelihoods of current data and external data, the commensurate prior, and hyperpriors of the corresponding parameter and precision parameter. Specification of the hyperprior for the precision parameter is challenging as it is typically poorly identified by the data. Hobbs et al.[20] recommend a "spike and slab" hyperprior, as it helps crystallize the choice between borrowing and not borrowing.

(h) Robust Mixture Priors

A robust mixture prior is the addition of a weakly informative (robust) component to the external data prior via a mixing weight. In the case of a single external data source, the prior for the parameter of interest is a weighted average of two components: (1) posterior on the parameter of interest using only the external data and (2) robust component of the prior. The weight, which takes values between 0 and 1, can be thought of as the prior probability that the current data differ systematically from the external data.[23] When the weight is at the extreme values of 0 or 1, the prior for the parameter of interest reduces to either one of the two components. Carlin and Louis[24] discuss how to update the prior weight to a posterior weight using standard Bayesian calculations. Robust mixture priors are sometimes seen as computationally easier than commensurate priors.

(i) The Discount Power Approach

The discount power approach aims to determine the power parameter in the power prior approach based on how dissimilar the external data are to the current data regarding the clinical outcome of interest.

A discrepancy measure, such as the *P* value of a pooling test (e.g., hypothesis test that means or proportions are different between the external and the current data) or the posterior probability of outcome difference between the external and the current data is first derived. A flexible, tweakable function of the discrepancy measure, termed the "discount function," maps the discrepancy measure to the power parameter (taking values between 0 and 1), enabling fine-tuning of design goals such as power and type I error across various simulated data scenarios.

For example, Haddad et al.[25] used the Weibull discount function that allows for an upper bound on the power parameter that is less than one (forcing partial borrowing only, regardless of how similar the data are); Liu[26] suggests a discount function in a simpler form. While typically all current data and external data are used to derive

the discount function, Thompson et al. (*under revision*) propose to specify the discount function at an interim look of the current study, using the full external data and only part of the current study data. In these proposed discount functions, there is a parameter (other than the discrepancy measure) that acts as a tuning parameter, used to achieve fixed type I error rates in power calculations.

4 USE OF EXTERNAL DATA FOR PREMARKET MEDICAL DEVICE REGULATORY DECISION-MAKING

Evidence derived from external data can potentially be used to support a wide range of premarket regulatory decision-making, such as approval/clearance of a new device or expansion of the indications for use of devices that are already on the market. The acceptable data quality, appropriate study design, and appropriate data analysis play significant roles in any successful use of the external data and require considerations from clinical, regulatory, and statistical perspectives.

There are many ways to use external data in a pivotal medical device study, including the following examples:
(1) Generate hypotheses to be tested in a prospective clinical study
(2) Establish performance goals or objective performance criteria for a single-arm study
(3) Augment a clinical study
(4) Construct an investigational device and/or control arm for a comparative study
(5) Generate a prior for the current study

Statistical methodologies discussed in this document are mostly related to and mainly targeted for examples (3) and (4). Note that, in these applications, clinical and regulatory considerations beyond statistical ones are often critical. For example, determining the maximal amount of borrowing from an external data source generally hangs on clinical and regulatory input.

Sometimes the opportunities to embed an RCT in a registry may be explored. As the infrastructure of a registry may be used as a platform to collect the data, some cost may be saved with such an approach when running an RCT. An example is the SAFE-PCI for Women (Study of Access Site for Enhancement of PCI for Women) trial.[27] Such studies may be denoted as randomized registry trials.[28]

For diagnostic devices, the pivotal clinical study is often a diagnostic clinical performance study. In such a study, diagnostic clinical performance of the diagnostic device is characterized by measure(s) that quantify how closely the diagnostic device output is associated with a clinical reference standard that is used to assess participants for the target condition. The evaluation can often be based on archived specimens obtained from completed studies, for example, archived samples from a disease registry or archived samples from a completed study with different study objectives. Similarly, stored participant data collected from devices that measure study participants directly (e.g., electrocardiogram [ECG], electroencephalogram [EEG], vital signs, image set) can also be used to support device diagnostic performance evaluation. In a prospective-retrospective study,[29,30] the study hypothesis, study design, and statistical analysis plan are prespecified. The archived data/samples/materials are then evaluated according to the prespecified study protocol and objective.

One important consideration when utilizing external data is how to avoid or alleviate bias. There are many potential sources of bias that may be introduced in the process. If the external data were collected years ago, then temporal biases may be present. Medical practice and technology, such as availability of adjunctive therapy and other treatment interventions, may evolve over time. With newly developed guidelines from clinical societies, the definition and adjudication of clinical outcomes may change over time. When the external data are collected from regions outside the United States, potential bias may be caused by regional differences in various factors. Attention may need to be paid to factors such as the clinical facility's equipment, levels of clinical skill, accessibility of care, standards of care, requirement of multiple specialties, and time when a medical intervention was available. Intensity

and rigor of monitoring patient outcome events may vary among different data sources (e.g., a registry database and a traditional clinical study). It is possible that some adverse events are undercounted in some data sources. External data may not help regulatory decision-making without successfully tackling issues related to bias.

Bias due to differences in patient baseline characteristics may be addressed by various statistical methodologies. When evaluating or attempting to mitigate such potential bias, it is vital that all necessary information on important baseline covariates is collected in the external data. In practice, oftentimes there are some missing data in these covariates, so a plan needs to be in place regarding how to handle these missing data and evaluate their impact on outcome analyses. Note that a substantial amount of missing data in one or more important covariates may make it unlikely to successfully address the issue of bias caused by differences in covariates.

In many applications involving the use of external data, some or all of the outcome data may already be available at the study design stage. Without careful considerations and proper implementation, this retrospective nature could often jeopardize the integrity of the study and thus cause difficulties in regulatory decision-making. When study design and data analysis are performed with outcome data in sight, the validity and interpretability of the analysis results may greatly suffer.

The solution to ease the concern is to adopt an "outcome-free" design. The main idea is to mimic a prospectively designed study by separating the study design and analysis, despite the fact that the outcome data are already known (at least to certain individuals). This can be accomplished, for example, by requesting an independent, third-party statistician who has no access to the outcome data to carry out the study design. Upon an agreement of the study design (with the regulatory body), the outcome data can be analyzed according to the concurred study design. The two-stage design, described in Section 3.2, is an example to accomplish such an objective.

To mitigate potential bias, careful study design is needed, and a study protocol and analysis plan should be created prior to accessing, retrieving, and analyzing external data, regardless of whether the external data are already collected or if they are to be collected in the future. Clinical protocols and analysis plans for how external data can be leveraged should address the same elements that a traditional clinical study protocol and statistical analysis plan would cover. In addition, they should include details such as external data source, discussion of relevance and reliability of data, whether the external data are already collected or if they are to be collected in the future, and statistical methodology that will be used to leverage external data to make statistical inference.

An upfront and frequent communication between the regulatory body and sponsor throughout the process has been proven to be an effective approach in using external data for premarket medical device regulatory decision-making.

5 EXAMPLES OF USE OF EXTERNAL DATA IN REGULATORY DECISION-MAKING

This section provides a few examples of past studies where external data were leveraged and some of the above-mentioned statistical methods were successfully applied to support the approval of medical devices, or the expansion of the indications for use of an already-approved device.

Example 1

P100047 – HeartWare

- Device approval: First-generation device
- External data source: Patient registry
- External data used to: Construct control arm
- Statistical method used to leverage external data: Propensity score analysis

https://www.accessdata.fda.gov/cdrh_docs/pdf10/P100047B.pdf

The subject device is the HeartWare Ventricular Assist System. The clinical study was a prospective, non-randomized, contemporaneous-controlled, open-label study. It used the Interagency Registry for Mechanically Assisted Circulatory Support (INTERMACS) as a contemporaneous control. The comparability between HeartWare and INTERMACS was to be evaluated using a propensity score analysis with prespecified baseline covariates. Propensity score stratification was used to test the primary endpoint hypotheses in terms of difference in success rates.

Example 2

P140031/S10 – Edwards Lifesciences

- Device approval: Indication expansion
- External data source: Historical clinical study
- External data utilized to: Construct control arm
- Statistical method used to leverage external data: Propensity score analysis

https://www.accessdata.fda.gov/cdrh_docs/pdf14/P140031S010b.pdf

The subject device is the Edwards SAPIEN 3 Transcatheter Heart Valve (THV). This premarket approval (PMA) supplement was submitted to expand the indication for the Edwards SAPIEN 3 THV to include patients with intermediate surgical risk for aortic valve replacement. The clinical study was a single-arm, non-randomized, historical controlled study to compare TAVR (transcatheter aortic valve replacement) with the Edwards SAPIEN 3 THV system with the SAVR (surgical aortic valve replacement) arm from the previous PARTNER II Trial Cohort A. Propensity score stratification was implemented to balance baseline covariates.

Example 3

P140010/S15 – Medtronic

- Device approval: Indication expansion
- External data source: Historical clinical study and patient registry
- External data utilized to: Construct investigational device and control arm
- Statistical method used to leverage external data: Propensity score analysis

https://www.accessdata.fda.gov/cdrh_docs/pdf14/P140010S015B.pdf

The subject device is the IN.PACT Admiral Paclitaxel-Coated Percutaneous Transluminal Angioplasty (PTA) Balloon Catheter. This supplement was submitted to expand the indication for the device to treat in-stent restenotic lesions in native superficial femoral or popliteal arteries. The investigational device arm was formed from participants in the IN.PACT Global Study conducted outside the United States in years 2012 and 2013, and the control arm was formed from patients in the Society of Vascular Surgery (SVS) Vascular Quality Initiative (VQI) Registry between 2011 and 2014. Due to the nature of this non-randomized, retrospective comparison, the bias of baseline differences was adjusted via propensity score methodology.

Example 4

P070015/S128 – Abbott Vascular

- Device approval: Indication expansion
- External data source: Historical clinical study and patient registry
- External data utilized to: Construct investigational device arm
- Statistical method used to leverage external data: Bayesian hierarchical model

https://www.accessdata.fda.gov/cdrh_docs/pdf7/P070015S128b.pdf

A prospective analysis was designed to expand the indication of the XIENCE Family of Stents to include patients with diabetes mellitus. A Bayesian hierarchical model was utilized to analyze the primary endpoint. The results from the four historical clinical trials (SPIRIT IV, SPIRIT PRIME, XIENCE V USA first-enrollment phase, and XIENCE V USA second-enrollment phase) were considered as prior information. The two external XIENCE databases (Cleveland Clinic and the Wake Forest Baptist Medical Center) were pooled as current data and served as the basis for statistical inference.

Example 5

P970003/S207 – Cyberonics (LivaNova)

- Device approval: Indication expansion
- External data source: Historical clinical studies
- External data utilized: Construct investigational device arm
- Statistical method used to leverage external data: Bayesian hierarchical model

https://www.accessdata.fda.gov/cdrh_docs/pdf/p970003s207b.pdf

The subject device is the VNS Therapy System. The original PMA (P970003) was approved on July 16, 1997, for the indication of an adjunctive therapy in reducing the frequency of seizures in adults and adolescents older than 12 years of age with partial-onset seizures that are refractory to antiepileptic medications. This supplement was submitted to expand the indication for the VNS Therapy System to include patients at least 4 years of age. The sponsor submitted data from four premarket studies, a postmarket study, and a database of clinical use to establish a reasonable assurance of safety and effectiveness of vagus nerve stimulation with the VNS Therapy System for use in the United States for patients aged 4 to 11 years. The main study was the Japanese post-approval study. The other studies served as prior information to be leveraged within a statistical Bayesian hierarchical model.

Example 6

P170030 – Biotronik

- Device approval: First-generation device approval
- External data source: Historical clinical studies
- External data utilized to: Augment both investigational device and control arms
- Statistical method used to leverage external data: Bayesian hierarchical model

https://www.accessdata.fda.gov/cdrh_docs/pdf17/P170030B.pdf

The subject device is the Orsiro Sirolimus Eluting Coronary Stent System. The sponsor conducted a prospective, international, multicenter, randomized controlled study (BIOFLOW-V). Data from BIOFLOW-V patients were combined with data from two historical clinical trials by employing a Bayesian hierarchical approach. The two historical clinical trials, BIOFLOW-II and BIOFLOW-IV, were both multicenter, randomized controlled clinical trials with the same treatment interventions as BIOFLOW-V.

Example 7

DEN140010 - Wallac Oy, a PerkinElmer subsidiary

- Device name: EnLite Neonatal TREC kit
- Device approval: First-generation device
- External data source: Lab test database
- External data utilized to: Validate device clinical performance

https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN140010.pdf

The subject device is an *in vitro* diagnostic device intended for the semi-quantitative determination of T-cell receptor excision circle (TREC) DNA in blood specimens dried on filter paper. The test is indicated for use as an aid in screening newborns for severe combined immunodeficiency disorder (SCID). The study involved testing of archived, retrospective dried blood spot (DBS) samples submitted to the laboratory for routine newborn screening and stored in the Danish biobank. All samples were tested by the EnLite Neonatal TREC kit. The clinical assessment of the study participants was obtained from their medical records to confirm that the newborn at one year of age or older had not been identified with SCID or was not deceased from SCID-related complications and was apparently healthy.

6 ADDITIONAL CONSIDERATIONS

Although the statistical methods cataloged in Section 3.2 represent those that have already been applied or, in the case of novel methods, that will likely be applied in the immediate future, they do not constitute all the methods that can possibly be used for leveraging external data. The latter, however, cannot be covered in any single document, especially since statistical innovations in this area are taking place at an unprecedented speed. Nevertheless, to broaden the horizon, a small sample of potentially applicable methodological approaches is provided below, as well as areas where a suitable statistical method may not have been established.

Cardinality Matching: Propensity score features prominently among the methods provided in Section 3.2. Another promising approach to accounting for baseline differences that is still in development is the use of cardinality matching achieved via mixed-integer programming.[31] Advantages of this approach include the ability to obtain matched samples that are both balanced and representative of a population of interest (facilitating the estimation of estimands such as the average treatment effect for certain populations) and the ability to perform matching efficiently even in data sets of sizes up to hundreds of thousands of individuals. Medical device applications may rarely push the upper limits of such extreme size boundaries, but the computational tools that make matching in such large groups feasible can also pay dividends if simulations need to be performed for the method in smaller patient data sets.

Meta-analysis: When there are multiple external data sources available with summary measures of outcomes, one may consider combining the summary measures using a fixed or random effects model, and then use the overall summary measures to construct an informative prior for the current study. The prior can be constructed by inflating the overall variance by using a scale parameter or by using a (known) power-prior approach. Sometimes, the scale parameter is assigned a prior in a hierarchical Bayesian setup.

Network Meta-analysis: Consider a scenario where the information on the current active control C is limited but there are historical clinical studies available where C was compared with C_1; and C_1 with C_2, …, C_{k-1} with C_k. The clinical evidence for C can be leveraged from external data sources involving these other controls C_1, …, C_k. Here, C was directly compared with C_1, but C is indirectly compared with C_2, …, C_k. Then, information on C can be indirectly extracted from C_1, …, C_k using a network meta-analysis.

Dirichlet Process Prior and Its Use in Semi-parametric Models: In situations where there are a large number of study sites and the patient-level outcomes vary across sites, one may, for example, assume that y_{ij}, the response of ith participant in jth site, is normally distributed, $N(\mu+ \alpha_j, \sigma_j^2)$, and then use a Dirichlet process to model the site-level effects; that is, $\alpha_j \sim DP(M_0, G_0)$, where G_0 is the baseline distribution representing the prior guess of the distribution of α_j, and M_0 is the precision parameter representing the confidence in G_0.

Modeling and Simulation: The use of modeling and simulation approaches for bench testing is well established, but their use in clinical trial applications is less so. Nevertheless, it is conceivable that combining modeling and simulation information with collected clinical study data may become more common someday. A previous MDIC working group made substantial strides in this area.[32]

Nonetheless, there are a few areas where further development would be of value. Some of these areas include determining the level and type of validation needed for the models used, how to appropriately propagate both epistemic and aleatory uncertainty from the mechanistic model into a synthesized analysis including patient data

and best practices for verification, validation, and uncertainty quantification of the mechanistic model (especially when combined with clinical data). In general, the community would benefit from the ongoing work to develop credibility assessment frameworks for these models. Creating additional opportunities for interactions between modelers and statisticians will enhance the familiarity of each group with the work of the other, and likely lead to better understanding and additional deployment of hybrid trial solutions.

7 RELATION TO OTHER WORK

The objective of leveraging external patient-level data is often to reduce the required number of prospectively enrolled patients, thereby saving time and bringing new safe and effective technologies to market sooner to help patients in need.

Sometimes, there are situations in device studies where randomization is impossible, difficult, or potentially inappropriate. For example, it may not be ethical to randomize participants to sham control, or investigators may face an ethical dilemma in recommending a randomized study to participants when they believe that there is clinical equipoise. In such situations, external data may be leveraged either to construct the control group and conduct a non-randomized comparative clinical study for an investigational device, or external data may be synthesized to produce a numerical value, such as an objective performance criterion (OPC) or performance goal (PG) and conduct a non-comparative clinical outcome study.

In other situations, it may be practically challenging to enroll a sufficient number of participants in a clinical study for the target population of interest, for example, when there is a lack of scientific evidence available to substantiate submissions for devices that are indicated for use in the diagnosis or treatment of pediatric patients. In such situations, it may be appropriate to leverage existing "adult data" to support demonstration of reasonable assurance of safety and effectiveness (RASE) of a medical device for pediatric use.[17] A Bayesian hierarchical model or Bayesian power prior may be applied to existing adult data to generate "external evidence" and to combine with pediatric data to increase the sample size of a prospective pediatric study.[33]

The current document also is in alignment with, and in some respects an extension of, framework documents developed by the National Evaluation System for health Technology Coordinating Center (NESTcc) and the MDIC framework for Real-World Clinical Evidence Generation: Advancing Regulatory Science and Patient Access for *In Vitro* Diagnostics (IVDs). The NESTcc data quality framework specifically focuses on data quality of EHR, and the methods framework contains a template for protocols of medical device studies using RWD. The RWE IVD framework focuses on issues pertinent to the use of RWE in clinical validation for both premarket and postmarket regulatory decision-making of IVD devices. Although none of the aforementioned documents introduce the concept of external data as the current document does, nor do they discuss specific statistical methods that can be used to leverage external data, all of them are related to the current document subject of RWD, and external data are often RWD. Note that the NESTcc methods and data quality frameworks are "living" documents in that they will be moving toward a more complete version in future iterations. It is expected that the introduction of the EEM framework will be a catalyst to the evolution of the two NESTcc frameworks.

Similar alignment exists with related work sponsored by different workstreams within the Medical Device Epidemiology Network,[34] with RWD/RWE teams sponsored by the International Society for Pharmacoepidemiology[35,36] and by the Biopharmaceutical Section of the American Statistical Association,[37] to name a few.

8 REFERENCES

1. *21st Century Cures Act.* U.S. Government Publishing Office; 2016. https://www.gpo.gov/fdsys/pkg/PLAW-114publ255/html/PLAW-114publ255.htm. Accessed December 17, 2020.
2. Yue LQ, Lu N, Xu Y. Designing premarket observational comparative studies using existing data as controls: challenges and opportunities. *J Biopharm Stat.* 2014;24(5):994-1010.
3. *Real-World Clinical Evidence Generation: Advancing Regulatory Science and Patient Access for In Vitro Diagnostics (IVDs) Framework.* Medical Device Innovation Consortium; 2020. https://mdic.org/resource/ivd-rwe-framework/. Accessed December 20, 2020.
4. *National Evaluation System for health Technology Coordinating Center (NESTcc) Data Quality Framework.* National Evaluation System for health Technology Coordinating Center 2020. https://nestcc.org/nestcc-data-quality-framework/. Accessed December 17, 2020.
5. *National Evaluation System for health Technology Coordinating Center (NESTcc) Methods Framework.* National Evaluation System for health Technology Coordinating Center 2020. https://nestcc.org/nestcc-methods-framework/. Accessed December 20, 2020.
6. *Code of Federal Regulations: 21CFR860.7 - Determination of Safety and Effectiveness.* U.S. Food and Drug Administration; 2019. https://www.ecfr.gov/cgi-bin/text-idx?SID=2ab340080e76eafcb089793d004bb949&mc=true&node=pt21.8.860&rgn=div5#se21.8.860_17. Accessed December 17, 2020.
7. *Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices: Guidance for Industry and Food and Drug Administration Staff.* U.S. Food and Drug Administration; 2017. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm513027.pdf. Accessed December 17, 2020.
8. ISO 14155:2020 - Clinical investigation of medical devices for human subjects — Good clinical practice. https://www.iso.org/standard/71690.html. Accessed December 17, 2020.
9. Agency for Healthcare Research and Quality. https://www.ahrq.gov/. Accessed December 17, 2020.
10. Patient-Centered Outcomes Research Institute. https://www.pcori.org/. Accessed December 17, 2020.
11. International Medical Device Regulators Forum (IMDRF) Registry. http://www.imdrf.org/about/about.asp. Accessed December 17, 2020.
12. Li H, Mukhi V, Lu N, Xu Y-L, Yue LQ. A note on good practice of objective propensity score design for premarket nonrandomized medical device studies with an example. *Stat Biopharm Res.* 2016;8(3):282-286.
13. 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):731-748.
14. Wang C, Lu N, Chen WC, et al. Propensity score-integrated composite likelihood approach for incorporating real-world evidence in single-arm clinical studies. *J Biopharm Stat.* 2020;30(3):495-507.
15. Chen WC, Wang C, Li H, et al. Propensity score-integrated composite likelihood approach for augmenting the control arm of a randomized controlled trial by incorporating real-world data. *J Biopharm Stat.* 2020;30(3):508-520.
16. *Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials.* U.S. Food and Drug Administration; 2010. https://www.fda.gov/downloads/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071121.pdf. Accessed December 17, 2020.
17. *Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices: Guidance for Industry and Food and Drug Administration Staff.* U.S. Food and Drug Administration; 2016. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/leveraging-existing-clinical-data-extrapolation-pediatric-uses-medical-devices. Accessed December 17, 2020.
18. Ibrahim JG, Chen M-H. Power prior distributions for regression models. *Stat Sci.* 2000;15(1):46-60.

19. Chen M-H, Ibrahim JG. The relationship between the power prior and hierarchical models. *Bayesian Anal.* 2006;1(3):551-574.
20. Hobbs BP, Carlin BP, Mandrekar SJ, Sargent DJ. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics.* 2011;67(3):1047-1056.
21. Hobbs BP, Sargent DJ, Carlin BP. Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Anal.* 2012;7(3):639-674.
22. Hobbs BP, Carlin BP, Sargent DJ. Adaptive adjustment of the randomization ratio using historical control data. *Clin Trials.* 2013;10(3):430-440.
23. Schmidli H, Gsteiger S, Roychoudhury S, O'Hagan A, Spiegelhalter D, Neuenschwander B. Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics.* 2014;70(4):1023-1032.
24. Carlin BP, Louis TA. *Bayesian methods for data analysis.* 3rd ed. Boca Raton, FL: Chapman & Hall/CRC; 2009.
25. 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):1089-1103.
26. Liu GF. A dynamic power prior for borrowing historical data in noninferiority trials with binary endpoint. *Pharm Stat.* 2018;17(1):61-73.
27. Rao SV, Hess CN, Barham B, et al. A registry-based randomized trial comparing radial and femoral approaches in women undergoing percutaneous coronary intervention: the SAFE-PCI for Women (Study of Access Site for Enhancement of PCI for Women) trial. *JACC Cardiovasc Interv.* 2014;7(8):857-867.
28. Registry Trials. https://www.ctti-clinicaltrials.org/projects/registry-trials. Accessed December 17, 2020.
29. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. *J Natl Cancer Inst.* 2009;101(21):1446-1452.
30. De A, Meier K, Tang R, et al. Evaluation of heart failure biomarker tests: a survey of statistical considerations. *J Cardiovasc Transl Res.* 2013;6(4):449-457.
31. Bennett M, Vielma JP, Zubizarreta JR. Building representative matched samples with multi-valued treatments in large observational studies. *J Comput Graph Stat.* 2020:1-29.
32. Medical Device Innovation Consortium (MDIC) Virtual Patient Model. https://mdic.org/project/virtual-patient-vp-model/. Accessed December 17, 2020.
33. Gamalo-Siebers M, Savic J, Basu C, et al. Statistical modeling for Bayesian extrapolation of adult clinical trial information in pediatric drug evaluation. *Pharm Stat.* 2017;16(4):232-249.
34. MDEpiNet Methodology Program. https://www.mdepinet.net/methodology. Accessed December 17, 2020.
35. International Society for Pharmacoepidemiology - Real-World Evidence Task Force. https://www.pharmacoepi.org/strategic-initiatives/rwe-task-force/. Accessed December 17, 2020.
36. International Society for Pharmacoepidemiology - Real-World Evidence for Regulatory Decision-Making. https://www.pharmacoepi.org/strategic-initiatives/rwe-for-regulatory-decision-making/. Accessed December 17, 2020.
37. American Statistical Association Biopharmaceutical Section. https://community.amstat.org/biop/home. Accessed December 17, 2020.

Appendix

APPENDIX – MDIC EXTERNAL EVIDENCE METHODS (EEM) SURVEY

In June 2020, MDIC conducted a survey to gather information from industry statisticians on the use of external data in clinical studies when making statistical inferences for primary safety and/or primary effectiveness endpoints. The survey collected information on the types of external data sources, how they are utilized, and the statistical methods (including Bayesian and Frequentist) used for leveraging/synthesizing external evidence to support regulatory decision-making.

The survey contained two sections, collecting information on:

1. Clinical studies that were associated with a medical device approved or cleared by the FDA via applications such as PMA, PMA supplement, 510k, Humanitarian Device Exemption (HDE), and De Novo.
2. Premarket investigational device clinical studies currently being conducted for FDA approval/clearance or completed but device not yet approved by the FDA (e.g., Investigational Device Exemption [IDE] and Q-submissions).

The following tables provide a highlight of responses to the survey.

MDIC EEM Survey Section I. Clinical Studies Associated with Approved/Cleared Medical Devices in Which External Data Sources Were Used

External data source	Utilization	Statistical method	SSED
Historical clinical study Patient registry	Construct device arm	Bayesian hierarchical model	P070015/S128b
Historical clinical study	Augment control arm	Bayesian hierarchical model	P100023
Historical clinical study Patient registry	Construct control arm	Propensity score methodology	P140010/S015
Historical clinical study	Construct control arm	Propensity score methodology	P090029

SSED, summary of safety and effectiveness data.

MDIC EEM Survey Section II. Ongoing/Not Yet Submitted for Approval/Clearance Medical Device Clinical Studies in Which External Data Sources Were Used

External data source	Utilization	Statistical method
Historical clinical study	Augment device arm	Bayesian hierarchical
Historical clinical study	Construct control arm	Propensity score method
Historical clinical study Patient registry	Augment control arm	Bayesian hierarchical
Laboratory data	Other: lab results database	Observational
Patient registry Electronic health records (EHRs)	Construct control arm Construct device arm	Propensity score method
Historical clinical study	Construct control arm Construct device arm	Propensity score method Bayesian power prior
Patient registry Historical clinical study	Construct control arm	Propensity score method
Concurrent external RCT data to be combined with the IDE RCT data	The external data will provide both control and investigational device data to be pooled with the IDE study data.	Pooling the patient level data by study with the IDE study data
Historical clinical study RWD from sites that participated in pivotal trial	Construct control arm Augment control arm	Propensity score method

IDE, investigational device exemption; RCT, randomized controlled trial; RWD, real-world data.

Contact information

For more information, please contact
Jithesh Veetil, PhD, Program Director, Data Science and Technology at
jveetil@mdic.org