Gruber *et al.*
BMC Medical Research Methodology (2023) 23:178
https://doi.org/10.1186/s12874-023-01998-2

BMC Medical Research Methodology

RESEARCH ARTICLE

Open Access

Evaluating and improving real-world evidence with Targeted Learning

Susan Gruber[1*], Rachael V. Phillips[2], Hana Lee[3], John Concato[4] and Mark van der Laan[2]

Abstract

Background The Targeted Learning roadmap provides a systematic guide for generating and evaluating real-world evidence (RWE). From a regulatory perspective, RWE arises from diverse sources such as randomized controlled trials that make use of real-world data, observational studies, and other study designs. This paper illustrates a principled approach to assessing the validity and interpretability of RWE.

Methods We applied the roadmap to a published observational study of the dose–response association between ritodrine hydrochloride and pulmonary edema among women pregnant with twins in Japan. The goal was to identify barriers to causal effect estimation beyond unmeasured confounding reported by the study's authors, and to explore potential options for overcoming the barriers that robustify results.

Results Following the roadmap raised issues that led us to formulate alternative causal questions that produced more reliable, interpretable RWE. The process revealed a lack of information in the available data to identify a causal dose–response curve. However, under explicit assumptions the effect of treatment with any amount of ritodrine versus none, albeit a less ambitious parameter, can be estimated from data.

Conclusions Before RWE can be used in support of clinical and regulatory decision-making, its quality and reliability must be systematically evaluated. The TL roadmap prescribes how to carry out a thorough, transparent, and realistic assessment of RWE. We recommend this approach be a routine part of any decision-making process.

Keywords Targeted learning, Real-world evidence, Causal inference, TMLE

Background

From a regulatory perspective, real-world evidence (RWE) arises from diverse sources, including randomized controlled trials (RCT) that involves analysis of real-world data (RWD), observational studies, and other study designs [1]. RWE can provide insight into treatments, outcomes, and populations beyond those that can be studied in traditional RCTs. Researchers studying causal inference have established a strong theoretical foundation for understanding when and how causal effects can be estimated from RWD and developed sophisticated tools for doing so [2–4]. Strategies for increasing acceptance of RWE by improving its quality and promoting transparency have appeared in the literature [5–8]. The adage "trust, but verify" reminds us that before RWE can be used in support of clinical and regulatory decision-making, its quality and reliability must be systematically evaluated, from study design and conduct through analysis and interpretation.

Originally introduced as a guide for statistical learning from data, the Targeted Learning (TL) roadmap is also invaluable for developing a statistical analysis plan and establishing the validity and interpretability of findings from a RWD study (Fig. 1) [9–12]. This paper

*Correspondence:
Susan Gruber
sgruber@putnamds.com
[1] Putnam Data Sciences, LLC, Cambridge, MA, USA
[2] Division of Biostatistics, University of California at Berkeley, Berkeley, CA, USA
[3] Office of Biostatistics, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA
[4] Office of Medical Policy, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA

demonstrates how to methodically step through the roadmap to expose weaknesses in causal claims. The roadmap provides a systematic way to evaluate the quality of RWE for both regulators and industry scientists. It can also inspire remediation strategies that strengthen the quality of the RWE.

Variants of the roadmap have been published in which Step 2 precedes Step 1. A preliminary description of the statistical model that relies on the time ordering of the covariates offers a starting point that doesn't rely on statistical knowledge. In some scenarios when little is known about the causal structure, e.g., rare diseases with little knowledge of the natural history of the disease, defining a statistical model based on time ordering and respecting known bounds on the data is a helpful starting point. Researchers with causal knowledge may prefer constructing a causal model first, and then working with statisticians to develop a statistical model that captures other elements of the data distribution that aren't represented in the causal model, such as bounds on continuous variables, monotonicity constraints, and known interactions. Refining the statistical and causal models can be an iterative process. Ultimately they must agree.

A published retrospective cohort study will serve to illustrate how to detect and overcome insufficiency for causal effect estimation. Our intent is not to provide a commentary on the published findings, but to discuss concepts in causality and present results from an alternative data analysis. *Shinohara, et. al.* studied the association between ritodrine hydrochloride and maternal pulmonary edema in twin pregnancy in Japan [13]. Ritrodine had previously been shown to increase risk of pulmonary edema in pregnant women [14]. Study authors wanted to establish this result in the sub-population of women pregnant with twins, who are at higher risk of pre-term labor. In Japan, ritodrine is a first line therapy for halting pre-term labor, although in the United States, it was withdrawn from the market in 1995 due to efficacy and safety concerns [13, 15].

The target of the primary analysis was the dose–response association. The odds ratio (OR) for developing pulmonary edema associated with a one unit increase in total ritodrine dosage was estimated as OR = 1.02, with a 95% confidence interval (CI) of (1.004, 1.03). The authors state that due to unmeasured confounding, a causal interpretation is not warranted. The finding was interpreted as a partially adjusted measure of the dose–response association, since certain pre-existing health conditions that confound the treatment-outcome association were not available to the study team.

The study provides a rich example of challenges to learning from data, well beyond unmeasured confounding. In the next section we follow the TL roadmap to identify additional barriers to evaluating a causal dose–response curve. Subsequently, we discuss potential solutions that are based on specifying a statistical model that respects the process that gave rise to the data, crafting a realistic definition of treatment consistent with real-world feasibility, and selecting an alternative, less ambitious, target parameter. Results of a modified data analysis support the conclusion that ritodrine treatment increases risk for pulmonary edema.

Methods

Evaluating real-world evidence

Data made publicly available on Dryad by the study authors consists of observations on $n = 225$ women in Japan pregnant between 2009 and 2016 [16]. Each observation contains baseline covariates, $L(0)$; ritodrine treatment administered over multiple time points, $A(t)$; time-varying covariates at multiple time points,$L(t)$; the pulmonary edema outcome, Y; and additional covariates measured up to 24 h post-delivery, $L(t+)$. Actual infusion

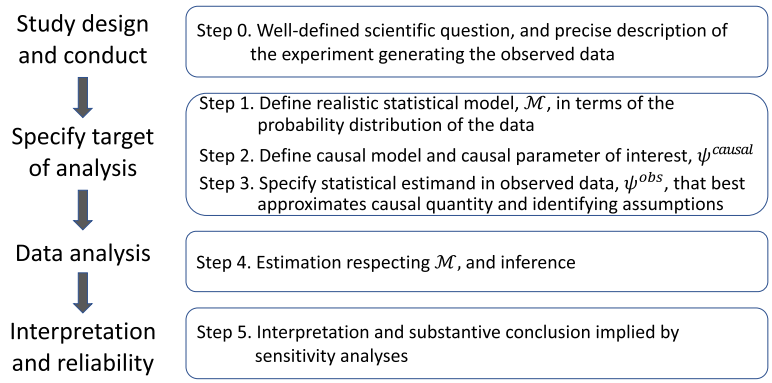

Fig. 1 The targeted learning roadmap

rates of the study drug varied from 50 mg per minute (mg/min) to 200 mg/min, over a variable period of time. Total dosage was converted to units of 72 mg/24 h. We step through the roadmap to better understand circumstances under which causal effect estimation might be possible, given the information available.

Step 1: Statistical model

Study authors posed a main terms logistic regression model, $logit\{P(pulmonaryEdema)\} = \beta_0 + \beta_1 ritodrine + \beta_2 PIH + \beta_3 BMI + \beta_4 PPH + \beta_5 corticosteroids + \beta_6 Mg + \beta_7 transfusion + \beta_8 term + \beta_9 bedrest$, where covariates are defined as *ritodrine (the treatment)*: total dosage of ritodrine; *PIH*: pregnancy-induced hypertension (Y/N); *BMI*: body mass index (kg/m^2); *PPH*: postpartum hemorrhage; *term*: term birth (Y/N); *bedrest*: bed rest > 6 weeks (Y/N).

This defines the statistical model narrowly, in a way that almost certainly precludes the true data distribution. Including treatment in the model as a continuous main term automatically imposes a linear and monotonic dose–response relationship. Making this restrictive modeling assumption at the outset for all situations is unwarranted. In fact, the paper provides the crude proportion of outcome events observed in the RWD, grouped by dosage levels [13]. Plotting these values suggests the dose–response relationship is, in fact, non-monotonic (Fig. 2). The crude risk increases as total dosage approaches 50 units, then decreases at larger doses. Although adjusting for measured confounders might explain away some of the crude dose–response association, a main terms logistic dose–response model appears to be unrealistic.

From a causal perspective, the timing of the outcome relative to other covariates included in the model is also problematic. The outcome event was measured from beginning of follow-up through 24 h postpartum [S. Shinohara, personal communication, December 2019]. That means that at least two covariates, *PPH* (postpartum hemorrhage) and *term* (term birth), occurred after the outcome, for some women. Including post-outcome covariates in a causal dose–response model violates the tenet that a cause must precede an effect.

Steps 2 and 3: Causal estimand and corresponding statistical parameter

Because the causal model must be contained within the statistical model, here it is identical to the statistical model. Both the causal estimand and the statistical parameter are given by β_1.

Steps 4 and 5: Estimation and inference

Maximum likelihood estimates of the model coefficients, standard error (SE) estimates, and 95% CIs were calculated using standard methodology.

Step 6: Interpretation

Mathematically, the model coefficients quantify the projection of the true dose–response curve onto the model. However, because the model is highly misspecified, as discussed in Step 1, the parameter estimate is not equivalent to the causal conditional log odds associated with a one unit increase in total dose.

Strategies for improving the quality of real-world evidence

Deficiencies in the study design, model specification, and available data undermine confidence in any causal interpretation of the study finding, and the RWE generated by the study is arguably not reliable. Designing and carrying out a new study may not be feasible, but instead we can revisit the roadmap to see if it may be possible to learn something relevant from the data we have.

Step 1: Statistical model

Parametric model misspecification can be avoided by defining a realistic, less restrictive, statistical model, $\mathcal{M}$ .In Step 1 We define the statistical model $\mathcal{M}$ non-parametrically as all distributions of the data consistent with

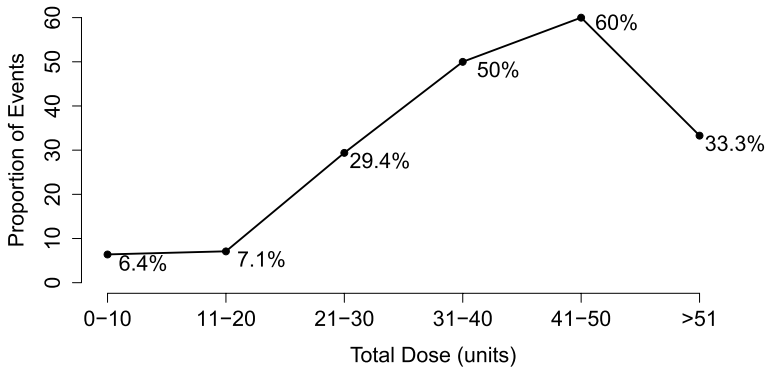

Fig. 2 Proportion of patients with pulmonary edema grouped by total dose of ritodrine

the process by which treatment, covariates, and the outcome arise over time. We can further restrict $\mathcal{M}$ to distributions that respect the study inclusion criteria. This specification of the statistical model includes some distributions of the data where the dose–response relationship is monotonic (e.g., a main terms parametric model) and some distributions where it is not.

Step 2: Causal estimand

The causal model makes conditional independence assumptions consistent with the time ordering of the data and assumes exogenous errors. The causal dose–response relationship can be defined in terms of a marginal multi-dimensional parameter. For example, given clinically meaningful groupings of total dosage administered, the mean risk for each treatment group can be targeted. Consider seven treatment categories: patients who receive no treatment ($A = 0$), or treatment at one of six levels ($A = 1$, …, 6, corresponding to > 0–10, 11–20, 21–30, 31–40, 41–50, 51 + units) (Fig. 2). The causal dose–response parameter is written as $\psi^{causal} = (\psi_0, \psi_1, \psi_2, \psi_3, \psi_4, \psi_5, \psi_6)$, where ψ_a is the counterfactual mean outcome observed if, contrary to fact, each patient received treatment at dosage level $A = a$. Furthermore, any causal contrast, such as the risk difference (RD), risk ratio (RR), or OR, can be easily calculated, e.g., $\psi_1 - \psi_0$ is the RD for treatment with up to 10 units of ritodrine vs. no treatment.

Step 3: Statistical estimand and assessment of identifiability

Next, we specify a statistical estimand in observed data, ψ^{obs}, that corresponds to our multi-dimensional ψ^{causal} under identifying assumptions. For each dimension, $\psi^{obs}_{A=a} = E(Y|A = a, \bar{L}(t))$, where $\bar{L}(t)$ is the complete covariate history from baseline ($t = 0$) through the time the event occurred, or 24 h post-delivery.

A problem is that the relative timing of $L(t)$ and $A(t)$, is not clearly recorded in the dataset. In other words, it is impossible to properly define $\bar{L}(t)$, thus we cannot specify any statistical parameter that corresponds to ψ^{causal}. Another complication is that clinicians may have slowed or stopped ritodrine infusion upon observing pulmonary edema in the patient. Under this scenario, the outcome partly causes the total dose, rather than the total dose causing the outcome. For these reasons, a causal dose–response curve is simply not identifiable from the data. Any study finding would rest entirely on a foundation of unrealistic modeling assumptions, and this RWE would not be appropriate to support decision-making.

Alternative formulation: In the absence of additional data that are fit for purpose, an alternative, unplanned analysis might provide insight into the causal relationship between ritodrine and pulmonary edema. Consider a simpler question: *does treatment with any dose of ritodrine vs. no treatment increase risk for pulmonary edema?* From this point treatment standpoint, the data consists of n independent and identically distributed observations $O = (Y, A, W)$, where Y is a binary outcome indicator, A is a binary treatment indicator ($A = 1$ for treated, $A = 0$ for no treatment), and W is a vector of baseline covariates. We are interested in a less ambitious causal parameter, the RD. Downstream covariates that affect treatment infusion over time are irrelevant, and time-dependent confounding is no longer an issue. In the next subsection we step through the roadmap with this revised clinical question in mind.

Determining a point treatment effect by following the Targeted Learning roadmap

Step 1: Statistical model

The statistical model is defined as all probability distributions of the data, with structure $O = (Y, A, W)$, consistent with study inclusion criteria.

Step 2: Causal estimand

The causal model makes no further assumptions, beyond exogenous errors. The causal parameter of interest is the marginal RD, defined in terms of counterfactual outcomes by $\psi^{causal} = E(Y_1 - Y_0)$, where Y_1 is the counterfactual outcome under any level of treatment and Y_0 is the counterfactual outcome under no treatment.

Step 3: Statistical estimand and assessment of identifiability

The statistical estimand, $\psi^{obs} = E[E(Y|A = 1, W) - E(Y|A = 0, W)]$, has a valid causal interpretation when underlying assumptions are met [17]. The *consistency* assumption states that for each observation, the outcome under the observed exposure is equivalent to the counterfactual outcome that would be seen had the observed treatment been assigned. It is satisfied under our simpler definition of treatment as any exposure to ritodrine, versus none.

The *positivity* assumption states that within all strata of confounders patients have a positive probability of receiving treatment at all levels considered. An outcome-blind look at the data shows that in some age groups no individuals were treated with ritodrine (Table 1), thus the parameter is not identifiable from the data. However, if coarser age categories are clinically justified, the violation can be eliminated by re-defining the age categories.

There is also another, more serious, violation of the positivity assumption. The prescribing information for ritodrine precludes administration when patients have serious pre-existing conditions, including maternal cardiac disease, hyperthyroidism, diabetes, and others [18]. If physicians adhere to these prescribing instructions, then no patients with these conditions would

Table 1 Number of subjects in control and treated groups by age in the original study age groupings (left), and re-defined age groupings (right)

Original Categories			Re-defined Categories		
Age	**Control**	**Treated**	**Age**	**Control**	**Treated**
16–20	2	0	**16—30**	48	33
21–25	9	7			
26–30	37	26			
31–35	50	29	**31–35**	50	29
36–40	38	19	**36–50**	45	20
41–45	4	1			
46–50	3	0			

receive ritodrine, and the causal contrast in this subpopulation of women cannot be evaluated. However, these covariates aren't in the publicly available dataset. If the information was also not known to clinicians, then women with these conditions could possibly receive treatment, and there would be no theoretical violation of the positivity assumption.

The final causal assumption, *coarsening at random* (CAR), is an assumption of no unmeasured confounders. With respect to the pre-existing conditions, if clinicians were unaware of patients' status, then these covariates could not have affected treatment decisions, so none are confounders. If clinicians were aware, then all of these covariates are unmeasured confounders.

Alternative formulation: One option would be to augment the exclusion criteria to rule out pregnant women who are ineligible to receive the study drug. The causal parameter would address a modified scientific question: *What is the marginal effect of ritodrine compared with no ritodrine on risk of pulmonary edema among women pregnant with twins to whom ritodrine may be prescribed?"* The RD would be identifiable from data, and interpretable as a subgroup-specific causal effect. Unfortunately, this approach isn't feasible with the available dataset because we cannot identify patients with the relevant pre-existing conditions.

A second option would be to modify the scientific question again. Suppose we were interested in understanding how incidence of pulmonary edema would be affected if ritodrine were withdrawn from the market. The target population includes all women pregnant with twins, even those who are ineligible to receive ritodrine. The following *realistic treatment rules* [19] can always be followed,

> Rule 1: Treat with ritodrine unless expressly contraindicated,
> Rule 2: Never treat with ritodrine

The marginal RD of pulmonary edema for following Rule 1 vs. Rule 2 can be estimated from observed data.

Step 4: Estimation and inference

Targeted minimum loss-based estimation (TMLE) with super learning (SL) was used to estimate the RD for the two realistic treatment rules. Potential baseline confounders included in the adjustment set were *age, height, weight, BMI,* and binary indicators of the following variables: *obesity* (*BMI* $\geq$ 25), *first pregnancy*, *single placenta*, *assistive reproductive technology use*, *magnesium administration, and corticosteroid administration*. Analyses were run using R (v4.0.2), and the *tmle* (v1.5.0–1) and *SuperLearner* (vx2.0–26) packages [20–22]. For SL, the number of cross validation folds, V, was set to 20 to account for the number of events [23]. The default library of algorithms for modeling the outcome included linear regression, Bayesian additive regression trees (BART, in *dbarts* v0.9–18)[24], and lasso (*glmnet* v4.0–2)[25]. The default library for modeling the propensity score (PS) included logistic regression, BART, and generalized additive models (*gam* v1.20) [26]. These library specifications allow us to explore more of the possible probability distributions contained in $\mathcal{M}$ than restricting to a parametric main terms model. TMLE uses the SL fit for the treatment assignment mechanism to update the initial SL estimate of the outcome prediction model to improve the bias variance trade-off for the target parameter [10]. TMLE requires the PS (1-PS) to be bounded away from zero for treated and untreated subjects, respectively. We set this lower bound to 0.06, based on the formula $5/[\sqrt{n}\ln(n)]$, with $n = 225$) [27]. Influence curve-based standard errors (SE) were reported by the software.

Results

PS diagnostics allow us to understand the overlap between treatment and control groups. The C-statistic associated with the predicted PS values (0.72), and the plot of the PS distributions within each treatment group (Fig. 3) indicate reasonable overlap of treated and comparator groups. No PS values were extreme, so truncation had no impact. The estimated RD was $\widehat{\psi}^{obs} = 0.21$ (SE, $\widehat{\sigma} = 0.062$; 95% CI = [0.09, 0.33]).

Step 5: Interpretation of the study finding and sensitivity analyses

Ritodrine was estimated to increase risk for pulmonary edema by 21%. Next, we assess if unexplained departures from the underlying causal assumptions could reverse that conclusion. If without our knowledge any of the three causal assumptions were violated, even at infinite sample size the estimated RD would not equal the true causal effect. The difference between the statistical

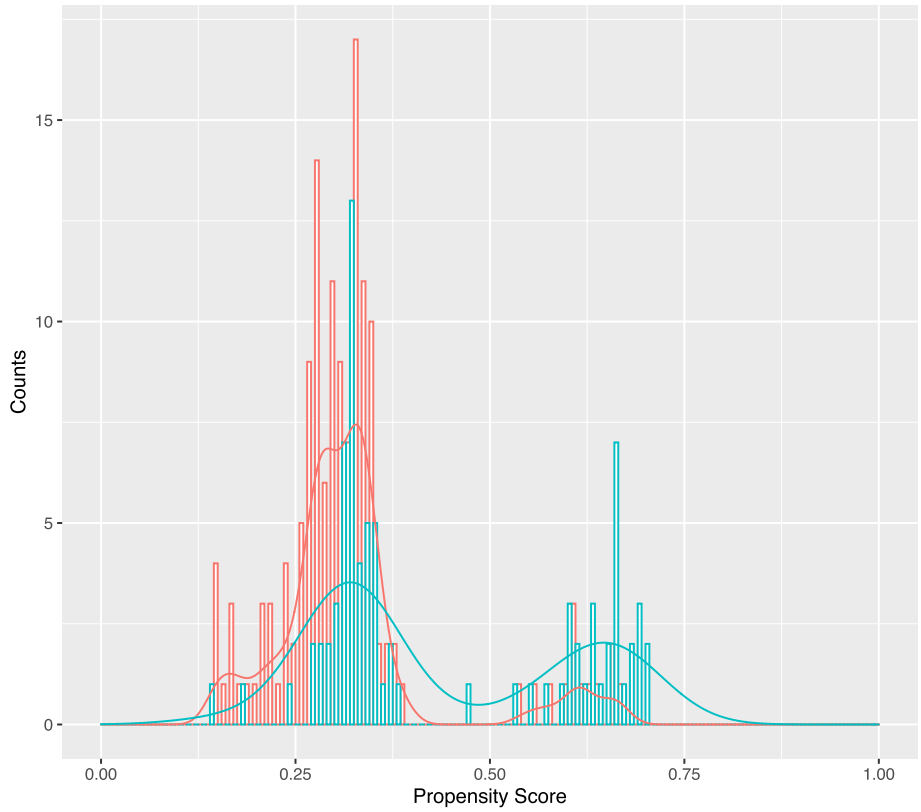

Fig. 3 Distribution of propensity scores in treated (blue) and comparator (red) groups

parameter and the causal parameter is termed the *causal gap*, $\delta = \psi^{stat} - \psi^{causal}$. A non-parametric sensitivity analysis illustrates how point estimates and CI bounds change at different assumed values of δ [28].

Figure 4 shows the estimated RD and 95% CI under different values of δ. For comparison, the size of the causal gap is also expressed relative to the SE of the effect estimate (SE units). The point estimate and 95% CI bounds determined from the data analysis are plotted at 0 on the x-axis. If there is a non-zero causal gap, then the estimate and CI would shift either left or right, depending on the direction and magnitude of the gap. Estimates and CIs plotted in gray correspond to different hypothetical causal gaps. If subject matter experts believe that any potential causal gap is likely to be negative then this sensitivity analysis reinforces a conclusion that ritodrine increases risk for pulmonary edema. If the causal gap is thought to be in the positive direction, unless the gap size is greater than approximately 0.1 the conclusion remains unchanged. The causal gap would have to be extremely large (> 0.325) to conclude that ritodrine is protective for pulmonary edema. Methods for establishing plausible values of the causal gap include expert knowledge, existence of a known external interpretable bound [28], data on negative controls, "worst case" imputation of missing outcomes, and analyses of data with key confounders omitted.

The recently proposed *G-value* calls attention to the gap size that would be needed to negate the finding from the current study (cause the CI to include the null if it is currently excluded, or exclude the null if it currently lies within the CI). For a 95% CI $G\text{-}value = min\left(\left|\psi_n - 1.96\sigma - null\right|, \left|\psi_n + 1.96\sigma - null\right|\right)$, where ψ_n is the estimated effect size, σ is the SE, and *null* is the null value for the parameter (0 for the ATE, 1 for the RR, etc.)[12]. The *G-value* takes both bias and variance into account to help determine an appropriate level of confidence in conclusions drawn from the study. Here the $G\text{-}value = 0.09$ (1.5 SE units).

Discussion

Although the findings suggest that exposure to ritodrine increased risk for pulmonary edema among women pregnant with twins in Japan, in the absence of information on pre-existing conditions that affect risk for pulmonary edema the study finding cannot be interpreted as an unbiased estimate of the true causal effect.

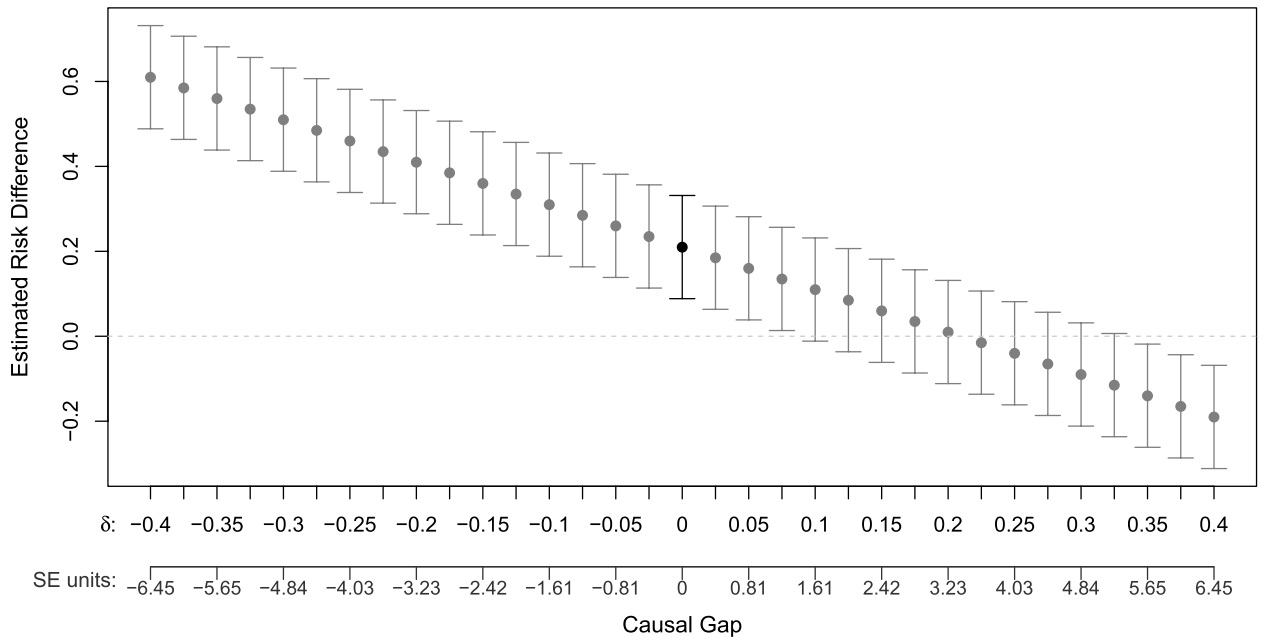

Fig. 4 Non-parametric sensitivity analysis showing the risk difference and 95% confidence intervals under different presumed values of the causal gap, δ, and also relative to the standard error (SE-units)

Conclusions

For regulatory purposes, a well-designed study and good quality data are of paramount importance. Following the TL roadmap allowed us to systematically evaluate the suitability of these data for estimating a causal dose–response curve. Outside of a regulatory environment, the roadmap pointed us to explore alternative formulations of the causal question to produce more reliable, interpretable RWE.

Steps 1–3 of the roadmap crystallized the statistical learning task by defining the statistical model ($\mathcal{M}$), causal parameter (ψ^{causal}), and statistical estimand (ψ^{obs}), that can answer the corresponding clinical question of interest. The original study's overly restrictive $\mathcal{M}$ was ill-suited for modeling the true causal dose–response relationship. Our alternative formulation of a multidimensional ψ^{causal} addressed that problem. However, given the uncertainty around the time ordering of treatment, covariates, and outcome, we were unable to describe a corresponding statistical estimand that could be identified from the data. Even without looking at the actual data, we were able to identify structural barriers that preclude evaluating a causal dose–response curve. This situation motivated targeting a point treatment parameter, defined in terms of realistic treatment rules.

Step 4, estimation of the statistical parameter, should go beyond fitting the coefficients in a single parametric model. If $\mathcal{M}$ is sufficiently general, i.e., realistic, then flexible machine learning (SL) is required. TMLE tailors the procedure for unbiased, efficient estimation of the statistical parameter, and provides influence curve-based inference.

Step 5, interpretation of the study finding, should incorporate a non-parametric sensitivity analysis that avoids imposing unwarranted parametric constraints. If a small, hypothetical but clinically plausible causal gap is sufficient to nullify or reverse the substantive conclusion, then the study findings are not a dependable guide decision making. On the other hand, when findings are robust in the face of plausible values of the causal gap, confidence is reinforced.

RWE can fulfill needs for information beyond that generated by RCTs. However, trust must be earned, not assumed. The TL roadmap provides a systematic process for establishing whether the RWE provides transparent, reliable, and actionable support for decision-making. A thorough, honest, realistic assessment of RWE can be a routine part of any decision-making process. The TL roadmap prescribes how this can be accomplished.

Abbreviations

BART	Bayesian additive regression trees
BMI	Body mass index
CI	Confidence interval
GAM	Generalized additive model
OR	Odds ratio
PS	Propensity score
RCT	Randomized controlled trial
RD	Risk difference
RR	Relative risk
RWD	Real-world data
RWE	Real-world evidence
SE	Standard error
SL	Super learner or super learning
TMLE	Targeted minimum loss-based estimation
TL	Targeted learning

Acknowledgements
Not applicable.

Authors' contributions
SG, RVP, HL, JC, MvdL contributed equally to study focus, design, and interpretation. SG and RVP contributed to the data analysis. All authors read and approved the final manuscript.

Funding
This project was funded by the United States Food and Drug Administration (US FDA) pursuant to Contract 75F40119C10155. The content is the view of the author(s), and does not necessarily represent the official views of, nor an endorsement, by FDA/HHS, or the U.S. Government.

Availability of data and materials
The dataset supporting the conclusions of this article was made publicly available by the original study team and is available in the Dryad repository, https://doi.org/10.5061/dryad.1v8v6.

Declarations

Ethics approval and consent to participate
This project involved only secondary analysis of publicly available de-identified data, thus does not constitute human subjects research and is exempt under the Federal U.S. Policy for the Protection of Human Subjects, 45 CFR, part 46. This manuscript is original work that reflects the views of all the authors. It should not be construed to represent the view of the FDA.

Consent for publication
Not applicable.

Competing interests
The authors report no conflicts of interest.

Received: 24 August 2022 Accepted: 20 July 2023
Published online: 02 August 2023

References

1. Framework for FDA's real-world evidence program, December 2018, US-FDA. Accessed 28 Feb 2022. https://www.fda.gov/media/120060/download.
2. Pearl J. Causality: Models, Reasoning and Inference 2nd. ed. Cambridge University Press; 2009.
3. Hernán MA, Robins JM. Causal Inference: What If. Chapman & Hall/CRC. 2020.
4. Rosenbaum PR. Modern Algorithms for Matching in Observational Studies. Annu Rev Stat Appl. 2020;7(1):143–76.
5. Levenson M, He W, Chen J, Fang Y, Faries D, Goldstein BA, Ho M, Lee K, Mishra-Kalyani P, Rockhold F, Wang H, Zink RC. Biostatistical Considerations When Using RWD and RWE in Clinical Studies for Regulatory Purposes: A Landscape Assessment, Statistics in Biopharmaceutical Research. Stat Biopharm Res. 2021. https://doi.org/10.1080/19466315.2021.1883473.
6. Patorno E, Schneeweiss S, Wang SV. Transparency in real-world evidence (RWE) studies to build confidence for decision-making: Reporting RWE research in diabetes. Diabetes Obes Metab. 2020;22(Suppl 3):45–59. https://doi.org/10.1111/dom.13918.
7. Martina R, Jenkins D, Bujkiewicz S, et al. The inclusion of real world evidence in clinical development planning. Trials 2018;19(468). https://doi.org/10.1186/s13063-018-2769-2.
8. Klonoff DC. The Expanding Role of Real-World Evidence Trials in Health Care Decision Making. J Diabetes Sci Technol. 2020;14(1):174–9. https://doi.org/10.1177/1932296819832653.
9. Gruber S, Lee H, Phillips R, Ho M, van der Laan M. Developing a Targeted Learning-Based Statistical Analysis Plan. Stat Biopharm Res. 2022. https://doi.org/10.1080/19466315.2022.2116104.
10. van der Laan MJ, Rose S. Targeted Learning: Prediction and Causal Inference for Observational and Experimental Data. New York: Springer; 2011.
11. van der Laan MJ, Rose S. Targeted Learning: Causal Inference for complex longitudinal studies. New York: Springer; 2018.
12. Gruber S, Phillips RV, Lee H, Ho M, Concato J, van der Laan MJ. Targeted learning: Toward a future informed by real-world evidence. Stat Biopharm Res. 2023. https://doi.org/10.1080/19466315.2023.21823562022.
13. Shinohara S, Sunami R, Uchida Y, et al. Association between total dose of ritodrine hydrochloride and pulmonary oedema in twin pregnancy: A retrospective cohort study in Japan. BMJ Open. 2017;7:e018118. https://doi.org/10.1136/bmjopen-2017-018118.
14. Gezginç K, Gül M, Karatayli R, et al. Noncardiogenic pulmonary edema due to ritodrine usage in preterm labor. Taiwan J Obstet Gyneco l. 2008;47:101–2.
15. Von Der Pool BA. Preterm labor: diagnosis and treatment. Am Fam Physician. 1998;57(10):2457–64.
16. Shinohara S, et al. Data from: Association between total dose of ritodrine hydrochloride and pulmonary edema in twin pregnancy: a retrospective cohort study in Japan, 2017. Dryad, Dataset, https://doi.org/10.5061/dryad.1v8v6.
17. Cole SR, Frangakis CE. The Consistency Statement in Causal Inference, A Definition or an Assumption? Epidemiology. 2009;20(1):3–5.
18. Ritodrine. Family Practice Notebook January 21, 2022. Accessed 25 Jan 2022. https://fpnotebook.com/ob/Pharm/Rtdrn.htm.
19. van der Laan MJ, Petersen ML. Causal effect models for realistic individualized treatment and intention to treat rules. Int J Biostat. 2007;3(1):3–3. https://doi.org/10.2202/1557-4679.1022.
20. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing; 2019.
21. Gruber S, van der Laan MJ. An R Package for Targeted Maximum Likelihood Estimation. J Stat Softw. 2012;51(13):1–35. https://doi.org/10.18637/jss.v051.i13 (http://CRAN.R-project.org/package=tmle).
22. Polley EC. Super Learner in Prediction, v2.0–24. http://CRAN.R-project.org/package=SuperLearner.
23. Phillips RV, van der Laan MJ, Lee H, Gruber S. Practical considerations for specifying a super learner, International Journal of Epidemiology, 2023; dyad023, https://doi.org/10.1093/ije/dyad023.
24. Dorie V. dbarts v0.9–11 https://CRAN.R-project.org/package=dbarts.
25. Friedman J, Hastie T, Tibshirani R. Regularization Paths for Generalized Linear Models via Coordinate Descent. J Stat Softw. 2010;33(1):1–22.
26. Hastie TJ. Generalized Additive Models, v1.16. http://CRAN.R-project.org/package=gam.
27. Gruber S, Phillips RV, Lee H, van der Laan MJ. Data-adaptive selection of the propensity score truncation level for inverse probability weighted and targeted maximum likelihood estimators of marginal point treatment effects. Am J Epidemiol. 2022;191(9):1640–51. https://doi.org/10.1093/aje/kwac087.
28. Díaz I, van der Laan MJ. Sensitivity analysis for causal inference under unmeasured confounding and measurement error problems. Int J Biostat. 2013;9(2):149–60.

Publisher's Note