

The importance of having a conceptual stage when reporting non-randomized studies

M.-A. C. Bind & D.B. Rubin

To cite this article: M.-A. C. Bind & D.B. Rubin (2021) The importance of having a conceptual stage when reporting non-randomized studies, *Biostatistics & Epidemiology*, 5:1, 9-18, DOI: [10.1080/24709360.2021.1913707](https://doi.org/10.1080/24709360.2021.1913707)

To link to this article: <https://doi.org/10.1080/24709360.2021.1913707>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 30 Apr 2021.



Submit your article to this journal [↗](#)



Article views: 612



View related articles [↗](#)



View Crossmark data [↗](#)

The importance of having a conceptual stage when reporting non-randomized studies

M.-A. C. Bind^a and D.B. Rubin^{b,c}

^aBiostatistics Center, Massachusetts General Hospital, Boston, MA, USA; ^bYau Center for Mathematical Sciences, Tsinghua University, Beijing, People's Republic of China; ^cDepartment of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, USA

ABSTRACT

Formal guidelines for statistical reporting of non-randomized studies are important for journals that publish results of such studies. Although it is gratifying to see some journals providing guidelines for statistical reporting, we feel that the current guidelines that we have seen are not entirely adequate when the study is used to draw causal conclusions. We therefore offer some comments on ways to improve these studies. In particular, we discuss and illustrate what we regard as the need for an essential initial stage of any such statistical analysis, the *conceptual stage*, which formally describes the embedding of a non-randomized study within a hypothetical randomized experiment.

ARTICLE HISTORY

Received 12 January 2020
Accepted 4 April 2021

KEYWORDS

Observational studies;
statistical analysis;
embedding studies

1. Introduction

According to Cochran [1]:

Dorn (1953) recommended that the planner of an observational study should always ask himself¹ the question, ‘How would the study be conducted if it were possible to do it by controlled experimentation?’

By proposing to begin by conceptualizing a related controlled experiment, Dorn’s advice set the stage for the estimation of the causal effect of an exposure, whether randomized or not.

Recently, there have been relatively extensive discussions about how to report empirical evidence from non-randomized studies in applied fields such as medicine, epidemiology, and social science. Even though we agree with many aspects of these published guidelines,² we think that (i) some of their points should be stated more forcefully and (ii) some important points have been omitted. More specifically, we feel that the guidelines that we have seen are inadequate in at least one important aspect: A non-randomized study that estimates causal effects using Fisherian *p*-values or other common measures of uncertainty, such as Neymanian confidence intervals for estimands, should always be *explicitly*

CONTACT M.-A. C. Bind  ma.bind@mail.harvard.edu  Biostatistics Center, Massachusetts General Hospital, 50, Staniford Street, Boston, MA 02114, USA

embedded within hypothetical randomized experiments; without this step, any probabilistic statements generally have little formal Fisherian or Neymanian foundation. Moreover, the results from a particular proposed hypothetical randomized experiment should be challenged by plausible deviations from the assumed randomized assignment mechanism using sensitivity analyses, which influence the discussion of results. Our recommendations apply to any non-randomized study estimating the effects of causal (i.e. manipulable) factors.

More explicitly, we present a sequence of, what we regard as essential steps (i.e. conceptual, design, statistical analysis, sensitivity analysis, and summary stages) that we believe need to be undertaken before summarizing empirical evidence of a non-randomized study

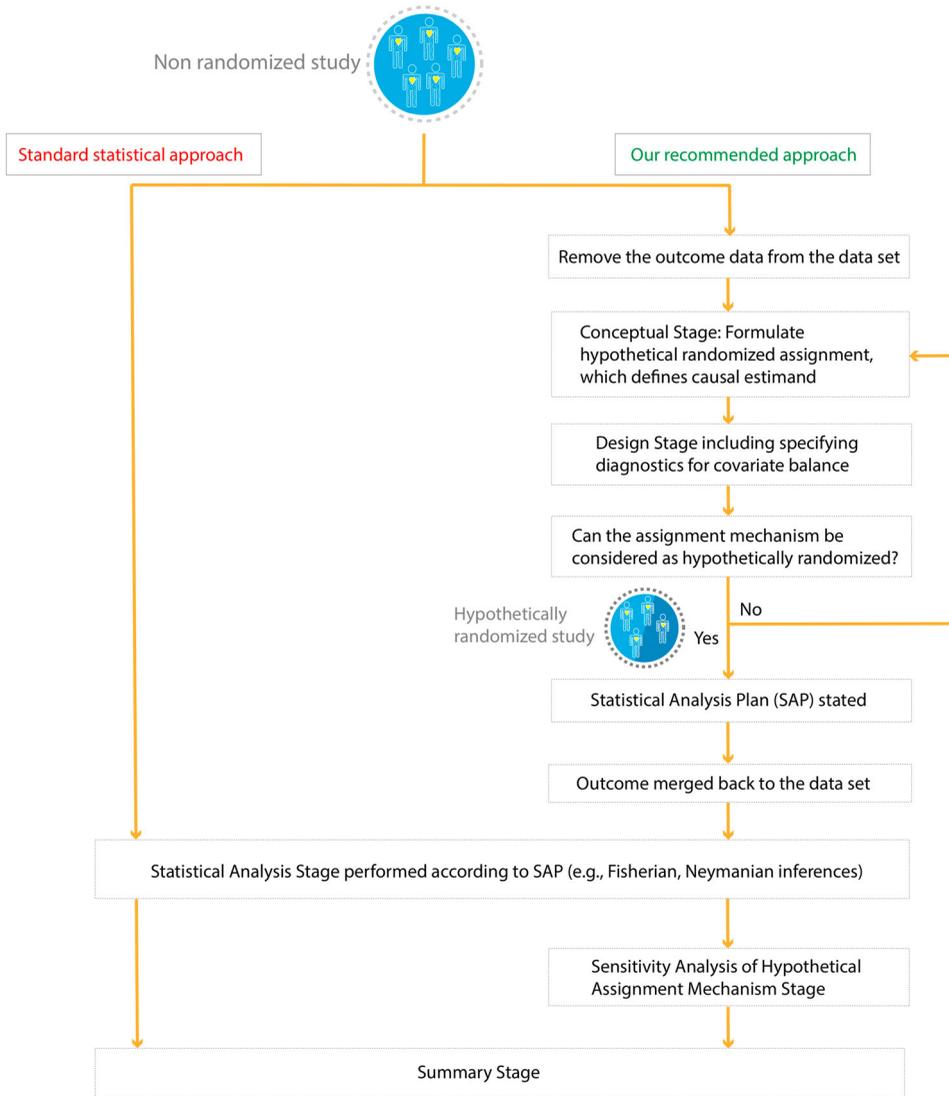


Figure 1. Contrasting two approaches to the analysis of non-randomized studies of causal effects.

(Figure 1); here, we expand especially on the conceptual stage because it is critical yet often omitted. Note that the standard statistical approach, which immediately plunges into the statistical analysis stage (i.e. regressing the observed outcome on the observed treatment and background variables), effectively proceeds as if the study were randomized with the statistical analyses reported as if the authors were following a pre-specified statistical analysis plan (SAP).

Moreover, one of the negative aspects of the recent focus on machine learning contributions to causal inference is the subsequent focus on mechanical details for creating balance in background covariates (i.e. design stage), without any real concern for the conceptual stage, e.g. Tarr and Imai [2].

2. The need of a conceptual stage when reporting the statistical analysis of a non-randomized study that estimates causal effects

A prespecified SAP is required for essentially all clinical trials to be approved by the US Food and Drug Administration [3]. This systematic requirement and its public availability together force the validity and transparency of the resultant studies, and enhance the chance for the production of reproducible results. For a specific example of the possible consequences when this step is omitted, consider a recent controversial non-randomized clinical study examining the effect of hydroxychloroquine treatment on COVID-19 [4]. This study was approved by the French National Agency for Drug Safety and the French Ethic Committee, and although we could find some published information on the primary and secondary endpoints and their evaluation timepoints, we could not find any prespecified SAP on the EU Clinical Trial Register (Number 2020-000890-25). Consequently, we have little way to access whether the results in the publication should be regarded as simply exploratory.

Although *some* journals (e.g. NEJM, JAMA) wisely emphasize the need for a prespecified SAP, not only for randomized trials but also for non-randomized studies, we feel that this feature should be considered essential for *all* applied studies to ensure constructive debates of their results and conclusions. We too have previously advocated this critically important point, but even more extensively and forcefully including the explicit embedding aspect [5–7], as have others before us [8, 9]. In particular, applied journals should demand that the SAP of any non-randomized study that estimates causal effects using standard frequentist statistics include the specification of an assumed hypothetical randomized experiment that generates the basis for Fisherian p -values or Neymanian confidence intervals.

Of course, the study of an intervention is not the topic of all scientific questions: most simply, there are two types of variables describing exposures, the *modifiable* ones, also called (treatment) factors, and the *non-modifiable* ones, also called blocks or strata. Causal relationships focus on treatment factors, whereas blocks allow us to understand differing susceptibility of different kinds of units to various treatments by examining how treatment effects vary across strata of units.

The causal inference framework illustrated in Figure 1 has been successfully implemented in recent observational studies examining whether air pollution exposure triggers multiple sclerosis relapses [10] or changes in human gut microbiome [11]. We now provide some additional examples.

Example 2.1: There is no need for the above conceptual stage for studies descriptively examining the role of non-modifiable variables, for instance, sex. Observing outcome differences between blocks (e.g. observing that males compared to females from the same population have greater risks of dying from COVID-19) is a *descriptive* statement, as opposed to a *causal* one. We believe the term ‘causal effect’ is ideally reserved to describe the result of a possible intervention versus the absence of that intervention, i.e. the involvement of a (modifiable) factor at two or more levels, which *can* be studied within blocks. In this context, non-randomized studies with both modifiable and non-modifiable variables should be embedded within hypothetical randomized block experiments.

Example 2.2: Consider studies that focus on the effects of meteorological or natural conditions, such as heat waves, rainfall [12], earthquake [13], on an outcome (e.g. daily number of crimes, deaths in a city, depression). To formulate the causal question in terms of a hypothetical randomizable intervention, we have to posit that ‘Mother Nature’ randomizes the occurrence of heat waves (or rainfall, or earthquake) given background covariates defining blocks, which usually we do not find entirely satisfactory, especially because such formulations do not envision any concrete policy intervention that would encourage or discourage the natural event. However, the effect of public ‘cooling centers,’ implemented during the summer 2020 in Boston during a heat alert, could be an interesting question to investigate, e.g. hypothetically by randomly assigning the locations to have or not have some centers, within the city. We have only seen such a concrete proposal being actually investigated, in a few contexts, e.g. a needle exchange program, Frangakis et al [14].

Example 2.3: Suppose we wish to estimate the causal effect of the creation of ‘bicycle lanes on streets’ on air pollution concentrations. The goal would be to formulate the causal question using potential outcomes described in terms of hypothetical randomized bicycle-lane-related interventions. For instance, suppose that, each year, city mayors decide, at random within levels of covariates (e.g. region, city size, average socio-economic status), to create new bicycle lanes within their cities to examine whether annual air pollution concentrations would decrease with more bicycle lanes. Such urban bicycle-friendly policies have been differentially implemented across cities worldwide, e.g. Rojas-Rueda et al [15–17]. Table 1 reveals the arguable plausibility of this hypothetical intervention.

We denote the bicycle-related intervention by W_i , which equals 1 if it is implemented by January 1st, 2019 in a city i , and 0 otherwise. First, assume that all interventions were equally randomized given background covariates, denoted X , which are pre-2019 city characteristics, e.g. $P(W_i = 1|X) = 0.5$. Let $Y_i(W_i = 1)$ and $Y_i(W_i = 0)$ be the potential outcomes (e.g. 2019-averaged $PM_{2.5}$ concentrations) had the city created bicycle lanes as of January 1st, 2019 or not. To assess the effect of this hypothetical intervention on air pollution concentrations, the goal is to estimate the following causal estimand under the Stable Unit Treatment Value Assumption [18]:

$$\tau := \frac{1}{N} \sum_{i=1}^N (Y_i(W_i = 1) - Y_i(W_i = 0)).$$

Of course, the randomness of each mayor’s decision seems implausible. We could instead propose using a randomized encouragement design [19–21] as the template and

Table 1. Top twenty most bicycle-friendly cities in the world from 2013 to 2019 (Copenhagenize design company).

| Ranking | 2013 | 2015 | 2017 | 2019 |
|---------|----------------|--------------|------------|------------|
| 1 | Amsterdam | Copenhagen | Copenhagen | Copenhagen |
| 2 | Copenhagen | Amsterdam | Utrecht | Amsterdam |
| 3 | Utrecht | Utrecht | Amsterdam | Utrecht |
| 4 | Seville | Eindhoven | Strasbourg | Antwerp |
| 5 | Bordeaux | Malmoe | Malmoe | Strasbourg |
| 6 | Nantes | Nantes | Bordeaux | Bordeaux |
| 7 | Antwerp | Bordeaux | Antwerp | Oslo |
| 8 | Eindhoven | Strasbourg | Ljubljana | Paris |
| 9 | Malmoe | Antwerp | Tokyo | Vienna |
| 10 | Berlin | Seville | Berlin | Helsinki |
| 11 | Dublin | Barcelona | Barcelona | Bremen |
| 12 | Tokyo | Ljubljana | Vienna | Bogota |
| 13 | Munich | Dublin | Paris | Barcelona |
| 14 | Montreal | Buenos Aires | Seville | Ljubljana |
| 15 | Nagoya | Berlin | Munich | Berlin |
| 16 | Rio de Janeiro | Minneapolis | Nantes | Tokyo |
| 17 | Barcelona | Paris | Hamburg | Taipei |
| 18 | Budapest | Hamburg | Helsinki | Montreal |
| 19 | Paris | Munich | Oslo | Vancouver |
| 20 | Hamburg | Montreal | Montreal | Hamburg |

assume that mayors are randomly encouraged or not to create new bicycle lanes in their city (Z_i). This template seems more plausible to us than an experiment where mayors are forced to comply with the assigned randomization, as we now describe.

The encouragement effect on the bicycle line creation (e.g. Z_i on W_i) is a causal question that can be of practical interest for urban planners and policy makers. However, without extraneous assumptions, the ‘intention-to-treat’ (ITT) causal question, where urban planners are forced to comply with their assignment, provides only indirect insights about the effectiveness of the encouragement intervention itself on the outcome. Therefore, in this finite population of mayors, the embedding should be extended to consider a hypothetical randomized experiment with two-sided non-compliance. Let $W_i(Z_i = 0)$ and $W_i(Z_i = 1)$ be the potential outcome for creation of bicycle lanes that would be implemented by January 1st, 2019 had the mayor of city i been encouraged to do so or not. Then, we are interested in estimating the ‘intention-to-treat’ effect, not among all cities, but rather in the compliers indicated by ‘co,’ which are units defined by their potential outcomes $W_i(Z_i = 1) = 1$ and $W_i(Z_i = 0) = 0$, because under the common monotonicity assumption (i.e. no defier, Angrist et al. [22]), this set of cities is the only set of cities that can be observed with and without bicycle lanes:

$$\tau_{co} := \frac{1}{N_{co}} \sum_{i \in co} (Y_i(Z_i = 1, W_i(Z_i = 1) = 1) - Y_i(Z_i = 0, W_i(Z_i = 0) = 0)).$$

Finally, suppose the ultimate goal is to estimate the causal effect of air pollution on cardiovascular disease. A non-randomized study could examine the effect of the creation of bicycle lanes that will proxy for encouraging the reduction of vehicle traffic (Z_i), and thereby decreasing air pollution concentrations (W_i). In this case, the embedding would again involve the formulation of a hypothetical randomized experiment with two-sided non-compliance.

Recent studies have actually capitalized on meteorological ‘instrumental variables’ (i.e. indicator variables for which the ‘unconfoundedness’ assumption [23] is assumed to hold), such as height of the planetary boundary layer, wind speed, and air pressure, to estimate the effect of air pollution on health outcomes, e.g. Schwartz et al. [24]; here, the description of a hypothetical intervention modifying such meteorological factors would involve ‘Mother Nature.’

Example 2.4: Suppose we are interested in estimating the role that air pollution plays in COVID-19. There are different ways to formulate this imprecise question. We could focus on the conditional causal effects of ambient air pollution ($W_i = 1$ for high and $W_i = 0$ for low concentrations) on mortality among the blocks of units diagnosed to have COVID-19 or not, $B_i = B^+$ if unit i is COVID-19 positive, $B_i = B^-$ otherwise, and consider a hypothetical randomized block experiment for the embedding, where pollution is the randomized treatment factor, where $\tau(B^+)$ and $\tau(B^-)$ are the average causal effects among the B^+ and B^- blocks.

Or, we could examine whether air pollution and COVID-19 are interacting (causal) factors for death. Interestingly, similar causal questions have been addressed in randomized *in vitro* studies [25, 26]. However, randomizing humans to coronavirus (SARS-CoV-2) infections is unethical. So instead, we could ‘mimic’ a two-factor randomized experiment and examine whether air pollution ($W_{i,1}$) and SARS-CoV-2 vaccination ($W_{i,2}$) are interacting causal factors for death (Y_i). Conducting a randomized clinical trial would be the ‘gold standard’ to estimate this interaction. But, a more common epidemiological approach when the outcome is rare is to conduct a ‘case-control’ study. We prefer calling such a study, a ‘case-noncase’ study, to reserve the term ‘control’ for the baseline level of treatment factor. A standard statistical approach simply regresses logit $P(Y_i = 1)$ on a possibly complicated function of air pollution, SARS-CoV-2 vaccination, and background covariates (X_i). We view this strategy as inappropriate, at least without elaboration. Instead, we prefer considering the target study population, from which the cases and non-cases arose, for which we want to draw causal inferences [27].

More specifically, to do so, we would proceed in four major steps conducted by two distinct and isolated analysts, or teams:

- (1) [Analyst 1] With access to all observed data, multiply-impute all missing potential outcomes (i.e. Y_i : death) under both treatments (i.e. $W_{i,1}$: high air pollution and $W_{i,2}$: SARS-CoV-2 vaccination) given covariates in the target population. This will also involve multiply-imputing X values for the unsampled units (i.e. indexed by $N_{cn}+1$ to N in Table 2).
- (2) [Analyst 1] Strip the outcome data from this multiply-imputed non-randomized data sets and provide them to other analyst.
- (3) [Analyst 2] Embed the resulting multiply-imputed data sets into a hypothetical two-factor (i.e. $W_{i,1}$: high vs. low air pollution and $W_{i,2}$: SARS-CoV-2 vaccination vs. not) randomized experiment, including a SAP.
- (4) [Analyst 2] Analyze the resulting data from the hypothetical experiment following the *a priori* defined SAP formulated in Step 3. Causal estimands of interest are the main

Table 2. Observed and missing data table (case-noncase study, units indexed by 1 to N_{cn} , target population study: units indexed by 1 to N).

| Unit i | $W_{i,1}$ | $W_{i,2}$ | X_i | $Y_i(W_{i,1} = 0, W_{i,2} = 0)$ | $Y_i(W_{i,1} = 0, W_{i,2} = 1)$ | $Y_i(W_{i,1} = 1, W_{i,2} = 0)$ | $Y_i(W_{i,1} = 1, W_{i,2} = 1)$ |
|------------|-----------|-----------|--------------|---------------------------------|---------------------------------|---------------------------------|---------------------------------|
| 1 | 0 | 1 | X_1 | ? | Y_1^{obs} | ? | ? |
| ... | ... | ... | ... | ... | ... | ... | ... |
| N_{cn} | 1 | 1 | $X_{N_{cn}}$ | ? | ? | ? | $Y_{N_{cn}}^{obs}$ |
| $N_{cn}+1$ | ? | ? | ? | ? | ? | ? | ? |
| ... | ? | ? | ? | ? | ? | ? | ? |
| ... | ? | ? | ? | ? | ? | ? | ? |
| N | ? | ? | ? | ? | ? | ? | ? |

effect of high vs. low air pollution, the main effect of SARS-CoV-2 vaccination vs. not, and their interaction. Respectively, these are defined by:

- $\tau_{W_1} := \frac{1}{2N} \sum_{i=1}^N (Y_i(W_{i,1} = 1, W_{i,2} = 0) - Y_i(W_{i,1} = 0, W_{i,2} = 0) + Y_i(W_{i,1} = 1, W_{i,2} = 1) - Y_i(W_{i,1} = 0, W_{i,2} = 1))$
- $\tau_{W_2} := \frac{1}{2N} \sum_{i=1}^N (Y_i(W_{i,1} = 0, W_{i,2} = 1) - Y_i(W_{i,1} = 0, W_{i,2} = 0) + Y_i(W_{i,1} = 1, W_{i,2} = 1) - Y_i(W_{i,1} = 1, W_{i,2} = 0))$
- $\tau_{W_1*W_2} := \frac{1}{2N} \sum_{i=1}^N ((Y_i(W_{i,1} = 1, W_{i,2} = 1) - Y_i(W_{i,1} = 0, W_{i,2} = 1)) - (Y_i(W_{i,1} = 1, W_{i,2} = 0) - Y_i(W_{i,1} = 0, W_{i,2} = 0)))$

3. Conclusions

We strongly believe that embedding non-randomized studies in hypothetical randomized experiments, as implied by Dorn’s advice conveyed by Cochran, is an essential step to obtain transparent, though hypothetical, randomization-based p -values or confidence intervals. We believe that a non-randomized study lacking such a conceptual stage cannot generate statistical statements that have any real scientific meaning, even hypothetical; we must remember that a randomized assignment mechanism, actual or posited, is required for valid Fisherian or Neymanian inferences. And these inferences change as the specification of the assignment mechanism changes, as illustrated in Bind and Rubin [5], the results based on a particular, but hypothetically, assignment generally should be accompanied with a sensitivity analysis, e.g. Rosenbaum [28], Bind and Rubin [29], exposing how answers might change as assumptions about the hypothetical assignment mechanism change.

Acknowledgements

Research reported in this publication was supported by the John Harvard Distinguished Science Fellow Program within the FAS Division of Science of Harvard University, by the Office of the Director, National Institutes of Health under Award Number DP5OD021412 and NIH RO1-AI102710, and by the National Science Foundation under Award Number NSF IIS ONR 1409177. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors thank the Editor-in-Chief, the Associate Editor, and the reviewers for providing helpful comments that improved our original manuscript and pointed out related medical perspectives (Tu et Pratt, 2020).

Notes

1. himself/herself (although himself was grammatically standard at that time)
2. Vandembroucke et al. [31] wrote the STROBE statement, which has been adopted by many journals (e.g., *Epidemiology*, *PLoS Medicine*, *Annals of Internal Medicine*), and Harrington et al. [30] provided new guidelines for the *New England Journal of Medicine* (NEJM).

Notes on contributors

M.-A. C. Bind is an Instructor of Investigation at the MGH Biostatistics Center. Her research interests focus on developing causal inference methods for quantifying the effects of environmental exposures on health outcomes and understanding the mechanisms explaining these health effects. Her current research is funded by the NIH Early Independence Award program. She completed her joint PhD in Biostatistics and Environmental Health at the Harvard School of Public Health, working with Professors Joel Schwartz and Brent Coull. She then became a Ziff postdoctoral Fellow at the Harvard University Center for the Environment. In 2016, she was awarded an Early Independence Award (NIH High-Risk High-Reward research grant) and became Research Associate in the Department of Statistics. From 2017 to 2021, she became a John Harvard Distinguished Science Fellow.

D.B. Rubin is currently Professor in the Yau Center for Mathematical Sciences, Tsinghua University; Murray Schusterman Senior Research Fellow, Fox Business School, Temple University; and Professor of Statistics Emeritus, Harvard University. He has been elected to be a Fellow/Member/Honorary Member of: the Woodrow Wilson Society, Guggenheim Memorial Foundation, Alexander von Humboldt Foundation, American Statistical Association, Institute of Mathematical Statistics, International Statistical Institute, American Association for the Advancement of Science, American Academy of Arts and Sciences, European Association of Methodology, the British Academy, and the U.S. National Academy of Sciences. As of 2021, he has authored/coauthored nearly 500 publications (including ten books), has four joint patents, and for many years has been one of the most highly cited authors in the world, with currently over 300,000 citations, and over 20,000 per year in recent years (Google Scholar). Of his many publications with over 1,000 citations each, over ten of them are solely authored by Rubin. He has received honorary doctorate degrees from Otto Friedrich University, Bamberg, Germany; the University of Ljubljana, Slovenia; Universidad Santo Tomás, Bogotá, Colombia; Uppsala University, Sweden; and Northwestern University, Evanston, Illinois. He has also received honorary professorships from the University of Utrecht, The Netherlands; Shanghai Finance University, China; Nanjing University of Science & Technology, China; Xi'an University of Technology, China; and University of the Free State, Republic of South Africa. He is a widely sought international lecturer and consultant on a variety of statistical topics.

References

- [1] Cochran WG. The planning of observational studies of human populations (with discussion). *J Royal Stat Soc Series Gen.* 1965;128:234.
- [2] Tarr A, Imai K. Estimating average treatment effects with support vector machines. *arXiv:2102.11926.* 2021.

- [3] US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for industry E9 statistical principles for clinical trials (section 5.1). 1998.
- [4] Gautret P, Lagier JC, Parola P, et al. Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open-label non-randomized clinical trial. *Int J Antimicrob Agents*. 2021;57(1):106243. doi:10.1016/j.ijantimicag.2020.106243.
- [5] Bind MA, Rubin DB. Bridging observational studies and randomized experiments by embedding the former in the latter. *Stat Methods Med Res*. 2019;28(7):1958–1978. doi:10.1177/0962280217740609.
- [6] Rubin DB. For objective causal inference, design trumps analysis. *Ann Appl Stat*. 2008;2:808.
- [7] Rubin DB. The design versus the analysis of observational studies for causal effects: parallels with the design of randomized trials. *Stat Med*. 2007;26:20–36.
- [8] Freedman DA. Statistical models for causation: what inferential leverage do they provide? *Eval Rev*. 2006;30:691–713.
- [9] Freedman DA. Linear statistical models for causation: a critical review. *Encyclopedia of statistics in behavioral science*. 2005.
- [10] Sommer A, Leray E, Lee Y, et al. Assessing environmental epidemiology questions in practice with a causal inference pipeline: an investigation of the air pollution-multiple sclerosis relapses relationship. *Stat Med*. 2021a;40(6):1321–1335.
- [11] Sommer A, Peters A, Cyrys J, et al. A randomization-based causal inference framework for uncovering environmental exposure effects on human gut microbiota. *bioRxiv*. 2021b. doi:10.1101/2021.02.24.432662
- [12] Sommer A, Lee M, Bind MA. Comparing apples to apples: an environmental criminology analysis of the effects of heat and rain on violent crimes in Boston. *Palgrave Commun*. 2018;4:138. doi:10.1057/s41599-018-0188-3.
- [13] Zubizarreta J, Cerdá M, Rosenbaum P. Effect of the 2010 Chilean earthquake on posttraumatic stress: reducing sensitivity to unmeasured bias through study design. *Epidemiology*. 2013;24(1):79–87.
- [14] Frangakis CE, Brookmeyer RS, Varadhan R, et al. Methodology for evaluating a partially controlled longitudinal treatment using principal stratification, with application to a needle exchange program. *J Am Stat Assoc*. 2004;99:239–249.
- [15] Rojas-Rueda D, de Nazelle A, Teixido O, et al. Health impact assessment of increasing public transport and cycling use in Barcelona: a morbidity and burden of disease approach. *Prev Med*. 2013;57:573–579.
- [16] Rojas-Rueda D, de Nazelle A, Teixido O, et al. Replacing car trips by increasing bike and public transport in the greater Barcelona metropolitan area: a health impact assessment study. *Environ Int*. 2012;49:100–109.
- [17] Rojas-Rueda D, de Nazelle A, Tainio M, et al. The health risks and benefits of cycling in urban environments compared with car use: health impact assessment study. *Br Med J*. 2011;343:d4521.
- [18] Rubin DB. Randomization analysis of experimental data: the fisher randomization test comment. *J Am Stat Assoc*. 1980;75:591–593.
- [19] Holland PW. Causal inference, path analysis, and recursive structural equations models. *Sociol Methodol*. 1988;18:449–484.
- [20] Rubin DB. Bayesian analysis of a two-group randomized encouragement design. In: Dorans N, editor. *Looking back. lecture notes in statistics*. New York, NY: Springer; 2011. p. 55–65.
- [21] Zell E, Wang X, Yin L, et al. Bayesian causal inference: approaches to estimating the effect of treating hospital type on cancer survival in Sweden using principal stratification. 2013.
- [22] Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables (with discussion). *J Am Stat Assoc*. 1996;91:444–472.
- [23] Rubin DB. Formal mode of statistical inference for causal effects. *J Stat Plan Inf*. 1990;25:279–292.
- [24] Schwartz J, Fong K, Zanobetti A. A national multicity analysis of the causal effect of local pollution, NO₂, and PM_{2.5} on mortality. *Environ Health Perspect*. 2018;126:87004–87004.

- [25] Groulx N, Urch B, Duchaine C, et al. The pollution particulate concentrator (PoPCon): a platform to investigate the effects of particulate air pollutants on viral infectivity. *Sci Total Environ.* [2018](#);628-629:1101–1107.
- [26] Wang S, Zhang X, Chu H, et al. Impact of secondary organic aerosol exposure on the pathogenesis of human influenza virus (H1N1). *AGU Fall Meeting Abstracts 2018:GH13B-0943.* 2018.
- [27] Andric N. Exploring objective causal inference in case-noncase studies under the rubin causal model. 2015.
- [28] Rosenbaum PR. *Design of observational studies.* New York (NY): Springer New York; 2010.
- [29] Bind MA, Rubin DB. When possible, report a fisher-exact p -value and display its underlying null randomization distribution. *Proc Natl Acad Sci USA.* [2020](#);117(32):19151–19158. doi:[10.1073/pnas.1915454117](#).
- [30] Harrington D, D’Agostino RB S, Gatsonis C, et al. New guidelines for statistical reporting in the journal. *N Engl J Med.* [2019](#);381:285–286.
- [31] Vandembroucke JP, Erik VE, Altman DG, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *Epidemiology.* [2007](#);18:805–835.