STRengthening Analytical Thinking for Observational Studies (STRATOS): Estimands – Summary from the "STRATOS Accelerated Guidance for Real World Data Analysis" Workshop

Nicholas Bakewell 1 , Jonathan Bartlett 2 , Kelly Van Lancker 3 , Els Goetghebeur 3 , Pamela Shaw 4 , Doranne Thomassen 5 , Emmanuelle Boutmy 6 , Suzanne M Cadarette 1 , Saskia le Cessie 3,5,7 , Nan van Geloven 5

1 Leslie Dan Faculty of Pharmacy and Dalla Lana School of Public Health, University of Toronto, Canada.
2 Department of Medical Statistics, London School of Hygiene and Tropical Medicine, UK.
3 Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium.
4 Biostatistics Division, Kaiser Permanente Washington Health Research Institute, USA.
5 Department of Biomedical Data Sciences, Leiden University Medical Center, the Netherlands.
6 Merck KGaA, Germany.
7 Department of Clinical Epidemiology, Leiden University Medical Center, the Netherlands.

Introduction

In September 2024, the STRengthening Analytical Thinking for Observational Studies (STRATOS) initiative brought together 55 researchers for a one-week workshop at the Lorentz Center in Leiden, the Netherlands. Presentations and working sessions were organised around key themes of observational research methods where STRATOS aims to bring guidance to applied researchers. We summarise the Workshop's discussions on the key theme 'Estimands.'

Estimands as a bridge between statistics and its application

Successful application of statistics in empirical sciences requires appropriate study design, data collection procedures and statistical method(s), which should all be steered by the research question. A clearly specified research question grounded in subject-matter knowledge is of the utmost importance. The estimand translates the research question into a precise unequivocal quantity (or quantities) of interest that we aim to estimate from data [1] .

Clearly defined estimands serve as a communication tool among and between statisticians and non-statisticians, and reduce the risk of misinterpretation of results by providing clarity to the reported effect(s) and the population(s) for which they hold [2] .

Model-free estimands

Traditionally, estimands have been (implicitly) defined as parameters in parametric statistical models. However, as Breiman [3] highlighted over two decades ago, statistical modelling relies on a priori assumptions about the true data generating model that may not be satisfied, potentially leading to questionable or irrelevant conclusions.

Consequently, there is a growing interest in moving away from model-based estimands and defining model-free estimands in causal inference [4, 5] . While statistical modelling will remain needed for estimating estimands and delivering insight into complex data structures, it has been argued that the estimand itself should not be defined with reference to particular statistical assumptions, since these may often not be correct and cannot be assessed at the study design stage [6] .

Are estimands a new phenomenon? Defining estimands and assessing the plausibility of identification assumptions (i.e., conditions needed to estimate estimands) are not new to statistical practice. For example, randomized controlled trials (RCTs) have been using frameworks like Population, Intervention, Comparison, Outcome, (Time) (PICO(T)) for a long time. However, PICO(T) may lack the precision for an unequivocal estimand definition.

There are comprehensive textbooks on causal inference that describe how the estimand(s) may be defined using mathematical notation (e.g., potential outcomes or counterfactuals) and how to assess the plausibility of identification assumptions in practice [7-9] . However, textbooks and the statistically oriented causal inference literature may not be accessible to all applied researchers.

Several frameworks that may be used to define estimands with the applied researcher in mind have been developed. These frameworks may be useful for providing (1) guidance without requiring in-depth knowledge of the vast literature on causal inference that may be difficult to navigate for the novice; and (2) clarity and standardization in the reporting of empirical research, which may mitigate potential biases and misinterpretations.

Defining estimands: Insights from four frameworks

Among frameworks that may be used to define estimands, four were discussed at the Workshop (Box). In earlier work, the STRATOS Causal Inference Topic Group developed eight steps for the empirical evaluation of research questions [1] . They outline a principled approach encouraging researchers to think carefully about what they are estimating and to be transparent in reporting, including underlying assumptions.

The Causal Roadmap (CR) provides a seven-step process for causal inference from defining the estimand to identifying and estimating it [10]. A key objective of the CR is to ensure a well-specified estimand, which can be rewritten in terms of the observed data given appropriate identification assumptions, reflecting both the research question and challenges in the data.

Target Trial Emulation (TTE) provides a structured process for designing an observational study when comparing treatments by specifying the ideal "target trial" [11] . The "target trial" refers to a hypothetical RCT one would ideally design to answer the study question if there were no ethical or financial constraints.

Lastly, the addendum to the ICH E9 harmonised guideline on statistical principles for clinical trials (i.e., ICH E9(R1) addendum) presents five key attributes of the estimand in an RCT [12] . In particular, the ICH E9(R1) addendum emphasizes intercurrent events which are post-randomization events (e.g., treatment switching) that affect the outcome and/or the collection of the outcome.

There is a clear overlap between these frameworks, but they vary in scope and focus, and therefore, may result in estimands being

defined differently or may not be appropriate to address all types of research questions. For example, TTE and the ICH E9(R1) addendum focus on clinical studies comparing treatments, while the other two frameworks enable the definition of estimands for a wider range of causal questions.

It is of interest to further assess how these frameworks may complement, supplement, or conflict with each other. The STRATOS initiative aims to address this gap by synthesizing the extensive literature on these four frameworks (and potentially others) and by providing practical recommendations to applied researchers for effectively defining and using estimands.

Avoiding nonsensical estimands

The general recommended order in the estimand frameworks is to first define the estimand(s) and then delineate the assumptions needed to identify them; subsequently, specify how to estimate the estimand(s) from the data; and then start the actual data analysis. Identification assumptions include common assumptions such as consistency, no unmeasured confounding (i.e., (conditional) exchangeability) and positivity [13]. While identification assumptions may seem plausible by design or given a conceptual (structural) model, they are usually not empirically verifiable. Therefore, there must be a feedback loop with subject-matter experts assessing whether relevant identification assumptions are plausible for the setting at hand. This may ultimately avoid targeting nonsensical estimands; and reporting estimates with limited to no applicability in the real world [5].

Estimands are not just for causal effects

Having a well-defined inferential goal is not unique to the context of quantifying causal effects of treatments or exposures. The need for avoiding ambiguity and misinterpretation of results equally holds for studies with a descriptive, predictive or diagnostic aim. Therefore, estimands can also prove to be an extremely valuable tool in these types of studies.

For instance, recent work points out how age adjustment in descriptive studies may change the descriptive estimand from "What is the burden of disease in different racial/ethnic groups?" to "What would be the burden of disease in different racial/ethnic groups if they had the same age distribution as a chosen reference population?" [14]. Which one of these two is appropriate depends on the descriptive research question, and this research question needs to be unequivocal before we can decide on the appropriate analysis.

Similarly, misalignment between the intended use of prediction models and how the models handle treatments received by individuals in the development/training data has formed the basis of the prediction estimand framework [15-18]. For example, imagine a prediction model built on historical data where patients were treated according to a certain policy. If the treatment policy has evolved, the model might not be applicable in a contemporary setting.

In addition, the prediction estimand framework helps to clarify whether predictions are suitable for informing treatment decisions.

Lastly, in diagnostic studies, a well-defined estimand may ensure alignment between the target population (i.e., individuals for whom the diagnostic test is intended to be used in routine clinical practice) and the study population (i.e., individuals that are included in the study, and received both the diagnostic test under evaluation and the reference ("gold") standard test).

Conclusion

In conclusion, the size of the literature on estimands can be overwhelming even for the experienced statistician. We hope the summary of the Workshop's discussions on 'Estimands' and planned work from these discussions will chart a practical way forward through the literature and guide those conducting observational studies on the estimand frameworks. Interested readers are encouraged to visit the STRATOS initiative website for updates on this work (https://stratos-initiative.org).

Components of four frameworks to define estimands

STRATOS Causal Inference Topic Group

- Define the treatment that corresponds to the research question(s)
- Define the outcome that corresponds to the research question(s)
- Define the population(s) of interest
- Formalize the research question in terms of potential outcomes
- Specify the estimand as a contrast between potential outcome distributions
- State underlying assumptions validating the causal effect estimation
- Estimate the estimand
- Evaluate the validity of assumptions & perform sensitivity analyses

Causal Roadmap

- Causal question, causal model, and causal estimand
- Describe the observed data
- Assess identifiability: Can the proposed study provide an answer to our causal question?
- Define the statistical estimand
- Choose a statistical model and estimator that respects available knowledge and uncertainty based on statistical properties
- Specify a procedure for sensitivity analysis
- Compare alternative complete analytic study designs

Target Trial Emulation

- Eligibility criteria
- Treatment strategies
- Assignment procedures
- Follow-up period
- Causal contrasts of interest
- Analysis plan

ICH E9(R1) Addendum

- Population
- Treatment
- Variable [outcome]
- (Population-level) summary measure
- Intercurrent events

References

1. Goetghebeur, E., et al., Formulating causal questions and principled statistical answers. Statistics in Medicine, 2020. 39(30): p. 4922-4948.

2. Kahan, B.C., et al., Eliminating ambiguous treatment effects using estimands. American Journal of Epidemiology, 2023. 192(6): p. 987-994.

3. Breiman, L., Statistical modeling: The two cultures (with comments and a rejoinder by the author). Statistical science, 2001. 16(3): p. 199-231.

4. Vansteelandt, S. and O. Dukes, Assumption-lean Inference for Generalised Linear Model Parameters. Journal of the Royal

Statistical Society Series B: Statistical Methodology, 2022. 84(3): p. 657-685.

5. Vansteelandt, S. and K. VanLancker, Chasing Shadows: How Implausible Assumptions Skew Our Understanding of Causal Estimands. Statistics in Biopharmaceutical Research, 2024(just-accepted): p. 1-11.

6. Vansteelandt, S., Statistical modelling in the age of data science. Observational Studies, 2021. 7(1): p. 217-228.

7. Hernan, M.A. and J.M. Robins, Causal Inference: What If. 2023: Taylor & Francis.

8. Pearl, J., Causality: Models, reasoning, and inference. Causality: Models, reasoning, and inference. 2000, New York, NY, US: Cambridge University Press. xvi, 384-xvi, 384.

9. Imbens, G.W. and D.B. Rubin, Causal inference for statistics, social, and biomedical sciences: An introduction. Causal inference for statistics, social, and biomedical sciences: An introduction. 2015, New York, NY, US: Cambridge University Press. xix, 625-xix, 625.

10. Dang, L.E., et al., A causal roadmap for generating high-quality real-world evidence. J ClinTransl Sci, 2023. 7(1): p. e212.

11. Hernán, M.A. and J.M. Robins, Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol, 2016. 183(8): p. 758-64.

12. International Council For Harmonisation Of Technical Requirements For Pharmaceuticals For Human Use. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). 2019; Available from: https://database.ich.org/sites/default/files/E9-R1_Step4_Guideline_2019_1203.pdf.

13. Rosenbaum, P.R. and D.B. Rubin, The central role of the propensity score in observational studies for causal effects. Biometrika, 1983. 70(1): p. 41-55.

14. Zalla, L.C. and C.R. Lesko, RE: "Eliminating ambiguous treatment effects using estimands". Am J Epidemiol, 2023. 192(6): p. 1029-1030.

15. Prosepe, I., et al., The disconnect between development and intended use of clinical prediction models for Covid-19: a systematic review and real-world data illustration. Frontiers in Epidemiology, 2022. 2: p. 899589.

16. Sperrin, M., et al., Using marginal structural models to adjust for treatment drop-in when developing clinical prediction models. Statistics in medicine, 2018. 37(28): p. 4142-4154.

17. van Geloven, N., et al., Prediction meets causal inference: the role of treatment in clinical prediction models. Eur J Epidemiol, 2020. 35(7): p. 619-630.

18. Luijken, K., et al., Risk-Based Decision Making: Estimands for Sequential Prediction Under Interventions. Biom J, 2024. 66(8): p. e70011.