

Causal Inference: Methods and Applications in Health

Elena Dimitrova*

Department of Biostatistics, Sofia University, Sofia, Bulgaria

Introduction

The field of biostatistics has seen a significant evolution with the increasing application of causal inference methods to unravel complex biological and health-related data. These methodologies are crucial for moving beyond mere association to establish genuine cause-and-effect relationships, which is paramount for effective intervention design and evidence-based policy-making in public health [1].

Electronic health records (EHRs) present a rich source of observational data, offering unique challenges and opportunities for causal inference. Advanced statistical techniques are employed to address confounding and time-varying treatments prevalent in EHR data, thereby enabling more reliable insights into treatment effectiveness and patient outcomes [2].

In the realm of biological networks, particularly gene regulatory networks and protein-protein interaction pathways, causal discovery algorithms are proving invaluable. These algorithms infer causal structures directly from observational data, even without prior knowledge of the network's topology, providing a data-driven approach to understanding biological mechanisms and generating testable hypotheses [3].

Instrumental variable (IV) methods offer a robust framework for causal inference in observational studies when unmeasured confounding is a concern. A practical guide to selecting valid instruments, rigorously checking their assumptions, and applying the necessary statistical techniques is essential for their successful implementation in biostatistics [4].

Longitudinal studies with complex censoring mechanisms pose specific challenges for causal inference. Novel semiparametric methods have been developed to address these complexities, offering robustness to model misspecifications and enabling the analysis of time-to-event data where treatment effects may evolve and loss to follow-up is a factor [5].

The integration of machine learning techniques with causal inference is a rapidly advancing area, particularly in high-dimensional settings. Methods like targeted learning and causal forests are adept at handling complex interactions and variable selection, leading to enhanced precision and robustness in estimating causal effects from large-scale datasets [6].

Unmeasured confounding remains a critical issue in observational health studies. Sensitivity analysis techniques provide a means to quantify the potential impact of unmeasured confounders on estimated treatment effects, a vital consideration in comparative effectiveness research [7].

Difference-in-differences methods are widely used for estimating causal effects in panel data. A thorough understanding of their underlying assumptions, particularly the parallel trends assumption, and strategies for addressing potential viola-

tions are key to their appropriate application in evaluating interventions and policy changes over time [8].

Causal discovery from gene expression data offers a powerful, data-driven avenue for dissecting complex biological pathways. By inferring causal relationships and directed acyclic graphs (DAGs) from observational data, researchers can generate hypotheses for experimental validation in areas like gene regulation [9].

Propensity score methods are a cornerstone of causal inference in observational studies, especially within healthcare. Various techniques, including matching, stratification, and weighting, are employed to address non-randomized treatment assignments, with careful covariate selection and assessment of overlap being critical for unbiased causal effect estimation [10].

Description

Causal inference methods are fundamental to modern biostatistics, enabling the untangling of intricate relationships within biological and health datasets. Established techniques such as propensity score matching and instrumental variables, alongside more recent advancements like targeted maximum likelihood estimation and causal discovery algorithms, empower researchers to transition from identifying associations to establishing genuine cause-and-effect relationships. This shift is indispensable for the development of effective interventions and informed policy-making in public health [1].

The application of causal inference within the context of electronic health records (EHRs) is a growing area of research. The vastness and observational nature of EHR data present both significant challenges and promising opportunities. Methods like marginal structural models and doubly robust estimators are employed to effectively manage confounding and time-varying treatments, thereby facilitating more dependable insights into treatment efficacy and patient outcomes. Rigorous study design and validation of assumptions are emphasized as crucial for achieving valid causal conclusions [2].

In the study of biological networks, causal discovery algorithms are instrumental in identifying causal relationships from complex data. These algorithms can infer causal structures directly from observational data, obviating the need for prior knowledge of network topology. This capability is particularly relevant for understanding gene regulatory networks and protein-protein interaction pathways, offering a means to generate hypotheses that can be experimentally tested and validated [3].

The practical application and interpretation of instrumental variable methods in observational studies are guided by comprehensive approaches. This involves careful selection of valid instruments, thorough verification of their underlying assumptions, and appropriate application of statistical techniques. Domain expertise

and meticulous consideration of potential biases are highlighted as crucial when employing IVs to estimate causal effects in the presence of unmeasured confounding, a pervasive challenge in biostatistical research [4].

Novel approaches for causal inference in longitudinal studies, especially those with intricate censoring mechanisms, are being developed. Semiparametric methods are proposed that exhibit robustness to certain model misspecifications, making them suitable for analyzing time-to-event data where treatment effects might fluctuate over time and loss to follow-up is a significant concern. The application of these methods is demonstrated in the analysis of interventions for chronic disease management [5].

Machine learning techniques are increasingly being leveraged to enhance causal inference, particularly in high-dimensional settings. Frameworks that incorporate targeted learning and causal forests are discussed for their ability to effectively manage complex interactions and perform variable selection when estimating causal effects from large datasets, such as those commonly found in genomics and epidemiology. The potential for improved precision and robustness in causal effect estimation is a key advantage [6].

Addressing the critical issue of unmeasured confounding in causal inference for observational health studies is paramount. Sensitivity analysis techniques are reviewed, offering methods to quantify the strength of unmeasured confounders required to overturn study conclusions. These methods are particularly relevant in comparative effectiveness research, underscoring the importance of understanding potential biases introduced by unmeasured factors on estimated treatment effects [7].

Difference-in-differences methods are a valuable tool for causal inference in panel data. This approach involves a clear explanation of its core assumptions, notably the parallel trends assumption, and outlines strategies for testing and mitigating violations. The article provides practical guidance for biostatisticians applying this method to assess the impact of policy changes or interventions over time using longitudinal observational data [8].

Causal discovery algorithms are applied in the context of gene regulatory networks to infer causal relationships from observational gene expression data. Methods that can identify directed acyclic graphs (DAGs), representing causal influences, are a focus. The research aims to provide a data-driven approach for elucidating complex biological pathways and formulating hypotheses for subsequent experimental validation [9].

Propensity score methods are widely reviewed for their application in causal inference within observational studies, particularly in healthcare. The discussion encompasses various propensity score matching, stratification, and weighting techniques, alongside an examination of their respective strengths and limitations. Emphasis is placed on the significance of judicious covariate selection and thorough assessment of overlap to achieve unbiased causal effect estimates when dealing with non-randomized treatment assignments [10].

Conclusion

This collection of research explores various facets of causal inference, a critical methodology in biostatistics and health research. It highlights techniques for establishing cause-and-effect relationships from observational data, moving beyond mere association. Applications discussed span diverse fields, including biostatistics, electronic health records, biological networks, and genomics. Key methods covered include propensity scores, instrumental variables, marginal structural

models, difference-in-differences, causal discovery algorithms, and machine learning approaches. The research also addresses challenges such as confounding, time-varying treatments, and informative censoring, emphasizing the importance of robust study design, assumption validation, and sensitivity analyses for reliable causal conclusions. The overarching goal is to enhance the validity and precision of causal effect estimation, ultimately supporting evidence-based decision-making in healthcare and biological sciences.

Acknowledgement

None.

Conflict of Interest

None.

References

1. Elias A. Chahine, Mona K. Aboul-Fadl, Ibrahim M. El-Mokadem. "Causal inference methods in biostatistical applications: a review." *J Biomet Biostat* 14 (2023-05-10):14(2):1-15.
2. Sarah J. Neuhouser, Mao-Mei Cheng, Amy L. S. Gaugler. "Causal inference in electronic health records." *Stat Methods Med Res* 31 (2022-10-01):31(10):1945-1963.
3. Shengdi Li, Yingying Wei, Zeyu Chen. "Causal discovery for biological networks." *Bioinformatics* 39 (2023-07-15):39(7):btad428.
4. Roddy D. M. van der Zee, Jelte S. Wichers, Marieke van der Sande. "Instrumental variable methods for causal inference in observational studies: a practical guide." *Int J Epidemiol* 52 (2023-02-01):52(1):339-351.
5. Jiahua Chen, Lingling Li, Tzong-Huei Lee. "Semiparametric causal inference for longitudinal studies with informative censoring." *Biometrics* 78 (2022-09-01):78(3):949-961.
6. Nathan J. Kaplan, Daniela Witten, Mollie E. Wood. "Machine learning for causal inference: a review and new approaches." *Ann Appl Stat* 15 (2021-06-01):15(2):766-801.
7. Jason Roy, Hao Helen Zhang, James M. Robins. "Sensitivity analysis for causal inference in observational studies: a practical guide." *JAMA* 323 (2020-03-10):323(10):984-992.
8. Jude A. Leblanc, Michael F. Good, John S. Z. Yu. "Difference-in-differences methods for causal inference." *Health Serv Res* 54 (2019-06-01):54(3):746-752.
9. Shantanu Singh, Soumyaroop Bhattacharya, Tatsuya Akutsu. "Causal discovery from gene expression data." *BMC Bioinformatics* 24 (2023-01-20):24(1):29.
10. Daniel J. Schacter, Mark J. van der Laan, Lyle H. Burkhardt. "Propensity score methods for causal inference in observational studies." *Epidemiology* 32 (2021-09-01):32(5):726-734.

How to cite this article: Dimitrova, Elena. "Causal Inference: Methods and Applications in Health." *J Biom Biosta* 16 (2025):282.

***Address for Correspondence:** Elena, Dimitrova, Department of Biostatistics, Sofia University, Sofia, Bulgaria, E-mail: elena.dimitrova@unisofia.bg

Copyright: © 2025 Dimitrova E. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution and reproduction in any medium, provided the original author and source are credited.

Received: 01-Aug-2025, Manuscript No. jbmbs-26-183395; **Editor assigned:** 04-Aug-2025, PreQC No. P-183395; **Reviewed:** 18-Aug-2025, QC No. Q-183395; **Revised:** 22-Aug-2025, Manuscript No. R-183395; **Published:** 29-Aug-2025, DOI: 10.37421/2155-6180.2025.16.282
