On Counterfactual Explanations of Cardiovascular Risk in Adolescent and Young Adult Breast Cancer Survivors

Balordi, A., Bernasconi, A., Andreotti, A., Guzzinati, S., Cabañas De Paz, R., & Zanga, A

Published in:
Journal of Medical Systems

DOI (link to publication from Publisher):
https://doi.org/10.1007/s10916-025-02273-1

Publication date:
2025

Journal of Medical Systems (2025) 49:140
https://doi.org/10.1007/s10916-025-02273-1

RESEARCH

On Counterfactual Explanations of Cardiovascular Risk in Adolescent and Young Adult Breast Cancer Survivors

Antonio Balordi[1,2] · **Alice Bernasconi**[1,3] · **Alessandra Andreotti**[4] · **Stefano Guzzinati**[4] · **Rafael Cabañas De Paz**[5] · **Alessio Zanga**[1]

Received: 16 April 2025 / Accepted: 21 September 2025

Abstract

Cancer treatments might lead to several long-term effects. In this work we investigate their causal role on ischemic heart disease and their potential precursors (i.e. hypertension and dyslipidemia) of the ovarian suppression therapy in adolescent and young adult (AYA) breast cancer (BC) survivors. Additionally, we assess the external validity of our findings through comparative analysis of regional data. We take advantage of a causal network model that leverage on observational data on 1-year AYA BC survivors living the Lombardy region in Italy. Using a structural causal model (SCM) and counterfactual analysis within Pearl's causal inference framework, we estimate the Average Causal Effect (ACE), Probability of Necessity (PN), and Probability of Sufficiency (PS) for the cause-effect relationships. Data of a regional cohort of AYA BC patients living in the Veneto region were used to externally validate results. Ovarian suppression was found to be a necessary but not sufficient cause for ischemic heart disease (PN > 97.8%; PS < 1.97%). While PN is high for both hypertension and dyslipidemia, PS varied suggesting ovarian suppression alone could induce hypertension in about 30% of cases but was rarely sufficient for dyslipidemia onset. External validation confirmed the robustness of findings across regions. Our experimental results may be of interest for clinicians who aim at personalizing the follow-up of AYA BC survivors, with particular attention to be paid in monitoring the hypertension onset or in its prevention. The study demonstrates the value of counterfactual reasoning and causal inference when working with real-world data.

Keywords Causal networks · Counterfactual explanations · Breast cancer survivors · Treatment guidelines · Cardiovascular diseases · Adolescents and young adults

Introduction

In Italy, the leading causes of death among individuals under 40 years old are accidents or suicides [1]. While in young males the second major cause of death are cardiovascular diseases, like acute myocardial infarction, young women generally exhibit significantly lower cardiovascular morbidity and mortality rates compared to age-matched men. This protective advantage is largely attributed to endogenous estrogen's beneficial effects on vascular function, lipid metabolism, and inflammatory processes [2].

Differently from men, in young women a strong role in explaining premature mortality is played by breast cancer. Despite breast cancer is the most common cancer in Adolescents and Young Adults (AYAs, aged 15 to 39 years) as in older females, these two groups differs in terms of survival: thus, due to the most aggressive tumor

Antonio Balordi and Alice Bernasconi contributed equally to this article as co-first authors.

✉ Alice Bernasconi
a.bernasconi6@campus.unimib.it;
alice.bernasconi@istitutotumori.mi.it

1 Models and Algorithms for Data and Text Mining Laboratory (MADLab), University of Milano-Bicocca, Milan, Italy

2 CASD - Italian Defense University, Rome, Italy

3 Evaluative Epidemiology Unit, Department of Epidemiology and Data Science, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

4 Veneto Cancer Registry, Epidemiological Department, Azienda Zero, Padova, Italy

5 Department of Mathematics and CDTIME, University of Almería, Almería, Spain

Published online: 16 October 2025

case-mix, AYAs have lower survival rates compared to their older counterpart. Breast cancer treatment in these young patients is complex: most of the time it requires more intense schemes and different approaches, such as hormone therapies.

As of today, the standard neoadjuvant therapy (i.e. post-surgery) for the treatment of hormone receptor-positive breast cancer expects the use of endocrine therapies, like tamoxifen, sometimes combined together with CK 4/6 inhibitors [3] with the objective of reducing the probability of recurrence. Separately, in young patients especially when facing neoadjuvant (i.e. pre-surgical) chemotherapy, gonadotropin-releasing hormone (GnRH) agonists may be administered to preserve ovarian function and fertility from chemotherapeutic damages. This treatment induces a reversible menopause-like state that can cause significant cardiovascular implications as drawbacks. Thus, the natural hormonal protection in women diminishes considerably with the onset of menopause, when women's cardiovascular risk begins to approach that of men [4], both inducing ischemic heart diseases or increasing the risk of developing their precursors (i.e. hypertension and dyslipidemia).

Leveraging on a causal Bayesian network and the causal inference framework proposed by Pearl et al. [5], in our previous analysis we showed that the GnRH neoadjuvant hormonal therapy alone (if administered in combination with tamoxifen as neoadjuvant hormonal treatment) represents a necessary but not sufficient cause for developing ischemic heart disease within 5-years from diagnosis in young breast cancer survivors. Building upon our previous findings we aim to explore the potential mediating mechanisms underlying this relationship, focusing on two important considerations: i) the limited temporal span of our follow-up period and ii) the presence of potential unobserved risk factors that may be required to complete the causal pathway.

In this context, we extended our previous work by:

- Examining how GnRH neoadjuvant hormonal therapy influences the development of hypertension and dyslipidemia, which may serve as essential mediators between cancer treatment and ischemic outcomes in a longer time span (e.g. 10 years);
- Strengthening the external validity of our findings through comparative analysis of regional data (i.e. a different Italian region);
- Helping to contextualize our results and account for potential unmeasured confounders in the causal relationship that may differ between the internal and external validation cohort.

Methods

In this section we introduce the notation, together with the main concepts and the mathematical models needed to follow the rest of the paper. In particular, we give the definitions of Bayesian network, causal network and structural causal model, while also describing the three rungs of the *ladder of causation* [6] which are fundamental to understand our contributions.

Bayesian Networks

Bayesian networks (BNs) [7–9] are a type of probabilistic graphical model (PGM) used for reasoning under uncertainty. BNs are made of a qualitative component in the form of directed acyclic graph (DAG) encoding the independence relations between the variables in the problem, while the quantitative component is a set probability distributions measuring such relations. More formally, BNs can be defined as follows.

Definition 1 (Bayesian Network (BN)) A Bayesian network is a pair $\langle \mathcal{G}, \mathcal{P} \rangle$, where:

- $\mathcal{G} = \langle \mathrm{V}, \mathrm{E} \rangle$ is a DAG, with V a set of vertices and $\mathrm{E} \subset \mathrm{V} \times \mathrm{V}$ a set of directed edges,
- $\mathcal{P}$ is a probability distribution over the random vector X.

Each vertex $V_i \in \mathrm{V}$ is mapped to a variable $X_i \in \mathrm{X}$, so that the global probability distribution $\mathcal{P}$ is factorized over $\mathcal{G}$ into local probability terms $P(X_i \mid Pa(X_i))$, with $Pa(X_i)$ the parents[1] of X_i.
For each variable $V_i \in \mathrm{V}$, we define the *ancestors* of V_i to be the set of variables $V_j \in \mathrm{V} \setminus \{V_i\}$ such that there exists a directed path.[2] Similarly, we define the *descendants* of V_i to be the set of variables $V_j \in \mathrm{V} \setminus \{V_i\}$ such that there exists a directed path from V_i to V_j. Henceforth, we will refer to a vertex V_i and its corresponding variable X_i interchangeably.

Definition 2 (Causal Network (CN)) A *Causal Network* is a BN in which any edge from parents to children represents a cause-effect relationship.

Observational and Interventional Rungs

Standard probabilistic inference involves computing the posterior probability distribution for variables of interest given evidence about other variables, commonly referred

[1] A vertex V_j is said to be a parent of V_i if there exists a directed edge from V_j to V_i.

[2] A directed path from V_i to V_j is sequence of directed edges starting from V_i and ending in V_j

to as observational queries (e.g., "what if I see this?"). For instance, given states x and y of random variables X and Y, respectively, an observational query might involve computing the conditional probability $P(x \mid y)$. Here, x and y represent the presence of X and Y, while x' and y' denote their absence.

In causal reasoning, we focus on hypothetical scenarios where we calculate the probability of a variable given that we intervene on another. For example, the query $P(Y = y \mid \mathrm{do}(X = x))$ represents the probability that Y equals y when X is intervened to take the value x. The notation $\mathrm{do}(X = x)$ explicitly denotes an intervention, distinguishing it from mere observation. The difference between outcomes under different treatments or interventions is commonly measured through the *average treatment effect* (ACE), also known as the *average causal effect* (ACE), which can be defined as:

$$\mathrm{ATE}(X, Y) = P(Y = y \mid \mathrm{do}(X = x)) - P(Y = y \mid \mathrm{do}(X = x')). \tag{1}$$

To calculate an interventional query, a process often referred to as "surgery" is employed. This graphical operation involves removing the incoming arcs to the intervened variable X and setting the node to a specific value $X = x$. The model that results from this surgical intervention is known as the "post-intervention" model. This process is performed to restrict the natural tendency of the variable to change in response to other variables in the environment.

Performing graph surgery is the initial step required to distinguish the associative effect from the purely causal effect. However, a causal estimand cannot be directly estimated using a statistical estimator; it must first be translated into a statistical estimand by removing the intervention. This process is known as the *identification of the causal effect*.

If there exists a set of covariates Z that satisfies the *backdoor criterion* [6] in the model, then there exists a consistent estimator for the causal effect of X on Y:

$$P(Y = y \mid \mathrm{do}(X = x)) = \sum_{\mathrm{z}} P(Y = y \mid X = x, \mathrm{Z} = \mathrm{z}) P(\mathrm{Z} = \mathrm{z}). \tag{2}$$

Under the condition of exogeneity (also known as no-confounding) [10], the way Y would potentially respond to experimental conditions x or x' is independent of the actual value of X. This implies that $P(Y = y \mid \mathrm{do}(X = x)) = P(Y = y \mid X = x)$ and $P(Y = y \mid \mathrm{do}(X = x')) = P(Y = y \mid X = x')$, thus making $ACE(X, Y) = ATE(X, Y)$. A graphical criterion to identify the condition of exogeneity is the absence of a common ancestor of X and Y connected to Y through a directed path that does not include X. Apart from the exogeneity condition, performing Causal Inference requires a set of assumptions: i.e. no unobserved confounding, positivity [11] and Stable Unit Treatment Value Assumption [12].

Counterfactual Rung

Counterfactual queries [6] explore hypothetical scenarios, such as, "What would the outcome have been if the variable had taken a different value?" For example, $P(Y_x \mid X = x')$ represents the probability of Y if X had taken the value x instead of x'. Here, Y_x relates to the hypothetical scenario, while X is in the real scenario.

A key concept in this context is the *probability of necessity* (PN), which measures the extent to which one event is a necessary condition for another. The PN is defined as:

$$\mathrm{PN}(X, Y) = P(Y_{x'} = y' \mid X = x, Y = y). \tag{3}$$

Here, X is considered a necessary cause for Y if y would not have occurred without x, given that both x and y actually occurred. Therefore, PN represents our certainty about X being a necessary cause of Y.

Similarly, we may also be interested in determining whether an event is a sufficient condition. To address this, we define the *probability of sufficiency* (PS) as:

$$\mathrm{PS}(X, Y) = P(Y_x = y \mid X = x', Y = y'). \tag{4}$$

X is considered a sufficient cause for Y if y occurs whenever x occurs. Thus, PS represents the probability that X is a sufficient cause of Y. In other words, it is the probability that setting x would lead to y in a scenario where both x and y are currently absent.

Counterfactual queries cannot be directly computed from a CN. Instead, structural causal models (SCMs) [6], which can be viewed as an extension of CNs, are required. SCMs consist of endogenous variables, which represent internal elements of the model, and exogenous variables, which often lack a clear semantic interpretation. SCMs can be formally defined as follows [13].

Definition 3 (Structural Causal Model (SCM)) A structural causal model is defined as a 4-tuple $\langle \mathrm{U}, \mathrm{V}, \mathcal{F}, \mathcal{P} \rangle$, where:

- U is the set of exogenous variables;
- V is the set of endogenous variables;
- $\mathcal{F} = \{f_i : \mathrm{U}_i \cup Pa(V_i) \to V_i, \forall V_i \in \mathrm{V}\}$ is the set of *structural equations*;
- $\mathcal{P}$ is the set of exogenous probability distributions $P(U_i)$, for each $U_i \in \mathrm{U}$.

Note that the structural equations $\mathcal{F}$ actually define a DAG over the variables in $\mathrm{U} \cup \mathrm{V}$, with an edge from each variable in $\mathrm{U}_i \cup Pa(V_i)$ to V_i.

Experimental Results

Answering the Causal Query

The starting point for this analysis is the causal network that has been developed for the estimation of the risk of CVDs and is being described in [14] and in the Supplementary Material Section 1. Briefly, the model was developed fusing together prior medical knowledge and data coming from two different cohorts of AYA with BC: a population-based (N=1037) and a clinical cohort (N=339). Both cohorts have the same selection criteria, i.e. AYA BC 1-year survivors living in the Lombardy region in Italy. These two cohorts were complementary (in terms of information) even though they were not directly linkable. In both cohorts, we collected information on 24 categorical variables that included: cancer prognostic factors, cancer treatments, major cardiovascular risk factors (i.e, type 2 diabetes, hypertension and dyslipidemia) and cardiovascular diseases (including ischemic heart diseases and cardiotoxicities). This work focused on the population-based cohort data only. All the variable and sample selections made for this specific work are explained and motivated in the Supplementary Material Section 2; missing data were handled as explained in [15].

Taking advantage of this model, in this work we focused on a specific causal query extracted from BC treatment guidelines (see Supplementary Material Section 1 for more details). Briefly, our objective was to properly understand the impact of the neoadjuvant hormonal therapy `[hormons_neo]` (that for these patients corresponds to ovarian suppression induced by GnRH agonists) on the risk of developing ischemic heart diseases `[ischemic_heart_disease]` in a selected group of patients treated also with neoadjuvant hormonal therapy (i.e. tamoxifen); the associated risk has been computed considering a time horizon of 5-years from the diagnosis. According to our previous experimental results, we found that $P(D = yes|HN = yes) = 1.45\%$,[3] with 95% Confidence Interval[4] of [0.88%;2.37%] and $P(D = yes|HN = no) = 0.22\%$ [0.05%; 0.69%]; consequently, we obtain $ACE(HN, D) = 1.23\%$ [0.35%; 2.51%], this means that the probability of developing a CVD is almost 0 with reference to patients who do not receive ovarian suppression together with tamoxifen treatment and it increase of about one percentage point if the patient receives it.

In the context of counterfactual analysis, we investigated the likelihood that the variable `[hormons_neo]` is a necessary and sufficient cause for `[ischemic_heart_disease]`. To achieve this we ran the algorithm as described in Supplementary Material Section 3. For each run, we obtained a value for the causal query PN and another one for PS. The distribution of these values is depicted in Fig. 1. In the case of PN, all values exceeded 97.83%. Conversely, the values for PS remained below 1.97%. These results suggest that, with a high probability, the ovarian suppression treatment is a necessary but not a sufficient cause for the ischemic heart disease. That means that the latter factor is essential for the disease to occur, however, this treatment alone is not enough to cause the disease; other factors must also be present.

Impact of Hormonal Therapy on Hypertension and Dyslipidemia

Following the same methodological framework as our earlier analysis, we extended our counterfactual analysis to the two major mediators of ischemic heart diseases: hypertension (`[hypertension]`) and dyslipidemia (`[dyslipidemia]`).

The empirical distributions of PN and PS, related to `[hormons_neo]` development factors affecting the mediators `[hypertension]` and `[dyslipidemia]`, are displayed in Figs. 2 and 3.

Overall, the mean PN is high both for `[dyslipidemia]` and `[hypertension]` (about 94.45% and 77.70%, respectively) implying that nearly all patients who develop `[hypertension]` or `[dyslipidemia]` do so after receiving `[hormons_neo]`. Nevertheless, the mean PS for `[dyslipidemia]` is very low (approximately 1.78%), suggesting that `[hormons_neo]` is generally not enough to induce the condition without other contributing factors. On the opposite, the mean PS for `[hypertension]` is around 20%, indicating that about a fifth of patients who did receive `[hormons_neo]` would have developed `[hypertension]` had they been treated.

External Validation

As a further step, we implemented an external validation routine to strengthen the validity of our findings. We conducted a comparative analysis using an additional population-based cohort of N=952 patients identified in regional data from Veneto Cancer Registry (VCR), which helps to

[3] In terms of notations, D represents [hormons_neo]and HN stands for [ischemic_heart_disease]

[4] 95% Confidence intervals are calculated using a Wilson approximation due to low number of cases

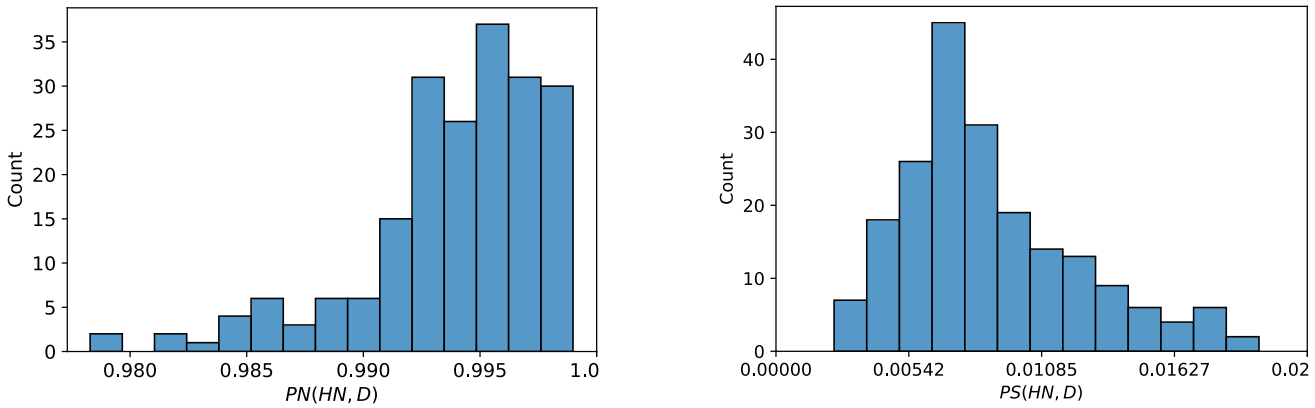

Fig. 1 Values of the probability of necessity (left) and probability of sufficiency (right) of variable `[hormons_neo]` for `[ischemic_heart_disease]`

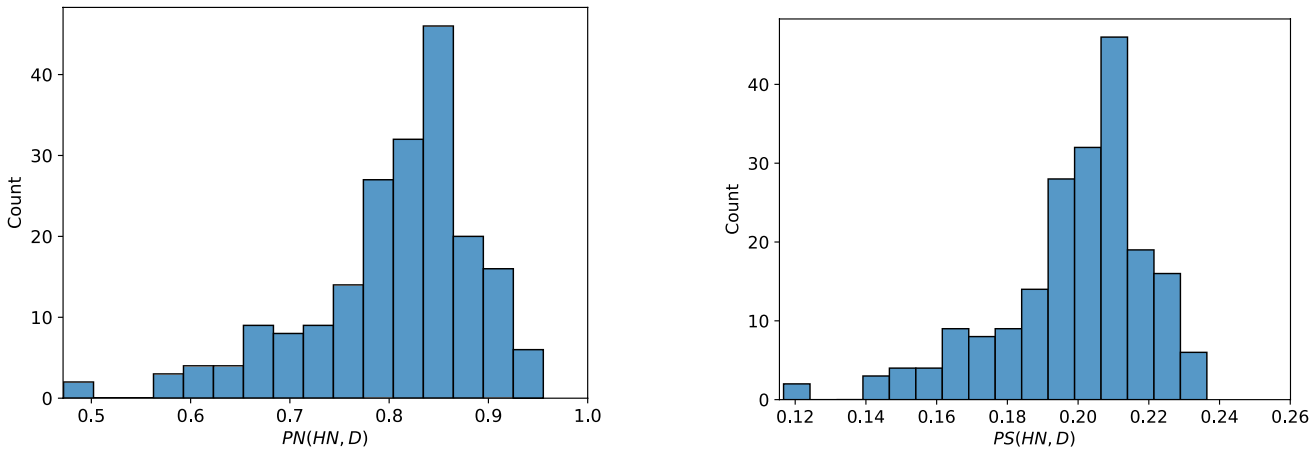

Fig. 2 Values of the probability of necessity (left) and probability of sufficiency (right) of variable `[hormons_neo]` for `[hypertension]`

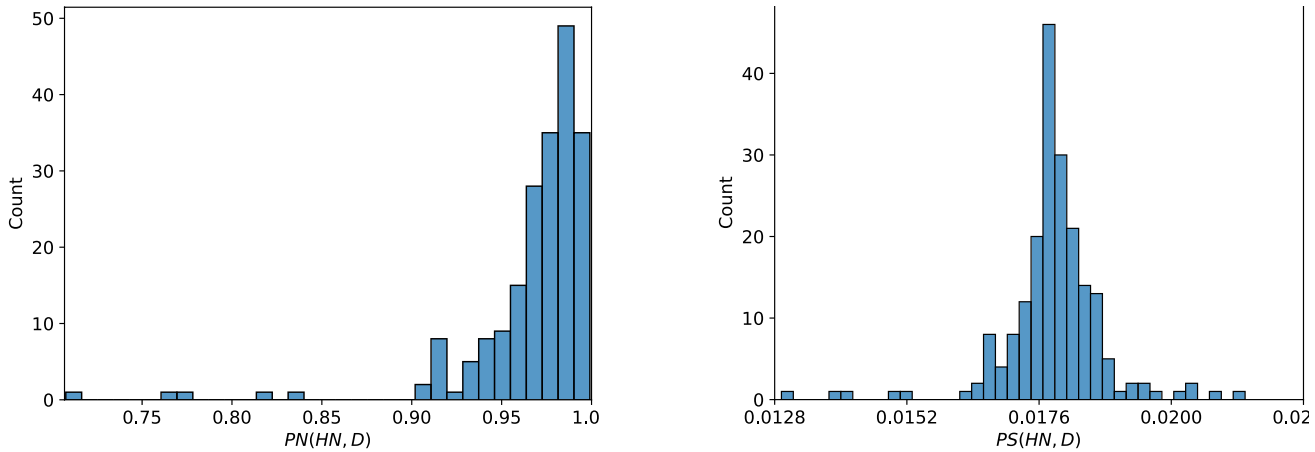

Fig. 3 Values of the probability of necessity (left) and probability of sufficiency (right) of variable `[hormons_neo]` for `[dyslipidemia]`

contextualize our results and account for potential unmeasured confounders.

Comparing the Veneto and the Lombardy cohorts, we observed a similar distribution pattern of all the baseline covariates, major cardiovascular risk factors and cardiovascular outcomes following hormone therapy with minor differences in the treatment decision only (Supplementary Material Table 1).

A workflow describing the data collection process and the patients selection for the analyses (with corresponding counts), both in Lombardy and Veneto region, is included in the Supplementary Material, Fig. 5.

Our comparative analysis demonstrates strong consistency between results obtained from the Lombardy and Veneto cohorts. As illustrated in Fig. 4, the boxplot visualization reveals that mean values from both regions fall within each other's standard deviation intervals.

Discussion

In this work, starting from a clinical question on the recommendability of the ovarian suppression addition to tamoxifen treatment, the counterfactual explanations made it evident that, the ovarian suppression treatment is a necessary but not a sufficient cause for the ischemic heart disease. That means that the latter factor is essential for the disease to occur, i.e., the disease cannot happen without the presence of this hormonal treatment. However, this treatment alone is not enough to cause the disease; other factors must also be present.

Moreover, we observed a different role of the ovarian suppression on the potential mediators of ischemic heart diseases. Thus, the low PS for dyslipidemia suggests that the addition of ovarian suppression is generally not enough to induce the condition without other contributing factors. On the opposite, the PS for hypertension of about 30%, underscores a more pronounced influence of the treatment choice on this condition.

It is important to say that it is very difficult to define a clinically meaningful tailored threshold of PS and PN. Considering the lack of evidence and the rarity of the disease in these patients, any observational study (like this one we conducted) could not be considered conclusive but they can serve as a starting point for further clinical and biological investigations and discussions with clinicians, oncologists and cardio-oncologists.

Knowing the causal role of the treatment choice on hypertension may be useful for clinicians to organize the follow-up of the patients i) by prescribing blood pressure exams more frequently for early detection or ii) by encouraging preventive strategies, whether pharmacological or through lifestyle interventions.

Our experimental results proved to be robust and valid even on the external validation cohort. This outcome is likely due to the similar distribution of features across the two cohorts, with minor differences in the treatment choices only. These differences may be attributed to the different incidence period covered by the two cohorts (thus, the Veneto cohort is more recent) or to differences in tumor staging at diagnosis that we could not verify because our data lack this information.

However, despite its relevance this work has some limitations too. First of all, while ACE is calculated using a statistical estimand, the computation of PN and PS is more complex and requires a SCM. In our proposed approach an exogenous variable was added as a parent to each endogenous variable. Given the unknown distributions of the exogenous variable we learn it by repeatedly applying a learning algorithm. This approach makes all the process extremely computationally expensive which allows to handle only

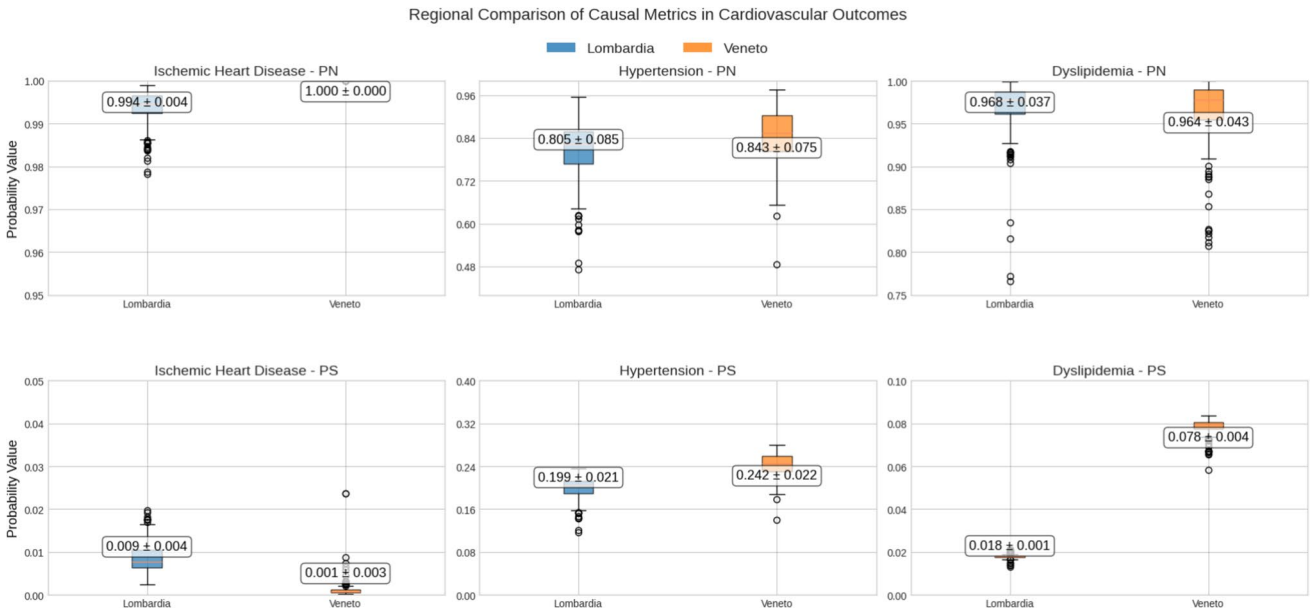

Fig. 4 Regional comparison of counterfactual metrics for hormonal therapy effects on cardiovascular outcomes in Lombardy and Veneto

queries in which the effect node can have maximum of three parents. From a methodological standpoint, this is an active area of research. Preliminary findings have been presented such as [16] and [17].

Finally, our experimental results are valid under the assumptions showed in "Methods" section.

Nevertheless, this assumptions are difficult to be valid especially considering the vast literature that describes the role of lifestyle factors (like smoking, physical inactivity, obesity and poor diet) on the development of cardiovascular diseases both directly and indirectly through type 2 diabetes, hypertension and dyslipidemia [18–21]. Hence, attention should be paid in the results interpretation because the exclusion of these unmeasured variables from the analyses may have led to an overestimation of the role of both treatments and the observed major cardiovascular risk factors on the outcomes. The distribution of these unmeasured factors looks similar between the Lombardy and Veneto regions [22], this might explain the absence of substantial differences in the experimental results. These considerations also support our conclusion that we approximate causal contrasts under stated assumptions, rather than emulate a randomized trial.

Finally, caution must be exercised when attempting to generalize our experimental results to external populations. In this context, one major factor that may limit the model's transportability is race. In Italy, the racial distribution is relatively homogeneous, with the vast majority of the population being White Caucasian. However, the literature clearly shows that racial and ethnic disparities exist in both breast cancer and CVDs, with non-Hispanic Black women often experiencing higher mortality rates from both conditions compared to White women. These survival gaps are attributable to a combination of biological differences, social determinants of health, and disparities in healthcare access — particularly in the United States [23].

Conclusions

In this work we showed how causal networks and counterfactual reasoning can be effectively used to disentangle uncertainty in treatment choices, while helping clinicians in better tailoring personalised follow-up guidelines for chronic patients, like cancer survivors. This is an important use case which concretely shows how effective observational real-world data can be to answer clinical questions. In these regards, the overlapping of association and intervention results support the idea that, despite confounding is present, clinicians are able to deal with it in the everyday clinical practice, even without ad hoc strict guidelines. This result makes even more important and relevant the evidence driven by the third ladder of causation (counterfactuals) where our analyses approximate causal contrasts under stated assumptions, rather than claiming to emulate a randomized trial. This approach can be particularly powerful when evidence from trials is not available nor ethical to be obtained [24].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10916-025-02273-1.

Acknowledgements This work was possible thanks to the data collected in the Ada project ("Adolescents and young adults with cancer in Italy. How to ensure access to the best care and quality of survival"). We thank the project working group: Manuel Zorzi, Carlotta Buzzoni, Sabrina Fabiano, Adele Caldarella, Stefania Giorgetti, Salvatore Sciacca, Rosa Angela Filiberti, Cinzia Gasparotti, Giuliano Carrozzi, Rosalba Amodio, Lucia Mangone, Silvia Iacovacci, Mario Fusco, Fabrizio Stracci, William Mantovani, Giuseppe Cascone, Sante Minerba, Giuseppe Sampietro, Anna Melcarne, Paolo Ricci, Federica Manzoni, Maria Letizia Gambino, Elisabetta Merlo, Rossella Bruni, Alessandra Sessa, Giancarlo D'Orsi, Anna Clara Fanetti, Lucia De Lorentis, Tiziana Scuderi, Fabio Savoia, Iolanda Grappasonni, Salvatore Bongiorno, Antonio Romanelli.

Author Contributions A.Be. contributed to the conceptualization and supervision of the project. The methodology was developed by A. Ba., A. Z. and R.C.. A.Ba. conducted the formal analysis and was in charge of the software curation. S.G. and A.A. contributed to the data curation. Moreover, all authors validated the results and contributed to the writing and editing of the manuscript. All authors approved the final version of the manuscript.

Funding Alice Bernasconi is funded by an *AIRC* 2020 project, grant number 24864, titled *"pRedicting cardiOvascular diSeAses iN adolescent and young breast caNcer pAtients (ROSANNA)*".
Grant PID2022-139293NB-C31 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe partially supported this work. Rafael Cabñas was also supported by "Plan Propio de Investigación y Transferencia 2024-2025" from University of Almería under the project P_LANZ_2024003.

Data Availability No datasets were generated or analysed during the current study.

Materials Availability Not applicable

Code Availability The code used for data analysis and to generate the results presented in this paper is available at https://github.com/AlessioZanga/cardiovascular-counterfactuals. The repository includes all necessary scripts, documentation, and example data to reproduce the findings reported in this study.

Declarations

Ethical Approval and consent to participate This study was conducted in accordance with the Declaration of Helsinki, and it was approved by the Institutional Review Board (or Ethics Committee) of Fondazione IRCCS Istituto Nazionale dei Tumori, study protocol number INT 0083/21.

Consent for publication Not applicable

Competing interests The authors declare no competing interests.

Clinical Trial Number Not applicable

References

1. Burgio, A., D'Errico, A.: Main causes of death in Italy. Italian Journal of Public Health. **1** (2003) https://doi.org/10.2427/6111
2. Marco, M., Luigi, T.P., Daniela, P., Giuseppe, A.M., Lucio, G., Donata, L., Federico, P., Stefano, P., Massimo, G.M., Domenico, G., *et al.*: Sex-related differences in demographics, diagnosis and management of patients with chronic coronary syndromes. Journal of Cardiovascular Medicine **25**(12), 845–853 (2024)
3. Shah, M., Nunes, R., Stearns, V.: Cdk4/6 inhibitors: Game changers in the management of hormone receptor–positive advanced breast cancer? Oncology (Williston Park, N.Y.) **32**, 216–22 (2018)
4. Lucà, F., Giuseppe, A.M., Iris, P., Angela, D.F.S., Simona, G., Massimiliano, R.C., Laura, P., Laura, C., Bruno, P., Francesco, G., Angelo, L., Maura, F.G., Carmine, R., Sandro, G., Furio, C., Massimo, G.M.: Update on management of cardiovascular diseases in women. Journal of Clinical Medicine **11**(5) (2022) https://doi.org/10.3390/jcm11051176
5. Pearl, J., Glymour, M., Jewell, N.P.: Causal Inference in Statistics: A Primer. John Wiley & Sons, United Kingdom (2016)
6. Pearl, J.: Causality: Models, Reasoning and Inference. Cambridge University Press, New York (2000)
7. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Elsevier, Los Angeles (1988). https://doi.org/10.1016/C2009-0-27609-4. https://linkinghub.elsevier.com/retrieve/pii/C20090276094
8. Jensen, F.V., Nielsen, T.D.: Bayesian Networks and Decision Graphs vol. 2. Springer, New York, NY (2007)
9. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles and Techniques. The MIT Press, Cambridge, Massachusetts (2009)
10. Tian, J., Pearl, J.: Probabilities of causation: Bounds and identification. Annals of Mathematics and Artificial Intelligence **28**(1), 287–313 (2000)
11. Hernán, M.A., Robins, J.M.: Estimating causal effects from epidemiological data. Journal of Epidemiology & Community Health **60**(7), 578–586 (2006)
12. Robins, J.M., Hernan, M.A., Brumback, B.: Marginal structural models and causal inference in epidemiology. Lww (2000)
13. Bareinboim, E., Correa, J.D., Ibeling, D., Icard, T.: On Pearl's hierarchy and the foundations of causal inference. In: Probabilistic and Causal Inference: the Works of Judea Pearl, pp. 507–556. ACM, New York, NY, USA (2022)
14. Bernasconi, A., Zanga, A., Lucas, P.J.F., Scutari, M., Stella, F.: Towards a Transportable Causal Network Model Based on Observational Healthcare Data, vol. 3578, pp. 122–129 (2023)
15. Bernasconi, A., Zanga, A., Lucas, P.J.F., Scutari, M., Di Cosimo, S., De Santis, M.C., La Rocca, E., Baili, P., Cavallo, I., Verderio, P., Ciniselli, C.M., Pizzamiglio, S., Blanda, A., Perego, P., Vallerio, P., Stella, F., Trama, A., Group, T.A.W.: From real-world data to causally interpretable models: A bayesian network to predict cardiovascular diseases in adolescents and young adults with breast cancer. Cancers **16**(21) (2024) https://doi.org/10.3390/cancers16213643
16. Bjøru, A., Cabañas, R., Langseth, H., Salmerón, A.: Divide and conquer for causal computation. International Journal of Approximate Reasoning **186**, 109520 (2025) https://doi.org/10.1016/j.ijar.2025.109520
17. Huber, D., Chen, Y., Antonucci, A., Darwiche, A., Zaffalon, M.: Tractable Bounding of Counterfactual Queries by Knowledge Compilation (2023). https://arxiv.org/abs/2310.03352
18. Ondimu, D.O., Kikuvi, G.M., Otieno, W.N.: Risk factors for hypertension among young adults (18-35) years attending in tenwek mission hospital, Bomet County, Kenya in 2018. Pan African Medical Journal **33** (2019) https://doi.org/10.11604/pamj.2019.33.210.18407
19. Saavedra, A., Rodrigues, E., Carvalho, D.: Dislipidemia secundária a hipotiroidismo e colestase. Acta Médica Portuguesa **33**, 204–207 (2020) https://doi.org/10.20344/amp.9944
20. Muhandiramge, J., Zalcberg, J.R., Londen, G.J., Warner, E.T., Carr, P.R., Haydon, A., Orchard, S.G.: Cardiovascular disease in adult cancer survivors: a review of current evidence, strategies for prevention and management, and future directions for cardio-oncology. Current Oncology Reports **24**, 1579–1592 (2022) https://doi.org/10.1007/s11912-022-01309-w
21. Berkman, A.M., Andersen, C.R., Roth, M.E., Gilchrist, S.C.: Cardiovascular disease in adolescent and young adult cancer survivors: Impact of sociodemographic and modifiable risk factors. Cancer **129**, 450–460 (2023) https://doi.org/10.1002/cncr.34505
22. Sanita, I.S.: PASSI - Progressi delle Aziende Sanitarie per la Salute in Italia. Available on line (2025). https://www.epicentro.iss.it/passi/
23. Troeschel, A.N., Liu, Y., Collin, L.J., Bradshaw, P.T., Ward, K.C., Gogineni, K., McCullough, L.E.: Race differences in cardiovascular disease and breast cancer mortality among us women diagnosed with invasive breast cancer. International Journal of Epidemiology **48**(6), 1897–1905 (2019) https://doi.org/10.1093/ije/dyz108
24. Hammerton, G., Munafò, M.R.: Causal inference with observational data: the need for triangulation of evidence. Psychological Medicine **51**, 563–578 (2021) https://doi.org/10.1017/S0033291720005127