Clinical Oncology xxx (xxxx) xxx

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Causal Inference in Oncology: Why, What, How and When

W.A.C. van Amsterdam [*], S. Elias [*], R. Ranganath [†]

[*] *Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Utrecht, the Netherlands*
[†] *Courant Institute of Mathematical Sciences, New York University, New York City, New York, USA*

Abstract

Oncologists are faced with choosing the best treatment for each patient, based on the available evidence from randomized controlled trials (RCTs) and observational studies. RCTs provide estimates of the average effects of treatments on groups of patients, but they may not apply in many real-world scenarios where for example patients have different characteristics than the RCT participants, or where different treatment variants are considered. Causal inference defines what a treatment effect is and how it may be estimated with RCTs or outside of RCTs with observational – or 'real-world' – data. In this review, we introduce the field of causal inference, explain what a treatment effect is and what important challenges are with treatment effect estimation with observational data. We then provide a framework for conducting causal inference studies and describe when in oncology causal inference from observational data may be particularly valuable. Recognizing the strengths and limitations of both RCTs and observational causal inference provides a way for more informed and individualized treatment decision-making in oncology.

Keywords: Causal inference; confounding; individualised cancer care; realworld data; research methodology; treatment effect heterogeneity

Why Causal Inference?

Everyday, oncologists make medical decisions that change the future outcome of a patient. Whether it is ordering a diagnostic test, prescribing medication, or scheduling a follow-up visit, each decision is an intervention that changes a patient's future outcome. The benefits and harms of such an intervention, or with a generic word *treatment*, determine the *treatment effect*, meaning the *causal effect* of giving that treatment. For treatment decision-making, clinicians appraise treatments by balancing their effects, preferably on a per-patient basis. Estimates of these treatment effects for groups of patients generally come from randomized controlled trials (RCTs). However, clinicians encounter patients that are not represented in the published RCTs, for instance because RCTs tend to include younger and fitter patients [1–3]. Also, clinicians may consider treatment variants that were not tested in RCTs or may have doubts whether a certain treatment that was shown to be effective *on average* is also beneficial for this *individual patient*. In these cases, estimates of the treatment effects are still required to make an informed treatment decision.

In addition, clinicians and policy makers fear that treatment effect estimates from RCTs may not translate directly to clinical care, for instance because treatments may be applied more strictly according to protocol in trials than in day-to-day care. Literature describes the discrepancy between outcomes in RCTs and standard clinical care as the efficacy-effectiveness-gap [4]. *Real-world evidence* [5] is promoted to evaluate what is called the *real-world* effectiveness of treatments. This approach aims to infer treatment effects outside of RCTs, namely in observational studies.

Because treatment effect estimates are needed for each treatment decision it is important to have a good understanding of what a treatment effect is. And as not all these decisions are covered by RCTs, we should know how treatment effects may be estimated inside and outside of RCTs. This is the topic of the field of *causal inference*.

In this review, we describe what causal inference is, what a treatment effect is, how to perform a causal inference study and in what situations in oncology causal inference from observational data may be particularly useful.

Author for correspondence: W.A.C. van Amsterdam, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Utrecht, the Netherlands.

E-mail address: w.a.c.vanamsterdam-3@umcutrecht.nl (W.A.C. van Amsterdam).

https://doi.org/10.1016/j.clon.2024.07.002

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

What is Causal Inference?

Causal inference is a field of science that defines exactly what a causal question is and how it may be answered from data. At the same time, causal inference is used as a verb for making an *inference* (i.e. an estimate or claim) about a *causal* quantity. We provide a glossary of terms important in causal inference in Table 1.

What is the Individual Treatment Effect?

To define a treatment effect, for example comparing treatment A with treatment B, we envision two possible future outcomes for a single patient: one after giving treatment A and one after giving treatment B. These future outcomes that would be observed after giving a certain treatment are called *potential outcomes*.

For an individual patient, the individual treatment effect *of treatment A versus treatment B is the difference between the potential outcomes corresponding with treatment A and treatment B.*

See Figure 1A for an illustration. In reality, we only observe the outcome that followed the actually given treatment but not the other potential outcome, so we cannot observe this individual treatment effect (Figure 1B). This lack of observing both potential outcomes in one individual is a fundamental issue in causal inference.

Estimating the Average Treatment Effect in an RCT

Whereas estimating individual treatment effects is generally impossible, we may be able to estimate the *average treatment effect* in a certain population, which is the average of the individual treatment effects in this population. The ideal way to estimate the average treatment effect is by conducting an experiment called a randomized controlled trial (RCT). In an RCT, we recruit a representative, random sample of the population of interest and randomly assign the patients to treatment A or B and later measure their outcome. For the patients in the RCT we never observe both the potential outcome of treatment A *and* treatment B so we cannot estimate their individual treatment effects.

However, the outcomes in the treatment A group have the same *distribution* as the outcomes that would have been observed if the treatment B group would have gotten treatment A instead, and vice-versa. The potential outcomes are said to be *exchangeable* between the treatment arms as a consequence of the randomization of the treatment allocation. Because of this exchangeability we can estimate the average treatment effect by comparing the average outcomes of group A with those of group B. Other common terms for 'exchangeability' are ignorability or unconfoundedness. See Figure 2 for an illustration of estimating the average treatment effect in an RCT.

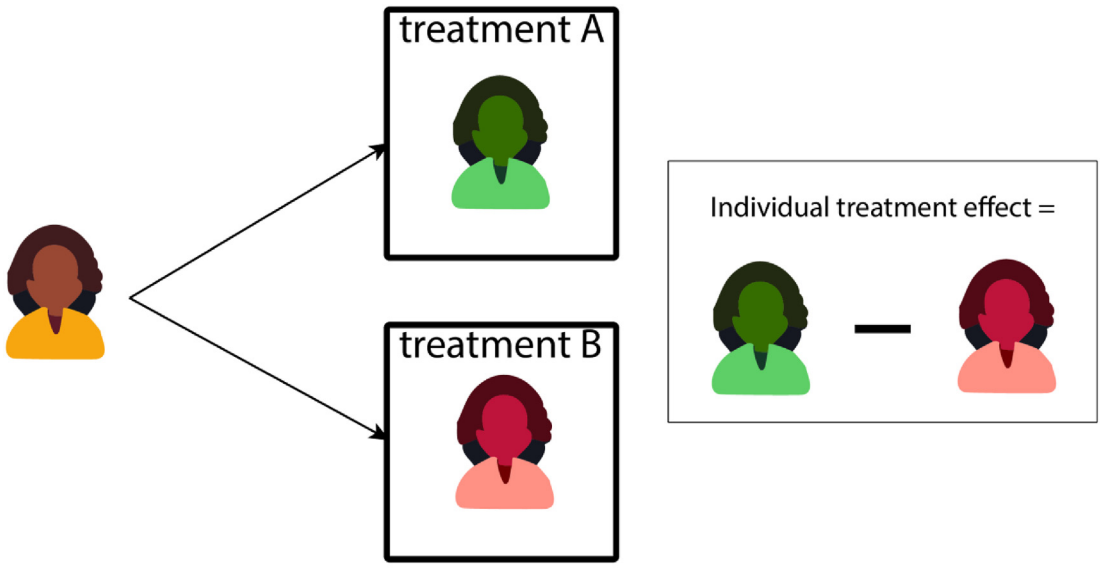
Estimating Treatment Effects Outside of RCTs

Now that we defined what a treatment effect is and how it is ideally estimated in an RCT, a natural question is whether it is possible to estimate treatment effects from non-RCT data such as historical patient cohorts (i.e. real-world data). These datasets are called *observational* to highlight the contrast with the experimental nature of RCTs. The fundamental difference between observational studies

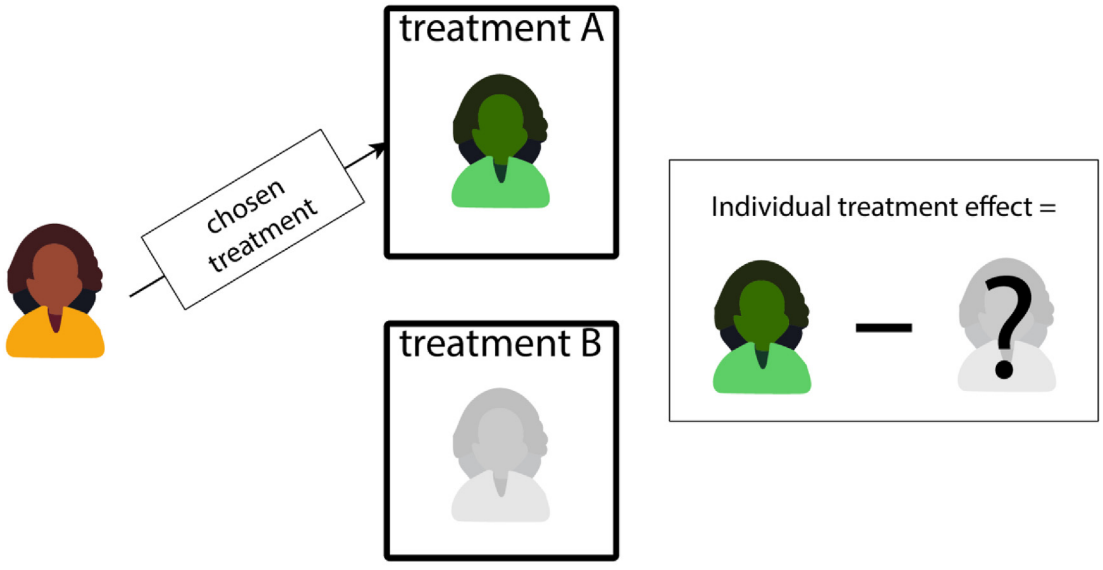
Table 1
Glossary of terms relevant in causal inference

Abbreviation	Term	Description
	causal inference (noun)	field of science that defines causal effects and studies how to estimate them from data
	causal inference (verb)	making an inference (estimate/claim) about a causal question
RCT	randomized controlled trial	experimental study where treatments are allocated by randomization
	observational study	study where treatments are not allocated through randomization but through standard clinical care
ITE	individual treatment effect	difference in potential outcomes between e.g. treatment A and treatment B for an individual patient
ATE	average treatment effect	average of individual treatment effects in a certain population
	exchangeability	when observed outcomes in a treatment group have the same distribution as when the treatment would have been assigned by intervention e.g. as in an RCT
	confounder	variable that influences both the treatment decision and the outcome
	valid adjustment set	set of variables that when adjusted for resolves any confounding and thus yields conditional exchangeability
	identifiability assumption	assumption needed for inferring causal treatment effects, such as "no unobserved confounders" in observational studies
	sensitivity analysis	a computational method to determine the effect of a hypothetical violation of an assumption on the treatment effect estimate
	target trial	the ideal randomized trial for estimating a certain treatment effect, the design of which guides the design of an observational causal inference study

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002



(a)



(b)

Fig 1. Illustration of the individual treatment effect. For an individual patient, the individual treatment effect of treatment A versus treatment B is the difference in potential outcomes that would be observed under treatment A or treatment B (1a). In practice only one potential outcome is observable: the one concordant with the actually given treatment, so the *individual treatment effect* cannot be estimated from the observed data (1b).

and RCTs is that in observational studies, treatments are allocated according to regular clinical care, whereas in RCTs the treatments are allocated by randomization. We explain treatment effect estimation from observational data with a stylized example of stage III non-small cell lung cancer.

Example: Concurrent Versus Sequential Chemoradiation in Lung Cancer

The primary treatment for stage III non-small cell lung cancer consists of both chemotherapy and radiotherapy (i.e. chemoradiation). This combination treatment can be given *sequentially* by first giving chemotherapy and then radiotherapy, or *concurrently* by giving chemotherapy and radiotherapy at the same time. Prior RCTs established that concurrent treatment leads to better overall survival than sequential treatment [6]. However, concurrent treatment is not endured well by patients with lower overall fitness, so for a patient with low overall fitness, sequential treatment may be preferred as the chance of successfully completing the treatment regimen is higher. Patients in higher overall fitness get concurrent chemoradiation more often in clinical practice, but at the same time higher overall fitness leads to better overall survival regardless of the given treatment.

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

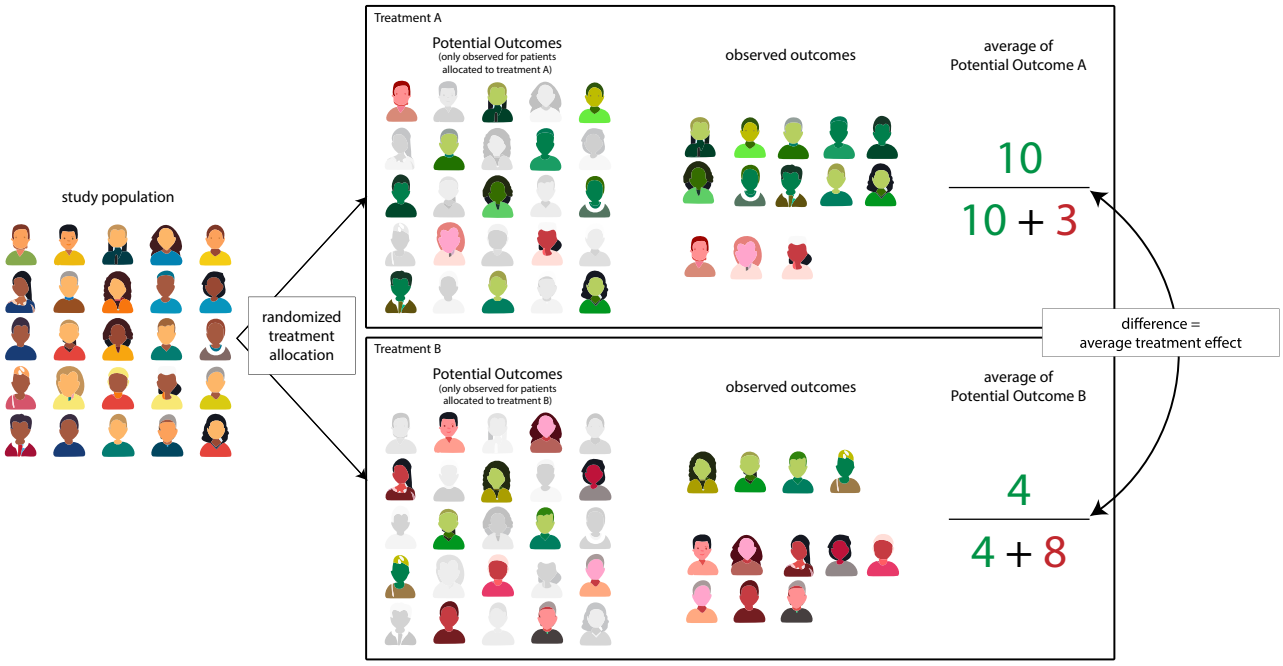


Fig 2. Average treatment effect estimation in a randomized controlled trial. Whereas individual treatment effects cannot be estimated from RCTs as only one potential outcome is observed per patient, the average of the individual treatment effects (i.e. the average treatment effect) can be estimated by comparing the average outcomes in each treatment arm.

This means that the potential outcomes of patients treated with concurrent treatment are not exchangeable with those who got sequential treatment. Imagine forcing all patients who would normally get sequential treatment to get concurrent treatment instead. Because this patient group has lower overall fitness on average, their outcomes would not have the same distribution as those who normally get concurrent treatment.

Confounding: The Central Problem in Observational Causal Inference

In the lung cancer example, overall fitness is a *confounder* of the treatment-outcome relationship because it influences both the treatment decision and the outcome. Confounders extinguish exchangeability and thereby hinder treatment effect estimation. Whereas, in RCTs, the average treatment effect may be estimated because the outcomes in the treatment groups are exchangeable due to the randomized treatment allocation, for observational data, this is not automatically the case.

There is, however, a potential solution. Exchangeability may still be achieved if we are willing to make assumptions about the confounders. In the lung cancer example, assume that overall fitness is the only confounder, and for simplicity that there are only two levels of overall fitness: high and low. Furthermore assume that in both levels of overall fitness, at least some patients got either treatment. With these assumptions, *within strata* of overall fitness, the potential outcomes of patients who had concurrent treatment or sequential treatment are exchangeable. The potential outcomes are thus said to be exchangeable *conditional* on the confounder overall fitness. The average treatment effect for the entire population can be estimated by first estimating the treatment effect in both strata of overall fitness and then taking the average of these estimates weighted by the prevalence of each level of overall fitness.

An important limitation with treatment effect estimates from observational data is that it is always possible that, in addition to overall fitness, there was another confounder that was unknown to the researchers or unmeasured. This unmeasured confounder breaks the exchangeability conditional on overall fitness, which means that the above procedure for estimating the treatment effect yields incorrect results. Unfortunately, in non-experimental data, it is always possible that unmeasured confounders exist and there is no way to test whether this occurs in a particular situation, so treatment effect estimation from observational data relies on *assumptions* regarding the confounders. The assumption of no unobserved confounding is called an *identifiability assumption*.

Graphical Tools in Causal Inference

The relationships between the variables of importance in our example can be depicted in a *causal directed acyclic graph* (DAG) as shown in Figure 3. In a DAG, variables are denoted as nodes and causal relationships between them are denoted with arrows that go from a cause variable (e.g. treatment) to an effect variable (e.g. overall survival).

Next to confounders, there are other types of variables important to treatment effect estimation such as *colliders*. Colliders are variables that are caused by two or more other variables. When stratifying a population based on the value

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

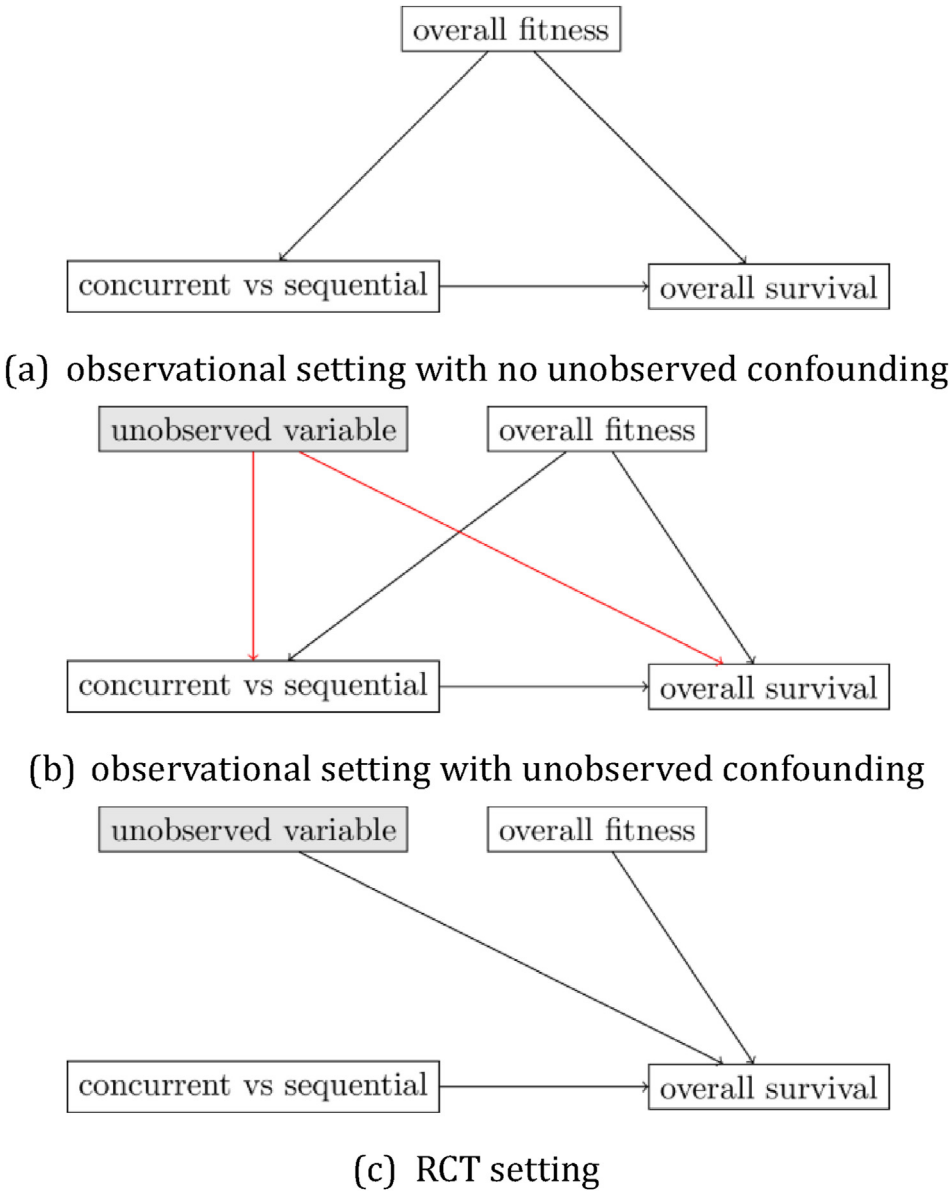


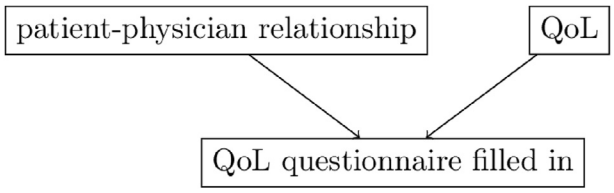
Fig 3. DAG for the stylized example of stage III non-small cell lung cancer, describing the observational setting without (3a) and with (3b) unobserved confounding, and the RCT regime (3c). In the first observational setting (3a), the observed confounder overall fitness influences the treatment decision and the outcome and exchangeability holds conditional on this observed confounder. In the second observational setting (3b) there is also an unobserved confounder so exchangeability no longer holds conditional on the observed confounder 'overall fitness'. In the RCT setting, no variables influence the treatment decision due to the randomization, ensuring exchangeability. DAG: causal directed acyclic graph; RCT: randomized controlled trial.

of a collider, for example by only considering patients with a certain value of the collider variable, or by including the collider variable in a regression analysis, colliders induce correlations between the variables that cause it. For example, consider investigating whether a good patient-physician relationship improves health-related quality of life in cancer patients [7]. It may be that both having a good relationship with the treating physician and having a good quality of life increase the willingness of a patient to fill in questionnaires. Then having a filled-in questionnaire is a collider with at least these two causes, see Figure 4. Let's assume there is no causal effect of relationship quality on quality of life. For some patients who filled in the questionnaire, the reason they filled it in is either having a good relationship or having a good quality of life. Thus in the subgroup of patients who filled in the questionnaire, a good relationship and good quality of life have a *negative* correlation regardless of the absent causal effect.

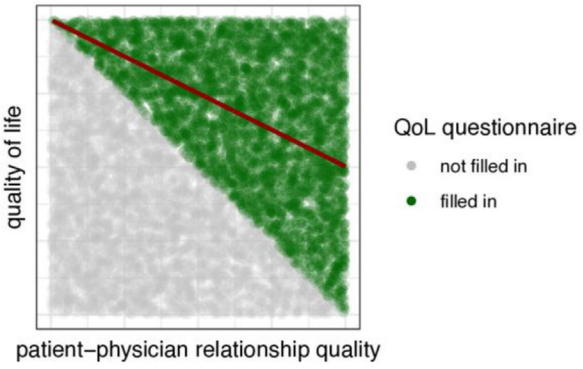
How to Conduct a Causal Inference Study?

Armed with the definition of a treatment effect and exchangeability, we now sketch a general framework for estimating treatment effects.

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002



(a) DAG



(b) simulated data

Fig 4. DAG and simulated data with a collider: QoL questionnaire filled in. 4a: Hypothetical DAG where the quality of the patient-physician relationship has no causal effect on quality of life, but both improve the willingness of patients to fill in QoL questionnaires. 4b: When only considering patients with a filled-in QoL questionnaire (i.e. conditioning on the collider), there appears to be a negative correlation between patient-physician relationship quality and QoL. DAG: causal directed acyclic graph; QoL: quality of life.

Step 1: Define Question and Target Trial

Each causal question includes an intervention. A useful framework for structuring causal inference studies is to imagine the *target trial* that randomizes this intervention of interest. The target trial is the ideal RCT that would allow to answer the research question. The design for this target trial should have a description of the trial protocol, including patient eligibility criteria, treatments/interventions, outcome, follow-up and statistical analysis [8].

Step 2: Gather Data

The next question is whether it is possible to run the target trial. When it is feasible to run an RCT, this is preferable for treatment effect estimation. If there is observational data available from patients who had the treatments of interest, observational causal inference studies may inform the design of the RCT. For example by using an estimate of the treatment effect from the observational study to inform sample size calculations. Researchers should gather previous evidence, most notably existing RCTs on the treatments in question, even if there is for example a slight variation of the treatments between the RCT and the current research question, or the patients from the published RCTs come from a different patient population. Further recommendations for the design and analysis of RCTs are out of scope for this review and for the remainder we assume running an RCT is not possible so the study relies on observational data. We assume access to observational data on patients who underwent the interventions under question.

Step 3: Formalize Assumptions

In the absence of an RCT, formalizing the *identifying assumptions* is crucial to causal inference. This includes carefully analyzing the decision process surrounding the intervention in clinical practice with domain experts and clinicians that actually make the treatment decisions. Relevant pre-existing evidence should also be incorporated in this stage. Preferably, the researchers and domain experts create a DAG with the relevant variables based on the background knowledge. There is no consensus on best practices for constructing a DAG, though some guidance exists. Specifically, “Evidencesynthetis for Constructing DAGs” is a framework for constructing a DAG, starting with a systematic review of studies relevant to the question, and then applying a structured process for creating a DAG based on these studies [9]. Whatever approach is taken, researchers may end up with multiple possible sets of identifying assumptions. Conclusions following these different alternatives should be compared as described later.

Step 4: Causal Inference Method

A crucial question is whether there is a set of observed variables that results in conditional exchangeability (i.e. a *valid adjustment set*), meaning these variables tackle confounding when adjusted for. Algorithms that implement the rules of *do-calculus* [10] can automatically determine whether a valid adjustment set exists from a DAG.

If a valid adjustment set exists, there are many approaches to treatment effect estimation. For low-dimensional discrete adjustment variables and treatment, this can be done by averaging the treatment effect in strata of the adjustment variables as described earlier in the lung cancer example. For other types of variables, often parametric assumptions need to be made. The average treatment effect can then be estimated through outcome modeling including all adjustment variables. Another approach is by creating a prediction model for the treatment based on these variables, called a *propensity score* model. By reweighting the population with the inverse propensity score the average treatment effect may be estimated.

Step 5: Sensitivity Analyses

The identifiability assumption of no unobserved confounding is fundamentally not testable. Instead, to assess the impact of potential violations of this assumption one can formulate *sensitivity analyses* [11]. An example of a sensitivity analysis is to assume there is an unobserved confounder with a certain association with the treatment

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

and the outcome. Using computational methods, the result of this unobserved confounder on the treatment effect estimate can be calculated. By defining a range of potential violations of the identifying assumptions, one can get a range of possible effect estimates. Note that sensitivity analyses again depend on assumptions, which must be grounded in background knowledge.

Step 6: Statistical Inference

Once the main analysis and sensitivity analyses are specified, the statistical inference can proceed. In this phase, "causal inference" is reduced to statistical inference on a causal question. All considerations relevant to statistical inference apply here, such as handling missing data, statistical significance and multiple testing. However, the requirement of conditional exchangeability can complicate the analysis. In the above example of quality of life questionnaires, only considering patients who filled in the questionnaire (i.e. the frequently chosen "complete-cases" approach) induces a negative correlation between the variables of interest. Without going into details, in this example, the (absence of a) causal effect can be recovered by reweighting observations by the inverse probability of having filled in the questionnaire. Such situations require careful causal and statistical consideration.

Step 7: Interpret

The results of the main analysis should be interpreted in the context of the sensitivity analyses. Is the direction of the treatment effect (favoring harm or benefit) consistent in all sensitivity analyses? When researchers had multiple possible sets of identifying assumptions the conclusions of the resulting analyses should be compared. As the identifying assumptions are typically not testable from the data, picking one approach (e.g. one of the possible DAGs) as the best answer is generally not possible. If the results from different approaches are qualitatively aligned this improves credibility of the estimates. Next, the results should be compared with existing RCT data if available in the light of available background knowledge. If the results deviate from the RCT data, is this expected by the background knowledge? For instance, a deviation may be expected when the populations of the observational study and the prior RCTs differ in aspects that may change the effect of the treatment. If the results indicate a stronger or weaker effect of treatment in some subpopulations, this may warrant conducting an RCT to confirm the results.

Opportunities and Challenges for Causal Inference in Oncology

We now turn to what are unique opportunities and challenges for causal inference in oncology. Though causal inference subsumes treatment effect estimation from RCTs, we focus here on causal inference from observational data.

Challenges for Causal Inference in Oncology

Treatment decisions in oncology depend on the overall fitness of a patient, but this is not accurately recorded for research. Thus for causal inference from observational data, methods based on proxy variables [12,13] or instrumental variables [14] may be required in many settings.

When Observational Causal Inference May Be of Particular Value in Oncology

Though RCTs have distinct advantages for treatment effect estimation, we highlight several settings where causal inference from observational data may augment RCT evidence.

A New Biomarker

RCTs are generally designed with the minimal sample size required to estimate the average treatment effect in a certain population. However, for clinical decision-making the treatment effect conditional on patient characteristics is of more value, but this cannot be estimated reliably from the RCTs because their sample size is too small. To individualize treatment decision-making in oncology, there is much interest in discovering *biomarkers* that help stratify patients into those who benefit more or less from certain treatments. Whether a biomarker is based on molecular tests, medical imaging, serum markers or clinical variables, its association with the treatment effect should be studied in observational data before a new RCT that incorporates the biomarker is warranted.

Generalization to Other populations

RCTs may be conducted in populations that do not mimic the real-world population of a practicing oncologist, for instance, because RCTs tend to include younger and fitter patients, or because the RCTs were done in a different country or health system. An important question is whether the RCT estimate is *transportable* to the real-world population [15]. Observational causal inference studies may provide evidence on whether this is likely the case and thus whether the RCT results apply in the real-world population.

Risk Based Treatment Decisions

Treatments are valued by weighing their benefits (e.g. improved overall survival) against the harms and costs associated with giving the treatment. The benefit a treatment can have is limited by the risk of an adverse outcome when deciding not to give (additional) treatment, sometimes called the *untreated risk*. One example where this is relevant is the prescription of adjuvant therapy to reduce the risk of recurrence following breast cancer surgery: some patients have a very low risk of recurrence after surgery and can thus expect very little benefit from adjuvant chemotherapy. Whereas RCTs often estimate the average treatment effect on a *relative* scale, for example with a risk ratio,

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

hazard ratio or odds ratio, the benefit of a treatment is determined by its reduction of the risk of adverse outcomes on an absolute probability scale [16]. A common assumption is that the relative treatment effect is constant in a population, but the untreated risk varies with patient characteristics [17]. With a constant relative treatment effect, the effectiveness of a treatment on an absolute probability scale will vary for patients who have different untreated risks. This means that the cost-benefit trade-off and thus the treatment decision may be different depending on a patients' untreated risk. This principle is applied in prescribing chemotherapy after breast cancer surgery using for example Predict [18], but also in cardio-vascular risk management [19]. Estimating this untreated risk is preferably done in the control arm of an RCT as this is what the untreated risk represents. However, RCTs are generally not sufficiently large to estimate this risk with much granularity depending on several patient characteristics. Also, the RCTs may not have measured all relevant pre-treatment patient characteristics determining the untreated risk. Most risk-prediction models such as Predict were therefore developed in large observational datasets but since they aim to predict outcomes under the hypothetical intervention of not giving (adjuvant) treatment, these approaches require causal inference [20].

Table 2
Table with additional potential applications of (observational) causal inference in oncology

Application	Description
iterative changes	Small changes in treatment protocols such as radiotherapy parameters can improve outcomes but it is infeasible to conduct an RCT for each possible change.
drug repurposing	Medications that have been on the market for different indications are sometimes suspected of having anti-cancer effectiveness. This should be studied in observational data before running a new RCT with the medication.
checking consistency	The experimental setting of RCTs is different from day-to-day clinical care and this may affect how treatments are given, which in turn changes the treatment effect between the RCT and the real-world setting. Observational causal inference studies can investigate the treatment decisions, treatment protocol violations and treatment effect directly in the 'real-world' setting.
rare and late events	As RCTs are bounded in sample size and follow-up time, rare events and late complications of treatments may not become evident from RCTs but require observational monitoring studies to make causal inferences about these effects.
counterfactual questions	*Counterfactual* questions are not answerable directly even from RCTs and require explicit causal modeling. A prototypical counterfactual question is "given that this patient got treatment A and died 1 year later, would they have lived longer if they would have gotten treatment B instead?"

RCT: randomized controlled trial.

Other Applications of Causal Inference in oncology

Prediction-Under-Intervention

Predicting the untreated risk of an outcome is a special case of prediction-under-intervention, where the goal is to predict outcomes under a hypothetical intervention. Instead of providing treatment effect estimates in terms of differences in expected outcomes between treatments, predicting absolute outcomes after each possible treatment may help counseling patients on the different treatment options. Building models for prediction-under-intervention requires large-scale datasets, which may be harder to attain from RCTs than from observational data, again motivating observational causal inference studies.

In addition to the above questions concerning treatment effect heterogeneity, we briefly mention several other use cases of observational causal inference in oncology in Table 2.

Conclusion

RCTs are the best study design for treatment effect estimation as randomization of treatment allocation ensures exchangeability, but RCTs also have important limitations for treatment decision-making in oncology. By understanding these limitations and the possibilities of causal inference with observational data, decision-making in oncology can go further than allowed by the classical distinction that "RCT estimates are causal but observational estimates are associations", and benefit from both RCT evidence and evidence from well-conducted causal inference studies in observational data.

Other Resources

We now provide some additional resources on causal inference. Excellent books introducing the field of causal inference are the Book of Why [21] and What If? [22].

The topic of causal inference is gaining attention in oncology [23,24]. Moodie describes methodological challenges in causal inference typical to oncology, such as having censored time-to-event outcomes, competing risks and time-varying treatments, though focuses the attention mainly on occupational hazards for developing cancer [23]. Van Amsterdam and colleagues describe the use of risk models for treatment decision-making in oncology [24]. They explain how deploying a risk model for treatment decision support is an intervention and thus these models should be developed, validated and deployed with causality in mind [24], and that the current acceptance criteria for risk models by the American Joint Committee on Cancer allow for the deployment of potentially harmful risk models because the criteria lack this causal understanding [25,26].

Finally, the European Society for Medical Oncology released reporting guidance for real-world evidence studies [5]. The reporting guidance contains points about many

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002

aspects of real-world studies for instance to assure data quality and comparability. As many real-world evidence studies target causal questions, our review is complementary by providing a methodological introduction to the fundamentals of causal inference.

Author Contributions

1 guarantor of integrity of the entire study; WA.
2 study concepts and design; all authors.
3 literature research; WA.
4 clinical studies; NA.
5 experimental studies/data analysis; NA.
6 statistical analysis; NA.
7 manuscript preparation; WA.
8 manuscript editing; all authors.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: Partners in the evolution of medical evidence. *Br J Cancer* 2014;110:551–555. https://doi.org/10.1038/bjc.2013.725.
[2] Lewis JH, Kilgore ML, Goldman DP, Trimble EL, Kaplan R, Montello MJ, *et al*. Participation of Patients 65 Years of Age or Older in Cancer Clinical Trials. *JCO* 2003;21:1383–1389. https://doi.org/10.1200/JCO.2003.08.010.
[3] Vinod SK. Decision making in lung cancer – how applicable are the guidelines? *Clin Oncol (R Coll Radiol)* 2015;27:125–131. https://doi.org/10.1016/j.clon.2014.10.008.
[4] Nordon C, Karcher H, Groenwold RHH, Ankarfeldt MZ, Pichler F, Chevrou-Severac H, *et al*. The "Efficacy-Effectiveness Gap": Historical Background and Current Conceptualization. *Value Health* 2016;19:75–81. https://doi.org/10.1016/j.jval.2015.09.2938.
[5] Castelo-Branco L, Pellat A, Martins-Branco D, Valachis A, Derksen JWG, Suijkerbuijk KPM, *et al*. ESMO Guidance for Reporting Oncology real-World evidence (GROW). *Ann Oncol* 2023:S0923753423040188. https://doi.org/10.1016/j.annonc.2023.10.001.
[6] Aupérin A, Le Pechoux C, Rolland E, Curran WJ, Furuse K, Fournel P, *et al*. Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol* 2010;28:2181–2190.
[7] Samuel CA, Mbah O, Schaal J, Eng E, Black KZ, Baker S, *et al*. The role of patient-physician relationship on health-related quality of life and pain in cancer patients. *Support Care Cancer* 2020;28:2615–2626. https://doi.org/10.1007/s00520-019-05070-y.
[8] The target trial. *Causal inference: what if*. Boca Raton: Champan & Hall/CRC; 2020. p. 37–40.
[9] Ferguson KD, McCann M, Katikireddi SV, Thomson H, Green MJ, Smith DJ, *et al*. Evidence synthesis for constructing directed acyclic graphs (ESC-DAGs): a novel and systematic method for building directed acyclic graphs. *Int J Epidemiol* 2020;49:322–329. https://doi.org/10.1093/ije/dyz150.
[10] Pearl J. Causal Diagrams for Empirical Research. *Biometrika* 1995;82:669–688. https://doi.org/10.2307/2337329.
[11] Greenland S. The Effect of Misclassification in the Presence of Covariates. *Am J Epidemiol* 1980;112:564–569. https://doi.org/10.1093/oxfordjournals.aje.a113025.
[12] Miao W, Geng Z, Tchetgen Tchetgen EJ. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika* 2018;105:987–993. https://doi.org/10.1093/biomet/asy038.
[13] van Amsterdam WAC, Verhoeff JJC, Harlianto NI, Bartholomeus GA, Puli AM, de Jong PA, *et al*. Individual treatment effect estimation in the presence of unobserved confounding using proxies: A cohort study in stage III non-small cell lung cancer. *Sci Rep* 2022;12:5848. https://doi.org/10.1038/s41598-022-09775-9.
[14] Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. *The Ann Math Stat* 1940;11:284–300. https://doi.org/10.1214/aoms/1177731868.
[15] Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci* 2016;113:7345–7352. https://doi.org/10.1073/pnas.1510507113.
[16] Murray EJ, Caniglia EC, Swanson SA, Hernández-Díaz S, Hernán MA. Patients and investigators prefer measures of absolute risk in subgroups for pragmatic randomized trials. *J Clin Epidemiol* 2018;103:10–21. https://doi.org/10.1016/j.jclinepi.2018.06.009.
[17] Kent DM, Hayward RA. Limitations of Applying Summary Results of Clinical Trials to Individual PatientsThe Need for Risk Stratification. *JAMA* 2007;298:1209–1212. https://doi.org/10.1001/jama.298.10.1209.
[18] Candido dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, *et al*. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res* 2017;19:58. https://doi.org/10/gbhgpq.
[19] Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, *et al*. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 2019;140. https://doi.org/10.1161/CIR.0000000000000678.
[20] van Amsterdam WAC, Ranganath R. Conditional average treatment effect estimation with marginally constrained models. *J Causal Inference* 2023;11:20220027. https://doi.org/10.1515/jci-2022-0027.
[21] Pearl J, Mackenzie D. *The book of why: the new science of cause and effect*, 1st edition. New York: Basic Books; 2018.
[22] Hernan MA, Robins JM. *Causal inference: what if* 2020.
[23] Moodie EEM. Causal inference for oncology: Past developments and current challenges. *Int J Biostat* 2022. https://doi.org/10.1515/ijb-2022-0056. 0.
[24] van Amsterdam WAC, de Jong PA, Verhoeff JJC, Leiner T, Ranganath R. From algorithms to action: Improving patient care requires causality. *BMC Med Inform Decis Mak* 2024;24. https://doi.org/10.1186/s12911-024-02513-3.
[25] Kattan MW, Hess KR, Amin MB, Lu Y, Moons KGM, Gershenwald JE, *et al*. American Joint Committee on Cancer acceptance criteria for inclusion of risk models for individualized prognosis in the practice of precision medicine. *CA Cancer J Clin* 2016;66:370–374. https://doi.org/10.3322/caac.21339.
[26] van Amsterdam WAC, van Geloven N, Krijthe JH, Ranganath R, Ciná G. *When accurate prediction models yield harmful self-fulfilling prophecies* 2024. https://doi.org/10.48550/arXiv.2312.01210.

Please cite this article as: van Amsterdam WAC et al., Causal Inference in Oncology: Why, What, How and When, Clinical Oncology, https://doi.org/10.1016/j.clon.2024.07.002