An Introduction to Causal Inference for Pharmacometricians
Presented by the ASCPT Family of Journals
Speaker: James A. Rogers, PhD
Moderator: Wei Gao, PhD
May 18, 2023

Audience Questions

1.) **Would using baseline tumor size as a covariate for PK not address the causal relationship issue for PK?**
    a. In the example that I presented, it would not suffice to include baseline tumor burden as a PK covariate only. Since it is the association between exposure and response that is confounded, the problem could only be corrected by including tumor burden as a covariate in the exposure response model.
2.) **Does this mean that a good exposure-response relationship could be just by chance and not necessarily imply a causal relationship to endpoints?**
    a. The point that I was making is not that exposure-response relationships can occur by chance (though that is also technically true), but rather that such relationships can arise purely due to confounding, even in the absence of a causal effect of exposure on the response.
3.) **If we have 2 dose levels and have some (but not perfect) separation of CL and AUC, we could keep them both in an exposure-response analysis where baseline CL and exposure effects both need to be estimated? Of course, with 1 dose level, it would not be possible as they would be perfectly correlated.**
    a. I think I understand the question better now that I have read it slowly – sorry I didn't pick up on the gist of it when you asked it live! Yes, I think what you say is true, but you would need to be able to "separate" the estimates of CL and AUC for every individual, which would seem to require something like a cross over design where each individual receives both dose levels.
4.) **Please give some intuition as to how early adverse events can cause high exposure.**
    a. To first clarify: this phenomenon is not true of exposure as such, but only true of particular summaries of the concentration versus time curve. In particular, the problem is when a summary metric is defined in a manner that depends on the event time (such as Cmax prior to the event). In any situation where the exposure metric (e.g., Cmax) has a decreasing trend (perhaps due to planned down-titration) and where – for "random" reasons totally unrelated to exposure – some subjects have early events and some subjects have later events, the subjects with earlier event times will have higher values of the exposure metric, and the subjects with later event times will have lower values. (The causal dependence of the exposure metric on the event time is an explicit functional dependence by definition.)
5.) **What are some of the common causes of selection bias in your experience like the hospitalization example?**
    a. I can't point to any regular themes in my personal experience, but perhaps it would help to just give another example. One such example is described in our paper

(https://ascpt.onlinelibrary.wiley.com/doi/full/10.1002/psp4.12894). See the subsection in that paper entitled "Detecting collider bias".

6.) **I have a question regarding the slide where you mentioned the potential effect of both eGFR and Age on renal function. Specifically, if the effect of Age is not totally mediated through eGFR, if I remember correctly, you said we should keep both variables regardless of whether their effects are statistically significant or not? Is it true? Could you elaborate on that?**
    a. Yes, that is an accurate reflection of my position. Even apart from any issues related to confounding, an important principle is that one should not confuse "absence of evidence" with "evidence of absence". When a covariate effect is not significantly different from zero, that is simply "absence of evidence" of an effect. "Evidence of absence" exists only when the confidence interval for the effect is sufficiently tight around zero. In scenarios where the simultaneous inclusion of both covariates results in wide intervals for both, the correct conclusion is that the data are insufficient to isolate the two causal effects, and that uncertainty / insufficiency of evidence is only adequately conveyed if the two effects are retained in the model (with wide intervals that include both the null effect and the large magnitude effects that would be clinically consequential).
7.) **For drugs like IOs which don't tend to have exposure-response relationships, how can DAG be used for dose optimization?**
    a. I will assume here that IO refers to immune oncology and not to intraosseous routes of administration. I don't think it is necessarily true in general that there is no exposure-response in immune-oncology. In any case, I think the trastuzumab example that I presented illustrates the value of using DAGs to think carefully about exposure-response in an immune-oncology context.
8.) **If you identify the minimally sufficient covariate set by the DAG in advance, do we need to test for the association between these covariates for the exposure and outcome before conditioning them in the final multivariate model? Or we can ignore the bivariate tests and condition on all of them?**
    a. I understand this question a bit better now that I have read it – sorry I didn't fully grasp the meaning at the time. In cases where there are relatively few candidate covariates, I would simply condition on as many of them as necessary to block the most likely backdoor paths in the DAG. In cases where inclusion of a large number of covariates results in model instability, I would favor sparse regression / regularization approaches (such as using Spike and Slab priors in the Bayesian setting). I would not advocate testing-based approaches such as the one you mention because failure to reject those tests might simply arise from insufficient data.
9.) **Can you expand on how the current thinking would have changed our thinking about the TOGO study and helped us better predict the outcome of the HELOISE study?**
    a. I don't want to claim that using causal inference framework would necessarily result in a different approach to those studies or analyses. As I said in the talk, I believe that the FDA authors in the paper had a very nuanced understanding of the issues. My goal was not to refute anything that was done, but rather to promote general conceptual frameworks that are likely to lead to that sort of nuanced thinking.

10.) **Thinking about using tumor growth kinetics (i.e., rate of regrowth after nadir) as a predictor of overall survival -- would that fit a causal relationship between tumor kinetics and overall survival? I'm finding it difficult to think of how to randomize a patient to rate of regrowth as it is not known as randomization.**

a. Good question! It is harder to go through this exercise than I let on in the talk :-), but I think it is worth trying.

Some general remarks: I would say that with respect to the "treatment" variable, the target trial principle is trying to ensure two things: 1. That we are not accidentally "borrowing information from the future" and 2. That we are being sufficiently precise (as precise as we might be in a protocol synopsis) in defining the explanatory variable of primary interest.

With regard to the first concern: Depending on how you do it, there can be a risk of "using data from the future" to predict an instantaneous rate and then falsely proceeding as if you could know that instantaneous rate at any given time. It depends on the details, but that is something worth thinking about.

With regard to specificity: it is worth thinking about the precise nature of the regrowth, e.g., is it regrowth as measured by size (SLD or volume) of target lesions vs non-target lesions vs new lesions, does the location of target/non-target lesions matter, are talking about average vs maximum rate of change across target lesions, etc. The point is that "rate of regrowth" could correspond to a variety of things. It may be as simple as rate of growth in SLD of target lesions, or it may not be. Again, it will depend on the particulars, but this is the sort of question that the target trial criteria is encouraging us to think through.

One thing you *don't* need to worry about is whether you could actually manipulate the relevant pathways in practice. In the hypothetical target trial world, we are free to disregard that mere "implementation detail" :-). The important thing is whether you know precisely what you would like to manipulate in that hypothetical trial.

For a humorous treatment of related issues, see Miguel Hernan's paper, "Does water kill? A call for less casual causal inferences".

11.) **It seems causal inference outperforms many standard statistical approaches for e.g., covariates selection, are there cases that are opposite? Any limitation of causal inference applications on PMX?**

a. As I indicated during the Q&A, I think the best way to understand the value of causal inference is not in terms of enhancing statistical performance. Statistical performance is a statistical issue and causal interpretability is a causal issue. In my view it is better to keep the two aspects separate. I don't think there is any limitation per se in PMx, but I there are a lot of problems in PMx where it is still very difficult to apply causal inference frameworks, for example when exposure varies dynamically, as in a PKPD model.

12.) **Do you think causal inference can be applied in the field of bacterial infections to obtain a definitive PK/PD target for antibiotics? Since most of these targets that are in use nowadays are derived from animal studies and we're not sure if it can be translated in humans.**

a. One thing that we should be clear about is that we can never "prove causality". Causal inference has a more modest goal: to help us realize the most likely sources of bias in

our analyses and to correct for those biases as much as possible. Or, in the context you mention, to help us think about the most likely reasons why the animal–human translation might fail, and to suggest ways of making that translation less fragile. It won't ever be definitive, only suggestive. (But suggestive evidence is better than nothing.)

13.) **Curious about the intersection of surrogacy and causality. There are some surrogates that you cannot randomize to, yet there are still ways to demonstrate surrogacy. What is the correct way to think about this?**

a. Good question. I can't claim any special expertise on this, but my sense is that for questions about surrogacy, the area of causal inference known as "causal discovery" would be especially helpful. The examples I covered are more like "causal deduction": if we hypothesize a particular set of causal connections, what does that imply about how we should interpret the associations in our data and what might lead to those associations being biased? Causal discovery addresses the reverse problem: given the observed joint distribution of some variables, what causal connections might be consistent with those observed associations? That seems more relevant for questions about surrogacy, which I think are ultimately arguments about what the causal pathways are that mediate between treatment and response. In "causal discovery", there is no need to distinguish between treatment / exposure and outcomes, and the target trial considerations are therefore irrelevant.

14.) **Are mechanistic QSP models also causal?**

a. Anything that is truly "mechanistic" will have causal interpretability, but just because a modeler has labeled a model as "mechanistic", that is no reason to take the modeler at his or her word. I take the word "mechanistic" to imply the possibility of control over the system: if we move the dial "X" this much, we know how much "Y" and "Z" will move, etc. That ability to understand the consequences of an intervention is the hallmark of a causal model. In that sense I would say that every mechanistic model is causal, but see a side note from Wei below. The reverse is not necessarily true, since "mechanistic" generally seems to refer specifically to *tight* control over the system, without much random variation in the outputs, and this is not necessarily implied for causal models.

(Side note: the term "mechanistic" originally arose – I think – back in the days of Newtonian "mechanics", which was perhaps a time when the machine analogy made more sense: not all our machines these days can be so tightly controlled!)

(Side note from Wei: Mechanistic QSP models intend to be causal. However, most existing QSP models have components where causal interpretability is not ensured, due to lack of biological knowledge or direct measurement (e.g., tissue biomarker concentrations) and many other reasons. A few model reduction techniques have been published to ensure the right level of balance of parsimony and granularity in mechanistic/QSP models.)

15.) **Yesterday in an ACCP MIDD webinar they referred to causal inference for approval based on biomarker, could you elaborate on how this is done?**

a. That is interesting but unfortunately, I did not see that webinar, so I am unable to comment.

16.) **I'm curious about the specific applications of causal inference in pharmacometrics? Is it a fundamental skill set for pharmetrician?**

a. As I indicated during the talk, I believe it is fundamental. I would teach basic causal concepts at around the same time that we teach the basics of conditional probability. It is just as fundamental, if not more. Richard McElreath's book, "Statistical Rethinking" (https://www.amazon.sg/Statistical-Rethinking-Bayesian-Course-Examples/dp/036713991X) is an excellent introductory text for weaving those concepts together. I think that text or something like it would make sense at the "100 level" in pharmacometrics programs.

17.) **How does the approach to causal inference change with model complexity?**

a. Depending on the type of complexity, it might not change at all. It's hard to say without considering a specific example.

18.) **The examples you provided don't deal with time-varying treatments or treatment - confounder feedback. How do DAGs accommodate problems with those features?**

a. Time-varying treatments don't really present a distinct challenge if they are static (non-adaptive) regimens. On the other hand, treatment confounder feedback (e.g., initial dose —> AE → reduced dose → lower risk of AE) is a more serious challenge. It is probably preferable in that case to use single-world intervention graphs (SWIGS) rather than simple DAGs. Personally, I haven't yet grappled seriously with these problems, but I am looking forward to attempting it in the coming years. (I would recommend getting comfortable with DAGs and simpler problems first.) Christian Bartels has published some good initial remarks on this: https://figshare.com/articles/online_resource/An_attempt_to_derive_g-computation_for_longitudinal_data_from_sequential_conditional_exchangeability/20406687.

19.) **Sometimes we aren't trying to estimate causal effects so much as we are trying to make predictions in new populations or contexts. Does causal inference have anything to say about that type of problem?**

a. This is the issue of transportability that I mentioned during Q&A. Pearl and Bareinboim 2014 (https://arxiv.org/pdf/1503.01603.pdf) is a good reference, albeit not an easy read.