Brief Causality Checklist for Pharmacometricians

Andrew M Stein, 2021
https://sites.google.com/site/andrewsteinphd/home

Table of Contents

Checklist

Introduction

1. Scoping: draw a causal graph

2. Explore Data

3. Plan analysis to adjust for confounding factors

4. Diagnostics and tests for confounding

Figures

References

Checklist

1. Scoping

- Draw a causal diagram by reviewing the literature and collaborating with the team.
    - If you have a *non-randomized dose-escalation study*, think about how cohorts may differ and how this might impact exposure and response
    - If you are *pooling multiple studies*, think about how the study could potentially impact exposure and response. For example, if patients in the different studies received different doses and had different baseline characteristics, this must be accounted for.
- Identify all forks and feedback loops.
    - *Forks*: plan to collect confounding covariate if possible. If not possible, consider how missing this variable could impact interpretation of analysis
    - *Feedback loops*: consider how feedback loops could impact interpretation of longitudinal PKPD models

2. Data Exploration

- Create a pairwise comparison plot for all covariates and response data.
- Enter your Causal Graph into DAGitty and generate exploratory plots that check the "Testable Implications" http://www.dagitty.net/dags.html#

3. Analysis Plan

- Stratify for all covariates that are forks
    - *Categorical Covariates*: use one regression coefficient for each category
    - *Continuous Covariates:* consider different functional forms for covariate (both linear and nonlinear)
- For feedback loops (including dose adjustment based on safety/efficacy response), identify a simpler exposure-response analysis that uses assigned dose or first-dose exposure metrics as a sensitivity analysis to see that you get similar results to a longitudinal PKPD model. If the results differ significantly, there may be confounding.
- Consider case-control analysis, if there is a control arm.

4. Diagnostics and tests for confounding

- *Continuous response data*: plot residuals for exposure vs residuals for response. If there are trends, there may be confounding.
- *Categorical response data*: plot exposure residuals (vs time, value and/or dose) stratified by response category. Any trends or biases may indicate confounding.
- *Visual predictive check.* Perform a visual predictive check of the exposure-response relationship, stratified by dose group.

Introduction

In pharmacometrics, when we build models of the relationship between exposure and response, we almost always assume our models can be interpreted in a causal framework, such that changing exposure (by changing the dose regimen) will change the response. However, there can be factors that correlate with both exposure and response that can confound this interpretation (c.f. Simpson's Paradox) and it's important we check for these factors as best we can. For example, many PD-1 and PD-L1 inhibitors demonstrated this issue, where if one only looks at steady state exposure and at a single dose, one may detect an exposure-response trend -- however the trend is due to the fact that responders exhibit a reduction in clearance, and therefore have greater exposure (Liu et al., 2017). Interpreting the exposure-response analysis as if the exposure caused a greater response rate can lead to erroneous conclusions and poor decisions.

Ultimately, biology is complex and there can always be an unmeasured gene or protein that confounds our analysis by independently affecting both exposure and response. It is impossible to prove there are no hidden confounding factors and that our analysis can correctly be interpreted as causal. In making decisions, there are tradeoffs between certainty and practicality (Nedelman et al, 2007). Practically, if we can be reasonably confident that there won't be any confounding factors that dramatically alter our findings and that the consequence of errors in our predictions are acceptable, then our teams may choose to act as if our model assumptions are correct and take the risk in using our model results to inform dose-regimen selection.

There are a few different schools of thought on causality:

- Potential Outcomes (PO) - Donald Rubin
- Direct Acyclic Graphs (DAGs) and do-Calculus - Judea Pearl

In additional, there is the Target Trial, which uses both DAGs and PO, developed by Miguel Herman and James Robins

Many descriptions of their work are fairly technical and not targeted for pharmacometricians and drug developers. In this document, we take the most practical lessons from the causality literature and propose how they can be applied to exposure-response modeling to support drug development. We cite specific examples when they are available.

1. Scoping: draw a causal graph

Drawing a causal graphical of how dose, exposure, biomarkers and response (both safety and efficacy) relate to each other can help with confounder identification (see Figure 1 for mini-examples). In the graph, arrows point in the direction of causality. It is particularly important to identify two network structures

- **Forks:** a covariate that affects both exposure and response through at least two paths.
- **Feedback loops**: where response can impact exposure or dose in the future.

Discussing the graph with team members is useful as a way to build shared understanding. It is also useful for teaching others about the drug in the context of the relationships between dose, exposure, biomarkers, and response.

In developing these graphs, it is critical to include not only all the known pharmacology, but also to think carefully about the study design and the pooling approach. For example, if you have a non-randomized dose escalation trial (as is typical for Phase I Oncology trials), it's important to consider whether the different dose cohorts may contain different patients. If signs of efficacy are observed early on, could that make investigators more likely to recruit a larger number of patients for the study? For pooling together multiple studies, the study should be treated as a covariate and one should think about how study might impact exposure, response, and other covariates. The study might be a fork, especially if the patient populations differ between studies.

In this guidance, we recommend drawing the causal graph and including all cycles. The causality literature focuses on directed acyclic graphs (DAGs) with no cycles because they permit a mathematical, causal analysis. However, we believe the greatest value of the graph is in building understanding and in identifying potential issues (e.g. forks, feedback loops), both of which can be identified whether or not the graph has cycles. Furthermore, acyclic graphs often do not represent the true underlying the biology; at equilibrium, it's possible for multiple elements of a causal graph to simultaneously affect each other (Imbens 2020, Sec 4.3)

2. Explore Data

Look at all pairwise correlations between covariates, and between the covariates and response.

- Lack of correlation means between two variables there is not a direct causal relationship between the two, and that there is not a shared, common cause for both.
- DAGitty does some automatic analysis of directed acyclic graphs to determine which relationships are testable. Generate exploratory plots that evaluate these relationships. http://www.dagitty.net/dags.html#

3. Plan analysis to adjust for confounding factors

Plan for how you'll account for any forks or feedback loops.

Forks: Collect data on this covariate and stratify the analysis by this covariate.

- Categorical covariates
    - Exploratory plots: facet by category
    - Analysis: one regression coefficient for each category
- Continuous covariates
    - Exploratory plots: scatter plots of covariate vs response

    - Analysis: can try linear regression, but should also explore other functional forms
        - Can we say anything more specific?
- If it's not possible to collect data on the confounding factor, consider simulating the study to understand how this factor might impact the interpretation of the analysis.

Feedback loops: There is risk of confounding in any longitudinal PKPD model, where response might impact dose (through dose reductions due to adverse events) or exposure (e.g. when response affects clearance).

- If you are confident you have the right causal model, a longitudinal PKPD model may be able to account for this confounding.
- However, there are always unknowns, and a good check of the analysis is to use the assigned dose or the exposure after the first dose to check that this gives similar results.

Case Match Control can be used to adjust for confounding factors

- This approach requires having a control arm to draw from.
- For example in Yang et al., 2013, they compared Kaplan-Meier curves, stratified by exposure quartiles, of metastatic gastric cancer patients receiving chemotherapy + trastuzumab. The patients with the lowest quartile of trastuzumab exposure had poorer survival. But these patients were also more likely to have lower ECOG scores, more metastatic sites, etc. The control arm of the study (chemotherapy alone) was used to identify an equivalent set of patients with similar baseline characteristics to the trastuzumab patients in the lowest exposure quartile. When the matched control subset was compared to the lowest quartile trastuzumab patients, the Kaplan-Meier curves overlapped, indicating that these baseline covariates confounded the exposure-response interpretation.
- In Yang et al., 2013, after accounting for all baseline covariates, they still saw an E-R and the hypothesis that a higher dose of trastuzumab could give better response was still possible (especially since higher doses were thought to be safe). Another trial was initiated (Shah et al., 2017) and in that study, no OS benefit of a higher dose was observed. This suggests that even after the case control analysis, there were additional confounders that were not identified by Yang et al., 2013.
- There are different ways to do the matching, such as Propensity Score Matching and Mahalanobis Score Matching

4. Diagnostics and tests for confounding

1. Visual predictive check. Stratify by dose group. Any mischaracterization at low or high doses is a sign that there may be general issues with the model.

2. Residual correlation

- *Continuous response variables:* plot residuals for exposure against residuals for response. If you see trends, there is risk of confounding
    - Example: Nedelman et al., 2007 on oxcarbazepine. No confounding found.

- *Categorical response variables:* look at exposure residuals (vs time, vs concentration and/or vs dose) stratified by response. If you see trends, there is risk of confounding
    - Example: Liu et al., 2017, on nivolumab. Response to drug led to higher exposure at steady state, but not at first dose

3. Two stage modeling

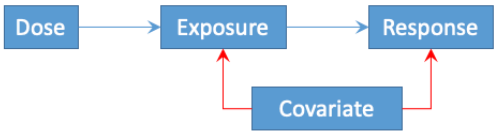
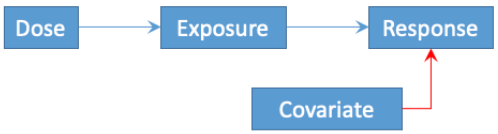
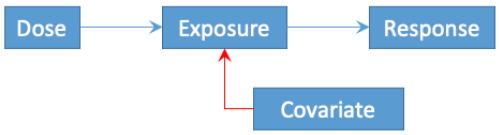
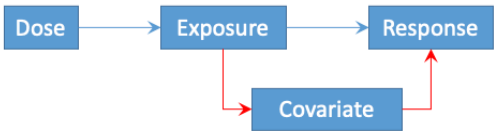
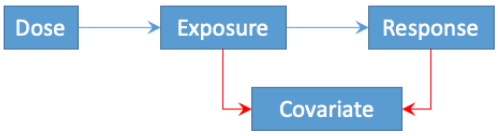
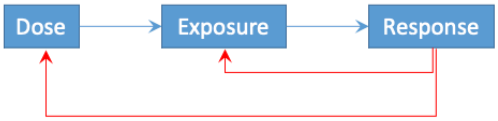
Fit two models. 1) ordinary least squares model for exposure and other covariates on response. 2) two stage model; first, fit model for dose-exposure relationship, then fit a second model for the exposure-response relationship. Compare the parameter estimates from the two analyses. If they're different, there is a risk of confounding. Use Hausman's test to assess this risk. [Nedelman et al., 2005.]

4. Simulate a confounder and refit

Simulate a confounding variable that correlates with both exposure and response. Refit the model and see if the results change. If the conclusions change a lot, then there is risk of an unknown confounder. [Nedelman et al., 2007]

Figures

Figure 1: Types of Covariates + Feedback Effect

Fork		You **must stratify** by covariate, otherwise the E-R relationship you estimate will not be causal
Independent Cause		It **may help to stratify** by covariate in improving the precision of the E-R relationship
Unnecessary		It is **useless to stratify** by covariate in estimating the E-R relationship. The covariate may help if exposure is imperfectly known (e.g. missing PK)
Mediator		You **must not stratify** by covariate if you are interested in the effect of exposure on response. But, if you want to establish the covariate as a surrogate, you must adjust for it.
Collider		You **must not stratify** by covariate
Feedback		You **must think** about the consequences of the feedback loop(s) and choose the appropriate analysis. Using dose or exposure only at day 1 may be done to avoid confounding.

References

- Imbens (2020). https://arxiv.org/abs/1907.07271
- Liu, C., et al. "Association of time-varying clearance of nivolumab with disease dynamics and its implications on exposure response analysis." *Clinical Pharmacology & Therapeutics* 101.5 (2017): 657-666.
- Nedelman, Jerry R. "On some “disadvantages” of the population approach." *The AAPS journal* 7.2 (2005): E374-E382.
- Nedelman, Jerry R., Donald B. Rubin, and Lewis B. Sheiner. "Diagnostics for confounding in PK/PD models for oxcarbazepine." *Statistics in medicine* 26.2 (2007): 290-308.
- Rogers, James A., Hugo Maas, and Alejandro Pérez Pitarch. "An introduction to causal inference for pharmacometricians." CPT: Pharmacometrics & Systems Pharmacology 12.1 (2023): 27-40.
- Shah, Manish A., et al. "HELOISE: Phase IIIb randomized multicenter study comparing standard-of-care and higher-dose trastuzumab regimens combined with chemotherapy as first-line therapy in patients with human epidermal growth factor receptor 2–positive metastatic gastric or gastroesophageal junction adenocarcinoma." Journal of clinical oncology 35.22 (2017): 2558-2567.
- Wiens, Matthew R., Jonathan L. French, and James A. Rogers. "Confounded exposure metrics." CPT: Pharmacometrics & Systems Pharmacology 13.2 (2024): 187.
- Yang, Jun, et al. "The combination of exposure-response and case-control analyses in regulatory decision making." *The Journal of Clinical Pharmacology* 53.2 (2013): 160-166.

Acknowledgements

This work was done in close collaboration with our Exposure-Response guidance team, including Guenter Heimann, Thomas Dumortier, Oliver Sander, Yu-Yun Ho, Xinrui Zhang, Alison Margolskee, and Maria Savelieva-Praz. Mike Perry provided an economist’s perspective on the PO vs the DAG approach, and explained the limits of all tools in estimating causal effects in a non-randomized setting.