Casual Methods in the Service of Good Epidemiological Practice: A Roadmap

Maya Petersen, University of California, Berkeley

Available at: https://works.bepress.com/maya_petersen/69/
Causal Methods in the Service of Good Epidemiological Practice: A “Road Map”

Maya Petersen
Divs. of Biostatistics and Epidemiology
School of Public Health
University of California, Berkeley

Society for Epidemiological Research, Boston, MA, June 21, 2013
We must ask causal questions

• Not all questions are causal, but many are...
  – How things work
  – How best to intervene to improve health

• Answering causal questions is HARD
  • Formal causal frameworks to the rescue...

DAGs!
Counterfactuals!
Marginal Structural Models!

Maya as Grad Student
Here Be Dragons....

• Overconfidence in assumptions
  – I can draw a causal graph
  -> My graph is useful! My graph represents reality!

• Misinterpretation of results
  – My observational analysis now replicating a trial...
  -> Why does my trial not match my analysis results?

• Complexity obscuring common sense
  \[
  E(Y(j + m)\overline{A(j-1)}_{\overline{a}(j)}|\overline{A}(j-1) = \overline{a}(j-1), S(j)) = E(Y(j + m)\overline{a}(j-1)_{a(j)}|S(j)\overline{a}(j-1))
  \]
  -> ???
Formal Causal Frameworks are a Tool

- What are they useful for?
  1. Make uncertainty explicit
  2. Frame better questions
  3. Understand assumptions
  4. Improve study design
  5. Design statistical analyses that come closer to answering our questions
  6. Interpret results appropriately
How? A roadmap...

1. Specify **Causal Model** representing **real** background knowledge
2. Specify **Observed Data** and link to causal model
3. Specify **Causal Question**
4. **Identify** - Knowledge + data sufficient?
5. Commit to an **estimand** as close to question as possible, and a statistical model representing **real** knowledge.
6. **Estimate.**
7. **Interpret** Results.
1. Specify Causal Model

• **Good Epidemiologic Practice**: Use background knowledge
  – Confounding, selection bias, measurement error....

• **Causal Framework**: A formal language for expressing it
  – Makes knowledge more powerful

• **Moving beyond intuition**...
  – We take for granted that intuition is not enough to help us make sense of complex data- that is why we have statistics
  – Same for causal reasoning. Intuition breaks down...
Causal Models Express Knowledge

• Example: Causal Graphs/Structural Causal Model

\[
\begin{array}{c}
\text{CD4 Count} \\
\text{Antiretroviral Therapy} \quad \text{Mortality}
\end{array}
\]

• Many formal causal frameworks:
  – Counterfactual/Potential Outcome, Non-Parametric Structural Equation Models, Single World Intervention Graphs, FFRCISTG, Decision Theoretic….
Causal Models Express Uncertainty

• We have real knowledge, **but it is limited**
  – Which variables affect each other?
  – Unmeasured factors?
  – Functional form of those relationships?

![Diagram showing causal relationships between variables including Homelessness, CD4 Count, Antiretroviral Therapy, and Mortality. An unmeasured factor (U) influences these variables.](image-url)
All causal models are not wrong...

- Causal models should represent real knowledge
  - We often need to make additional assumptions in order to make progress
  - Keep this process separate

- Beware the temptation to oversimplify
All causal models are not wrong...

• Causal models should represent real knowledge
  – We often need to make additional assumptions in order to make progress
  – Keep this process separate

• Beware the temptation to oversimplify

**Yuck!!**

• Use a simpler model? Not unless your knowledge justifies it
  – Can use a different language to express your knowledge and lack thereof...
2. Specify **Observed Data** and its link to the Causal Model

- Causal model describes the set of processes that may have given rise to the observed data
- Implies the set of distributions possible for the observed data: The *Statistical Model*
All statistical models are not wrong...

• Statistical Models should represent real knowledge

• **Good Epidemiologic Practice**: Choose an “appropriate” statistical model...

• **Causal Framework**: Can help us choose statistical models that reflect our uncertainty
  – Often put **no restrictions** on the joint distribution of our observed variables: Non-parametric
  – Sometimes we do have real knowledge... By all means use it!
1. **Causal Model**
   Representing background knowledge and uncertainty

2. **Observed Data**
   Process that generated the data described by the causal model

---

**Statistical Model**
Possible distributions for the Observed data
3. Specify Causal Question

- **Good Epidemiological Practice:** State your scientific question clearly
- **Causal Framework:** A language to translate your question into a formal query
  - Ex: Using counterfactuals
  - Forces you to be explicit about exactly which “experiment” (change(s) to the system of interest) would let you answer your question
Counterfactual Interventions

- Example: Simple static interventions

- Counterfactual Intervention: Antiretroviral Therapy
  - CD4 Count
  - Counterfactual Mortality with ART

- Counterfactual Intervention: Antiretroviral Therapy
  - No
  - CD4 Count
  - Counterfactual Mortality without ART
Very flexible - Make the target quantity match the question

- Example: Dynamic Regimes

If CD4 < 350

Antiretroviral Therapy

CD4 Count

Counterfactual Mortality

If CD4 < 500

Antiretroviral Therapy

CD4 Count

Counterfactual Mortality
Many more options!

- Longitudinal Effects
  - Effect of Antiretroviral Therapy (ART) over time
- Direct Effects
  - Effect of ART not mediated by CD4 changes
- Missing Data and Censoring
  - Interventions to prevent losses to follow up
- Stochastic Interventions
  - Shift the delay from diagnosis to ART start
- Can also summarize the distribution of the counterfactual outcomes in lots of ways...
4. Identify.
Knowledge + Data Sufficient?

• Our question is stated of in terms of counterfactual quantities
  – What would things have looked like under different conditions

• Can we translate our target causal quantity into a statistical parameter?
  – A function of the observed data distribution, or “estimand”

• If so, which estimand?
Intuition only gets you so far....

• **Good Epidemiological Practice: Control for confounding**

• **Causal Framework**
  – Which variables to “control for”
    • Adjusting for some pre-treatment variables is harmful!
  – Even when no single adjustment set is sufficient, your target parameter may still be identified
    • Longitudinal effects and time dependent confounding
    • Effect mediation and direct effects
    • Instrumental variables...
    • Many more.....

• **Many of these estimands are not intuitive**
A Roadmap....

1. Causal Model
2. Data
3. Question
   Translate the scientific question into a formal causal quantity (using counterfactuals)
4. Identified?
   Are data sufficient to answer causal question under model assumptions
5. Estimand
   Equal to the target causal quantity

Y
In many (most?) cases, data + model are NOT sufficient

• **Good Epidemiological Practice**
  – Get more data
  – Do the best job you can with data you have, and understand limitations

• **Formal Causal Framework:**
  – Which data/how to change design
  – What additional assumptions are needed: “convenience assumptions”?
  – Estimand that comes closest to answering our question with the data we have
A Roadmap....

1. Causal Model
2. Data
3. Question
4. Identified?
5. Estimand Equal or as close as possible to the target causal quantity

Convenience assumptions
6. Estimate

• Causal framework got us here...but this step is purely statistical
• One estimator is no more causal than another
  – IPW not “more causal” than logistic regression
• Estimators have different statistical properties
  – Bias, variance, robustness, consistency....
  – Given a statistical model and estimand, we can study these properties and pick the best for our problem
Causal inference can be hard, but so can statistics

• **Good Epidemiological Practice**: The question should drive the statistical analysis

• **Causal Framework**: Knowledge + data + question often = hard statistical problems

• Many estimands **do not** correspond to a coefficient in a single regression
  – Complexity is not an end in itself
  – Complex statistical methods are sometimes necessary to get the best possible answer to real world questions
A Roadmap....

1. Causal Model
2. Data
3. Question
4. Identified?
5. Estimand
6. Estimator

Statistical Model
Convenience assumptions

Y
N
7. Interpret Results

• Statistical Interpretation?
  – With correct statistical model and good estimator

• Causal Interpretation?
  – If causal model + convenience assumptions are true
  – Makes explicit what these are

• Impact of a real world intervention?
  – Many more assumptions...
    • Stability of conditions and population
    • Interference
    • Correspondence between hypothetical and real world intervention...
A Roadmap....

1. Causal Model
2. Data
3. Question
4. Identified? → 5. Estimand
   - Y
   - N
   → Statistical Model
   → 6. Estimator
   → Convenience assumptions
   → 7. Interpretation

A Roadmap...
"As we know, there are known knowns; there are things we know we know. We also know there are known unknowns; that is to say we know there are some things we do not know. But there are also unknown unknowns -- the ones we don't know we don't know...."

• Causal frameworks used well
  – Keep us uncomfortably aware of how little we know
  – While not freezing us into panicked inaction

• A systematic approach to using them can help epidemiology improve public health
Acknowledgments

• Mark van der Laan
• References too numerous to list
• Based on the work and insights of many
  – Judea Pearl
  – James Robins
  – Miguel Hernan
  – ....many others
• My work supported by
  The Doris Duke Charitable Foundation
  Clinical Scientist Development Award