"Causal Cascade": Not all causal relationships are simple and direct

Missing Data: Just ignore it…it’s Missing!

Donald E. Stull, PhD
RTI Health Solutions
dstull@rti.org

13 November 2018

Does Treatment Make a Difference/Have an Effect?

• Typically, analysis of RCT data will look at mean differences between treatment arms on some outcome (oftentimes controlling for other factors) or whether treatment has an effect on that outcome

• If there is no significant difference (or a small difference) between treatment arms on the outcome of interest, does that mean that treatment has no effect?

• There may be a “causal cascade” of effects such that treatment significantly affects one variable which affects another variable which affects the outcome of interest
  – Thus, treatment may have a significant *indirect* effect on the outcome of interest
Path Analysis

• A long history:
  – Sewall Wright (1918), a geneticist
• Recognizes and explicitly models indirect pathways between variables
• "Decomposes" correlations between variables (i.e., total effects of one variable on another) into direct and indirect causal paths

Example – Oncology

• Treatment vs placebo
• EORTC QLQ-C30
• N ~ 600 patients
• Q: Does overall QoL differ significantly by treatment?
• Simple t-test of QLQ-C30 QoL mean score by treatment:
  \( \text{Tx} = 59.6 \)
  \( \text{Plac} = 66.9 \)
  \( t = -3.71 \)
  \( p < 0.001 \)

  ➢ Small, statistically significant difference (but, there are >600 patients)*

  ➢ But why is treatment arm showing poorer QoL?

* From the EORTC website (accessed 11 September 2018): "When comparing scores, one should take into account that statistically significant differences do not necessarily imply clinically relevant differences and vice versa. For the QLQ-C30, a change in any scale of at least 10 points is considered to be clinically relevant (Ref: Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality of life scores. J Clin Oncol 1998; 16 (1): 139-144)."

Path Analytic/Causal Cascade Approach

• We recognize that treatment may influence intervening (mediating and/or moderating) variables

• We rely on logic, theory, mechanism of action, clinical and patient experience to consider how treatment plays out:

  – Treatment → diarrhea → fatigue → physical functioning → QoL
• Simple t-test of QLQ-C30 Diarrhea mean score (on scale of 0-100) by treatment:
  \[ \text{Tx} = 41.1 \]
  \[ \text{Plac} = 5.2 \]
  \[ t = 15.82 \]
  \[ p < 0.001 \]

Now we begin to see that there may be intervening variables between treatment and QoL that can modify the relationship.

Note: Controlling for baseline Eastern Cooperative Oncology Group status.
Note: All significant paths are \( P < 0.05 \).
Tx: 1 = treatment, 2 = comparator; n.s. = not significant.
Decomposition of Effects

<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>Explanatory Variable</th>
<th>Direct Effects</th>
<th>Indirect Effects</th>
<th>Total Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global QoL</td>
<td>Treatment</td>
<td>0.02 n.s.</td>
<td>0.12***</td>
<td>0.13***</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Treatment</td>
<td>0.001 n.s.</td>
<td>-0.15***</td>
<td>-0.15***</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>Treatment</td>
<td>-0.54***</td>
<td>--</td>
<td>-0.54***</td>
</tr>
<tr>
<td>Fatigue</td>
<td>Diarrhea</td>
<td>0.28***</td>
<td>--</td>
<td>0.28***</td>
</tr>
</tbody>
</table>

Conclusions

- Considering that many causal relationships are not direct (or that the effects of one variable on another are not entirely direct), anticipating this and explicitly modeling these more complex relationships – these “Causal Cascades” – may help demonstrate that a treatment has an effect on an outcome of interest when more “standard” analyses do not show a significant relationship.
- Path analysis, with a century-long history across many disciplines, can help tease out these more complex causal relationships, and requires only a minor change in the way analyses are typically done.
Challenge of (and potential solution for) Analysis of Missing Data

Missing Data

• Whether an RCT or observational study, participants commonly have missing observations
  – This becomes more common the longer the trial or due to the disease

• For PRO data, the likelihood of missing data is greater
  – Often secondary or tertiary endpoints
  – Participants are generally responsible for completing them

• Can’t we just ignore them?
  – “We still have plenty of data from other participants”
  – “We know death and adverse events were not a huge problem”
  – We’ll use last observation carried forward…we’ve always done that
Recent Acknowledgement and Advances

• In 2008, the FDA requested that the National Research Council commission a panel of experts to produce a report “on appropriate study designs and follow-up methods to reduce missing data and on appropriate statistical methods to address missing data for the analysis of results.”
• The US Institute of Medicine, FDA, and European Medicines Agency recommend pre-specification of a primary method to accommodate missing data, as well as sensitivity analyses that make alternative assumptions about the missing data.

• “Use of the LOCF method is therefore likely to misrepresent the results of a trial seriously, and so is not a good choice for primary analysis. In contrast, the MMRM method is unlikely to result in serious misinterpretation, unless the drop-out mechanism is missing not at random (MNAR) and there is substantially unequal drop-out…. Neither method is capable of dealing on its own with trials involving MNAR drop-out mechanisms, for which sensitivity analysis is needed using more complex methods.”

Use Pattern Mixture Models?

• Advantages
  – Can accommodate mixed-effects models with repeated measures (MMRM) and growth curve models
  – Accounts for distribution of missingness in data

• Disadvantages
  – Suitable for small number of missing data patterns
  – Makes assumptions about missing data patterns that cannot be tested
  – Does not account for heterogeneity across missing data patterns

However, data are seldom missing at random in clinical trials, especially in some therapeutic areas such as oncology AND there can be many missing data patterns
Solution?
Identify Latent Classes based on *Patterns of Change for every Individual Patient – Extended Pattern Mixture Models*

Can help us understand if those who have worsening PRO scores are also more likely to drop out. That is, the trajectory of change in PRO score helps ‘inform’ us about “why” the patient dropped out.

---

**Overall Survival – Standard Approach (Treatment Arm Only)**

---

Kaplan-Meier survival estimate
How does accounting for missing data patterns inform overall survival? (Treatment arm only)

Conclusions

- Examining patterns of missing data and identifying differential change in patient-reported outcomes scores can provide clues about differential survival
- We have now taken this differential missing data on the PROs and used it to inform us about the censoring of survival
Voting Questions

Instructions: Go into ISPOR app and select 'Take a Poll' or to myispor.cnf.io and select session W10

Can you think of instances when the path analytic way of thinking about or analyzing data could be beneficial?
   a) Yes
   b) No
Voting Questions

Instructions: Go into ISPOR app and select 'Take a Poll' or to myispor.cnf.io and select session W10

Given our discussion of addressing missing data in the analysis of clinical trial data, how likely is it that you would encourage alternative approaches to LOCF being built into analysis plans?

a) Very unlikely
b) Somewhat unlikely
c) Somewhat likely
d) Very likely