Decomposition models as a framework for thinking about heterogeneity of treatment effects

William Crown, PhD
Chief Scientific Officer
Overview

1. Randomization solves a lot of problems
2. Proper design is fundamental in observational studies
3. Decomposition of average treatment effects
4. Decomposition with propensity scoring
5. Looks like an RCT. What could go wrong?
Aspects of bias addressed by randomization

- Balances comparison groups on both observed and *unobserved* characteristics
- Greatly simplifies analysis
- Inclusion/exclusion criteria and intensive follow-up in trials introduce issues of generalizability of findings
- Addresses high variability of findings in small trials
RWD vs. RWE (and types of RWE)

The term “real-world evidence” is widely used by those who develop medical products or who study, deliver, or pay for health care, but its specific meaning is elusive. We believe it refers to information on health care that is derived from multiple sources outside typical clinical research settings, including electronic health records (EHRs), claims and billing data, product and disease registries, and data gathered through personal experience and practice. This shortage of researchers with adequate methodologic savvy could result in poorly conceived study and analytic designs that generate incorrect or unreliable conclusions. Accordingly, if we are to realize the full promise of such evidence, we must be clear about what it is and how it can be used most effectively, and we must have appropriate expectations about what it can tell us. It is important to distinguish two key dimensions...

Use of RWE vs. RCTs

Clinical Pharmacology & Therapeutics

When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?

Jessica M. Franklin¹ and Sebastian Schneeweiss¹

Regulators consider randomized controlled trials (RCTs) as the gold standard for evaluating the safety and effectiveness of medications, but their costs, duration, and limited generalizability have caused some to look for alternatives. Real world evidence based on data collected outside of RCTs, such as registries and longitudinal healthcare databases, can sometimes substitute for RCTs, but concerns about validity have limited their impact. Greater reliance on such real world data (RWD) in regulatory decision making requires understanding why some studies fail while others succeed in producing results similar to RCTs. Key questions when considering whether RWD analyses can substitute for RCTs for regulatory decision making are WHEN one can study drug effects without randomization and HOW to implement a valid RWD analysis if one has decided to pursue that option. The WHEN is primarily driven by externalities not controlled by investigators, whereas the HOW is focused on avoiding known mistakes in RWD analyses.

Guideposts that can improve reliability of database study results

- Active comparator, same treatment modality
- Control for medication adherence
- High-dimensional proxy adjustment
- New users
- Avoid depletion of ‘susceptibles’ and other design flaws
Increasing Levels of Restriction in Pharmacoepidemiologic Database Studies of Elderly and Comparison With Randomized Trial Results

Sebastian Schneeweiss, MD, ScD,* Amanda R. Patrick, MS,* Til Stürmer, MD, MPH,* M. Alan Brookhart, PhD,* Jerry Avorn, MD,* Malcolm MacIure, ScD,* Kenneth J. Rothman, DMD, DrPH,† and Robert J. Glynn, PhD, ScD*
Stand-alone dummy variable

\[ Y = B_0 + B_1 \text{Age} + B_2 \text{T} + \epsilon_1 \]
Stand-alone dummy variable and interaction

\[ Y = B_0 + B_1 \text{Age} + B_2 T + B_3 (\text{Age}^* T) + \epsilon_1 \]

Cost

TCA

SSRI

(\(B_1 + B_3\))

Age

B0 + B2

B0

B1
Simple stratification

Cost

TCA

SSRI

D0=(B0 + B2)

D1=(B1 + B3)

C0=B0

C1=B1

Age
Heterogeneity of treatment effects in observational studies

• **Heterogeneity of treatment response:** in observational studies treatment may interact with covariates resulting in differential patient response.

• Although the covariance between treatment and the covariates is zero in an RCT by design, it is still possible for patients in each treatment group to have heterogeneous treatment responses.
Sample Stratification Versus Interaction Terms

- **Sample Stratification** solves the multicollinearity problem often associated with interactions, assuming there is sufficient sample size to estimate the separate models.

  But how can the average treatment effects be reconstructed from the separate equations?
Blinder-Oaxaca Decomposition

\[ \bar{Y}_{T2} - \bar{Y}_{T1} = \left[ F(X_{T2}\beta_{T2}) - F(X_{T1}\beta_{T2}) \right] + \left[ F(X_{T1}\beta_{T2}) - F(X_{T1}\beta_{T1}) \right] \]

Composition Coefficients

The counterfactual concept:
- Substitute TCA sample through SSRI model (or vice versa)
- Perform many times with subsamples
- Statistical bootstrapping to obtain standard errors
Blinder-Oaxaca decomposition in RCTs

• Blinder and Oaxaca showed that, in observational studies, the difference in expectation between two groups was a function of differences in the sample composition of the groups as well as coefficient effects.

• RCTs, by design, balance sample composition across groups on both observable and unobservable variables. This causes the sample composition terms to drop from the decomposition equation.

• As a result, average treatment effects are equal to a weighted average of the coefficients (heterogeneity of treatment effects).
The Use of Decomposition Methods in Real-World Treatment Benefits Evaluation for Patients with Type 2 Diabetes Initiating Different Injectable Therapies: Findings from the INITIATOR Study

Lee Brekke, PhD1,*, Erin Buysman, MS2, Michael Grabner, PhD2, Xuehua Ke, PhD2, Lin Xie, MS3, Onur Baser, PhD3,4,5, Wenhui Wei, PhD6

1Optum, Eden Prairie, MN, USA; 2HealthCare, Inc., Wilmington, DE, USA; 3STATinMED Research, Ann Arbor, MI, USA; 4University of Michigan, Ann Arbor, MI, USA; 5School of Economy, Administrative and Social Sciences, MEF University, Istanbul, Turkey; 6Sanofi US, Bridgewater, NJ, USA
Overall decomposition results – HbA1c

<table>
<thead>
<tr>
<th>GLA (n = 1107)</th>
<th>LIRA (n = 1059)</th>
<th>Difference (GLA – LIRA)</th>
<th>Explained difference (attributed to differences in baseline characteristics)</th>
<th>Unexplained difference (attributed to differences in treatment effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean HbA1c change (%)</td>
<td>−1.387</td>
<td>−0.742</td>
<td>−0.645</td>
<td>−0.730</td>
</tr>
<tr>
<td>P value</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

GLA, glargine; HbA1c, glycated hemoglobin A1c; LIRA, liraglutide.

Components of treatment effects – HbA1c

Overall decomposition results – treatment persistence

Table 3 – Treatment persistence: Results of overall decomposition analysis at 1 y (N = 3010).

<table>
<thead>
<tr>
<th></th>
<th>GLA (n = 1572)</th>
<th>LIRA (n = 1438)</th>
<th>Difference (GLA – LIRA)</th>
<th>Explained difference (attributed to differences in baseline characteristics)</th>
<th>Unexplained difference (attributed to differences in treatment effects)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean persistence (%)</td>
<td>64.8</td>
<td>48.7</td>
<td>16.1</td>
<td>-1.8</td>
<td>17.9</td>
</tr>
<tr>
<td>P value</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>&lt;0.001</td>
<td>0.215</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

GLA, glargine; LIRA, liraglutide.

Components of treatment effects – treatment persistence

Blinder-Oaxaca decomposition in observational studies with prior matching

- If patients in comparison groups were first balanced via propensity score or other matching methods, the sample composition terms would also drop from the decomposition equation in observational studies (assuming no omitted variables, etc).

- As with RCTs, the average treatment effect would be a weighted average of the heterogeneity of patient response within the groups.

- For otherwise similar data measurement, the difference between results from an observational study and an RCT would result from omitted variables, measurement error, or other factors that might introduce bias in the coefficient estimates in the observational setting.
What can go wrong

- Estimation of treatment effects within subgroups has been shown to lead to high false-positive and false-negative estimates using RCT data (Brookes et al., 2001).

- Results can be different when the reference group is changed (Fortin, Lemieux, Firpo, 2010).

- Bias can be introduced by omitted variables, measurement error, and other issues common in observational analyses.
Summary

- **Decomposition methods** have been developed in the labor economics literature to estimate gender and racial discrimination in wages.

- The methods decompose total variation between groups into variation that is due to differences in sample characteristics and variation that is due to the relationships of each group’s characteristics to their outcomes.

- Randomization is intended to balance the characteristics of comparison cohorts on both observed and unobserved variables.

- Application of decomposition methods to RCT data leads to the conclusion that average treatment effects are a weighted average of structural coefficients within and across each arm of the trial.

- Comparison of RCT and observational studies within the decomposition analysis framework is helpful for identifying the conditions under which estimates of treatment effect would be expected to be similar for the two types of designs.

- Highlights the importance of variable completeness and measurement quality to minimize bias from residual confounding in observational studies.
References


Thank you.

Contact information:
William H. Crown, CSO OptumLabs
William.crown@optum.com