Enhancing Hybrid Study Designs for Comparative Effectiveness Research

Jennifer L. Lund, PhD
Department of Epidemiology, University of North Carolina at Chapel Hill

May 7, 2019
2019 Triangle Comparative Effectiveness Research Symposium
Disclosures

This work is funded by the Patient-Centered Outcomes Research Institute, PCORI (ME-2017C3-9337).

I receive salary support as a co-investigator from the National Institutes of Health (R01 AG056479 and R01 AG050733) and PCORI 0415500004.

My spouse is a full-time, paid employee of GlaxoSmithKline.
Overview

• What are hybrid study designs?

• Motivating example

• Hybrid study design features and analytic methods

• Implications for future research
Traditional hierarchy of evidence

Phase III clinical trials are the gold standard for determining the efficacy of interventions

- Designed to reduce threats to **internal validity** through:
  - Restriction (inclusion/exclusion)
  - Randomization
  - Masking
  - Intensive treatment monitoring and follow-up

![Diagram of the traditional hierarchy of evidence]

- Systematic reviews
- Clinical trials (randomized, masked)
- Observational studies (cohort, cross-sectional, case-control)
- Case reports, case series
- Anecdotal findings, opinions, ideas
But what about external validity/generalizability?

“The degree to which the results of an observation hold true in other settings.”

Fletcher and Fletcher, *Clinical Epidemiology: The Essentials, 4th Edition*

One of the primary motivators of comparative effectiveness research

- Understand the benefits and harms of alternative interventions in routine clinical practice settings.
Why would the effects of interventions differ between clinical trial and routine care settings?

- Treatment effect heterogeneity and differences in study populations
- Delivery of different “versions” of treatment
- Measurement of different outcomes
"Balancing" internal and external validity in CER

Internal validity: Confounding (-), Selection bias (-), Misclassification (-)

External validity: Diverse patients (+), Real-world monitoring and adherence (+), Clinical endpoints of interest (not surrogates, +)
Explanatory clinical trial

Hybrid study design

Pragmatic clinical trial

Quasi-experimental CER

Single site non-experimental CER

Population-based non-experimental CER

*Illustration demonstrates relative trade-offs among different approaches, but distance between each approach does not have quantitative meaning.
Hybrid study design: blending of trial and observational data

**Phase III Trial**
- Treatment
- Covariates
- Outcomes

**Routine Care Cohort / Target Population**
- Treatment
- Covariates

**Weighted Trial Population**
- Treatment efficacy relevant to the trial population
- Treatment effectiveness relevant to routine care setting

reflect routine care patient characteristics and treatment delivery
Motivating example

Comparative effectiveness of two adjuvant chemotherapy regimens for treating stage II and III colon cancer

FOLFOX versus 5-FU
Treatment for stage II and III colon cancer

Surgical resection + Adjuvant chemotherapy

1980–2004: 5-fluorouracil (5-FU) was the mainstay for stage III disease*
• 16% absolute reduction in mortality compared with surgery alone

Moertel, NEJM (1990)
Adjuvant Oxaliplatin for Stage II/III Colon Cancer

In the early 2000’s, Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC) trial

- Stage II and III colon cancer patients
- Adjuvant oxaliplatin + 5-FU (FOLFOX) vs. 5-FU
- Progression-free and overall survival

Major eligibility criteria: 18-75 years, good performance status, adequate laboratory measures of kidney and liver function

FDA Approval: November, 2004

RD=4%; HR=0.80
RD=0%; HR=1.0

Andre, NEJM (2004,2009)
Treatment effect heterogeneity in MOSAIC, Overall survival

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>LV5FU2, n</th>
<th>FOLFOX4, n</th>
<th>HR</th>
<th>LCL</th>
<th>UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>588</td>
<td>630</td>
<td>0.82</td>
<td>0.66</td>
<td>1</td>
</tr>
<tr>
<td>Female</td>
<td>535</td>
<td>493</td>
<td>0.89</td>
<td>0.7</td>
<td>1.13</td>
</tr>
<tr>
<td><strong>Age, years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 70</td>
<td>963</td>
<td>968</td>
<td>0.78</td>
<td>0.66</td>
<td>0.93</td>
</tr>
<tr>
<td>&gt; 70</td>
<td>160</td>
<td>155</td>
<td>1.19</td>
<td>0.83</td>
<td>1.7</td>
</tr>
<tr>
<td><strong>T</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>62</td>
<td>57</td>
<td>1.55</td>
<td>0.63</td>
<td>3.8</td>
</tr>
<tr>
<td>3</td>
<td>852</td>
<td>853</td>
<td>0.8</td>
<td>0.66</td>
<td>0.96</td>
</tr>
<tr>
<td>4</td>
<td>208</td>
<td>213</td>
<td>0.93</td>
<td>0.68</td>
<td>1.27</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>447</td>
<td>452</td>
<td>1.01</td>
<td>0.75</td>
<td>1.36</td>
</tr>
<tr>
<td>1</td>
<td>442</td>
<td>440</td>
<td>0.86</td>
<td>0.67</td>
<td>1.11</td>
</tr>
<tr>
<td>2</td>
<td>231</td>
<td>229</td>
<td>0.71</td>
<td>0.54</td>
<td>0.93</td>
</tr>
<tr>
<td><strong>Perforation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1,045</td>
<td>1,045</td>
<td>0.84</td>
<td>0.71</td>
<td>0.99</td>
</tr>
<tr>
<td>Yes</td>
<td>78</td>
<td>78</td>
<td>1.01</td>
<td>0.62</td>
<td>1.64</td>
</tr>
<tr>
<td><strong>ECOG PS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0,1</td>
<td>984</td>
<td>968</td>
<td>0.8</td>
<td>0.68</td>
<td>0.95</td>
</tr>
<tr>
<td>2,3</td>
<td>139</td>
<td>155</td>
<td>1.11</td>
<td>0.76</td>
<td>1.64</td>
</tr>
</tbody>
</table>

Andre, JCO (2015)
Adverse events in MOSAIC, Grade 3-4 events

<table>
<thead>
<tr>
<th>Adverse Event</th>
<th>FL plus Oxaliplatin (N=1108)</th>
<th>FL (N=1111)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All Grades</td>
<td>Grade 3</td>
<td>Grade 4</td>
</tr>
<tr>
<td>Paresthesia†</td>
<td>92.0</td>
<td>12.4</td>
<td>NA</td>
</tr>
<tr>
<td>Neutropenia</td>
<td>78.9</td>
<td>28.8</td>
<td>12.3</td>
</tr>
<tr>
<td>Thrombocytopenia</td>
<td>77.4</td>
<td>1.5</td>
<td>0.2</td>
</tr>
<tr>
<td>Anemia</td>
<td>75.6</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Nausea</td>
<td>73.7</td>
<td>4.8</td>
<td>0.3</td>
</tr>
<tr>
<td>Diarrhea</td>
<td>56.3</td>
<td>8.3</td>
<td>2.5</td>
</tr>
<tr>
<td>Vomiting</td>
<td>47.2</td>
<td>5.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Stomatitis</td>
<td>41.6</td>
<td>2.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Skin‡</td>
<td>31.5</td>
<td>1.4</td>
<td>0.6</td>
</tr>
<tr>
<td>Alopecia§</td>
<td>30.2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Allergic reaction</td>
<td>10.3</td>
<td>2.3</td>
<td>0.6</td>
</tr>
<tr>
<td>Thrombosis or phlebitis</td>
<td>5.7</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Neutropenia with fever or infection</td>
<td>1.8</td>
<td>1.4</td>
<td>0.4</td>
</tr>
</tbody>
</table>

Therapy completion
FOLFOX: 75%
5-FU: 87%

Andre, JCO (2004)
What is the comparative effectiveness of FOLFOX versus 5FU in clinical practice settings?

• Several studies have evaluated the comparative effectiveness of FOLFOX versus 5FU using observational data (e.g., claims data)

• Often limited by threats of unmeasured confounding (e.g. staging information, performance status), limiting internal validity

• Captured in disparate populations due to data availability (e.g., 65+ year olds, NY State Medicaid), limiting external validity
Develop tools for implementing hybrid study designs for CER

1: Design a process to evaluate and select target populations of interest for CER.

2: Generate guidance for characterizing differences in therapy adherence between trial and target populations.

3: Develop a flexible analytic approach to estimate the comparative effectiveness and safety of therapies for target populations of interest.
Data sources

MOSAIC Phase III Trial

**Trial sponsor:** Sanofi

**Third party data manager:**

![ClinicalStudy](data.png)

**Data access:** Remote server administered by SAS (Clinical Trial Data Transparency platform)

**US Oncology EHR: iKnowMed**

**US Oncology:** Network of community oncology practices

**Data system:** Oncology-specific, integrated, web-based EHR system, iKnowMed®

**Data capture:** Standardized fields for treatment (planned/received, dose reductions), performance status, tumor information, etc.
1. Evaluate target populations of interest for CER

1. Patients who would have met trial eligibility criteria
2. All treated patients
3. Patients treated with an approved indication

US population diagnosed with stage II or III colon cancer

Trial population
Propensity score diagnostics and techniques can be used to identify target populations of interest in hybrid study designs: e.g.,

- Areas of non-overlap (non-positivity)
- Standardized mean differences
- Weights to account for differences in populations

*
2. Characterize treatment adherence in trial and target populations

Clinical trials are generally designed to optimize treatment adherence

• Inclusion/exclusion criteria (e.g., run-in periods)
• Close monitoring via regular study visits

Treatment adherence in clinical practice is likely to differ, often with lower expected adherence
Hypothetical visualization plots

**Trial Population**

- **Group 1:** 70%
- **Group 2:** 10%
- **Group 3:** 20%

**Target Population #1**

- **Group 1:** 40%
- **Group 2:** 30%
- **Group 3:** 20%
- **Group 4:** 10%
3. Estimate the comparative effectiveness of specific treatment contrasts in target populations of interest

When estimating treatment effects in trials, the focus is generally on estimating intention-to-treat effects in the trial population.

What if we want to know the effect of a specific treatment protocol (e.g., 6 out of 12 FOLFOX cycles) in a different target population?

- Re-weight the trial population to reflect the characteristics of the target
- Model the effects of the specific treatment protocol in the weighted trial population
Re-weight the trial population to reflect the target

- Similar to a propensity score, model the probability trial enrollment,
  \[\ln \left( \frac{Pr(\text{trial}=1 \mid z)}{Pr(\text{trial}=0 \mid z)} \right) = \beta_0 + \beta_1 z_1 + \cdots + \beta_k z_k \]
  as a function of covariates, \( z \), that are potential effect measure modifiers.

- Create weights, as follows:
  \[W_i = \begin{cases} 
  \frac{Pr(\text{trial}_i=0 \mid Z_i)}{Pr(\text{trial}_i=1 \mid Z_i)} \times \frac{p(\text{trial}=1)}{p(\text{trial}=0)}, & \text{trial}=1 \\
  0, & \text{trial}=0 
  \end{cases} \]
Model the effects of specific treatment protocols

- Restructure data into person-cycle observations (12 per trial participant)
- Use the parametric g-formula to model outcome risk
  - Based on a set of pooled logistic regression models (by treatment arm) to predict time-varying confounders and outcomes

  \[ \ln \left[ \frac{p(\text{Grade 3-4})|Z_{i_k-1} = \text{death}_{k-1} = \text{censored}_{k-1} = 0)}{p(\text{Grade 3-4})|Z_{i_k-1} = Z_{i_k-1} = \text{censored}_{k-1} = 0)} \right] = \beta_0 + \beta_v \text{Baseline}_{i=0} + \beta_w \text{Ttrt}_{k-1} + \beta_x \text{Grade 3-4}_{k-1} + \beta_y \text{ECOG}_{k-1} \]

  \[ \ln \left[ \frac{p(\text{Death})|Z_{i_k-1} = \text{death}_{k-1} = \text{censored}_{k-1} = 0)}{p(\text{Death})|Z_{i_k-1} = Z_{i_k-1} = \text{censored}_{k-1} = 0)} \right] = \beta_0 + \beta_v \text{Baseline}_{i=0} + \beta_w \text{Ttrt}_{k-1} + \beta_x \text{Grade 3-4}_{k-1} + \beta_y \text{ECOG}_{k-1} \]

- Weighted Monte Carlo sampling and outcome estimation
- Repeat process for treatment protocols of interest
Flexible analytic approach for CER

**MOSAIC trial data**

- Weighted trial data
- Weighted trial data
- Weighted ITT effect
- Parametric G-formula
- Treatment effect of interest
- Trial ITT effect
- (1) Weighted ITT effect
- (2) Weighted per protocol (full adherence) effect
- (3) Weighted treatment duration protocol effect

**Can be repeated for alternative target populations**
Data visualization
- Target and trial patient composition
- Temporal changes evaluation
- Therapy adherence comparisons

Quantitative metrics
- Propensity-score based measures of trial and target similarity
- Weighted adherence comparisons

Key output for PCOR/CER:
(1) Inform selection of relevant target populations
(2) Inform selection of relevant treatment effect contrasts

Guidance for hybrid study implementation

Analytic methods
- Variable selection and specification for sampling model
- Weighted Monte Carlo methods to incorporate sampling weights
- Worked example of the weighted parametric g-formula implementation

Key output for PCOR/CER:
(1) Guide analysis of hybrid study using weighted parametric g-formula
Implications for future research

- Proliferation of real world data sources (e.g., EHRs) and increasing access to clinical trial data make hybrid designs increasingly feasible

- Application of these methods to the CER context is limited

- Guidance on the selection of relevant and appropriate target populations and treatment contrasts of interest

- Informative to a variety of stakeholders: regulators, payers, clinicians and patients – who want to know about the population-level benefits and harms of therapies in clinical practice
Acknowledgements

PCORI Study Team
Alan Brookhart
Alex Keil
Amy Lowman
Sharon Peacock Hinton
Hanna Sanoff
Til Stürmer
Michael Webster-Clark
Daniel Westreich
Questions?

Contact information:
Jennifer L. Lund, PhD
Department of Epidemiology, UNC
Email: Jennifer.Lund@unc.edu
Phone: 919-966-7440
Potential reasons why effects might differ between trial and routine care populations

Setting of the trial
- Healthcare system
- Country
- Recruitment from primary, secondary, or tertiary care
- Selection of participating centres
- Selection of participating clinicians

Characteristics of randomised patients
- Baseline clinical characteristics
- Racial group
- Uniformity of underlying pathology
- Stage in the natural history of their disease
- Severity of disease
- Comorbidity
- Absolute risks of a poor outcome in the control group

Selection of patients
- Methods of prierandomisation diagnosis and investigation
- Eligibility criteria
- Exclusion criteria
- Placebo run-in period
- Treatment run-in period
- Enrichment strategies
- Ratio of randomised patients to eligible non-randomised patients in participating centres
- Proportion of patients who declined randomisation

Differences between the trial protocol and routine practice
- Trial intervention
- Timing of treatment
- Appropriateness/relevance of control intervention
- Adequacy of non-trial treatment—both intended and actual
- Prohibition of certain non-trial treatments
- Therapeutic or diagnostic advances since trial was done

Outcome measures and follow-up
- Clinical relevance of surrogate outcomes
- Clinical relevance, validity, and reproducibility of complex scales
- Effect of intervention on most relevant components of composite outcomes
- Who measured outcome
- Use of patient-centred outcomes
- Frequency of follow-up
- Adequacy of the length of follow-up

Adverse effects of treatment
- Completeness of reporting of relevant adverse effects
- Rates of discontinuation of treatment
- Selection of trial centres and/or clinicians on the basis of skill or experience
- Exclusion of patients at risk of complications
- Exclusion of patients who experienced adverse effects during a run-in period
- Intensity of trial safety procedures

Rothwell PM, Lancet (2005); Barenboim et al, PNAS (2016); Westreich et al, AJPH (2016)