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).

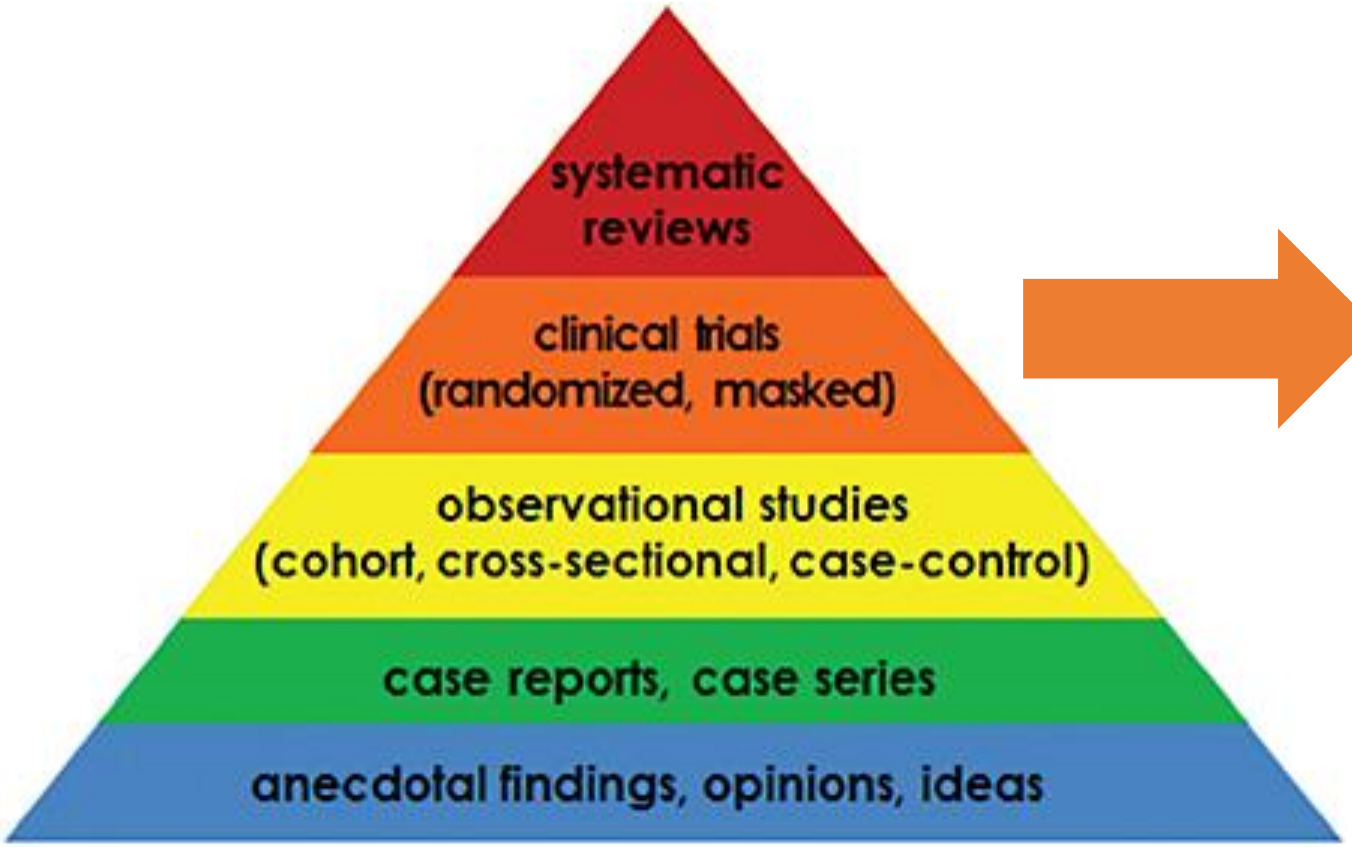
I receive salary support as a co-investigator from the National Institutes of Health (R01 AG056479 and R01 AG050733) and PCORI 0415500004.

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

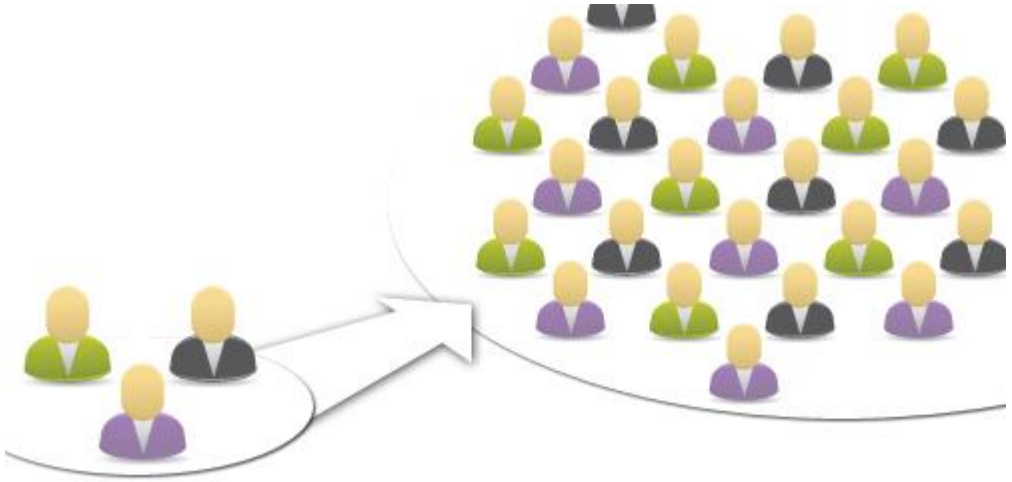
Traditional hierarchy of evidence



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

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, +)

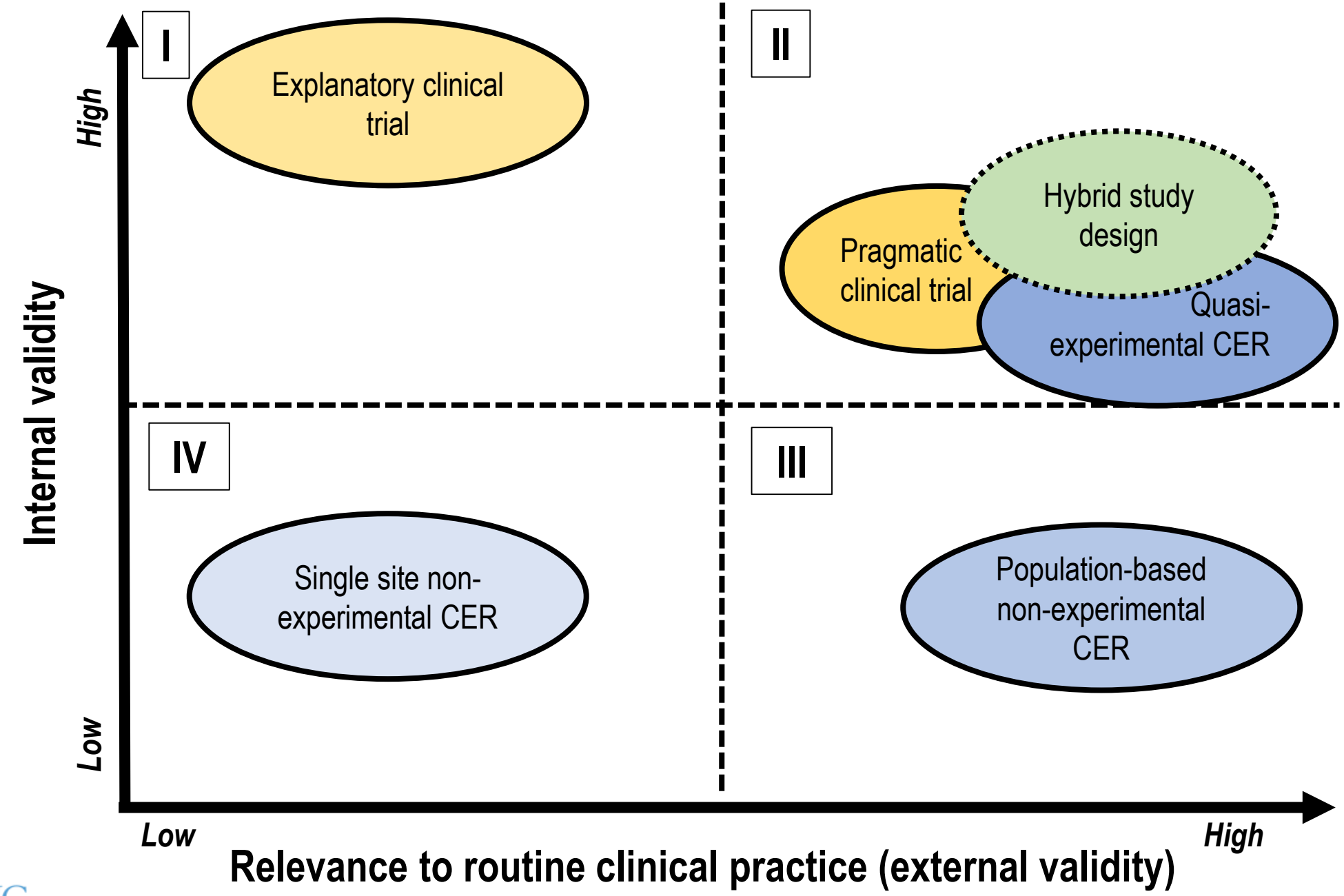
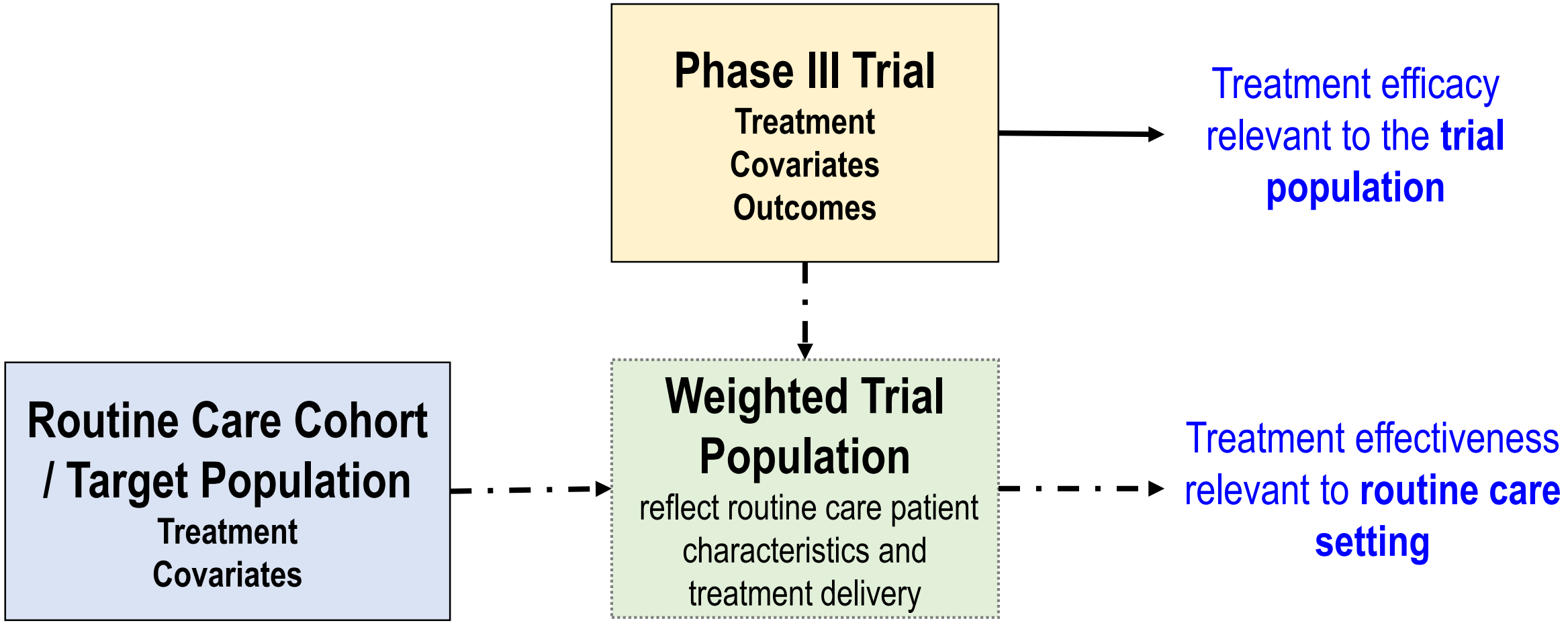


*Illustration demonstrates relative trade-offs among different approaches, but distance between each approach does not have quantitative meaning.

Hybrid study design: blending of trial and observational data

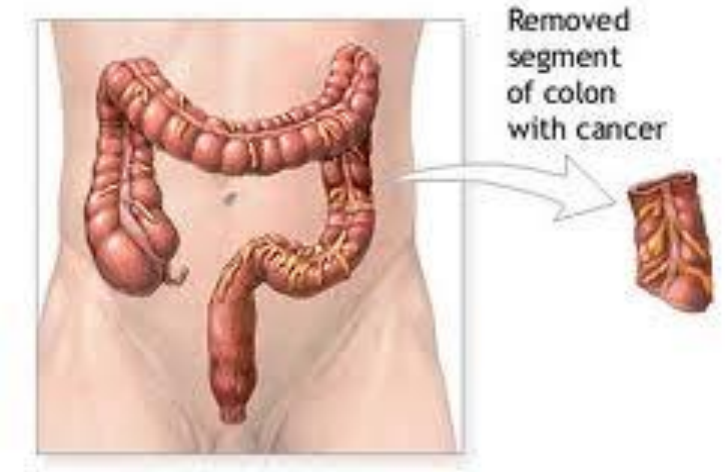


Motivating example

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*

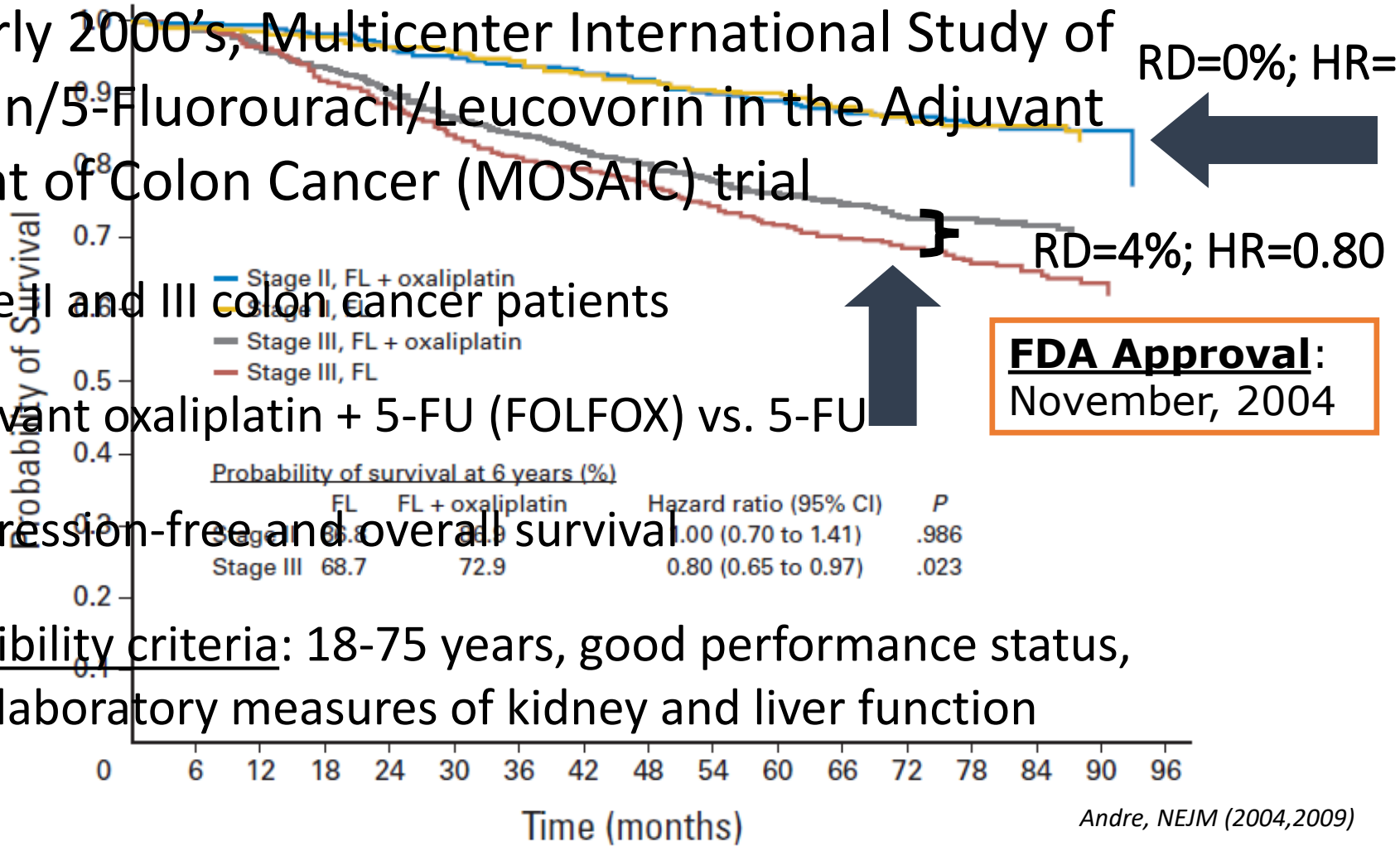
- 16% absolute reduction in mortality compared with surgery alone

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



RD=0%; HR=1.0

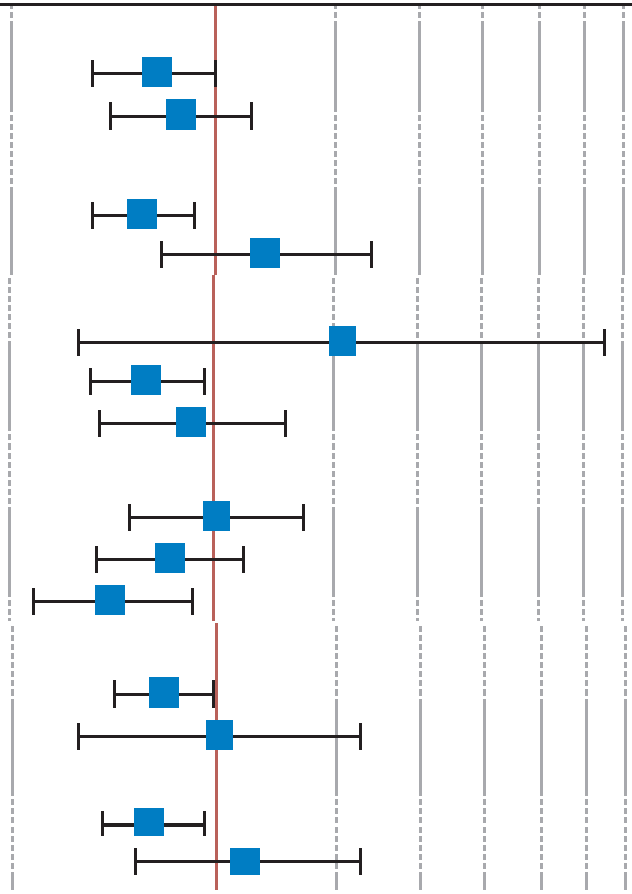
RD=4%; HR=0.80

FDA Approval: November, 2004

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

Andre, NEJM (2004,2009)

Treatment effect heterogeneity in MOSAIC, Overall survival



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

Andre, JCO (2015)

Adverse events in MOSAIC, Grade 3-4 events

Adverse Event	FL plus Oxaliplatin (N=1108)			FL (N=1111)			P Value	
	All Grades	Grade 3	Grade 4	All Grades	Grade 3	Grade 4	All Grades	Grades 3 and 4
	percent							
Paresthesia†	92.0	12.4	NA	15.6	0.2	NA	<0.001	0.001
Neutropenia	78.9	28.8	12.3	39.9	3.7	1.0	<0.001	<0.001
Thrombocytopenia	77.4	1.5	0.2	19.0	0.2	0.2	<0.001	0.001
Anemia	75.6	0.7	0.1	66.9	0.3	0.0	<0.001	0.09
Nausea	73.7	4.8	0.3	61.1	1.5	0.3	<0.001	<0.001
Diarrhea	56.3	8.3	2.5	48.4	5.1	1.5	<0.001	<0.001
Vomiting	47.2	5.3	0.5	24.0	0.9	0.5	<0.001	<0.001
Stomatitis	41.6	2.7	0.0	39.6	2.0	0.2	0.34	0.41
Skin‡	31.5	1.4	0.6	35.5	1.7	0.7	0.05	0.67
Alopecia§	30.2	NA	NA	28.1	NA	NA	0.28	NA
Allergic reaction	10.3	2.3	0.6	1.9	0.1	0.1	<0.001	<0.001
Thrombosis or phlebitis	5.7	1.0	0.2	6.5	1.7	0.1	0.48	0.29
Neutropenia with fever or infection	1.8	1.4	0.4	0.2	0.1	0.1	<0.001	<0.001

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 DataRequest.com

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

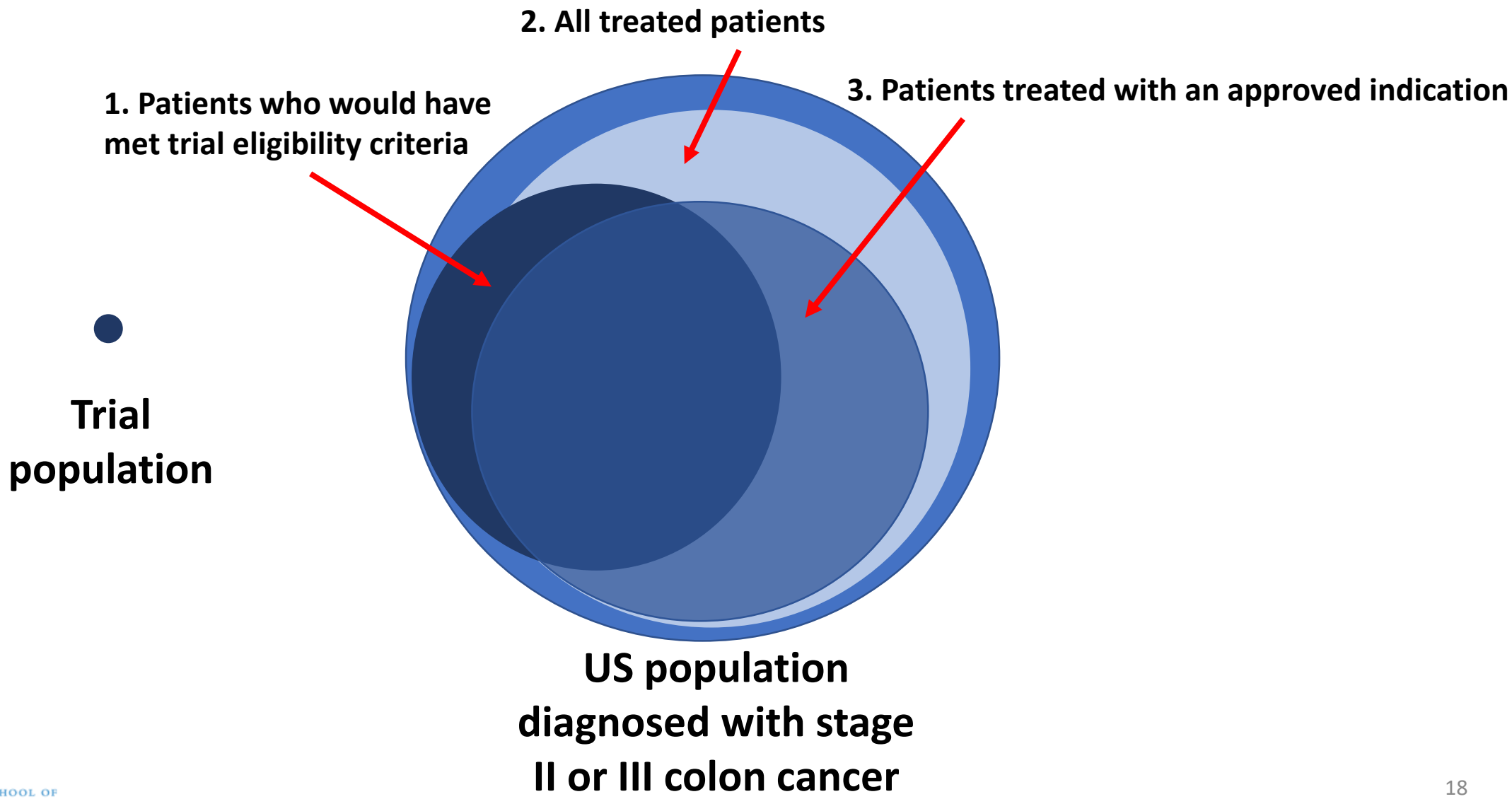
US Oncology EHR: iKnowMed

US Oncology: Network of community oncology practices

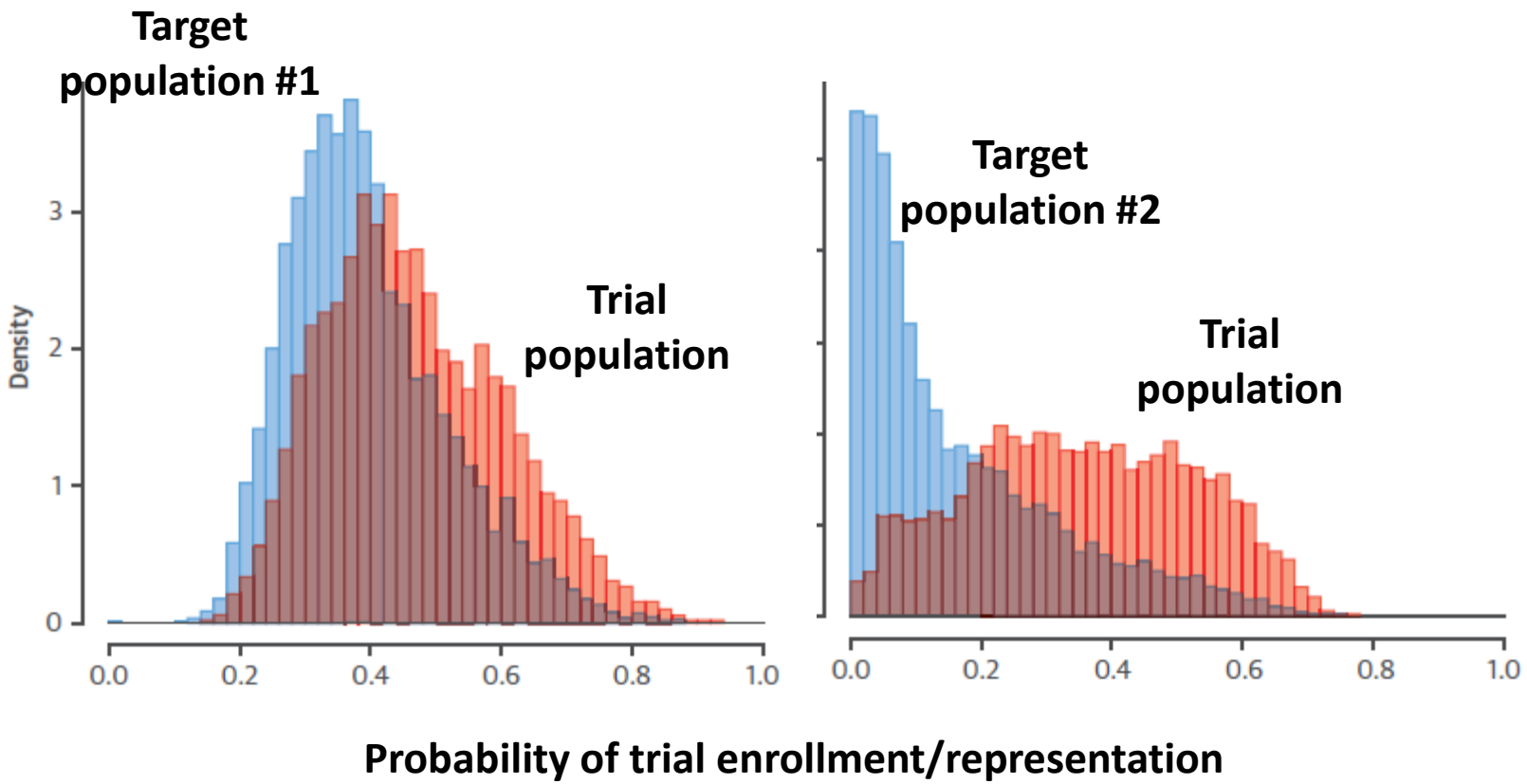
Data system: Oncology-specific, integrated, web-based EHR system, iKnowMed®

Data capture: Standardized fields for treatment (planned/received, dose reductions), performance status, tumor information, etc.

1. Evaluate target populations of interest for CER



Hypothetical visualization plots



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*

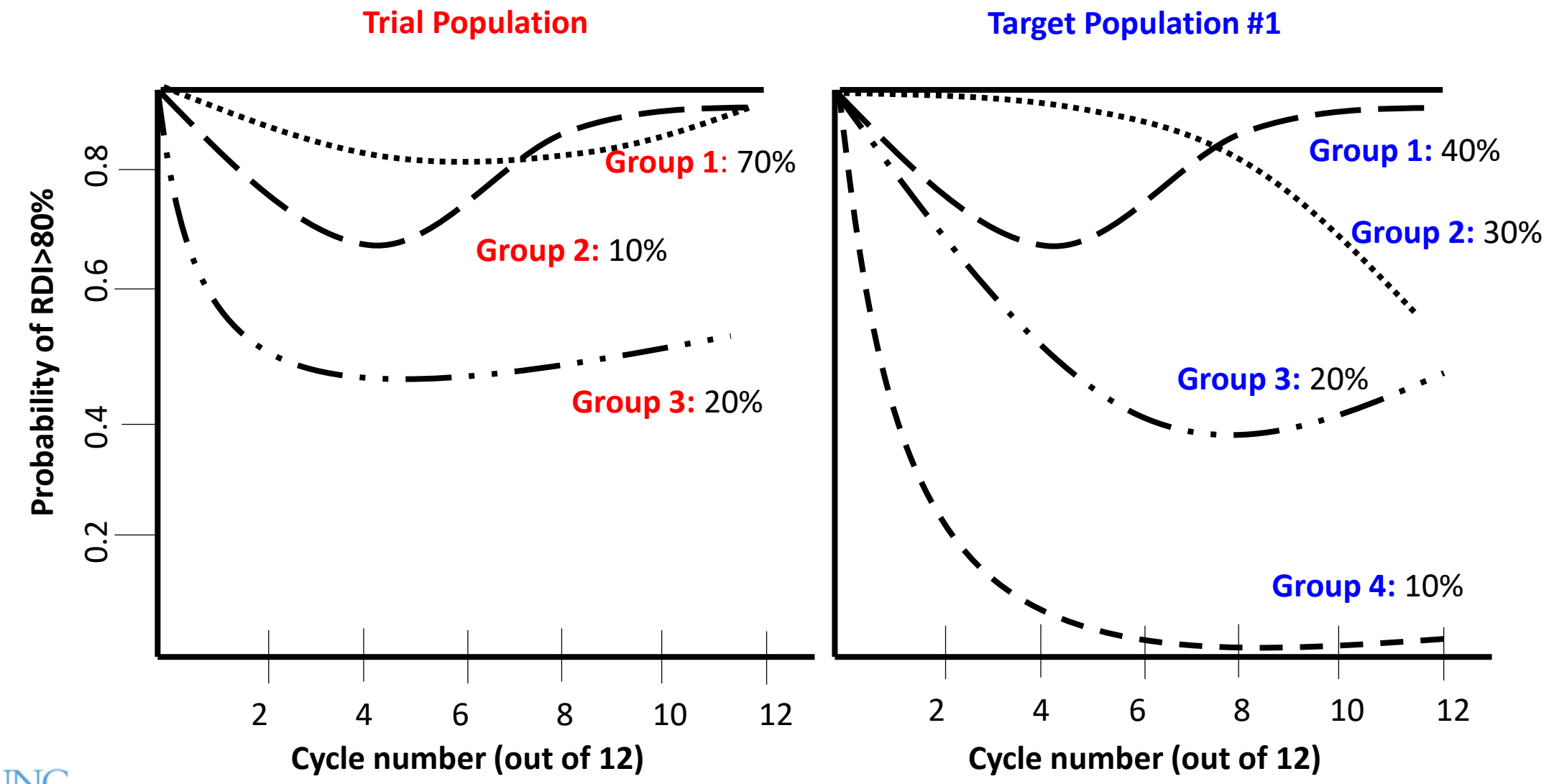
2. Characterize treatment adherence in trial and target populations

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



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,

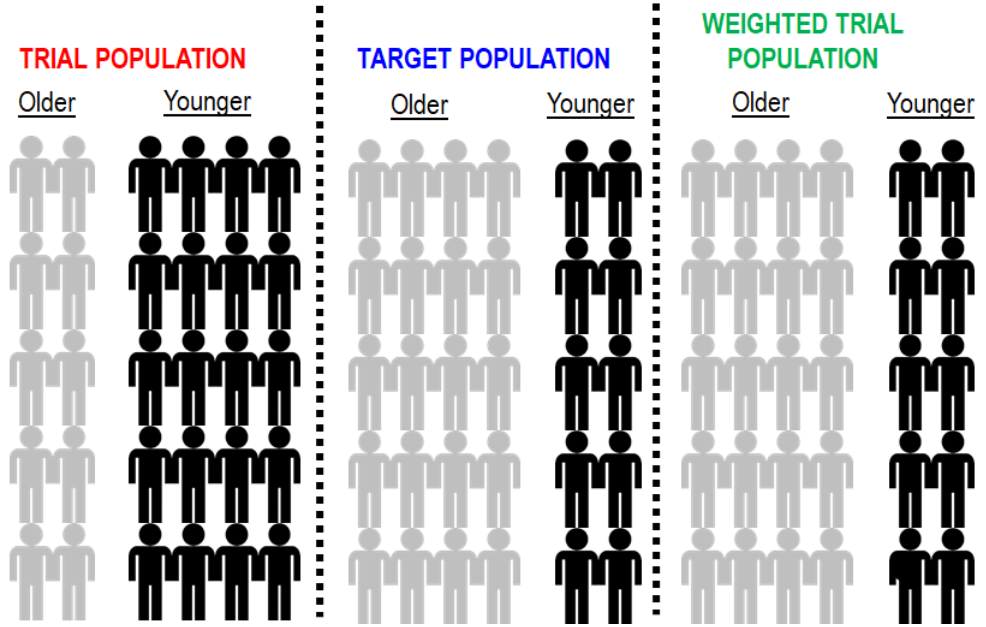
$$ln\left[\frac{Pr(trial{=}1 \mid z)}{Pr(trial{=}0 \mid z)}\right] = \beta_0 + \beta_1 z_1 + \cdots + \beta_k z_k$$

, as a function of covariates, z,

that are potential effect measure modifiers.

- Create weights, as follows:

$$W_i = \begin{cases} \frac{Pr(trial_i{=}0 \mid Z_i)}{Pr(trial_i{=}1 \mid Z_i)} \times \frac{p(trial{=}1)}{p(trial{=}0)}, & trial{=}1 \\ 0, & trial{=}0 \end{cases}$$



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

  Time-varying confounder (e.g., grade 3-4 adverse event):

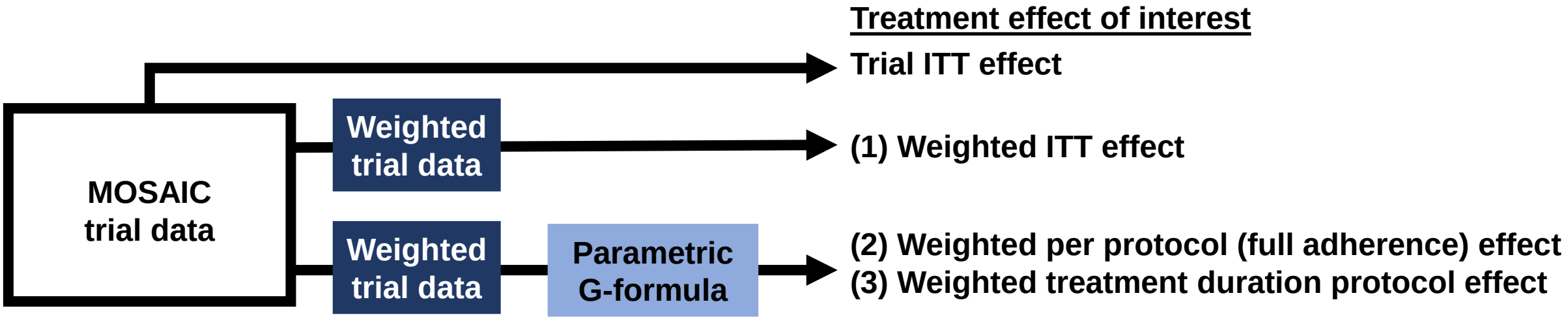
$$ln\left[\frac{p(Grade34_{k)}=1\,|\bar{Z}_{i_{k-1}},death_{k-1}=censored_{k-1}=0)}{p(Grade34_{k)}=0\,|\bar{Z}_{i_{k-1}},death_{k-1}=censored_{k-1}=0)}\right]=\beta_0+\beta_v Baseline_{i=0}+\beta_{\bar{w}}\overline{Trt}_{k-1}+\beta_{\bar{x}}\overline{Grade34}_{k-1}+\beta_{\bar{y}}\overline{ECOG}_{k-1}$$

  Time-varying outcome (e.g., all-cause mortality):

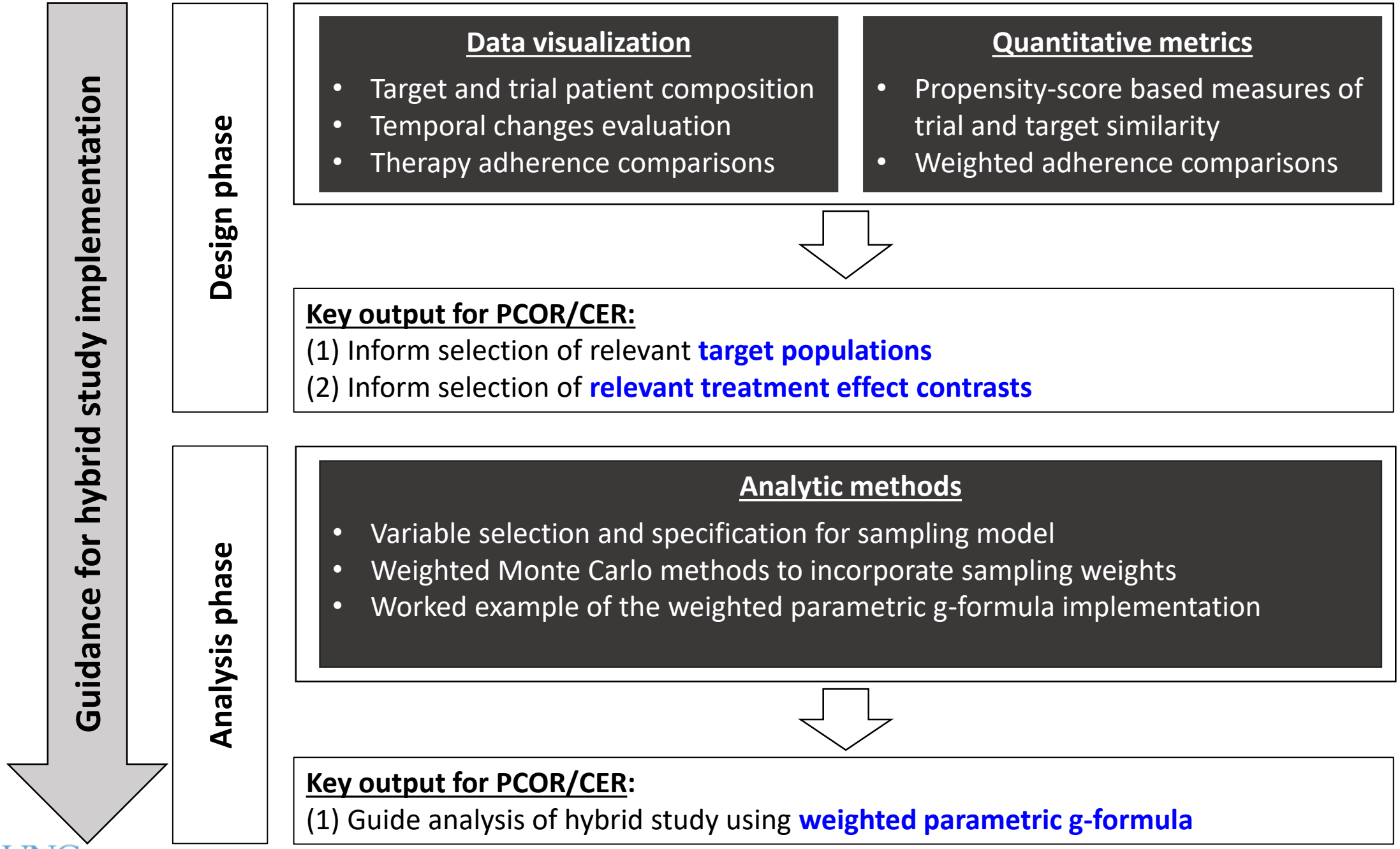
$$ln\left[\frac{p(Death_{k)}=1\,|\bar{Z}_{i_{k-1}},death_{k-1}=censored_{k-1}=0)}{p(Death_{k)}=0\,|\bar{Z}_{i_{k-1}},\ death_{k-1}=censored_{k-1}=0)}\right]=\beta_0+\beta_v Baseline_{i=0}+\beta_{\bar{w}}\overline{Trt}_{k-1}+\beta_{\bar{x}}\overline{Grade34}_{k-1}+\beta_{\bar{y}}\overline{ECOG}_{k-1}$$

- Weighted Monte Carlo sampling and outcome estimation
- Repeat process for treatment protocols of interest

Flexible analytic approach for CER



** Can be repeated for alternative target populations


Guidance for hybrid study implementation
Design phase
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
Analysis phase
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

Selection of patients
- Methods of prerandomisation 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

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

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)