



Clinical Data Wrangling
Session 1: Introduction

The Added Value of Clinical Data

Nicole G Weiskopf, 8/21/18

Defining “Clinical” Data

- Clinical documentation
 - Unstructured: Progress notes, Medical Hx, etc
 - Structured: Labs, Medications, Orders, Dx, etc
- Administrative data (often included in this definition)
 - Billing
 - CMS, Insurance
- **Primary purpose is *not* research**

Benefits of Clinical Data in Research

1. Decrease costs (time and money)
2. Enable recruitment and retention
 - Underrepresented populations
 - Rare diseases
3. Volume, variety, and velocity of data
<https://www.gartner.com/it-glossary/big-data>
4. Increased representativeness (aka, generalizability or external validity)

Cost of traditional trials

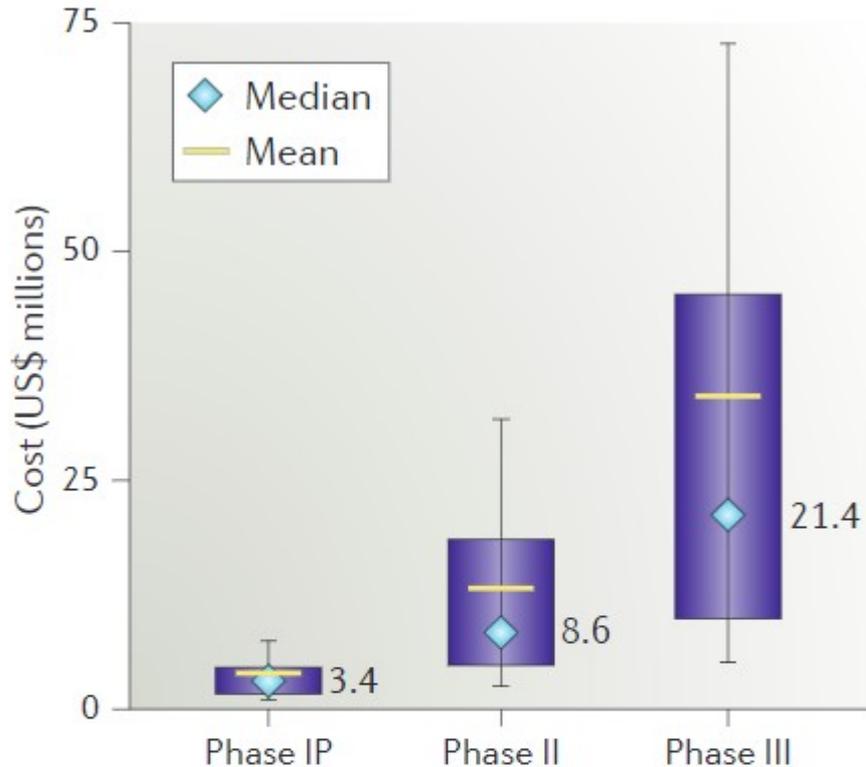
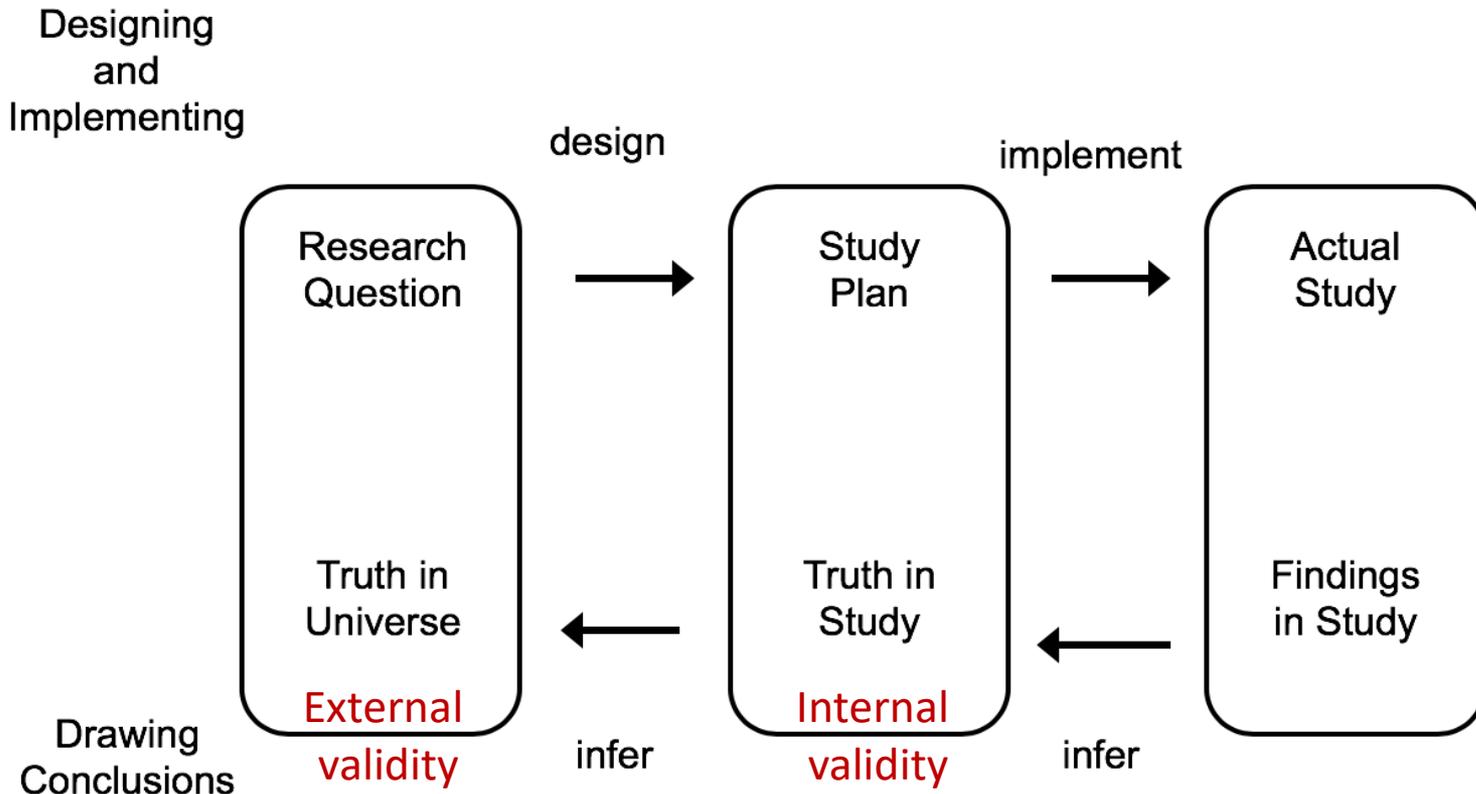


Figure 2. **Costs of clinical trials.** Data are from 726 studies conducted in patients from 2010-2015. Medians and means are indicated with diamonds and lines. Boxes indicate the 25th and 75th percentiles and whiskers indicate the 10th and 90th percentiles.

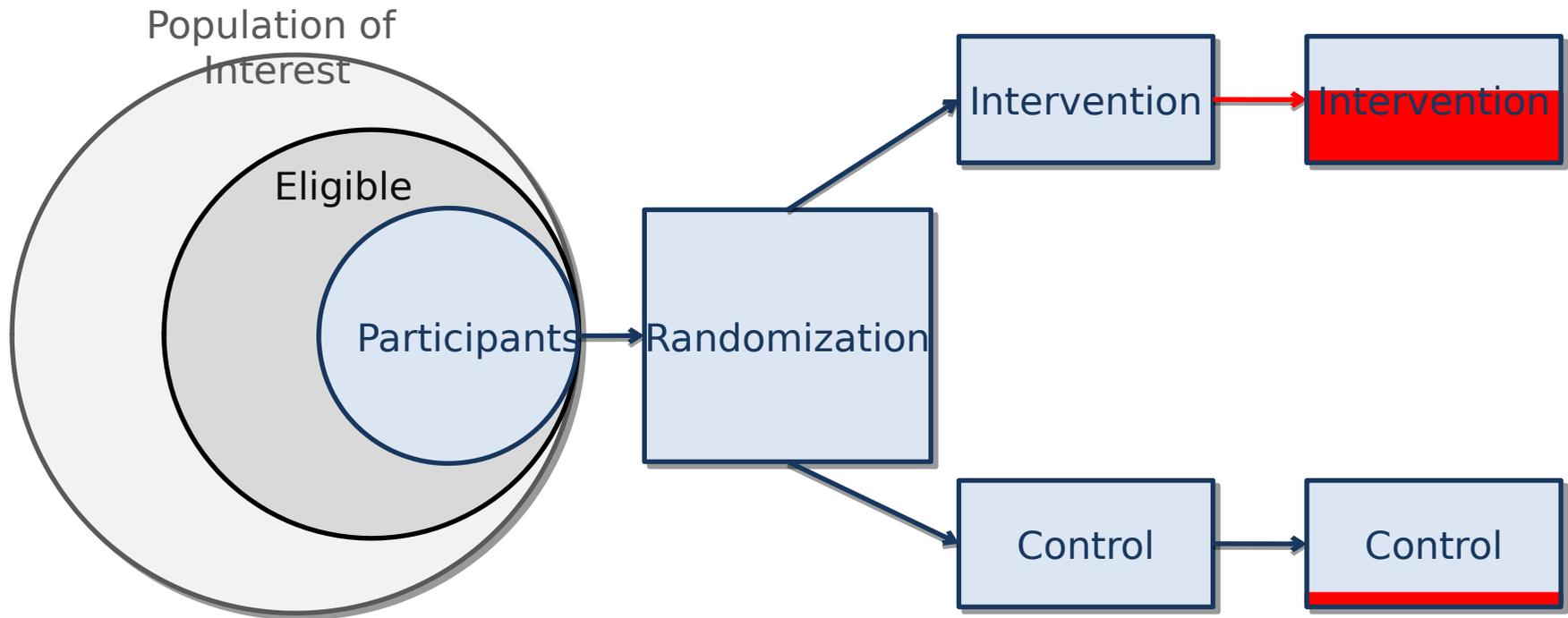
Good research should give us broadly applicable truths



Thanks to Adam Wilcox

Credit to Designing Clinical Research: An Epidemiologic Approach, Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Third edition, 2006

Randomized controlled trials are the gold standard of clinical research...



...but external validity may be limited

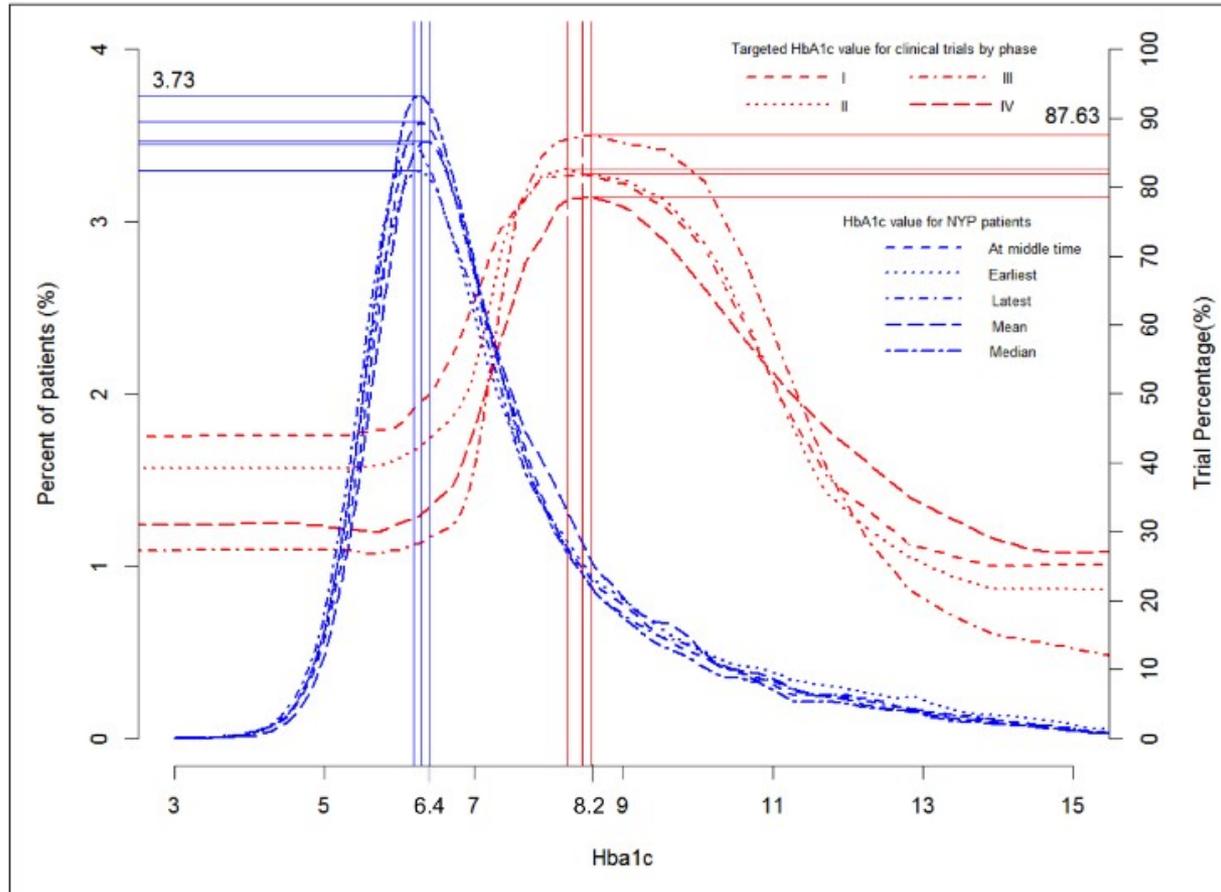


Fig. 1 For each HbA_{1c} value on X, the red lines indicate the percentage of trials of different phases whose eligibility criteria include that value and the blue lines indicate the percentage of patients whose earliest, latest, mean, median, or middle- collection-time HbA_{1c} was that value.

...but external validity may be limited

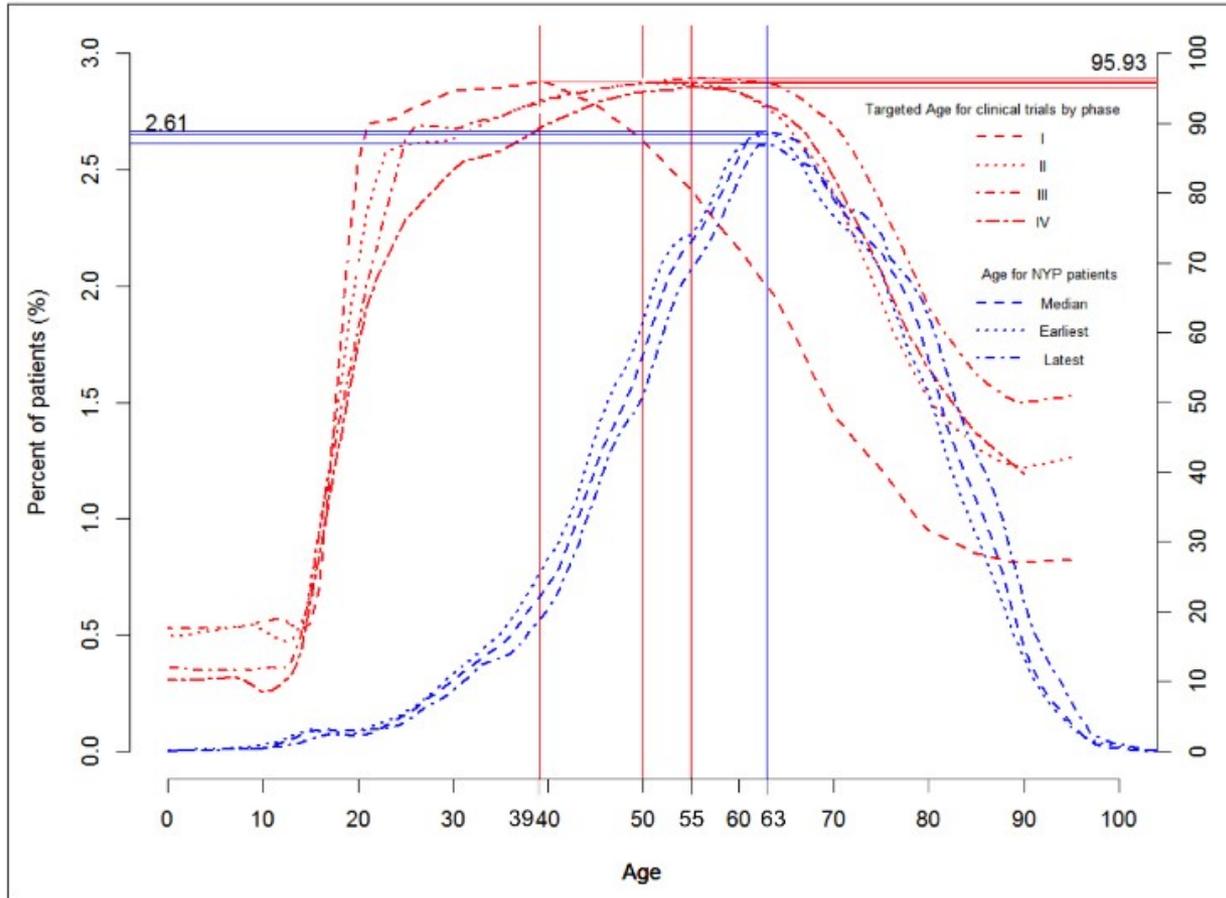
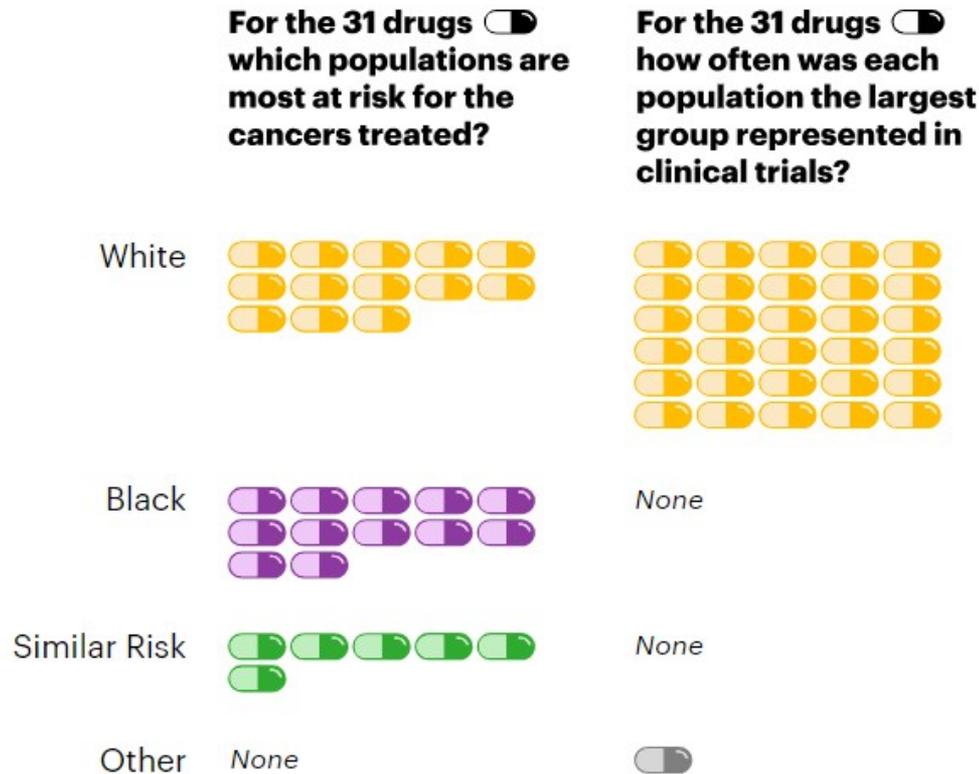


Fig. 2 For each age value on X, the red solid line indicates the percentage of trials of different phases whose eligibility criteria include that value and the blue lines indicate the percentage of patients who were that age at the earliest, median, or latest HbA_{1c} measurements.

Lack of representativeness of clinical research can lead to healthcare

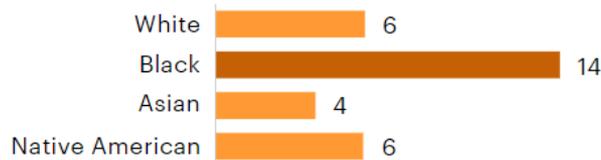


Note: Drugs are labeled "Similar Risk" if black Americans are at least 80 percent as likely as white Americans to be diagnosed with the cancer treated.

Lack of representativeness of clinical research can lead to healthcare

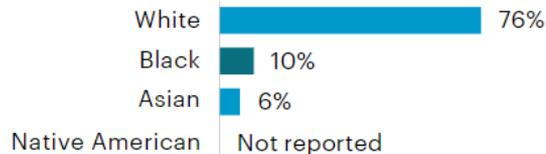
Black Americans Face the Highest Risk of Multiple Myeloma but Are Underrepresented in Trials Treating the Cancer

New Cases of Multiple Myeloma, Per 100,000 People

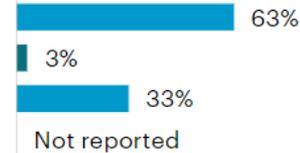


Representation in Clinical Trials

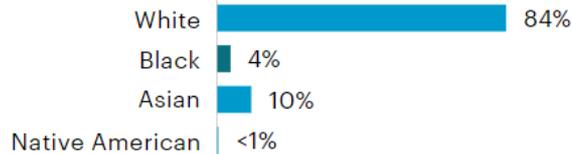
Darzalex



Farydak



Empliciti



Ninlaro



W

of EHR data and other sources of clinical or administrative data has the potential to enable more efficient research with improved generalizability.