Causal estimands

Should we ask different causal questions in randomized trials and in the observational studies that emulate them?

Miguel Hernán

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

No

- If a causal question is important enough to be asked in an observational study
- then we should also ask it in a randomized trial
- and vice versa

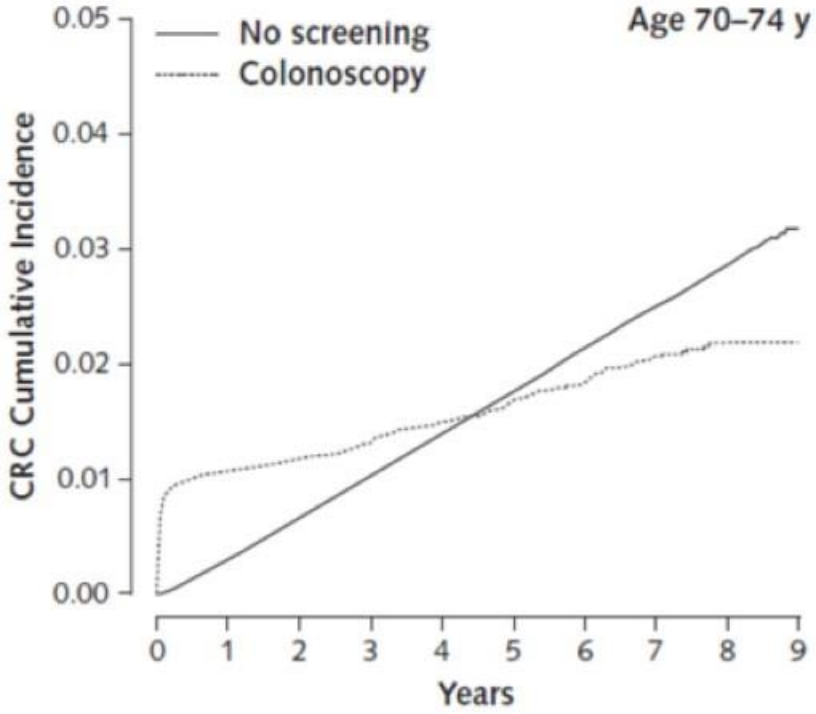
EXAMPLE
Screening colonoscopy and colorectal cancer

- **Screening colonoscopy at age 50 (and then every 10 years) has been recommended for decades**
  - without direct evidence from randomized trials

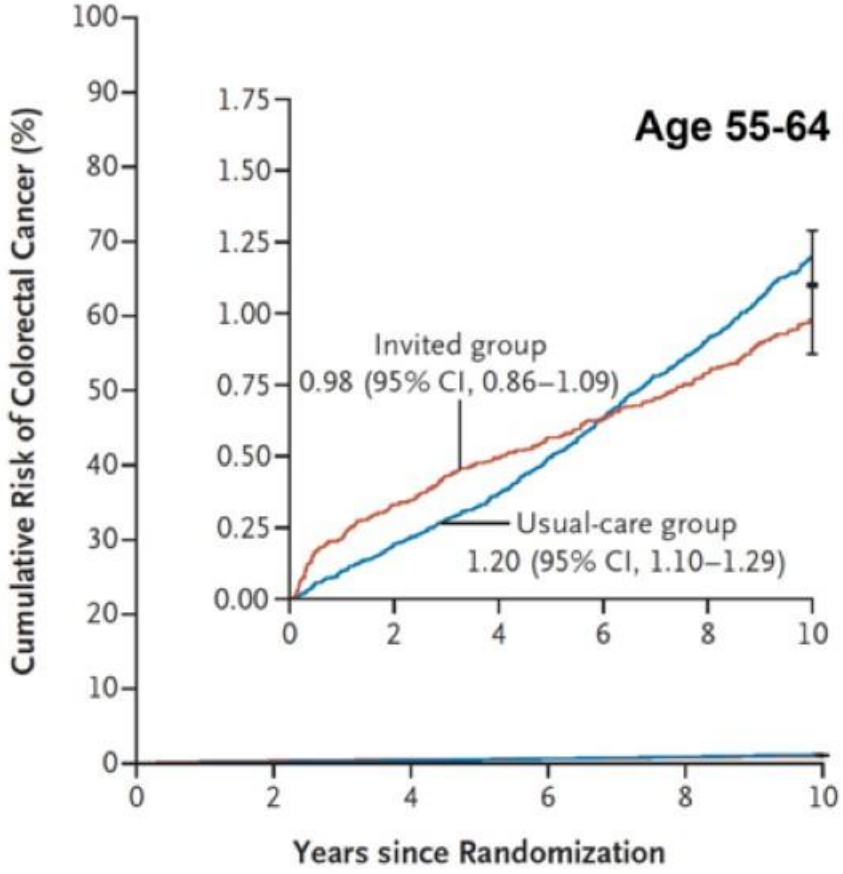
- **We launched a randomized trial**
  - ~100,000 individuals followed for 15 years
    - Preliminary results in Bretthauer et al. *N Engl J Med* 2022; 387:1547-1556
- **In the meantime, we conducted an observational study**
  - to emulate a target trial in older patients
    - Garcia-Albeniz et al. *Ann Int Med* 2017; 166(1):18-26

Results were similar (accounting for higher risk in older patients)

Observational study (2017)



Randomized trial (2022)



But these studies asked different causal questions

Observational study

- What is the effect of undergoing screening?
  - as per the protocol of the hypothetical target trial
- Per-protocol effect

Randomized trial

- What is the effect of being assigned to screening?
  - regardless of whether participants follow the trial's protocol
    - 42% of invited participants participated in screening
- Intention-to-treat effect

Why?

What is the justification for asking different causal questions?

- If the intention-to-treat effect is of interest
  - then we should also estimate it in the observational study
- If the per-protocol effect is of interest
  - then we should also estimate it in the randomized trial

- Which effect would you prefer if you were considering having a colonoscopy?
  - The effect of being invited to have a colonoscopy, even though 58% of invitees did not undergo a colonoscopy
  - The effect of having a colonoscopy

Question is:
What causal questions we want to ask?

- Or, if we wish to sound really sophisticated, what **causal estimands** do we want to estimate?

- **Estimand: the quantity we want to estimate**
  - A trial protocol defines a causal estimand
- **Estimator: the method used to estimate the estimand**
  - In a trial protocol, look for it under "Statistical analysis plan"
- **Estimate: the number obtained when applying the estimator to the data**

Components of the Randomized Trial protocol

- Eligibility criteria
- Treatment strategies
- Assignment
- Time zero and follow-up
- Outcomes
- Causal contrasts
- Data analysis

Causal estimand

Estimator

Traditionally, randomized trials were allowed to ask only one causal question

- **What is the effect of assigning people to a treatment, regardless of whether they actually take the treatment?**
- The intention-to-treat effect
- Increasing concern that this may be
  - a waste of societal resources
  - malpractice (for safety studies)

The intention-to-treat effect is the effect of assignment to treatment

- Not the effect of treatment
- In fact, the intention-to-treat effect may differ across trials that study the same treatment
  - If adherence patterns vary across studies
  - Even if there is zero bias in each of the studies

- So why do we use intention-to-treat effects?
  - Let's review some arguments for ITT effects
    - Hernán, Hernández-Díaz. *Clinical Trials* 2012

1. A commonly heard argument for ITT effect: It preserves the null

- Placebo-controlled double-blind trial
  - if treatment has a null effect on the outcome
  - then the ITT effect will be null

- Presented as a key advantage of the ITT analysis
  - it correctly estimates the effect of treatment under the null, regardless of the adherence pattern

2. A commonly heard argument for ITT effect: It is conservative

- Placebo-controlled double-blind trial
  - If treatment has a nonnull effect
  - If some participants do not adhere to their assigned treatment
- **ITT effect will be closer to the null that the actual effect of treatment**
  - in between the effect of treatment and the null value
  - we say that the ITT effect is conservative because it is "biased towards the null"

3. A commonly heard argument for ITT effect: It measures effectiveness in real world

- Magnitude of ITT effect is affected by non-adherence
  - "As it should be", some say
  - Because there is no perfect adherence in the real world

- The ITT effect quantifies the realistic treatment effects in clinical settings outside the trial

Demystifying intention-to-treat effects: Null preservation is not guaranteed

- Consider a non-blinded trial
- ITT effect may not be null even if treatment has a null effect
  - Assignment may make patients and doctors alter their behavior in ways that affect the outcome

- *Most pragmatic trials are not blinded*

Demystifying intention-to-treat effects: Not necessarily biased towards the null

- Even if the treatment effect is monotonic
- Trial of 2 active treatments with differential adherence
  - due to a mild, easily palliated side effect
- An ITT analysis may indicate a beneficial effect of the less effective treatment

- *Many randomized trials are head-to-head trials*

Demystifying intention-to-treat effects: Bias towards the null is often undesirable

- Safety trials
- Non-inferiority trials

- **In these trials, a "conservative" ITT analysis is statistical malpractice**
  - *Randomized cynical trial*: A trial designed to quantify harm and whose protocol considers only an ITT analysis

- ***Many randomized trials are for safety, non-inferiority***

Demystifying intention-to-treat effects: Not necessarily a measure of effectiveness

- Degree of adherence outside the trial may differ from that in the trial
  - because doctors and patients learn of the trial's findings
  - because procedures for assignment and enforcement of adherence vary
- Effectiveness in the community may differ from ITT effect estimate from trial

Demystifying intention-to-treat effects: Not of primary interest for doctors/patients

- Remember the screening example
- Another example: A couple trying to decide whether to use a contraceptive method would want to know
  - the effectiveness of the method when used as indicated
  - not the estimated effectiveness in a population in which, say, 40 of couples failed to use the method properly

- *Many trials are designed to guide clinical decisions by patients and doctors*

The intention-to-treat effect is NOT the only question of interest

- For clinicians, patients, regulators...

- We actually asked people with chronic diseases
  - in focus groups, after explaining in nontechnical terms
- They wanted **both** intention-to-treat & per-protocol effects
  - absolute risks in subgroups
  - superiority designs (forget noninferiority)

  - Murray et al. *J Clin Epidemiol* 2018; 103:10-21

The intention-to-treat effect is NOT the only question of interest

- But it is a question that can be easily answered
  - via an intention-to-treat analysis

- So we convinced ourselves that's the only question we wanted to answer

Another causal question that we'd like to answer in randomized trials

- **What is the effect of adhering to the treatment strategy that people were assigned to?**

- The per-protocol effect
  - Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. *Clinical Trials* 2012
  - Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed like observational studies. *Annals of Internal Medicine* 2013
  - Murray EJ, Hernán MA. Getting the most out of randomized clinical trials: A call for better per-protocol effect estimates. *Clinical Trials* 2016

Problem: valid estimation of the per-protocol effect requires adjustment

- Naïve per-protocol analyses don't adjust for confounding
  - but adherence is not randomly assigned!

- Correct per-protocol analyses need to adjust for confounding
  - because they are observational analyses of the trial data
- Using the same methods and assumptions as observational analyses
  - Hernán et al. *Ann Int Med* 2013

So can we ever be sure that per-protocol effect estimates are unbiased?

- No
- Historically, trialists have been suspicious of per-protocol analyses
- A key reason for those suspicions was the Coronary Drug Project…

Coronary Drug Project (CDP)

- A double-blind randomized trial
  - U.S. men with history of myocardial infarction
  - 53 study centers, NIH funded
- Enrollment: 1966-1969; visits every 4 months
- Randomly assigned to placebo or
  - one of 5 active treatments (all but clofibrate discontinued)
  - 1103 patients assigned to clofibrate; 2789 to placebo
- Null ITT effect of clofibrate on mortality

INFLUENCE OF ADHERENCE TO TREATMENT AND RESPONSE OF CHOLESTEROL ON MORTALITY IN THE CORONARY DRUG PROJECT

THE CORONARY DRUG PROJECT RESEARCH GROUP

1038 THE NEW ENGLAND JOURNAL OF MEDICINE Oct. 30, 1980

- In the placebo group, 5-year mortality risk was higher among those who did not adhere to the placebo pills than among those who did

- This finding was taught in courses around the world
  - a cautionary tale about the dangers of deviating from the intention-to-treat principle.
  - chilling effect on subsequent attempts to conduct per-protocol (observational) analyses in randomized trials

A 21st century update of the CDP analysis

Article

CLINICAL TRIALS

Clinical Trials
1–7

DOI: 10.1177/1740774516634335
ctj.sagepub.com
SAGE

Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials

Eleanor J Murray[1] and Miguel A Hernán[1,2]

(with thanks to Paul Canner)

Difference in 5-year mortality between adherers and nonadherers to placebo

- Replication of 1980 analysis
  - Unadjusted: 14.3% (95% CI 10.8 to 17.8)
  - Adjusted: 10.9% (95% CI 7.5 to 14.4)
    - for baseline variables only
- 2015 update
  - Unadjusted: 11.0% (95% CI 6.5. to 15.6)
  - Adjusted: 2.5% (95% CI -2.1 to 7.0)
    - for baseline and post-baseline variables

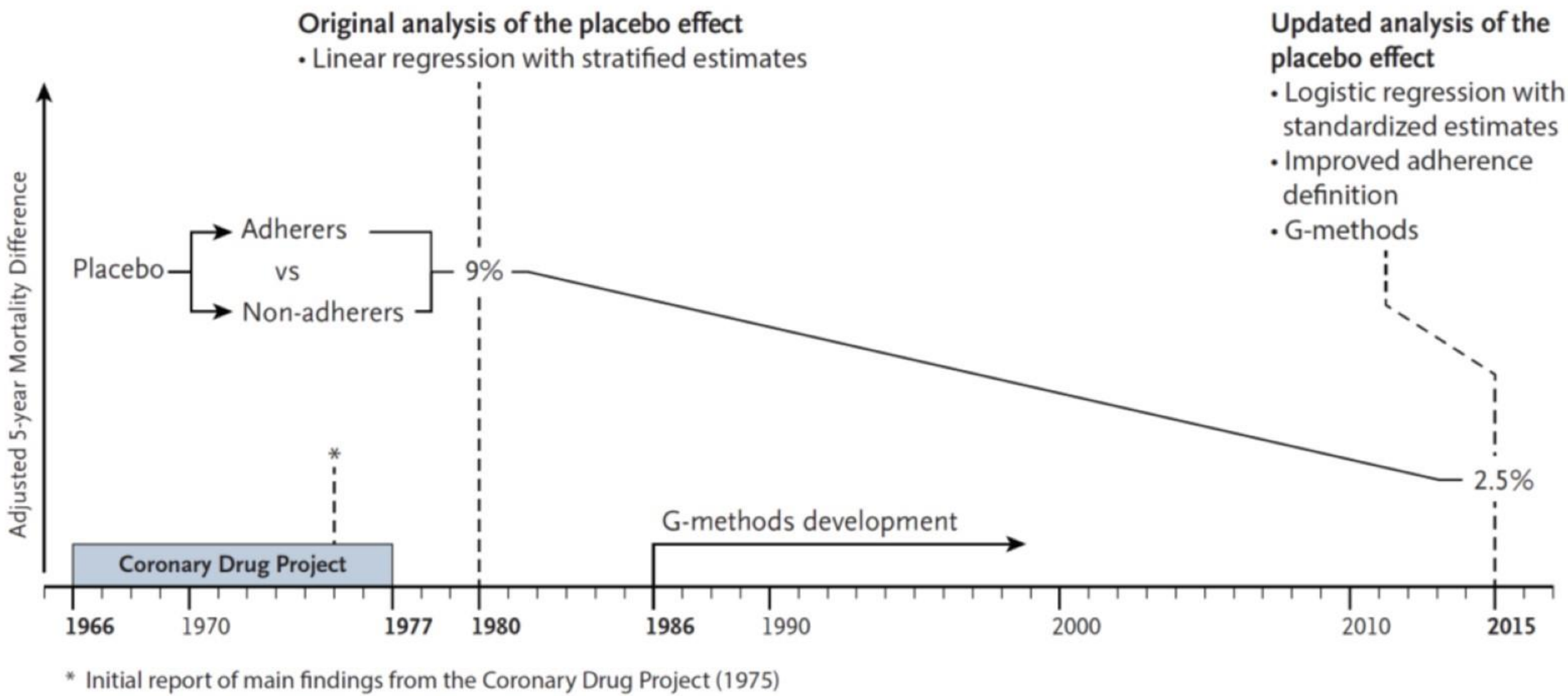


Figure from Hernán, Robins. *NEJM* 2017

So there is some hope for estimation of per-protocol effects

- If there is sufficient information on pre- and post-randomization prognostic factors
  - and sound causal inference methods are used

- Further, in placebo-controlled trials
  - we can indirectly test the assumptions of the per-protocol analysis
  - As we did for CDP and CHARM randomized trials
    - Murray et al. *Contemporary Clinical Trials* 2020

Per-protocol effect is generally a contrast of dynamic treatment strategies

- Not a comparison of
  - "treatment A" vs. "treatment B"
- but a comparison of treatment strategies like
  - "start taking A, continue taking A until toxicity arises or disease progresses, then switch to C (or B)" vs similar for B
    - Hernán, Scharfstein. *Ann Int Med* 2018
- This has 2 important implications

Implication #1: People who stop treatment may be adhering to the protocol

- **The trial's protocol is expected to define the treatment strategies under study**
  - e.g., take assigned treatment until disease progression or toxicity
- **Adherence means "adherence to a treatment strategy"**
  - e.g., a person who stops treatment because of toxicity is adhering to the protocol
- **Censoring people because they stop treatment is incorrect**
  - Need to determine why they are discontinuing treatment
  - If discontinuation is consistent with protocol, do not censor
    - Hernán, Robins. *N Eng J Med* 2017

Implication #2: Conventional adjustment methods may be inadequate for per-protocol analyses

- Because per-protocol effects generally compare treatment strategies that are sustained over time
  - Not point interventions
- Need post-randomization data on adherence and confounders
- G-methods are generally necessary
  - Robins 1986 and later

The unification of causal inference methodology after Robins (1986)

- **A common framework for time-varying treatments**
  - The causal estimands were explicitly defined
- **Methods to estimate per-protocol effects are the same for randomized trials and observational studies**
  - To articulate causal questions when using observational data, we can specify the protocol of the target trial
    - Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *New England Journal of Medicine* 2017
    - Murray EJ, Hernán MA. Improved adherence adjustment in the Coronary Drug Project. *Trials* 2018

And then the Addendum to ICH E9 happened

- Historically, regulators had been reluctant to consider causal effects other than the intention-to-treat effect
- Increasing pressure from industry to consider other effects
- A group of methodologists from the pharmaceutical industry and the regulatory agencies worked together to find a consensus

- The "estimand framework" was born
  - Regulators were willing to move beyond the intention-to-treat effect at last!

The Addendum to ICH E9 made it respectable to talk about non-intention-to-treat effects

- A great step forward for more rigorous causal inference
- Kudos
- But…
  - Hernán MA, Scharfstein D. Cautions as regulators move to end exclusive reliance on intent-to-treat. *Annals of Internal Medicine* 2018

The ICH E9 addendum has some room for future improvement

- Choice of standard terminology (minor issue)
- Choice of estimands relevant for decision making
- Lack of emphasis on treatment strategies
- Lack of emphasis of trial design

ICH E9 addendum
Choice of standard terminology

- **Effect of assignment (intention-to-treat effect) renamed as "treatment policy effect":**
  - all effects are about treatment policies (or strategies)
  - e.g., the per-protocol effect is the effect of adhering to the policy
- **Intercurrent events**
  - Conceptual confusion: "Use of rescue medication", "treatment discontinuation" are components of the treatment policy
  - Death is a competing event (huge literature, why a new name?)
- **Estimand (rather than causal estimand)**
  - Hernán. The C-word. *American Journal of Public Health* 2018

ICH E9 addendum
Choice of estimands

- **While on treatment: Not relevant for patients, clinicians, regulators**
  - "what would happen if people were always treated, regardless of toxicity and efficacy?"
  - Not possible to estimate correctly because of lack of positivity
- **Principal stratum: Not relevant for patients, clinicians, regulators**
  - "what would happen to an unidentified subset of the population?"
  - For point interventions
- **Composite**
  - For rescue medication, conflates definition of treatment strategy with outcome
  - For death, just one possible way of dealing with competing events
    - Young et al. *Statistics in Medicine* 2020
    - Stensrud et al. *Journal of the American Statistical Association* 2020

ICH E9 addendum
Lack of emphasis on treatment strategies

- **From ICH E9(R1) Training materials: the Hypothetical estimand is the effect in**
  - "a hypothetical scenario … in which the intercurrent event would not occur. For example, when rescue medication must be made available for ethical reasons, a treatment effect of interest might concern the outcomes if rescue medication had not been available"
- **This is strange**
  - Aren't we more interested in the effect in an scenario in which people adhere to the strategies described in the protocol?
  - Otherwise, why did we approve the protocol?

ICH E9 addendum
Lack of emphasis on trial design

- Estimation of causal estimands requires adequate data
- We cannot keep designing trials in the same way if we are serious about causal estimands other than intention-to-treat effect
  - For example, to estimate the per-protocol effect we need post-randomization data on adherence and prognostic factors that affect adherence
  - Also, implications for sample size calculations
- Causal estimands as an afterthought?

Causal estimands in randomized trials
Arguably, the most relevant contrasts are

1. Intention-to-treat effect,
   - "treatment policy effect" in ICH E9 jargon

2. Per-protocol effect
   - a type of "hypothetical effect" in ICH E9 jargon

- Observational studies that emulate target trials try to estimate observational analogs of both causal estimands

Key message: We do not compare treatments but treatment strategies

- **Therefore, the causal contrast needs to specify the treatment strategies of interest, including**
  - Treatment initiation
  - Valid reasons for treatment discontinuation
  - Use of concomitant therapies
- **Typically, these will be the treatment strategies specified in the protocol**
  - Observational studies for causal inference already try to estimate well-defined per-protocol effects when emulating target trials
  - It's the turn for randomized trials now

Our research is funded by

- National Institute of Allergy and Infectious Diseases
- National Institute of Mental Health
- National Institute of Child Health and Development
- National Cancer Institute
- PCORI – Patient-Centered Outcomes Research Institute
- U.S. Department of Veterans Affairs

Thank you
@_MiguelHernan

CAUSALab
A Center to Learn What Works

HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH