

Causal Guidelines for Randomized Trials

Beyond intention-to-treat effects

Miguel Hernán

**DEPARTMENTS OF EPIDEMIOLOGY
AND BIostatISTICS**



HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH



@_MiguelHernan

We often talk about “randomized trials”

in opposition to “observational studies”

- As if “randomized trials” were a monolithic group
- But the term “randomized trial” encompasses a great variety of studies
 - With different susceptibility to bias

Typical early randomized trials

- ❑ Highly-controlled experiments
 - ❑ Stringently selected participants
 - ❑ Short duration
 - ❑ Small sample size
 - ❑ No long-term clinical outcomes
 - ❑ For drugs, mostly pre-market trials
-
- ❖ Little deviation from study protocol, high adherence, no losses to follow-up

Typical pragmatic randomized trials

- ❑ Loosely controlled experiments
 - ❑ Typical patients
 - ❑ Long duration
 - ❑ Large sample size
 - ❑ Long-term clinical outcomes
 - ❑ For drugs, mostly post-market trials
-
- ❖ Greater deviations from protocol, low adherence, losses to follow-up

Very different types of randomized trials, should we use the same analysis?

- Highly controlled trials resemble laboratory experiments
- Pragmatic trials resemble observational studies
 - except for baseline randomization of treatment strategies
- The default analysis for all randomized trials is an intention-to-treat analysis
 - Patients assigned to a treatment strategy are kept in that group during the analysis, even if they deviated from their assigned strategy after randomization

Intention-to-treat **analysis** (estimator) estimates intention-to-treat **effect** (estimand)

- Intention-to-treat (ITT) effect
 - The effect of being assigned to a treatment strategy, regardless of treatment received
- ITT effect: agnostic about post-randomization decisions
 - Changes in studied treatment: discontinuation, switching...
 - Use of therapies prohibited by the study protocol
- ITT effect: inadequate as only effect measure in pragmatic trials

For example, consider 2 randomized trials of new treatment vs. standard of care

- In neither study did patients assigned to standard of care take active treatment, but...
- In 1st trial, half of patients assigned to treatment took it
- In 2nd trial, all patients assigned to treatment took it
- An intention-to-treat analysis may show an effect in 1st trial but not in 2nd trial
 - Even if biological effect of treatment is the same in both trials
- Strange, so why do we use intention-to-treat effects?
 - Hernán, Hernández-Díaz. *Clinical Trials* 2012

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

- Consider a non-blinded trial
- The ITT effect may not be null even if treatment has a null effect on the outcome
 - Patients and doctors may just alter their behavior in ways that affect their outcome

- *Most pragmatic trials are not blinded*

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

- When the treatment effect is not monotonic
 - not in the same direction for all individuals
- Trial of active treatment vs placebo
 - 30% of the individuals assigned to treatment did not adhere to treatment
 - direction of the effect in adherers opposite to that in non-adherers

- An ITT analysis may misleadingly indicate a beneficial effect of the less efficacious treatment

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 misleadingly indicate a beneficial effect of the less efficacious treatment

- ❑ *Many pragmatic 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
 - A trial designed to quantify harm and whose protocol foresees only an ITT analysis could be referred to as a ‘randomized cynical trial’
- *Many pragmatic trials are for safety, non-inferiority*

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

- Degree of adherence outside the trial may change drastically after doctors and patients learn of the trial's findings
- Actual effectiveness in the community may differ from ITT effect estimate from trial

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

- A couple trying to decide whether to use a contraceptive method wants to know
 - the effectiveness of the method when used as indicated
 - not the ITT effect in a population in which, say, 40% of couples failed to use the method properly
 - See also Murray et al. *J Clin Epidemiol* 2018; 103:10-21

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

Need a complement to the ITT effect:

- An effect measure (a causal estimand)
 - not affected by the degree of adherence
 - usable in safety, noninferiority trials
 - clinically relevant, patient-centered

- How about the per-protocol effect?
 - the effect of implementing the treatment strategies as described in the protocol

A big difference between ITT effect and per-protocol effect

- Universally accepted way of estimating ITT effects
 - ITT analysis
 - Almost uncontroversial
- No universally accepted way of estimating per-protocol effects
 - Many types of per-protocol analysis
 - Including the common unadjusted, naïve per-protocol analysis

Intention-to-treat effect

Analysis plan

- Simple
- Compare outcome distribution between group assigned to different strategies
 - Regardless of whether individuals actually followed the strategies
- Often overlooked problem:
 - ITT analysis cannot be conducted if there are losses to follow-up
 - Potential selection bias due to informative censoring

Intention-to-treat effect

Analysis plan

- Estimating ITT effect requires adjustment for selection bias due to loss to follow-up
 - Adjustment for baseline and post-baseline covariates
 - Little et al, NEJM 2012
- In fact, intention-to-treat effect is more precisely defined as
 - the effect of being assigned to a strategy, regardless of strategy received, *while staying under follow-up throughout the study*

Per-protocol effect

Analysis plan

- Not so simple
- Treatment decisions after baseline are not randomized
 - Potential post-randomization confounding and selection bias
- Example
 - In a statins trial, statin use after baseline may depend on post-baseline cholesterol levels; dropout may depend on side effects and prognosis

Per-protocol effect Analysis plan

- Estimating the per-protocol effect requires adjustment for confounding
 - Adjustment for baseline and post-baseline covariates
- In addition to adjustment for selection bias
 - same as for ITT effects

Effects (estimands) vs. *analyses* (estimators)

The elephant in the room

- Typical ITT and per-protocol **analyses**
 - adjust for neither pre- nor post-randomization variables
 - Potentially biased estimates of ITT and per protocol **effects**
- This is a problem for all randomized trials
 - because treatment choices and participation decisions after baseline are not randomly assigned
- But especially for pragmatic trials
 - with lots of room for non-adherence and loss to follow-up

What type of strategies are compared in causal estimands of randomized trials?

- Intention-to-treat effect
 - A contrast of point interventions (assignment)

- Per-protocol effect
 - In general, a contrast of sustained treatment strategies (adherence during the follow-up)
 - Further...

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

- Not a comparison of continuous treatment A vs. continuous treatment B
- But a comparison of strategies of the sort
 - “start taking A, continue taking A until toxicity arises, then switch to B”
- Deep implications for definition of per-protocol effect and definition of adherence

Choice of statistical adjustment method depends on type of strategies

- Comparison of strategies involving point interventions only
 - All adjustment methods work
 - if all confounders are measured or the instrumental variable conditions hold
- Comparison of sustained strategies
 - Because of treatment-confounder feedback, generally only g-methods work
 - Developed by Robins and collaborators since 1986

G-methods

- Parametric g-formula
 - Robins 1986
- G-estimation of nested structural models
 - Robins 1989, 1991
- Inverse probability weighting of marginal structural models
 - Robins 1998

- Doubly-robust versions
 - Robins, Vanderlaan, Rotnitzky...
 - e.g., collaborative targeted maximum likelihood estimation

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

- No
 - Even using g-methods, there may be residual post-randomization confounding

- Historically, trialists have been suspicious of per-protocol analyses

- A key reason for those suspicions was the Coronary Drug Project...

The Coronary Drug Project (CDP): a randomized trial in men with MI

- Enrollment: 1966-1969; visits every 4 months
- Double-blind, placebo-controlled
 - 5 active treatments; all but clofibrate discontinued
- 53 study centers, NIH funded
 - 1103 patients assigned to clofibrate; 2789 to placebo
 - three pills daily at randomization and dosage was increased to nine daily pills by the fifth visit
- 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

- Among individuals assigned to placebo, the 5-year mortality risk was higher among those who did not adhere than among those who did adhere to the placebo pills

- This finding is taught in courses around the world
 - often quoted as a reminder of the dangers of analyses that deviate from the intention-to-treat principle.
 - chilling effect on subsequent attempts to conduct per-protocol (observational) analyses in trials

A 21st century update of the CDP analysis

**CLINICAL
TRIALS**

Article

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

Clinical Trials

1-7

© The Author(s) 2016

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/1740774516634335

ctj.sagepub.com

 SAGE

Eleanor J Murray¹ 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)

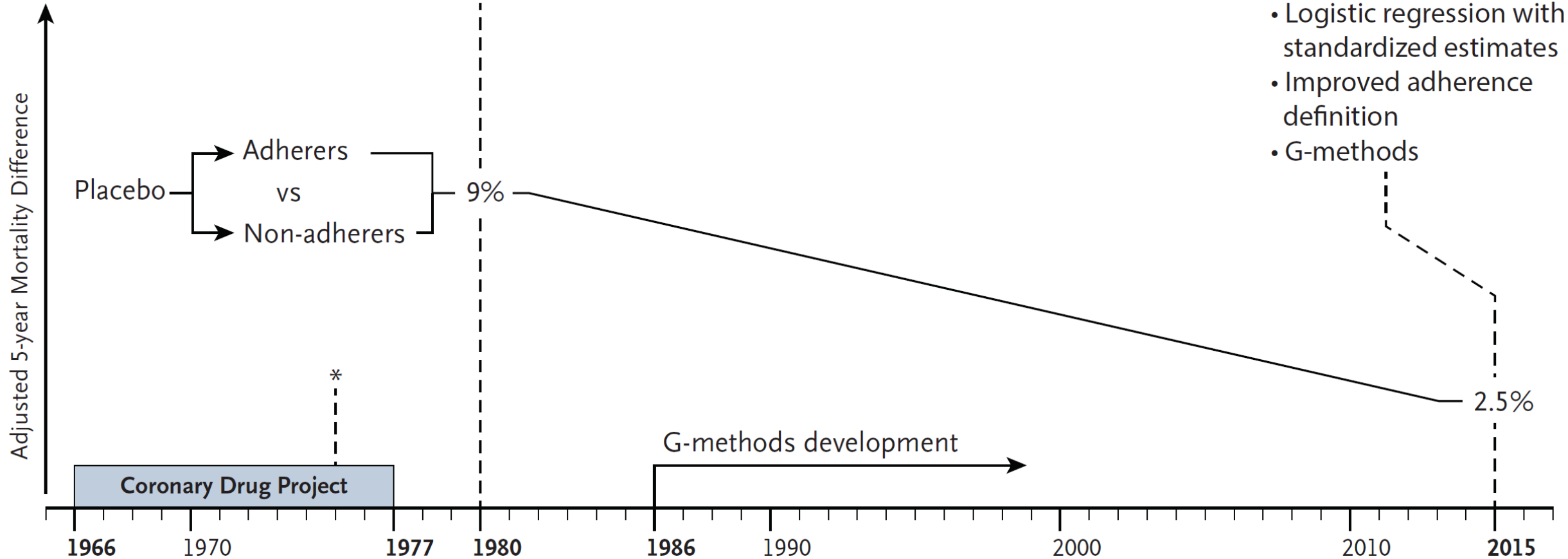
□ for baseline and post-baseline variables

Original analysis of the placebo effect

- Linear regression with stratified estimates

Updated analysis of the placebo effect

- Logistic regression with standardized estimates
- Improved adherence definition
- G-methods



* Initial report of main findings from the Coronary Drug Project (1975)

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

- when there is sufficient information on post-randomization prognostic factors
 - and sound causal inference methods are used
-
- Implications for the design of randomized trials
 - Protocols need to specify the collection of data on prognostic factors after randomization
 - Protocols need to explicitly specify the per-protocol effect
 - Protocols need to pre-specify the per-protocol analysis

Case study

Antiretroviral therapy in HIV-positive patients

Question:

- What is the effect of early initiation of antiretroviral therapy on clinical outcomes of HIV-positive individuals?

Data

- INSIGHT Strategic Timing of AntiRetroviral Treatment (START) randomized trial

START trial: HIV-positive individuals with CD4 cell counts > 500 cells/ml

- 4685 participants randomized to
 - start antiretroviral therapy, ART, immediately after randomization (immediate initiation group) or
 - wait until the CD4 cell count dropped below 350 cells/ml or an AIDS diagnosis (deferred initiation group)
- Composite clinical outcome
- ITT effect estimates for immediate vs. deferred
 - Hazard ratio: 0.43 (0.30, 0.62)
 - 5-year risk difference estimate: -3.1% (-5.2, -0.8)

The per-protocol effect of immediate versus deferred antiretroviral therapy initiation

Sara Lodi^a, Shweta Sharma^b, Jens D. Lundgren^c, Andrew N. Phillips^d,
Stephen R. Cole^e, Roger Logan^a, Brian K. Agan^{f,g}, Abdel Babiker^h,
Hartwig Klinkerⁱ, Haitao Chu^b, Matthew Law^j, James D. Neaton^b,
Miguel A. Hernán^{a,k,l}, on behalf of the INSIGHT Strategic Timing of
AntiRetroviral Treatment (START) study group

AIDS 2016, **30**:2659–2663

Per-protocol effect of immediate vs deferred ART initiation

- Adjustment for pre- and post-baseline covariates via the parametric g-formula
 - Because of treatment-confounder feedback
- Per-protocol effect estimates for immediate vs. deferred
 - Hazard ratio: 0.34 (0.21, 0.52)
 - 5-year risk difference estimate: -3.8% (-6.7, -1.5)

Per-protocol effect estimates about 20% greater than ITT effect estimates

- Not earth-shattering
- But per-protocol estimates are a welcome supplement to ITT estimates
 - For patients, doctors, modelers...
- Interestingly, a naïve per-protocol analysis resulted in approx. same estimates as ITT effect
 - Hazard ratio: 0.41 (0.28, 0.61)

Wouldn't some causal guidelines for randomized trials be helpful?

- Oh, wait
- **Guidelines for estimating causal effects in pragmatic randomized trials**
 - Murray et al. (2019)

Available for public comment at
www.hsph.harvard.edu/causal/pragmatictrials/

Causal Guidelines for Pragmatic Trials

Choice of effect measure

1. Report estimates of both the intention-to-treat effect and the per-protocol effect
 - as well as conditions underlying the estimation
2. Report absolute risks and their differences
 - as well as their ratios (for discrete outcomes)
3. Use the additive scale to report heterogeneity of treatment effects

Causal Guidelines for Pragmatic Trials

Adjustment in intention-to-treat analysis

4. Pre-specify prognostic factors and the maximum acceptable difference in the distribution of these factors between groups
 - adjust via standardization, inverse probability weighting or, preferably, doubly-robust methods
5. In sensitivity analyses, adjust for large imbalances in any important prognostic factors, regardless of whether they have been pre-specified

Causal Guidelines for Pragmatic Trials

Survival analysis with competing events

6. Report both the risk of the competing event by treatment group and the risk of the event interest among those who survived the competing event
7. Specify the intention-to-treat effect as the total effect of treatment assignment on the outcome of interest (the simplest analysis)
 - justify interest in any additional effects that are estimated

Causal Guidelines for Pragmatic Trials

Loss to follow-up

8. Ensure that the trial protocol specifies the collection of post-randomization time-varying prognostic factors that predict loss to follow-up
 - and describes how to adjust for these factors to reduce selection bias

Causal Guidelines for Pragmatic Trials

Per-protocol effect of point interventions

9. Ensure that the trial protocol specifies the collection of baseline prognostic factors that predict adherence
 - and describes how to adjust for these factors to reduce confounding
10. Estimate bounds for the per-protocol effect of point interventions when the instrumental conditions are expected to hold
11. When the three instrumental conditions and monotonicity are expected to hold, discuss whether the effect in the “compliers” is of interest

Causal Guidelines for Pragmatic Trials

Per-protocol effect of sustained strategies

12. Specify a protocol that incorporates real world clinical decision-making, including discontinuation, switching, or dose-reduction rules
13. Ensure that sufficient data are collected to determine whether participants adhered to their assigned strategies throughout the follow-up, and to adjust for time-varying prognostic factors that predict adherence
14. Use g-methods to adjust for time-varying confounders when there is treatment-confounder feedback

Thank you

This work was supported by PCORI ME-1503-28119

- Hernán MA, Scharfstein D. Cautions as regulators move to end exclusive reliance on intent-to-treat. *Annals of Internal Medicine* 2018; 168(7):515-516
- Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *New England Journal of Medicine* 2017; 377(14): 1391-1398
- Hernán MA, Hernández-Díaz S, Robins JM. Randomized trials analyzed like observational studies. *Annals of Internal Medicine* 2013; 159(8): 560-562
- Hernán MA, Hernández-Díaz S. Beyond the intention to treat in comparative effectiveness research. *Clinical Trials* 2012; 9(1):48-55.

Wouldn't some causal guidelines for randomized trials be helpful?

- Oh, wait
- **Guidelines for estimating causal effects in pragmatic randomized trials**
 - Murray et al. (2019)

Available for public comment at
www.hsph.harvard.edu/causal/pragmatictrials/