Evaluating RWE from Observational Studies in Regulatory Decision-Making: Lessons Learned from Trial Replication Analyses

February 16 & 17, 2021

FDA U.S. FOOD & DRUG ADMINISTRATION | Duke MARGOLIS CENTER for Health Policy

Welcome & Overview | Day 1

Mark McClellan

Duke-Robert J. Margolis, MD, Center for Health Policy

Meeting Agenda

Day One

- Session 1: Principles for Using Rigorous Observational Study Designs to Replicate the Results of Clinical Trials
- Session 2: Session 2: Presentations from Trial Replication Projects
- Session 3: Reactions to Replication Results

Day Two

- Session 4: Key Themes Emerging from Replication Efforts
- Session 5: Observational Studies: Opportunities, Limitations, and Next Steps

Virtual Meeting Reminders

- Visit the Duke-Margolis website (https://healthpolicy.duke.edu/events) for meeting materials, including the agenda, speaker biographies, and discussion topics.

- Questions for our panelists? Feel free to submit questions via Zoom's Q&A function.

- Join the conversation @Duke-Margolis #TrialReplication

Opening Remarks from FDA

Jacqueline Corrigan-Curay
U.S. Food and Drug Administration

Thank you to the planning committee!

- **Josie Briggs**, PCORI
- **Bill Crown**, Brandeis University
- **Jessica Franklin**, formerly with Brigham and Women's Hospital/Harvard University
- **Frank Harrell**, Vanderbilt University
- **Adrian Hernandez**, Duke University
- **Joseph Ross**, Yale University
- **Sebastian Schneeweiss**, Brigham and Women's Hospital/Harvard University
- **Nilay Shah**, Mayo Clinic

Session 1: Principles for Using Rigorous Observational Study Designs to Replicate the Results of Clinical Trials

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Miguel Hernán

Harvard T.H. Chan School of Public Health

Causal inference from observational data
Emulating a target trial

Miguel Hernán

DEPARTMENTS OF EPIDEMIOLOGY
AND BIOSTATISTICS

HARVARD T.H. CHAN
SCHOOL OF PUBLIC HEALTH

How do we learn what works and what harms?
(How do we estimate causal effects?)

- The standard scientific answer:
  - Conduct a randomized experiment

- A relevant randomized trial would, in principle, answer each causal question about comparative effectiveness and safety
  - Interference/scaling up issues aside

But we rarely have randomized trials

expensive unethical impractical untimely

- And deferring decisions is not an option
  - no decision is a decision: “Keep status quo”
- What do we do?
  - We analyze observational data

We analyze observational data

because we cannot conduct a randomized trial

Observational analyses are **not** our preferred choice

- For each observational analysis for causal inference, we can imagine a hypothetical randomized trial that we would prefer to conduct
  - If only it were possible

The Target Trial

- The (hypothetical) randomized trial that we would like to conduct to answer a causal question
  - To learn what works and what harms

- A causal analysis of observational data can be viewed as an attempt to emulate some target trial
  - If we cannot translate our causal question into a target trial, then the question is not well-defined

The Target Trial

- Suggested more or less explicitly by many authors
  - Dorn (1953), Cochran, Rubin, Feinstein, Dawid…
  - for simple settings with a time-fixed treatment and a single eligibility point

- Explicit generalization to time-varying treatments and multiple eligibility points
  - Robins (1986)
  - Hernán, Robins. *Am J Epidemiol* 2016

The Target Trial concept leads to a simple algorithm for causal inference

1. Ask a causal question (point at the Target)
   - Specify the protocol of the Target Trial
2. Answer the causal question (shoot the Target)
   - Option A
     - Conduct the Target Trial
   - Option B
     - Use observational data to **explicitly** emulate the Target Trial
     - Apply appropriate causal inference analytics

Step 1
Specify Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- Analysis plan

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- Analysis plan

Not explicitly describing our causal goal is like shooting without a target

AJPH PUBLIC HEALTH OF CONSEQUENCE

Am J Public Health. **2018**;108: 616–619

The C-Word: Scientific Euphemisms Do Not Improve Causal Inference From Observational Data

Causal inference is a core task of science. However, authors and editors often refrain from explicitly acknowledging the causal goal of research projects; they refer to causal effect estimates as associational estimates.

Miguel A. Hernán, MD, DrPH

See also Galea and Vaughan, p. 602; Begg and March, p. 620; Ahern, p. 621; Chiolero, p. 622; Glymour and Hamad, p. 623; Jones and Schooling, p. 624; and Hernán, p. 625.

You know the story:

Dear author: Your observational

Confusion then ensues at the most basic levels of the scientific process and, inevitably, errors are

glass of red wine per day versus no alcohol drinking. For simplicity, disregard measurement error and

Ok, so why is this a big deal?

- Why do we need to explicitly emulate a target trial for causal inference from observational data?

- Because not doing so leads to bias
  - Deviations from the target trial are sources of bias in observational analysis

- Let's review 3 examples

Important
Target trial must be a pragmatic trial

- Observational data cannot be used to emulate
  - a placebo-controlled trial
    - at most a trial with a “usual care” group
  - a trial with blind design
    - individuals are generally aware of the treatment they receive
  - treatment strategies that do not exist in the real world
  - enforcement of adherence to the protocol
  - tight monitoring that doesn’t happen in the real world

EXAMPLE #1 (a classic)
Postmenopausal hormone therapy and heart disease

- Observational epidemiologic studies
  - >30% **lower risk** in current users vs. never users
    - e.g., hazard ratio: 0.68 in Nurses' Health Study
      - Grodstein et al. *J Women's Health* 2006
- Randomized trial
  - >20% **higher risk** in initiators vs. noninitiators
    - hazard ratio: 1.24 in Women's Health Initiative
      - Manson et al. *New England J Med* 2003

EXAMPLE #1
What was the problem?

- The randomized trial compared
  - initiators (incident users) vs. noninitiators
- Observational studies compared
  - Current (prevalent) users vs. nonusers
  - Current users were depleted of susceptibles so current use became a marker of not being susceptible

- Solution: observational re-analysis that compared
  - initiators (incident users) vs. noninitiators
  - Hernán et al. *Epidemiology* 2008

EXAMPLE #2
Statins and cancer

- Observational studies reported an association between statins and lower cancer risk
  - some studies found an implausible 50-65% lower risk
- Subsequent analyses of randomized trials: No effect

- Confounding bias due to lack of randomization?
  - Unlikely because cancer was not an intended effect of treatment

EXAMPLE #2
Statins and cancer

- We explicitly emulated a target trial of statins and cancer using electronic health records
  - Linked CPRD primary care electronic health records accessed through the CALIBER resource
  - Dickerman et al. *Nature Medicine* 2019

- First, we specified the protocol of the target trial

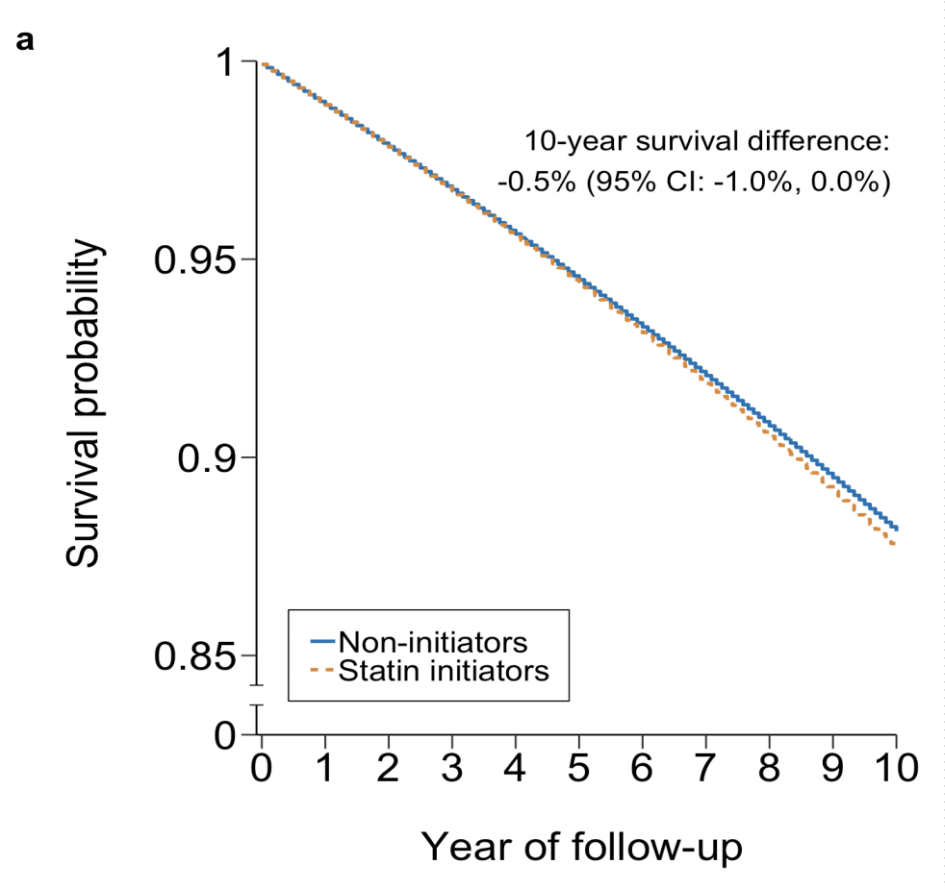
Summary of Protocol of Target trial Statins and cancer	
Eligibility criteria	Individuals aged ≥30 in January 1998-February 2016 with no history of cancer; no statin use in previous year; no statin contraindication (hepatic impairment, myopathy) LDL cholesterol <5 mml/L; at least 1 year of up-to-standard data in a CPRD practice.
Treatment strategies	1. Initiation of any statin therapy at baseline and continuation over follow-up until the development of a contraindication 2. No initiation of statin therapy over follow-up until the development of an indication
Assignment procedures	Participants are randomly assigned to either strategy at baseline, and are aware of the strategy they have been assigned to.
Follow-up period	Starts at randomization and ends at diagnosis of cancer, death, loss to follow-up, or February 2016, whichever occurs earlier.
Outcome	Total cancer and 7 site-specific cancers
Causal contrasts	Intention-to-treat effect, per-protocol effect
Analysis plan	Intention-to-treat analysis, non-naïve per-protocol analysis

CALIBER emulation: Hazard ratio estimates for statin vs. no statin

- Total cancer: 1.02 (0.99, 1.05)

- Breast cancer: 1.00 (0.92, 1.09)
- Colorectal: 1.04 (0.95, 1.13)
- Lung: 1.08 (0.99, 1.17)
- Prostate: 1.02 (0.95, 1.09)
- …
  - these are intention-to-treat HRs, per-protocol HRs are similar

CALIBER emulation: Survival estimates for statin vs. no statin



No beneficial effect of statins?
What about previous observational studies?

Previous study:
Odds ratio of lung cancer: 0.23 (0.20, 0.26)

for long-term users (>4 years) vs nonusers

- Two key deviations from the target trial:.
  1. included prevalent users at baseline
  2. using postbaseline information (observed duration of statin therapy) to assign baseline treatment status)
- When we did this in our data, the hazard ratio was:
  - 0.23 (0.22, 0.24) for total cancer
  - 0.27 (0.25, 0.29) for lung cancer

Aside: Case-control studies also benefit from emulating a target trial

- To emulate a target trial with case-control data
  1. specify the protocol of the target trial
  2. define the cohort study that explicitly emulates it
  3. sample cases and controls from that cohort
- Case-control analyses that deviated from this approach found strong inverse associations between statins and cancer
  - For details, see Dickerman et al. *Int J Epidemiol* 2020

In these examples, the problem with the observational studies was **not** confounding

(similarly in many other examples)

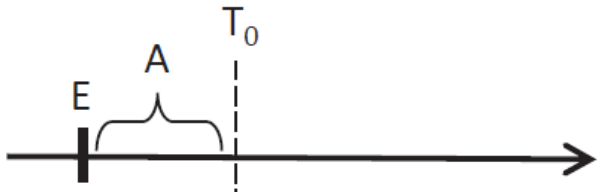
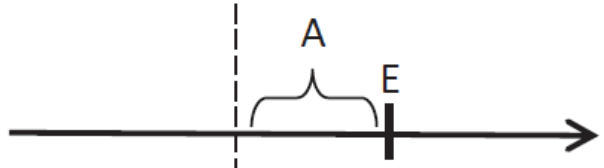
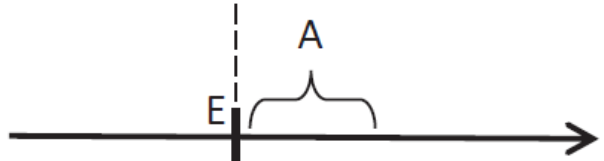
- Yet criticisms of observational analyses often focus on lack of randomization
  - even if the problem has nothing to do with lack of randomization
- Many observational analyses have a more fundamental problem
  - Failure to choose a correct time zero

Time zero of follow-up in the Target Trial

- For each person, the time when 3 things happen
  - eligibility criteria are met
  - treatment strategies are assigned
  - study outcomes begin to be counted
- The same applies to observational analyses

- Misalignment of eligibility criteria and treatment assignment leads to selection bias / immortal time bias
  - Hernán et al. *J Clin Epidemiol* 2016; 79:70-75.

Misalignment of eligibility (E) and treatment assignment (A) prevents correct emulation

Type of emulation failure		Selection of…	Immortal time
1. T_0 after E and A		eligible individuals who initiate a treatment strategy and remain under follow-up through reset T_0	No
2. T_0 at E but before A		individuals who initiated a treatment strategy before, and remained under follow-up until, eligibility (specified at T_0)	No
3. T_0 before E and A		individuals who initiated a treatment strategy before, and remained under follow-up until, eligibility (specified after T_0)	Yes
4. T_0 at E but before A		eligible individuals at T_0 who remained under follow-up until completing a treatment strategy	Yes

Hernán et al.
J Clin Epidemiol
2016; 79:70-75

2 key components of the emulation of the target trial

1. Randomized assignment
   - Emulation requires adjustment for confounding
2. Specification of time zero
   - Time zero must be synchronized with determination of eligibility and assignment of treatment strategies

- Lack of randomization is usually blamed for the failings of observational analyses, but…
  - we have seen that incorrect specification of time zero is often the actual culprit

Step 1
Specify Target Trial protocol

- Eligibility criteria
- Outcomes
- Causal contrast
- Analysis plan

Choosing time zero correctly: The low-hanging fruit for causal inference

Step 2
Emulate Target Trial protocol

- Eligibility criteria
- Treatment strategies
- Randomized assignment
- Start/End follow-up
- Outcomes
- Causal contrast
- Analysis plan

So does that mean that lack of randomization is Ok?

- No

- Confounding due to lack of randomization always possible when using observational real world data
- Explicitly emulating the target trial only eliminates self-inflicted injuries
  - Selection bias, immortal time bias…
  - Confounding is not a self-inflicted injury

Failures in the emulation of randomized assignment

- Treatments that are proxies for prognostic factors that remain unmeasured
  - Example: Preventive interventions (e.g., screening colonoscopy) and mortality
    - Garcia-Albeniz et al. *Am J Epidemiol* 2019
  - Unmeasured confounding: biased effect estimate
- Treatments that are universally administered to individuals with certain prognostic factors
  - Example: antihypertensives vs no antihypertensives
    - Danaei et al. *J Clin Epidemiol* 2018
  - Intractable confounding: biased effect estimate

Observational data to emulate target trials similar to actual trials?

- Not very useful in itself
  - if we already know the answer from actual trials, why spend time trying to replicate them?

- But attempting to replicate trials helps us understand under which circumstances target trial emulation is possible
  - e.g., can't use claims data for preventive interventions that reduce mortality
- and to design better trials

EXAMPLE #3
Tocilizumab for COVID-19

- Tocilizumab
  - humanized monoclonal antibody against interleukin 6 (IL-6) receptor
- Early observation from China
  - Increased death risk in COVID-19 patients with elevated IL-6 levels
- Spring 2020: No randomized trials
  - Off-label use common in many hospitals for COVID-19 patients with evidence of hyperinflammation

Emulation of target trial of tocilizumab STOP-COVID Observational Study

- 3924 individuals with COVID-19 admitted to ICU
  - 68 U.S. hospitals
  - Gupta et al. *JAMA Internal Medicine* 2020

JAMA Internal Medicine | Original Investigation

Association Between Early Treatment With Tocilizumab and Mortality Among Critically Ill Patients With COVID-19

Shruti Gupta, MD, MPH; Wei Wang, PhD; Salim S. Hayek, MD; Lili Chan, MD, MSCR; Kusum S. Mathews, MD, MPH, MSCR; Michal L. Melamed, MD, MHS; Samantha K. Brenner, MD, MPH; Amanda Leonberg-Yoo, MD, MS; Edward J. Schenck, MD, MS; Jared Radbel, MD; Jochen Reiser, MD, PhD; Anip Bansal, MD; Anand Srivastava, MD, MPH; Yan Zhou, MD; Diana Finkel, DO; Adam Green, MD, MBA; Mary Mallappallil, MD; Anthony J. Faugno, MD; Jingjing Zhang, MD, PhD; Juan Carlos Q. Velez, MD; Shahzad Shaefi, MD, MPH; Chirag R. Parikh, MD, PhD; David M. Charytan, MD, MSc; Ambarish M. Athavale, MBBS, MD; Allon N. Friedman, MD; Roberta E. Redfern, PhD; Samuel A. P. Short, BA; Simon Correa, MD, MMSc; Kapil K. Pokharel, MBBS; Andrew J. Admon, MD, MPH, MSc; John P. Donnelly, PhD; Hayley B. Gershengorn, MD; David J. Douin, MD; Matthew W. Semler, MD; Miguel A. Hernán, MD, DrPH; David E. Leaf, MD, MMSc; for the STOP-COVID Investigators

Emulation of target trial of tocilizumab Findings

- 30-day mortality
  - 27.5% in the tocilizumab group
  - 37.1% in the non-tocilizumab group
  - Risk difference: 9.6% (95% CI 3.1%-16.0%)
- Hazard ratio: 0.71 (95% CI 0.56-0.92)
  - If admitted to the ICU within 3 days of symptom onset: 0.41 (95% CI: 0.23-0.74)
  - If admitted to the ICU after 3 days of symptom onset: 0.85 (95% CI: 0.65-1.11)

This observational study emulated a target trial that didn't exist yet

- It wasn't taken seriously by many journal editors
  - First submitted to a journal in May
  - A round of rejections
    - Journal 1: "I am sorry to say it was not accepted for publication. This was an editorial decision […] the decision was to wait for actual trials."
    - Journal 2: "there was concern that there was a high risk of residual confounding. None of the randomized clinical trials that are beginning to report out have found such an effect. You did an excellent job analyzing the observational data, but in the end, there was a credibility problem."
  - Published in October

This observational study emulated a target trial that didn't exist yet

- It wasn't taken seriously by many guidelines writers
  - recommended against use during much of the pandemic
- National Institutes of Health
  - COVID-19 Treatment Guidelines Panel's Statement on the Use of Tocilizumab for the Treatment of COVID-19
  - "Brief Summary of Evidence" didn't even mention the observational studies (as of February 3, 2021)
    - https://www.covid19treatmentguidelines.nih.gov/statement-on-tocilizumab/

Randomized trials published in 2021 confirmed the findings from the target trial emulation

- Surprising?
  - Not at all
- Tocilizumab was a poster child for when target trial emulation can work
  - Large causal effect
  - Data with rich information on confounders
  - Residual confounding in a "conservative" direction
  - (+ sound design of data analysis and time zero handling)

Yet observational analyses were ignored

- By journal editors, guideline writers, regulators

- Without considering the studies on its own merits
  - just because they were observational
- During a public health emergency
  - with an alarming scarcity of effective treatments

- That's how biased we are against observational studies
  - The legacy of so many bad observational analyses that didn't even try to emulate a target trial

The Target Trial concept leads to a simple two-step algorithm for causal inference

1. Ask a causal question (point at the Target)
   - Specify the protocol of the Target Trial
2. Answer the causal question (shoot the Target)
   - Option A
     - Conduct the Target Trial
   - Option B
     - Use observational data to **explicitly** emulate the Target Trial
     - Apply appropriate causal inference analytics

Every time someone presents observational estimates to estimate causal effects, ask

"What is the target trial?"

- If they look puzzled, help them specify the target trial
- If no target trial can be identified, ask them to start over

- Only after we know the question, we can evaluate the methods used to obtain the answer

Adrian Hernandez

Duke University

Nandita Mitra

University of Pennsylvania

Jennifer Graff

National Pharmaceutical Council

Gerald Dal Pan

U.S. Food and Drug Administration

Break — 20 Minutes

We will be back momentarily.

Session 2 will begin at 2:45 pm (U.S. Eastern).

Duke-Margolis is hiring!

Are you interested in real-world evidence, payment for medical products, or antimicrobial resistance?

We have multiple openings. Please visit https://healthpolicy.duke.edu/careers to learn more about opportunities.

Session 2: Presentations from Trial Replication Projects

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Sebastian Schneeweiss

Harvard Medical School

Emulating randomized clinical trials with non-randomized real-world evidence studies

Results from the RCT DUPLICATE* initiative

Sebastian Schneeweiss, MD, ScD
Professor of Medicine and Epidemiology

Jessica Franklin, PhD
Associate Professor of Medicine

Shirley Wang, PhD
Assistant Professor of Medicine

Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine
Brigham and Women's Hospital, Harvard Medical School, Boston

*Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology

This study was funded by FDA under contracts HHSF223201710186C and HHSF223201710146C

Disclosures Dr. Schneeweiss

- PI, Sentinel Innovation Center (FDA)
- Co-Chair, Partners Center for Integrated Healthcare Data Research
- PI of grants and contracts from NIH, AHRQ, PCORI, FDA, IMI, Arnold Foundation
- Investigator of research grants awarded to BWH by Boehringer Ingelheim
- Consulting fees from Aetion, Inc. (incl. equity)

Circulation
JOURNAL OF THE AMERICAN HEART ASSOCIATION
American Heart Association. Learn and Live.

Jessica M. Franklin[1], Elisabetta Patorno[1], Rishi J. Desai[1], Robert J. Glynn[1], David Martin[2], Kenneth Quinto[2], Ajinkya Pawar[1], Lily G. Bessette[1], Hemin Lee[1], Elizabeth M. Garry[3], Nileesa Gautam[1], Sebastian Schneeweiss[1]

1. Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA
2. Office of Medical Policy, Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, MD, USA
3. Scientific Research, Aetion, Inc., Boston, MA, USA

Dec. 18, 2020

Harvard study team:

Faculty: Drs. Schneeweiss, Franklin, Wang, Glynn, Patorno, Desai, Choudhry, Huybrechts, Fischer, Feldman, Gagne, Bykov

Research Staff: Dr. Pawar, Besette, Lee, Gautham, Chin, Dr. D'Andrea, Dr. Gopalakrishna, Jawaid, Jin, Lee, Dr. Mahesri, Sears, Tesfaye, Umarje, York, Zabotka, Zakoul

Aetion team: Drs. Garry, Rassen, and Isaman, Gibbs, Gilpin

Much thanks to our colleagues from FDA: Drs. Martin, Quinto, Concato, Corrigan-Curay, Paraoan

Expert advisor panel:*

Drs. Steve **Goodman**, Stanford; Miguel **Hernan**, Harvard; Wayne **Ray**, Vanderbilt; Samy **Suissa**, McGill; Alan **Brookhart**, Duke

*While we are most grateful for the advice we received, the authors are solely responsible for the presented work

Real-World Evidence (RWE) studies

Randomized controlled trials are an accepted design to establish the efficacy of medical products

RWE is based on data produced by the routine operation of the healthcare system

It is thought to complement and expand the evidence generated by RCTs and often expands the line of inquiry into

- Different populations
- Different treatment patterns
- Different endpoints
- Different comparators

Can RWE studies estimate causal treatment effects?

We wish to calibrate RWE findings against the true causal treatment effect

-> Can we ever know the true treatment effect in a given population?

If not, what is the next best thing?

- Relying on expert opinion – no!
- Statistical simulation studies – no!
- Comparisons against RCT findings:

  Based on the assumption that a well-planned and well executed RCT is accepted as having a causal interpretation – possibly?

Why is this so important?

If RWE cannot estimate causal treatment effects, what is the point of doing RWE?

What some RWE proponents say:

"RWE studies answer different questions than RCTs and therefore you should never expect the same findings," "you should not compare; it may backfire"

Translates to: "We can never test the validity of RWE because we don't have an agreeable gold standard to test against"

> So where does that leave us? With the conclusion that, for RWE, there is no real upside to the RCT replication endeavor—only downside.
>
> *David Thompson, Value Health 2021*

Karl Popper noted that if a hypothesis evades testability it is not a viable hypothesis.

What some RCT proponents say:

"RWE studies have never been able to convincingly demonstrate that they have causal conclusions like RCTs have"

Translates to: "The bar is set high and we are open to listen but doubt that RWE will ever be trusted"

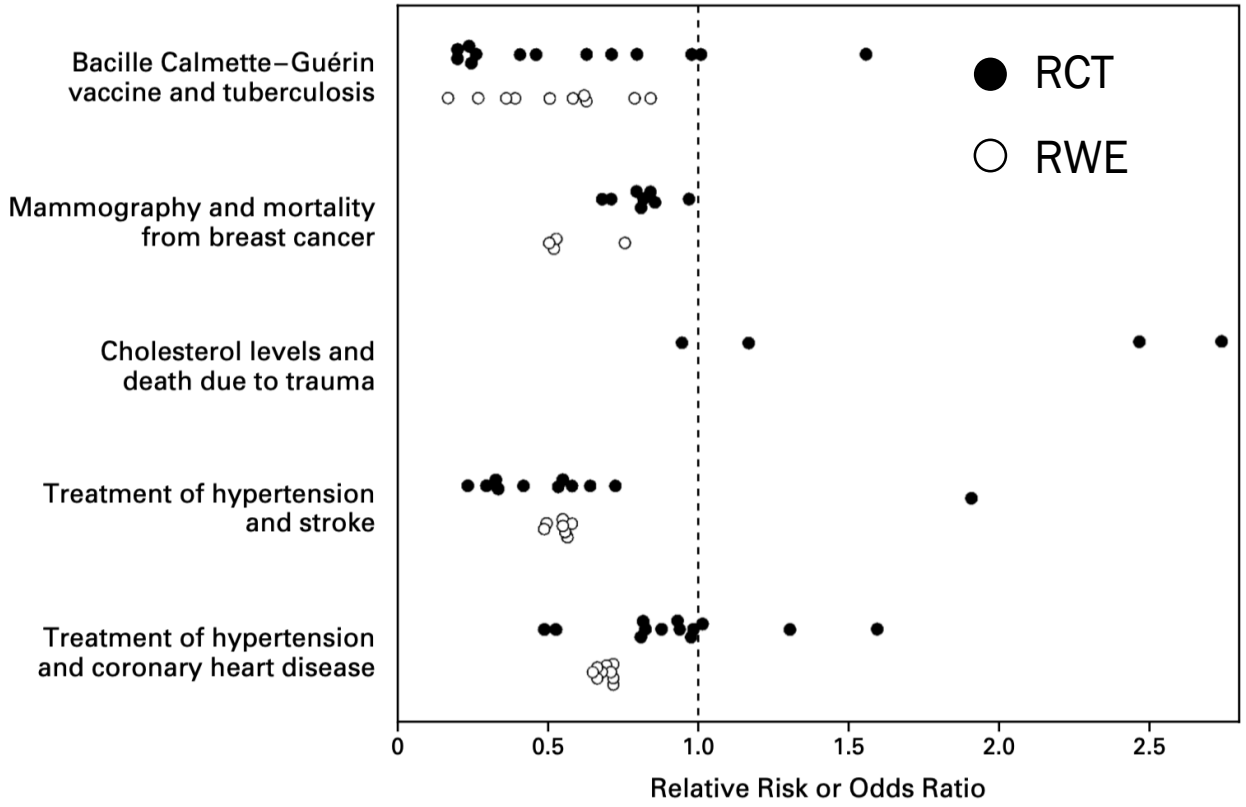
What we don't mean:

We don't want to imply that all RWE studies need to calibrate against an RCT – that would defeat the purpose of RWE as it is meant to complement RCT evidence

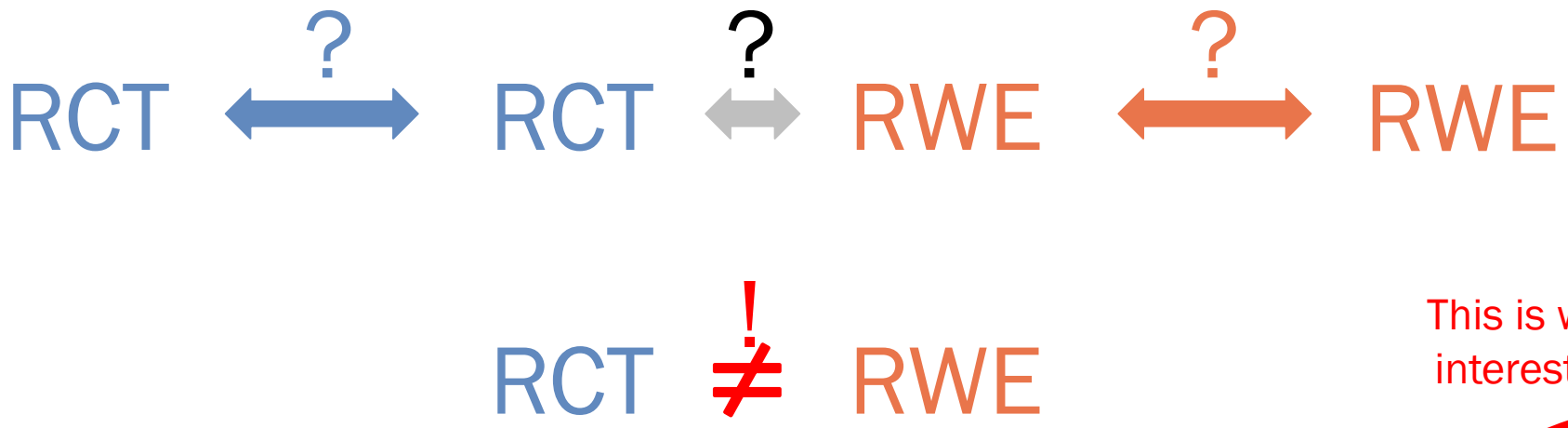
Variability in RCT-to-RCT and in RWE-to-RWE comparisons

RANDOMIZED, CONTROLLED TRIALS, OBSERVATIONAL STUDIES, AND THE HIERARCHY OF RESEARCH DESIGNS

JOHN CONCATO, M.D., M.P.H., NIRAV SHAH, M.D., M.P.H., AND RALPH I. HORWITZ, M.D.



Considerations of RCT emulation



This is what we are really interested in quantifying

1) Agreement with what?

- How variable are RCT results?
- What is the true treatment effect in the study population?

…

2) Emulation failure?

- Different population
- Different treatment pattern
- Different outcome measure
- Different follow-up duration

…

3) Bias?

- Confounding
- Differential surveillance
- Time-related biases

…

Franklin JM, Glynn RJ, Suissa S, Schneeweiss S. CPT 2020

Range of RCT emulation successes by RWE studies

Retrospective emulations

Targeted emulations

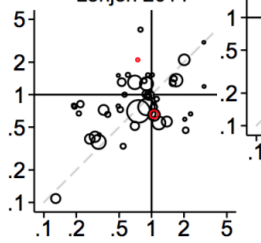
Post-hoc re-weighting, Double-randomized

Best

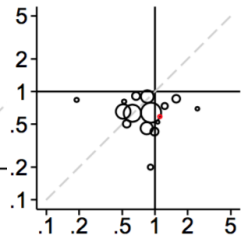
Worst

Expected Agreement

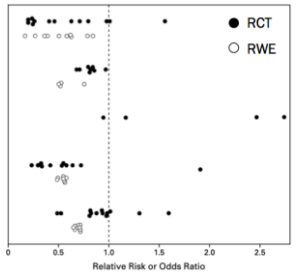
Lonjon 2014



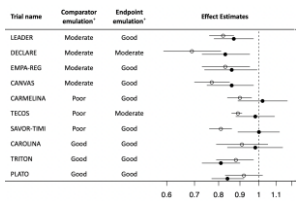
Hemkens 2016



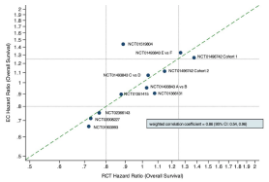
Concato 2000



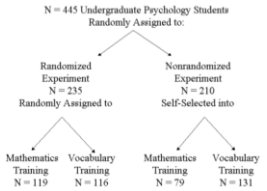
RCT-DUPLICATE 2020



Carrigan 2020



Shadish 2008



Less

Attention to RWE study quality and emulation

More

RCT-DUPLICATE objectives

Aimed to understand and improve the validity of RWE studies for regulatory decision making

Clinical Pharmacology & Therapeutics 2020 Apr;107(4):817-826

Nonrandomized Real-World Evidence to Support Regulatory Decision Making: Process for a Randomized Trial Replication Project

Jessica M. Franklin[1,*], Ajinkya Pawar[1], David Martin[2], Robert J. Glynn[1], Mark Levenson[3], Robert Temple[4] and Sebastian Schneeweiss[1]

1

Replicate **30 RCTs** and predict **7 RCTs** considered by FDA

Learnings:
Had we replaced an RCT with a single RWE study would we have come to the same decision?

2

Test a **process** with FDA to conduct and submit RWE studies

Learnings:
Can we successfully enable transparent and reproducible RWE and enable regulators to re-analyze data?

3

Factors that predict replication success, causal estimates

Learnings:
Identify factors that predictably increase validity of RWE studies.

Franklin, Pawar, Martin, Glynn, Levenson, Temple, Schneeweiss. CPT 2020

Data sources

U.S. longitudinal claims data

- Enrollment and disenrollment dates
- Patient-level information on visits, hospitalizations, pharmacy fills, death
- Including service date, diagnoses, procedures, and drug ingredients

- Optum Clinformatics: Commercial, incl. Medicare Advantage
- IBM MarketScan: Commercial, incl. Medicare Advantage
- Medicare FFS: Beneficiaries 65 years and older

RCT selection strategy: Breadth

1. Mix of regulatory submissions:
    1. Primary approvals
    2. Supplemental approvals
    3. Negative trials
    4. FDA special interest
2. Mix of therapeutic areas
3. Mix of comparator: Placebo, active
4. Mix of hypothesis testing intention: Superiority, non-inferiority

RCT selection strategy: Data fit-for-purpose

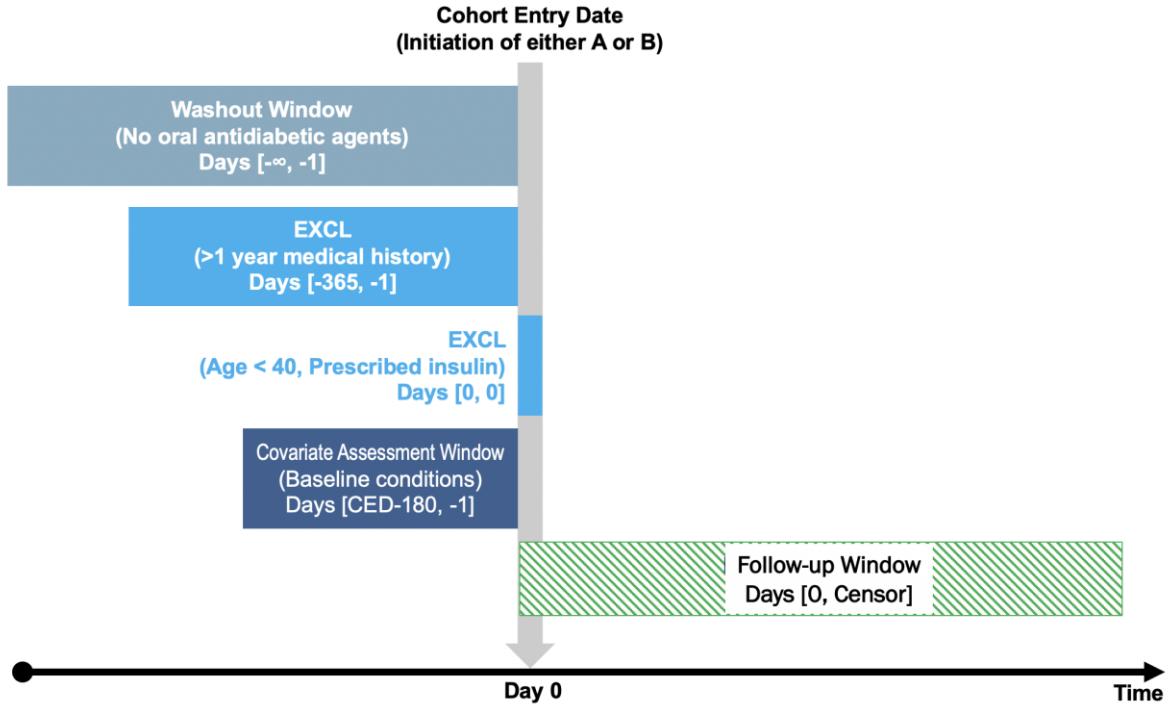
5. Outcome observable?
6. Treatment observable?
7. Key inclusion criteria observable?
8. Key exclusion criteria observable?
9. Key pre-exposure outcome predictors observable?

RWE study design and analysis strategy

1. Emulate the target trial -> new-user active-comparator cohort study
2. Emulate inclusion/exclusion criteria as best as possible given the data
3. Adjustment for baseline imbalances using 1:1 propensity score matching on >100 pre-exposure covariates
4. Validated outcome definitions when available w/ focus on highly specific definitions
5. We wanted to emulate an RCT ITT analysis with perfect compliance (>90%); in light of suboptimal real-world adherence we used an on-treatment analysis
6. One single pre-defined analysis
7. A single investigator team plus clinical and methodological advisors
8. Few sensitivity analyses if any for this iteration

Process and feasibility

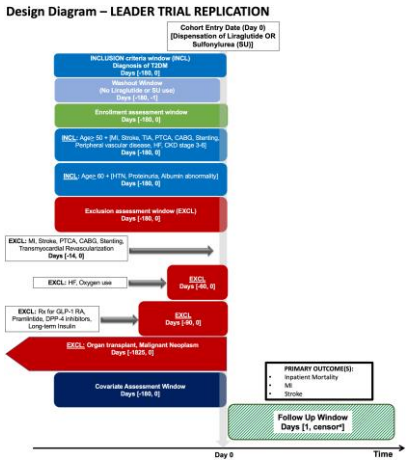
- Trial Design
  - Treatment arms
  - Population and exclusions
- **RWE Emulation Study Design**
- Feasibility: power
- Feasibility: baseline balance



Target trial emulation:

New-user active-comparator cohort study

Clear temporality, no adjustment for intermediates, no immortal time bias, no depletion of susceptibles



This is an example;
all details are on clinicaltrials.gov

Process and feasibility

- Trial Design
  - Treatment arms
  - Population and exclusions
- RWE Emulation Study Design
- Feasibility: power
- Feasibility: baseline balance

This is an example;
all details are on clinicaltrials.gov

	MarketScan		Optum		Medicare	
	Sacubitril/ Valsartan	ACEi	Sacubitril/ Valsartan	ACEi	Sacubitril/ Valsartan	ACEi
Unmatched						
N Patients	1,476	2,218	2,729	4,217	1,738	6,293
N Outcomes	592		1,435		1,992	
Follow Up	111	118	92	99	86	81
Matched						
N Patients	743	743	1278	1,278	1,008	1,008
N Outcomes	--		--		--	
Follow Up	137	126	109	118	107	102

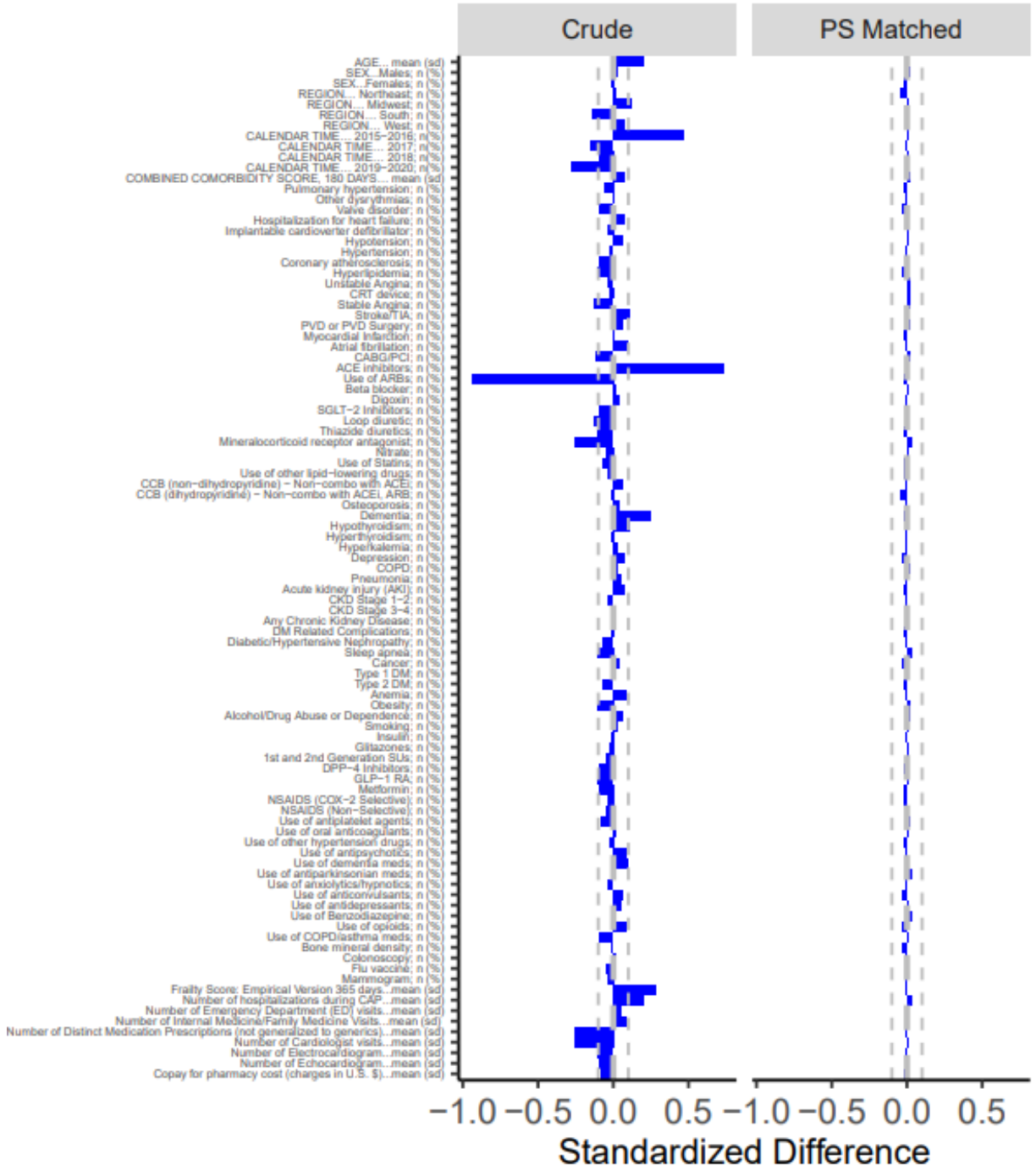
	MarketScan	Optum	Medicare	Pooled
# Matched patients	1,486	2,556	2,016	6,058
Risk per 1,000 patients	160.3	206.6	248.3	215.3
Desired HR from RCT	0.8	0.8	0.8	0.8
Alpha (2-sided)	0.05	0.05	0.05	0.05
Number of events expected	238	528	501	1,304
Power	0.41	0.73	0.70	0.98

Process and feasibility

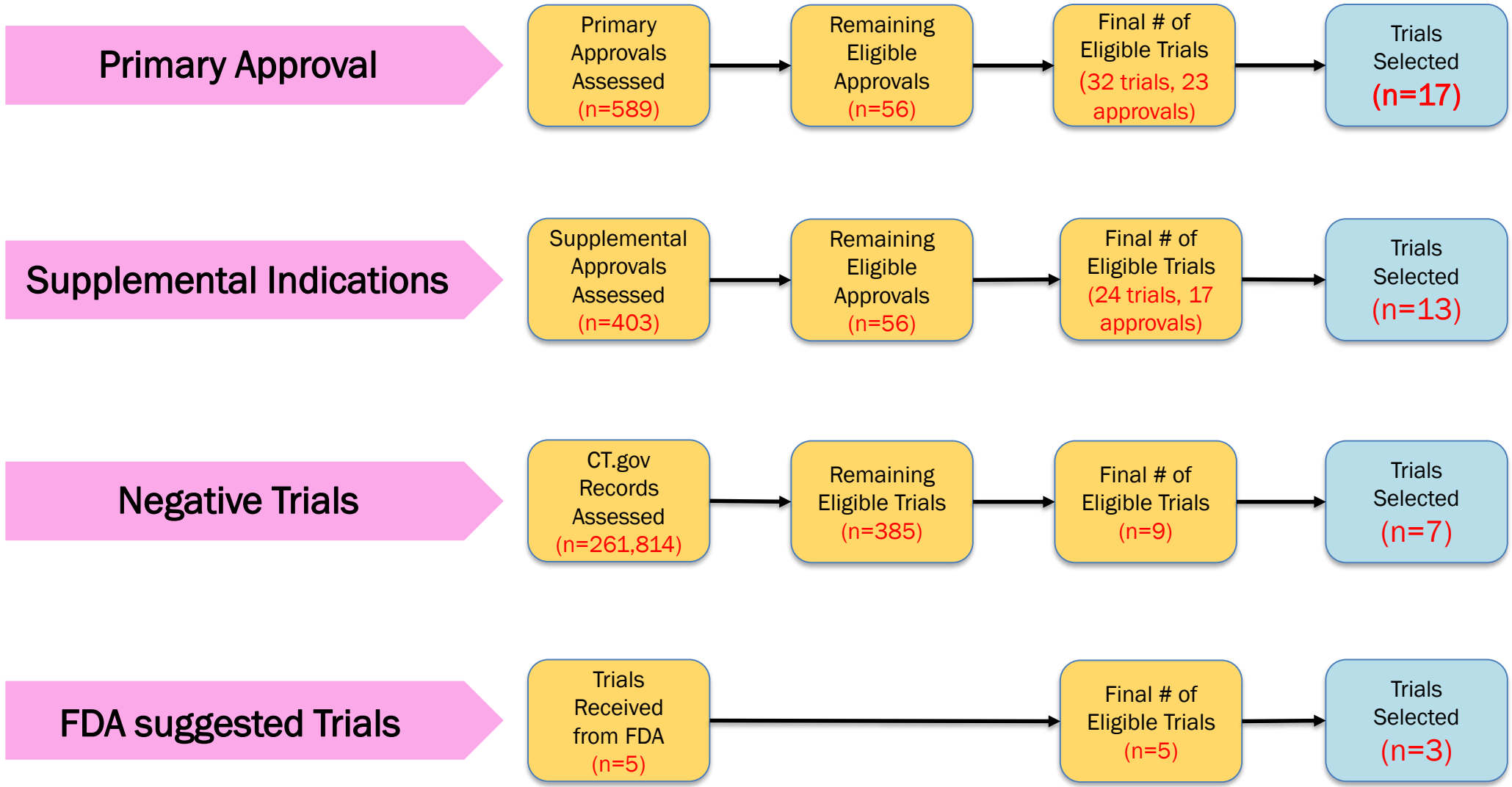
- Trial Design
  - Treatment arms
  - Population and exclusions
- RWE Emulation Study Design
- Feasibility: power
- Feasibility: baseline balance

Adjusted for >100 pre-exposure covariates:

- Demographics, region, calendar time, disease risk score
- CVD and non-CVD comorbidities
- CVD and non-CVD medications
- Proxies of healthcare utilization, SES

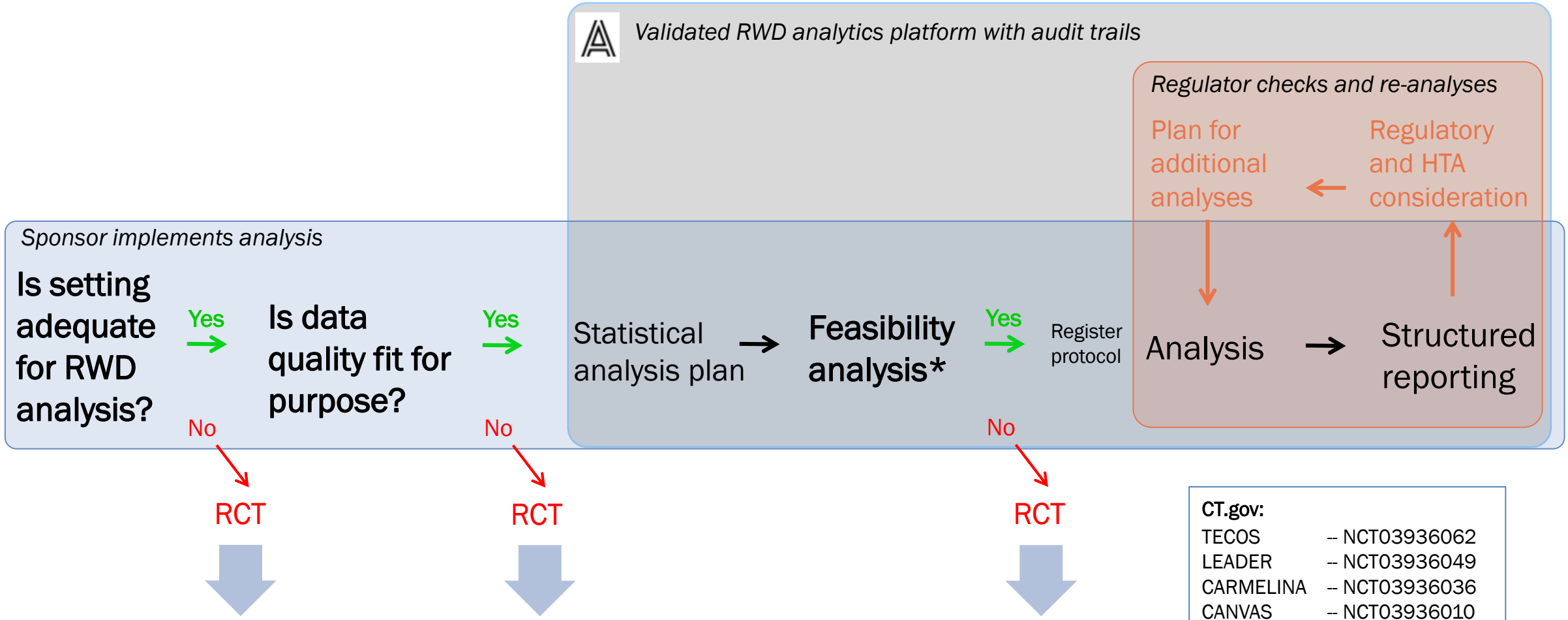


Selecting 30 regulatory-standard RCTs for replication



Franklin, Pawar, Martin, Glynn, Levenson, Temple, Schneeweiss. CPT 2020

A pathway with regulatory validation



CT.gov:

TECOS	-- NCT03936062
LEADER	-- NCT03936049
CARMELINA	-- NCT03936036
CANVAS	-- NCT03936010
SAVOR-TIMI	-- NCT03936023

Franklin, Glynn, Martin, Schneeweiss. CPT 2019

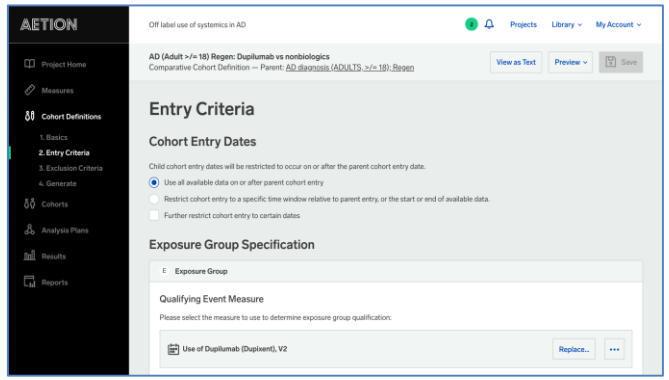
* Feasibility analysis can include 1) checking covariate balance after applying the chosen confounding adjustment strategy, 2) checking statistical power, 3) evaluating positive or negative control outcomes, and 4) other analyses, *without* evaluating the study outcomes in the two treatment groups.

Examples for dropping RCTs during feasibility check

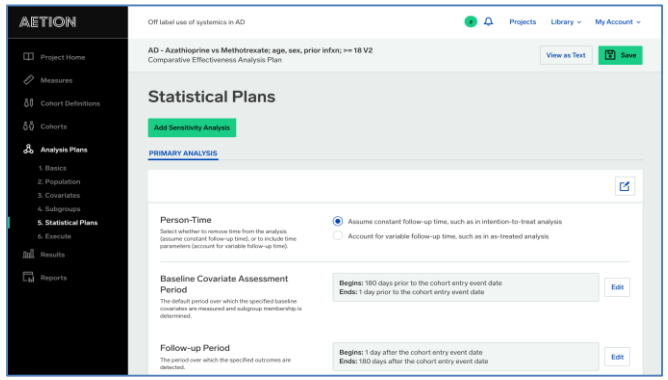
Trial Group	Trial Name	Reason for Dropping
Antiplatelet	CLARITY-TIMI 28	Assessed treatments given during hospitalization and cannot be emulated with outpatient dispensing data
Antiplatelet	COMMIT	Assessed treatments given during hospitalization and cannot be emulated with outpatient dispensing data
Antiplatelet	TRA 2P - TIMI 50	Low number of vorapaxar users
Antiplatelet	PROFESS	Low number of aspirin/dipyridamole users
Antiplatelet	PEGASUS-TIMI	Low number of patients using ticagrelor beyond 1 year after myocardial infarction

Study implementation with the Aetion Evidence Platform to reduce error, increase transparency

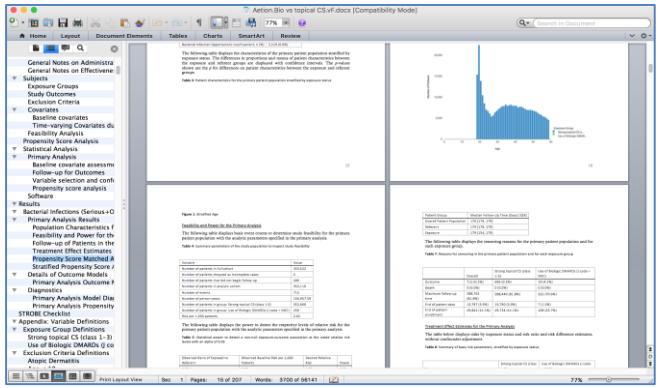
Select **patients** in reproducible ways



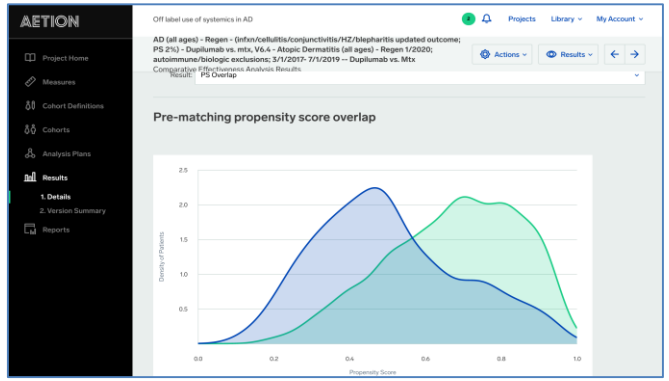
Select **comparison groups**



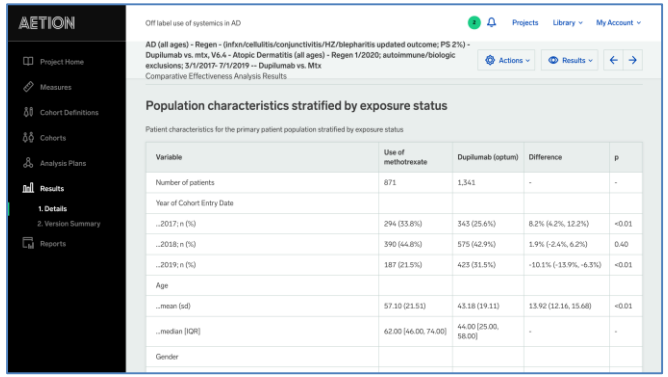
Select **treatment strategy,** follow-up



Document **study results and audits**



Preview **feasibility & diagnostics**



Select **risk adjustment** method

AETION

Transparency

CT.gov registration:

- Complete protocol of each emulation

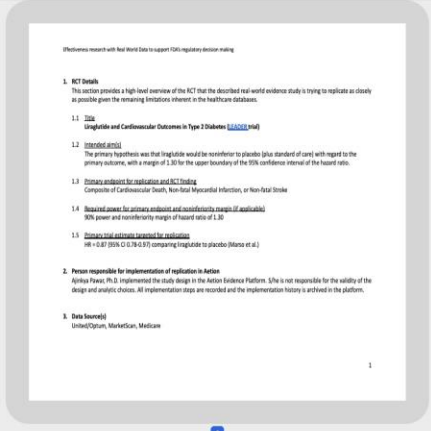
1. **RCT Details**
2. **Person responsible for implementation of replication in Aetion**
3. **Data Source(s)**
4. **Study Design Diagram**
5. **Cohort Identification**
   - 5.1 Inclusion/exclusion criteria for cohort entry
   - 5.2 Flowchart of the study cohort assembly
6. **Variables**
   - 6.1 Exposure-related variables:
   - 6.2 Preliminary Covariates:
   - 6.3 Outcome variables and study follow-up:
7. **Initial Feasibility Analysis**
   - Aetion report name:
   - Date conducted:
8. **Initial Power Assessment**
9. **Balance Assessment after PS matching**
   - Aetion report name:
   - Date conducted:
10. **Final Power Assessment**
    - Aetion report name:
    - Date conducted:
11. **Study Confidence and Concerns**
12. **Register study protocol on clinicalTrials.gov**
13. **Comparative Analyses**
    - Aetion report name:
    - Date conducted:
14. **Requested Results**
15. **References**

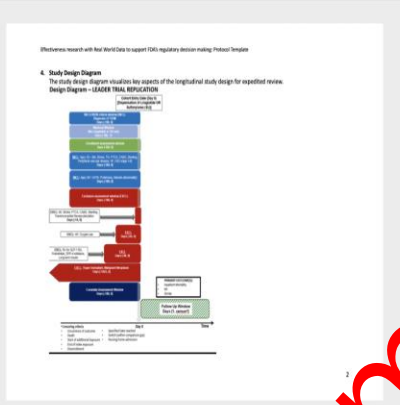
Comparative analysis starts after registration

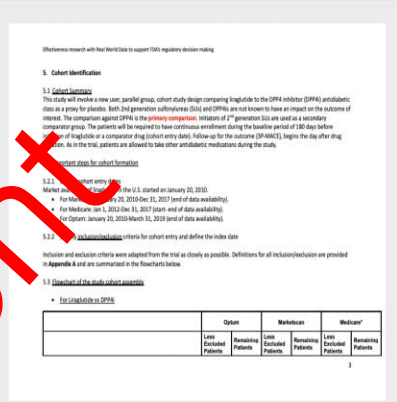
Transparency

ClinicITrials.gov registration:

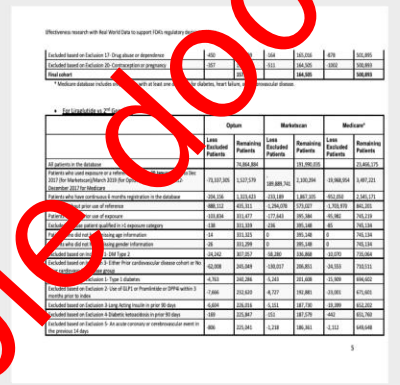
- Complete protocol of each emulation
- Incl. hotlinks to the Aetion Evidence Platform:
  - Inspect definitions
  - Inspect audit trails
  - Reproduce analyses
  - Make changes and run sensitivity analyses
  - Produce additional reports

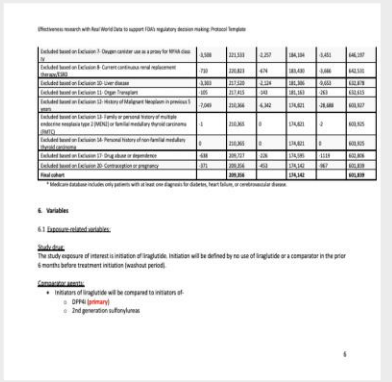

1

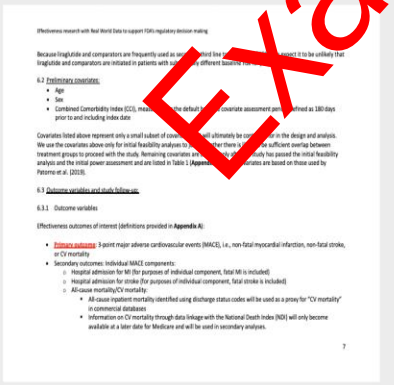

2

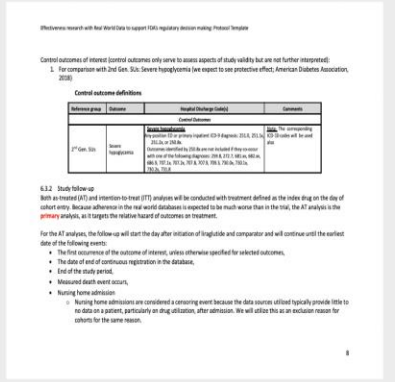

3

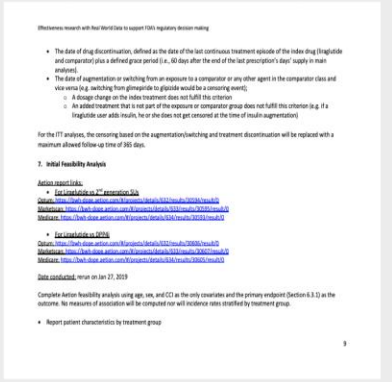

4


5


6





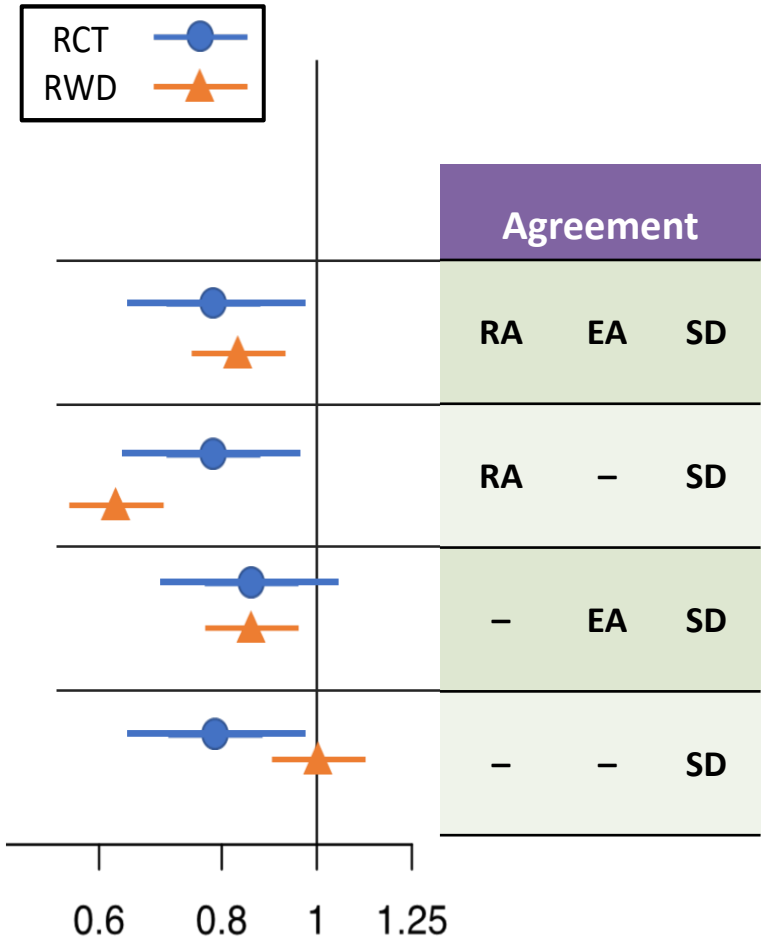


Pre-defined agreement assessment

Regulatory agreement (RA)
Interpretation of the RWE and RCT results would lead to equivalent regulatory decisions based on $p<0.05$

Estimate agreement (EA)
Estimates for RWE fell within the 95% confidence interval of the RCT results

Numeric difference in estimate (SD)
Difference between the RWE and RCT estimates, on a standardized scale



Franklin, Pawar, Martin, Glynn, Levenson, Temple, Schneeweiss. CPT 2020

Emulation quality assessment

Comparator emulation:

Good	if RCT has active comparator
Moderate	if RCT has placebo comparator that was emulated by other drug (unrelated to outcome) and used in similar patients
Poor	if RCT has placebo comparator that was emulated by other drug (unrelated to outcome) and used in different patients

Endpoint emulation:

Good	if endpoint measurement has high specificity
Moderate	if endpoint measurement has moderate specificity

Overview 1-10

	Trial name	RCT Exposure	RCT Comparator	RWE emulation Exposure	RWE emulation Comparator	RCT Outcome	RWE Emulation
1	LEADER	Liraglutide (GLP1)	Placebo	Liraglutide	DPP4i	3p MACE	
2	DECLARE	Dapagliflozin (SGLT2)	Placebo	Dapagliflozin	DPP4i	HHF + CV death	
3	EMPA-REG	Empagliflozin (SGLT2)	Placebo	Empagliflozin	DPP4i	3p MACE	
4	CANVAS	Canagliflozin (SGLT2)	Placebo	Canagliflozin	DPP4i	3p MACE	
5	CARMELINA	Linagliptin (DPP4i)	Placebo	Linagliptin	Sulfonylureas	3p MACE	
6	TECOS	Sitagliptin (DPP4i)	Placebo	Sitagliptin	Sulfonylureas	3p MACE+ angina	
7	SAVOR-TIMI	Saxagliptin (DPP4i)	Placebo	Saxagliptin	Sulfonylureas	3p MACE	
8	CAROLINA	Linagliptin (DPP4i)	Glimerpiride	Linagliptin	Glimerpiride	3p MACE	
9	TRITON	Prasugrel	Clopidogrel	Prasugrel	Clopidogrel	3p MACE	
10	PLATO	Ticagrelor	Clopidogrel	Ticagrelor	Clopidogrel	3p MACE	

2nd-line Antidiabetics

Antiplatelets

MACE = Major adverse cardiovascular events

Comparator emulation:

Good | Moderate | Poor

		RCT		RWE emulation		RCT	RWE
	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
1	LEADER	Liraglutide (GLP1)	Placebo	Liraglutide	DPP4i	3p MACE	
2	DECLARE	Dapagliflozin (SGLT2)	Placebo	Dapagliflozin	DPP4i	HHF + CV death	
3	EMPA-REG	Empagliflozin (SGLT2)	Placebo	Empagliflozin	DPP4i	3p MACE	
4	CANVAS	Canagliflozin (SGLT2)	Placebo	Canagliflozin	DPP4i	3p MACE	
5	CARMELINA	Linagliptin (DPP4i)	Placebo	Linagliptin	Sulfonylureas	3p MACE	
6	TECOS	Sitagliptin (DPP4i)	Placebo	Sitagliptin	Sulfonylureas	3p MACE+ angina	
7	SAVOR-TIMI	Saxagliptin (DPP4i)	Placebo	Saxagliptin	Sulfonylureas	3p MACE	
8	CAROLINA	Linagliptin (DPP4i)	Glimerpiride	Linagliptin	Glimerpiride	3p MACE	
9	TRITON	Prasugrel	Clopidogrel	Prasugrel	Clopidogrel	3p MACE	
10	PLATO	Ticagrelor	Clopidogrel	Ticagrelor	Clopidogrel	3p MACE	

2nd-line Antidiabetics

Antiplatelets

MACE = Major adverse cardiovascular events

Endpoint emulation:

Good | Moderate

	Trial name	RCT Exposure	RCT Comparator	RWE emulation Exposure	RWE emulation Comparator	RCT Outcome	RWE Emulation
2nd-line Antidiabetics	1 LEADER	Liraglutide (GLP1)	Placebo	Liraglutide	DPP4i	3p MACE	
	2 DECLARE	Dapagliflozin (SGLT2)	Placebo	Dapagliflozin	DPP4i	HHF + CV death	HF IP any position, no cause of death
	3 EMPA-REG	Empagliflozin (SGLT2)	Placebo	Empagliflozin	DPP4i	3p MACE	
	4 CANVAS	Canagliflozin (SGLT2)	Placebo	Canagliflozin	DPP4i	3p MACE	
	5 CARMELINA	Linagliptin (DPP4i)	Placebo	Linagliptin	Sulfonylureas	3p MACE	
	6 TECOS	Sitagliptin (DPP4i)	Placebo	Sitagliptin	Sulfonylureas	3p MACE+ angina	Angina non-specific
	7 SAVOR-TIMI	Saxagliptin (DPP4i)	Placebo	Saxagliptin	Sulfonylureas	3p MACE	
	8 CAROLINA	Linagliptin (DPP4i)	Glimerpiride	Linagliptin	Glimerpiride	3p MACE	
Antiplatelets	9 TRITON	Prasugrel	Clopidogrel	Prasugrel	Clopidogrel	3p MACE	
	10 PLATO	Ticagrelor	Clopidogrel	Ticagrelor	Clopidogrel	3p MACE	

MACE = Major adverse cardiovascular events

Overview 11-20

		Trial name	RCT Exposure	RCT Comparator	RWE emulation Exposure	RWE emulation Comparator	RCT Outcome	RWE Emulation
DOAC AF	11	ARISTOTLE	Apixaban	Warfarin	Apixaban	Warfarin	Stroke/Systemic Embolism	
DOAC AF	12	RE-LY	Dabigatran	Warfarin	Dabigatran	Warfarin	Stroke/Systemic Embolism	
DOAC AF	13	ROCKET-AF	Rivaroxaban	Warfarin	Rivaroxaban	Warfarin	Stroke/Systemic Embolism	
DOAC DVT	14	EINSTEIN-DVT	Rivaroxaban	Enoxaparin/ VKA	Rivaroxaban	Warfarin	VTE	
DOAC DVT	15	RE-COVER II	Dabigatran	Warfarin	Dabigatran	Warfarin	VTE / VTE Related Death	
DOAC DVT	16	AMPLIFY	Apixaban	Enoxaparin/ warfarin	Apixaban	Warfarin	VTE / VTE Related Death	
Heart failure	17	PARADIGM-HF	Sacubitril/ Valsartan	Enalapril	Sacubitril/ Valsartan	ACEi	HHF/ Mortality	
Anti-HTN	18	TRANSCEND	Telmisartan	Placebo	Telmisartan + Loop/CCB /TZ	Loop/CCB/ TZ	3p MACE + HHF	
Anti-HTN	19	ON-TARGET	Telmisartan	Ramipril	Telmisartan	Ramipril	3p MACE + HHF	
Osteoporosis	20	HORIZON	Zoledronic Acid	Placebo	Zoledronic Acid	Raloxifene	Hip Fracture	

Comparator emulation:

Good | Moderate | Poor

		Trial name	RCT Exposure	RCT Comparator	RWE emulation Exposure	RWE emulation Comparator	RCT Outcome	RWE Emulation
DOAC AF	11	ARISTOTLE	Apixaban	Warfarin	Apixaban	Warfarin	Stroke/Systemic Embolism	
DOAC AF	12	RE-LY	Dabigatran	Warfarin	Dabigatran	Warfarin	Stroke/Systemic Embolism	
DOAC AF	13	ROCKET-AF	Rivaroxaban	Warfarin	Rivaroxaban	Warfarin	Stroke/Systemic Embolism	
DOAC DVT	14	EINSTEIN-DVT	Rivaroxaban	Enoxaparin/ VKA	Rivaroxaban	Warfarin	VTE	
DOAC DVT	15	RE-COVER II	Dabigatran	Warfarin	Dabigatran	Warfarin	VTE / VTE Related Death	
DOAC DVT	16	AMPLIFY	Apixaban	Enoxaparin/ warfarin	Apixaban	Warfarin	VTE / VTE Related Death	
Heart failure	17	PARADIGM-HF	Sacubitril/ Valsartan	Enalapril	Sacubitril/ Valsartan	ACEi	HHF/ CV death	
Anti-HTN	18	TRANSCEND	Telmisartan	Placebo	Telmisartan + Loop/CCB /TZ	Loop/CCB/ TZ	3p MACE + HHF	
Anti-HTN	19	ON-TARGET	Telmisartan	Ramipril	Telmisartan	Ramipril	3p MACE + HHF	
Osteoporosis	20	HORIZON	Zoledronic Acid	Placebo	Zoledronic Acid	Raloxifene	Hip Fracture	

Endpoint emulation:

Good | Moderate

		Trial name	RCT Exposure	RCT Comparator	RWE emulation Exposure	RWE emulation Comparator	RCT Outcome	RWE Emulation
DOAC AF	11	ARISTOTLE	Apixaban	Warfarin	Apixaban	Warfarin	Stroke/Systemic Embolism	
	12	RE-LY	Dabigatran	Warfarin	Dabigatran	Warfarin	Stroke/Systemic Embolism	
	13	ROCKET-AF	Rivaroxaban	Warfarin	Rivaroxaban	Warfarin	Stroke/Systemic Embolism	
DOAC DVT	14	EINSTEIN-DVT	Rivaroxaban	Enoxaparin/ VKA	Rivaroxaban	Warfarin	VTE	May include some rule-out Dx
	15	RE-COVER II	Dabigatran	Warfarin	Dabigatran	Warfarin	VTE / VTE Related Death	May include some rule-out Dx
	16	AMPLIFY	Apixaban	Enoxaparin/ warfarin	Apixaban	Warfarin	VTE / VTE Related Death	May include some rule-out Dx
Heart failure	17	PARADIGM-HF	Sacubitril/ Valsartan	Enalapril	Sacubitril/ Valsartan	ACEi	HHF/ CV death	HF IP any position, no cause of death
Anti-HTN	18	TRANSCEND	Telmisartan	Placebo	Telmisartan + Loop/CCB /TZ	Loop/CCB/ TZ	3p MACE + HHF	HF IP any position, no cause of death
	19	ON-TARGET	Telmisartan	Ramipril	Telmisartan	Ramipril	3p MACE + HHF	HF IP any position, no cause of death
Osteoporosis	20	HORIZON	Zoledronic Acid	Placebo	Zoledronic Acid	Raloxifene	Hip Fracture	Shorter follow-up

Event rates 1-10

Good | Moderate

Generally lower event rates in RWE studies

	Trial name	Outcome	RCT Exposure			RCT Comparator			RWE Exposure			RWE Comparator		
			Events	N	Rate*	Events	N	Rate	Events	N	Rate	Events	N	Rate
1	LEADER	3p MACE	608	4,668	3.4	694	4,672	3.9	1,352	84,346	2.1	1,955	84,346	2.6
2	DECLARE	HHF +CV death	417	8,582	1.2	496	8,578	1.5	242	24,895	1.6	367	24,895	2.4
3	EMPA-REG	3p MACE	490	4,687	3.7	282	2,333	4.4	416	51,875	1.5	478	51,875	1.9
4	CANVAS	3p MACE	564	5,795	2.7	496	4,347	3.2	772	76,099	1.5	990	76,099	1.9
5	CARMELINA	3p MACE	434	3,494	5.8	420	3,485	5.6	1,540	50,913	4.6	1,826	50,913	5.2
6	TECOS	3p MACE+ angina	839	7,257	4.1	851	7,266	4.2	8,106	174,739	7.3	9,692	174,739	8.3
7	SAVOR-TIMI 53	3p MACE	613	8,280	3.6	609	8,212	3.6	1,662	91,064	2.4	2,390	91,064	3.1
8	CAROLINA	3p MACE	356	3,023	2.1	362	3,010	2.1	373	24,131	2.7	458	24,131	3.0
9	TRITON-TIMI 38	3p MACE	643	6813	7.9	781	6795	9.7	718	21,932	3.8	960	24,446	3.9
10	PLATO	3p MACE	864	9333	9.8	1014	9291	11.7	649	13,980	8.0	858	13,980	7.1

3P MACE = 3-point major adverse cardiovascular events (myocardial infarction, stroke, or cardiovascular death); HHF = hospitalization for heart failure

* Incidence rate per 100 person-years.

Higher event rates in RWE studies: Less specific endpoint definitions

Event rates 11-20

Good | Moderate

Similar event rates

Higher event rates in RWE studies: Less specific endpoint

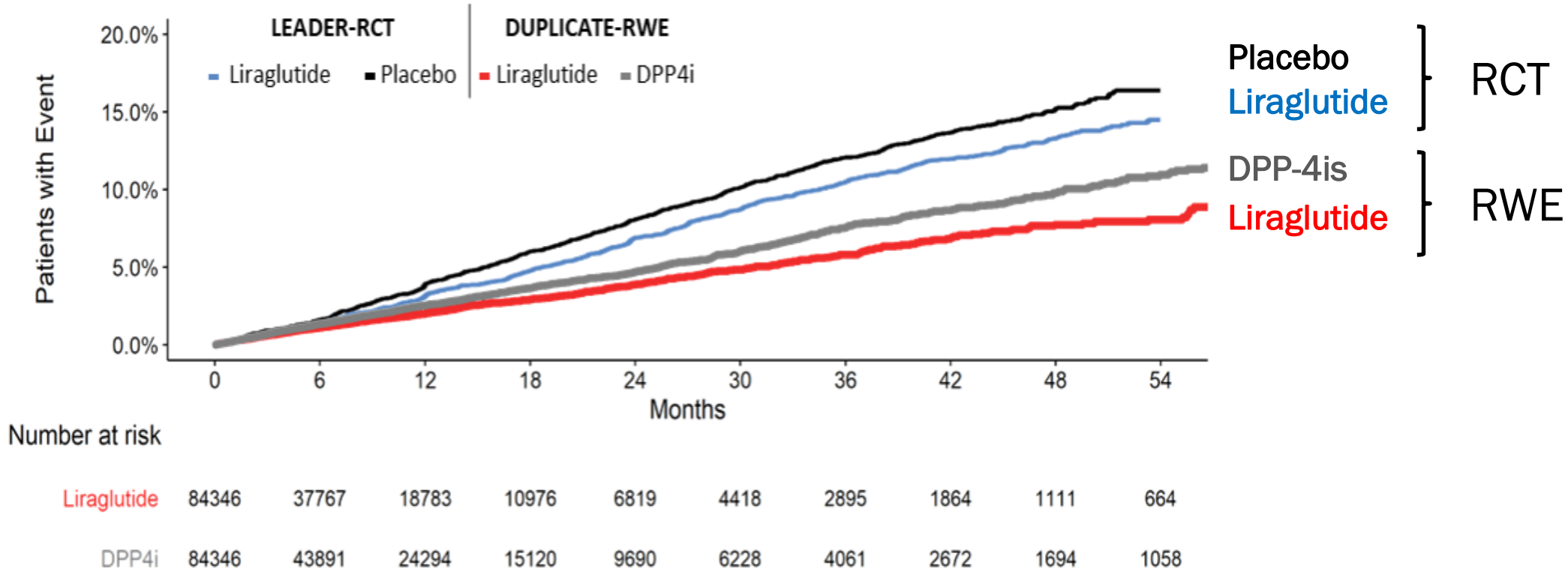
	Trial name	Outcome	RCT Exposure			RCT Comparator			RWE Exposure			RWE Comparator		
			Events	N	Rate	Events	N	Rate	Events	N	Rate	Events	N	Rate
11	ARISTOTLE	Stroke/ Sys Embol	212	9,12	1.3	265	9,08	1.6	545	110,259	0.9	694	110,259	1.5
12	RE-LY	Stroke/ Sys Embol	134	6,07	1.1	199	6,022	1.7	172	39,070	0.9	221	39,070	1.3
13	ROCKET-AF	Stroke/ Sys Embol	188	6,95	1.7	241	7,004	2.2	419	51,318	1.5	518	51,318	2.4
14	EINSTEIN-DVT	VTE	36	1,73	2.1	51	1,718	3.0	207	12,985	4.9	271	12,985	6.2
15	RECOVER II	VTE / VTE Death	30	1,27	2.3	28	1,289	2.2	46	2,671	5	48	2,671	5.1
16	AMPLIFY	VTE / VTE Death	59	2,60	2.3	71	2,635	2.7	155	3,570	11.6	99	3,570	8.2
17	PARADIGM-HF	HHF/ Mortality	914	4,18	21.8	1,117	4,212	26.5	645	3,033	46.4	636	3,033	44.6
18	TRANSCEND	3p MACE + HHF	465	2,95	15.7	504	2,972	17.0	826	20,024	7.4	1,383	20,024	7.6
19	ON-TARGET	3p MACE + HHF	1,412	8,57	16.5	1,423	8,542	16.7	874	17,626	6.4	1,306	17,626	8.2
20	HORIZON-PIV	Hip Fracture	88	3,87	2.5	52	3,861	1.4	78	9,003	0.7	97	9,003	0.9

3P MACE = 3-point major adverse cardiovascular events (myocardial infarction, stroke, or cardiovascular death); Sys Embol = systemic embolism; HHF = hospitalization for heart failure

* Incidence rate per 100 person-years.

Lower event rates in RWE studies: Lower sensitivity endpoint definitions

Example of a K-M plot: LEADER trial and emulation



Second-line anti-diabetics: SGLT2-is and GLP-1 RAs

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
1	LEADER	Liraglutide (GLP1)	Placebo	Liraglutide	DPP4i	3p MACE	
2	DECLARE	Dapagliflozin (SGLT2)	Placebo	Dapagliflozin	DPP4i	HHF + CV death	HF IP any position, no cause of death
3	EMPA-REG	Empagliflozin (SGLT2)	Placebo	Empagliflozin	DPP4i	3p MACE	
4	CANVAS	Canagliflozin (SGLT2)	Placebo	Canagliflozin	DPP4i	3p MACE	

	Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
1	LEADER	Moderate	Good	**0.87** (0.78, 0.97)	**0.82** (0.76, 0.87)	0.90	NI	**RA**	**EA**	**SD**
2	DECLARE	Moderate	Moderate	**0.83** (0.73, 0.95)	**0.69** (0.59, 0.81)	1.76	NI	**RA**	**–**	**SD**
3	EMPA-REG	Moderate	Good	**0.86** (0.74, 0.99)	**0.83** (0.73, 0.95)	0.35	NI	**RA**	**EA**	**SD**
4	CANVAS	Moderate	Good	**0.86** (0.75, 0.97)	**0.77** (0.70, 0.85)	1.34	NI	**RA**	**EA**	**SD**

Second-line anti-diabetics: DPP4is

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
5	CARMELINA	Linagliptin (DPP4i)	Placebo	Linagliptin	Sulfonylureas	3p MACE	
6	TECOS	Sitagliptin (DPP4i)	Placebo	Sitagliptin	Sulfonylureas	3p MACE+ angina	Angina non-specific
7	SAVOR-TIMI	Saxagliptin (DPP4i)	Placebo	Saxagliptin	Sulfonylureas	3p MACE	
8	CAROLINA	Linagliptin (DPP4i)	Glimerpiride	Linagliptin	Glimerpiride	3p MACE	

	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
5	CARMELINA	Poor	Good	**1.02** (0.89, 1.17)	**0.90** (0.84, 0.96)	1.61	NI	*	**EA**	**SD**
6	TECOS	Poor	Moderate	**0.98** (0.88, 1.09)	**0.89** (0.86, 0.91)	1.71	NI	*	**EA**	**SD**
7	SAVOR-TIMI	Poor	Good	**1.00** (0.89, 1.12)	**0.81** (0.76, 0.86)	3.16¶	NI	*	–	–
8	CAROLINA	Good	Good	**0.98** (0.84, 1.14)	**0.91** (0.79, 1.05)	0.70	NI	**RA**	**EA**	**SD**

Note:
Positive interpretation of CAROLINA; very similar to TECOS yet no Reg Agreement

Antiplatelets: Prasugrel and Ticagrelor

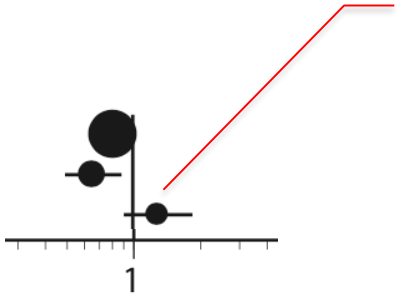
	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
9	TRITON	Prasugrel	Clopidogrel	Prasugrel	Clopidogrel	3p MACE	
10	PLATO	Ticagrelor	Clopidogrel	Ticagrelor	Clopidogrel	3p MACE	

	Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
9	TRITON	Good	Good	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
10	PLATO	Good	Good	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**

1) PLATO's treatment effect was not established among US participants possibly due to high aspirin dosing in the US compared to Europe

PLATO and regional variation:*

	N	Ticagrelor	Clopidogrel	HR (95% CI)
All Countries	18624	9.8	11.7	0.84 (0.77, 0.92)
	2666	7.5	10.8	0.69 (0.53, 0.90)
USA	1413	12.6	10.1	1.27 (0.92, 1.75)



Note: RCT-DUPLICATE used U.S. data sources only

* *Mahaffey KW et al. Circ 2011*

DOAC treatment for Afib

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
11	ARISTOTLE	Apixaban	Warfarin	Apixaban	Warfarin	Stroke/Systemic Embolism	
12	RE-LY	Dabigatran	Warfarin	Dabigatran	Warfarin	Stroke/Systemic Embolism	
13	ROCKET-AF	Rivaroxaban	Warfarin	Rivaroxaban	Warfarin	Stroke/Systemic Embolism	

	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
11	ARISTOTLE	Good	Good	**0.79** (0.66, 0.95)	**0.65** (0.59, 0.72)	1.81	NI	**RA**	–	**SD**
12	RE-LY	Good	Good	**0.66 (**0.53, 0.82)	**0.69** (0.57, 0.83)	-0.31	NI	**RA**	**EA**	**SD**
13	ROCKET-AF	Good	Good	**0.79** (0.66, 0.96)	**0.77** (0.69, 0.86)	0.22	NI	**RA**	**EA**	**SD**

DOAC treatment for VTE

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
14	EINSTEIN-DVT	Rivaroxaban	Enoxaparin/ VKA	Rivaroxaban	Warfarin	VTE	Can't measure clinical parameter
15	RE-COVER II	Dabigatran	Warfarin	Dabigatran	Warfarin	VTE / VTE Related Death	Can't measure clinical parameter
16	AMPLIFY	Apixaban	Enoxaparin/ warfarin	Apixaban	Warfarin	VTE / VTE Related Death	Can't measure clinical parameter

	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
14	EINSTEIN-DVT	Good	Moderate	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	**EA**	**SD**
15	RE-COVER II	Good	Moderate	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	**RA**	**EA**	**SD**
16	AMPLIFY	Good	Moderate	**0.84** (0.60, 1.18)	**0.76** (0.53, 1.09)	0.40	NI	**RA**	**EA**	**SD**

Heart failure: Sacubitril/ Valsartan (Entresto)

17	PARADIGM-HF	Sacubitril/ Valsartan	Enalapril	Sacubitril/ Valsartan	ACEi	HHF/ CV death	HF IP any position, no cause of death

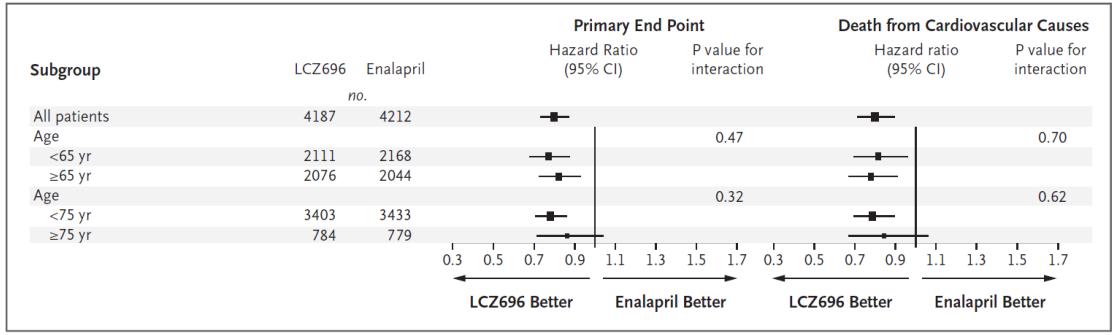
	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
17	PARADIGM-HF	Moderate	Moderate	**0.80** (0.73, 0.87)	**0.97** (0.87-1.08)	-3.17	Sup	–	–	–

1) HR by data source

2) Treatment effect reduced in those 75+

Optum	0.98 (0.84, 1.16)
MarketScan	0.85 (0.67, 1.08)
Medicare FFS	1.02 (0.85, 1.22)
pooled	0.97 (0.87-1.08)
<= 75 yrs	0.89 (0.77-1.02)
> 75 yrs	1.04 (0.89-1.23)

PARADIGM-HF effect estimates by age:



Antihypertensives: Telmisartan

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
18	TRANSCEND	Telmisartan	Placebo	Telmisartan + Loop/CCB /TZ	Loop/CCB/ TZ	3p MACE + HHF	HF IP any position, no cause of death
19	ON-TARGET	Telmisartan	Ramipril	Telmisartan	Ramipril	3p MACE + HHF	HF IP any position, no cause of death

	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
18	TRANSCEND	Moderate	Moderate	**0.92** (0.81, 1.05)	**0.88** (0.81, 0.96)	0.55	Sup	*	**EA**	**SD**
19	ON-TARGET	Good	Moderate	**1.01** (0.94, 1.09)	**0.83** (0.77, 0.90)	3.46	NI	*	–	–

1) We investigate subtle differences in exposure, outcome, inclusion-exclusion criteria, covariates, follow-up

ON-TARGET	1.0 (0.9-1.1)
Fralick et al. RWE JAMA-IM	1.0 (0.9-1.1)
RCT-DUPLICATE	0.8 (0.8-0.9)

Osteoporosis: Zoledronic acid

	Trial name	Exposure	Comparator	Exposure	Comparator	Outcome	Emulation
20	HORIZON	Zoledronic Acid	Placebo	Zoledronic Acid	Raloxifene	Hip Fracture	Time-varying hazard

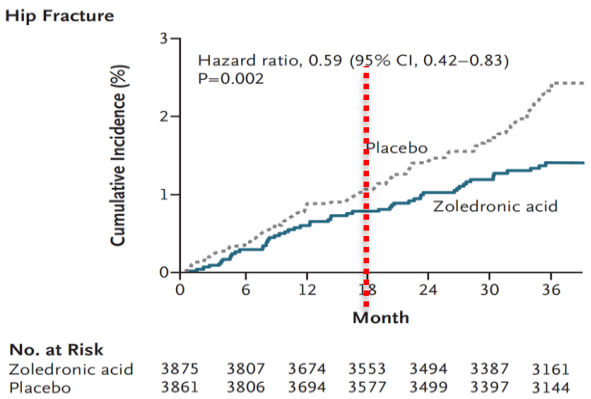
	Trial name	Comparator emulation	Endpoint emulation⁺	RCT result	RWE results	Stand. Diff.	Test	Agreement		
20	HORIZON	Moderate	Moderate	**0.59** (0.42, 0.83)	**0.75** (0.58, 0.97)	-1.10	Sup	**RA**	**EA**	**SD**

1) Time-varying treatment effects

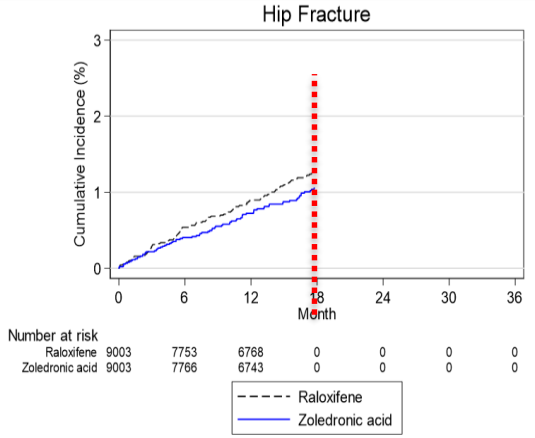
RCT: HR_{36mo} = 0.59 RWE: HR_{36mo} = ?? Emulation mismatch

HR_{18mo} = 0.75 HR_{18mo} = 0.75 Calibration success

KM from Horizon-pivotal trial

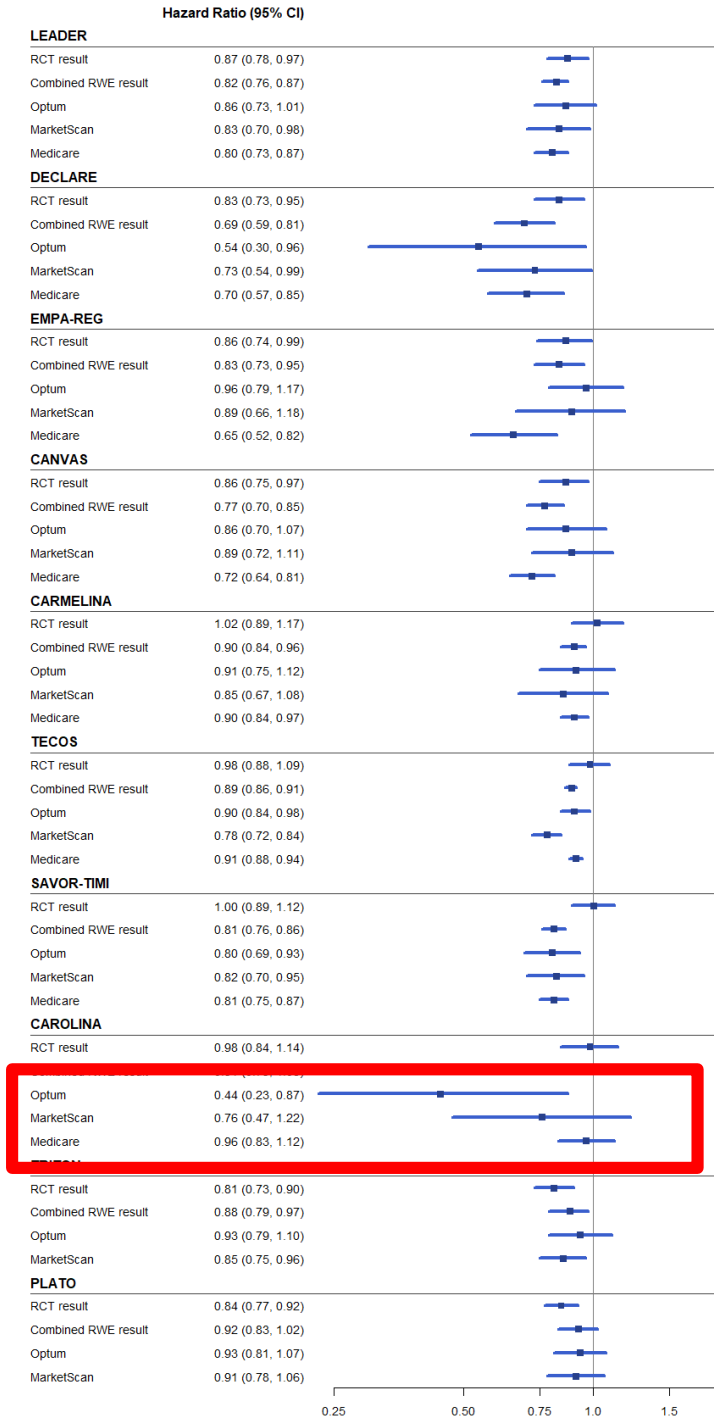


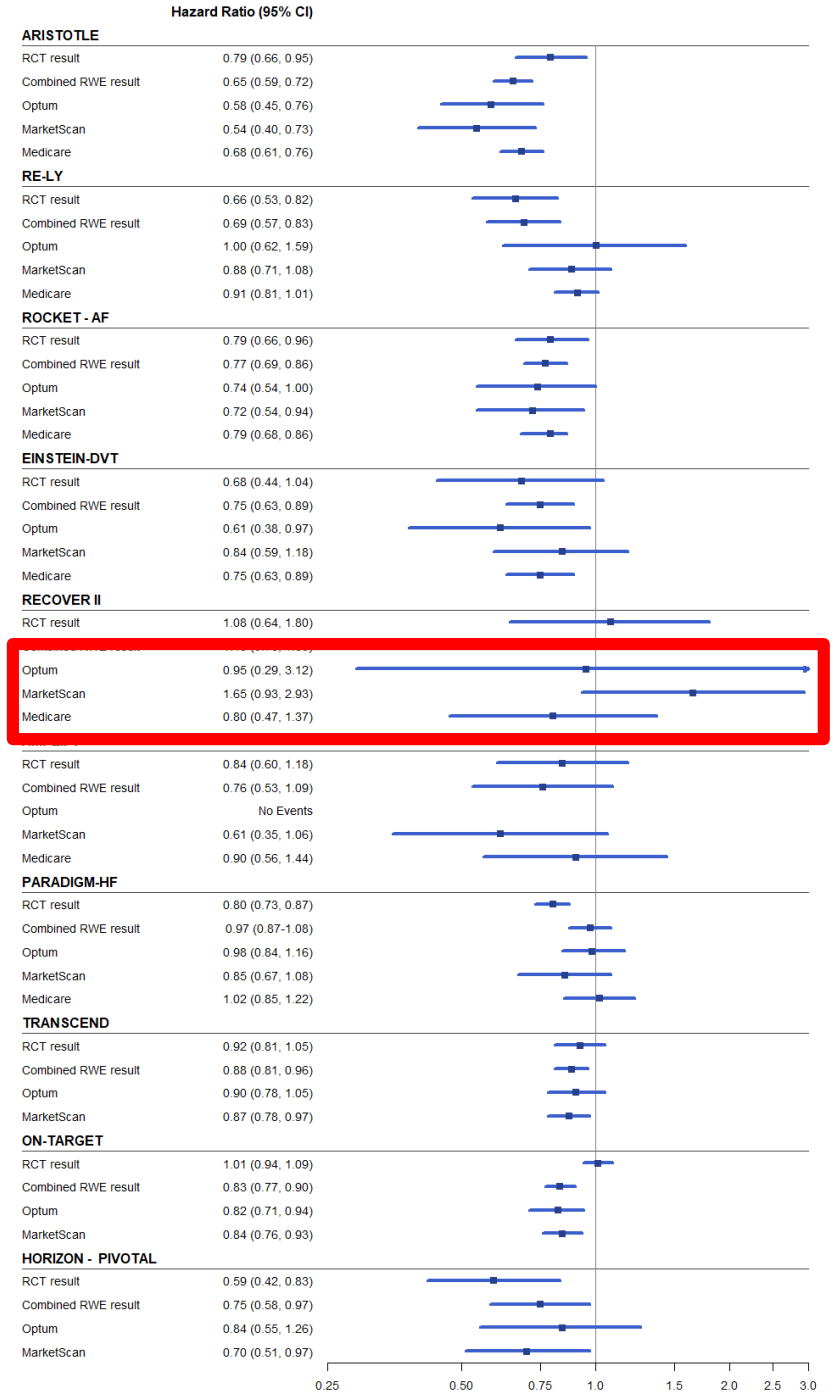
KM - Optum+Marketscan



Variation between data sources

Limited variation between US commercial claims data sources





Conclusion

- With data that are fit-for-purpose and proper design and analysis, non-randomized real-world evidence studies usually come to the same conclusion about a drug's treatment effect as randomized trials

- These initial findings of the RCT-DUPLICATE program indicate circumstances when RWE may offer causal insights in situations where RCT data is either not available or cannot be quickly or feasibly generated.

Some learnings

- We need to take into account the uncertainty inherent in any single RCT
- One wouldn't likely take only the primary result of a single RCT in isolation
  - It is important to have planned sensitivity analyses to help interpret findings as a whole
- A single binary success metric will not do justice
- In any emulation, despite best efforts, there will remain differences in population, measurement, and drug use:
  - For our emulation success most critical seemed:
    - Population, comparator, and outcome emulation
    - Data fit-for-purpose and study design choices are most important considerations
  - We remain concerned about 3 emulations with an opportunity for more learnings:
    - PARADIGM-HF: some emulation differences, effect modification,
    - ON-TARGET: ??? (we are investigating multiple issues)
    - SAVOR-TIMI: Residual confounding by correlates of soc-econ factors?

Calibrating our RWE tool kit

- Repository of well-documented studies that illustrate the agreement between RCTs and RWE, in specific situations when the RWE study is explicitly designed to answer the same question as the RCT.

- May serve as reference points to assess validity in RWE:
  - By therapeutic area
  - By data source
  - By type of comparator
  - By type of outcome
  - Further categorization:
    - Population
    - Follow-up

- A repository of case studies would
  - Increase predictability of future RWE studies
  - Increase the use of common methodological approaches emulating target trials
  - Point out areas that are currently difficult to address with RWE and highlight the need to improve data sources

William Crown

Brandeis University

Trial Emulation Studies and OPERAND

William H. Crown, PhD

Distinguished Research Professor

Brandeis University

OPERAND

- Study Objective:
  - better understand sources of variability in treatment effect estimates from observational health care data through comparisons with RCTs
  - examine heterogeneity in treatment effect estimates as the inclusion/exclusion criteria of the RCTs are relaxed to reflect the real world patient population

OPERAND Overview

Teams and Approach	Two research teams independently attempt to emulate the same two trials: 1. ROCKET AF 2. LEAD-2 Diabetes
Data	OptumLabs® Data Warehouse. (1) claims data alone and (2) claims plus EMR. Initial analyses restricted to inclusion/exclusion criteria of the trials. Followed by relaxation of inclusion/exclusion criteria but within approved indication
Approach	1. Each team used study design documentation provided in the original pivotal publications of the trial results. 2. Given a prescribed set of methods. 3. Allowed to use methods of their own choosing
Decision-making of researchers	Each team documented analytic decisions in research design

Co-Leads

OPTUM Labs

MULTI-REGIONAL CLINICAL TRIALS
THE MRCT CENTER of BRIGHAM AND WOMEN'S HOSPITAL and HARVARD

Sponsors

Amgen
AstraZeneca
Merck
Optum
Pfizer
Sanofi
UCB BioSciences, Inc.

Research Partners Selected

Brown University

Harvard Pilgrim Health Care Institute

Technical Expert Panel

Sponsor representatives

+

9 representatives from academia, pharmaceutical companies, professional societies, etc.

FDA participant as observer

The Trials

- The **ROCKET Atrial Fibrillation Trial** was a double-blind study that randomly assigned 14,264 patients with nonvalvular atrial fibrillation to either rivaroxaban (daily dose of 20 mg) or dose-adjusted warfarin. The trial was intended to evaluate whether rivaroxaban was noninferior to warfarin for the primary endpoint of stroke or systemic embolism.
    - *Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, Breithardt G, Halperin JL, Hankey GJ, Piccini JP, Becker RC, Nessel CC, Paolini JF, Berkowitz SD, Fox KA, Califf RM; ROCKET AF Investigators Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. New England Journal of Medicine 2011 Sep 8;365(10):883-91. https://www.nejm.org/doi/full/10.1056/NEJMoa1009638*
- The **LEAD-2 diabetes trial** was a double-blind, double-dummy, placebo- and active-controlled, parallel-group trial where 1,091 participants with type 2 diabetes were randomly assigned to once-daily liraglutide, placebo, or glimepiride. All treatments were in combination with metformin. Efficacy (as measured by HbA1c levels) and safety of adding liraglutide to metformin was assessed.
    - *Nauck M, Frid A, Hermansen K, Shah N, Tankova T, Mitha I, Sdravkovic M, During M, Matthews D, LEAD-2 Study Group. Efficacy and Safety Comparison of Liraglutide, Glimepiride, and Placebo, All in Combination with Metformin, in Type 2 Diabetes. Diabetes Care, 32(1):84-90, 2009. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2606836*

Emulation Agreement Measures

- Regulatory Agreement—statistically significant result with directional equivalence between RCT and observational estimate
- Statistical Agreement—defined as the point estimate from the observational study falling within the 95% confidence interval of the ATE of the RCT using the reported standard errors of the RCT to define the confidence interval

Transparency

- Both teams registered their study protocols on the EU-PAS registry before they were given access to any data.
  - http://www.encepp.eu/encepp_studies/indexRegister.shtml

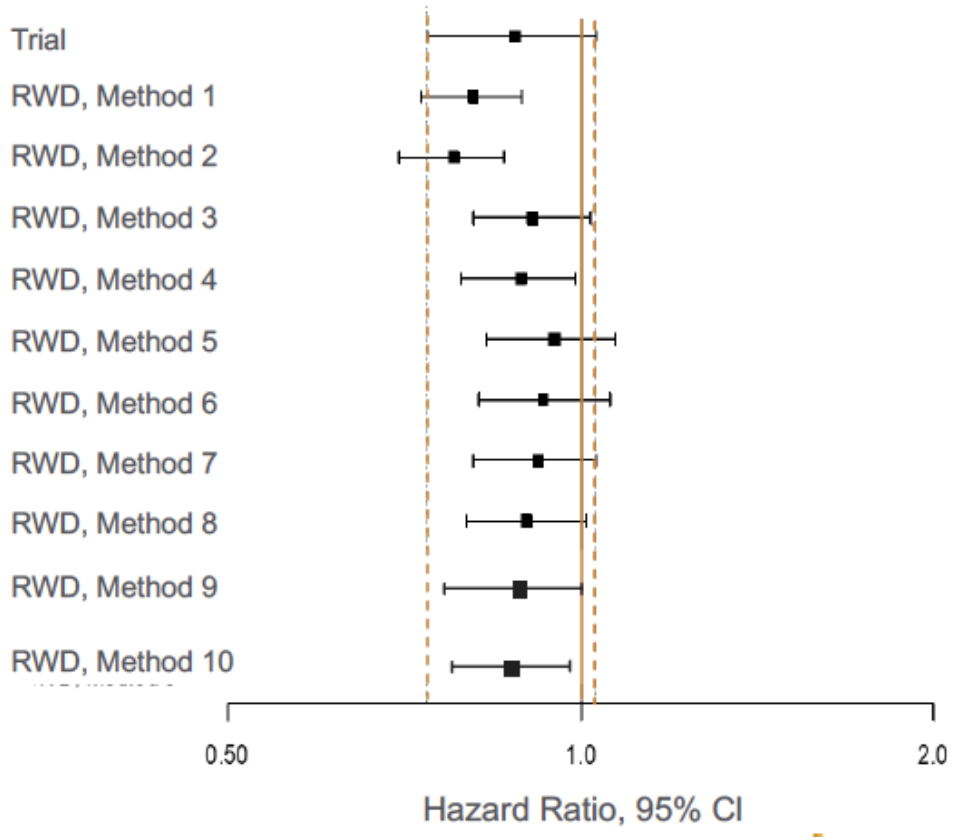
Cohort Identification and Sample Characteristics

- Similar for ROCKET AF. One group was slightly more restrictive than the other
- Substantial differences for LEAD2. Due mainly to differences in how the research teams addressed availability of follow-up HbA1c.
  - Interpretation of the exercise: Target Trial versus Actual Trial
- Similarity of observational samples and differences with respective trials

High Level Treatment Effect Estimate Results

- Both teams generated treatment effect estimates similar to the respective trials (using both regulatory and statistical comparison methods)
- Little variation in estimates by statistical methods
- Little variation when inclusion/exclusion criteria were loosened
- These results are not necessarily generalizable to other disease states or clinical interventions

ROCKET AF Preliminary Trial Results



Crown W. RCT Replication with Observational Data. FDA/Duke Margolis Annual RWE Summit. National Press Club, Washington, DC, October 2, 2019

What Have We Learned From Clinical Trials Emulation Efforts?

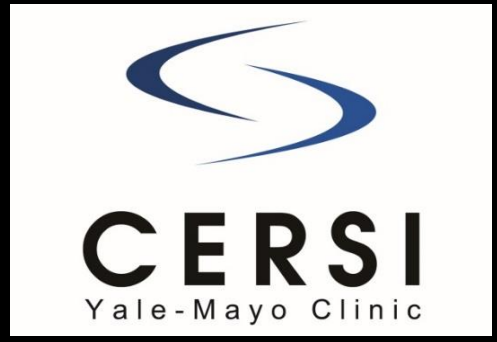
- In emulation efforts, the target trial is strongly guided by the actual trial but may differ in important ways to deal with significant data shortcomings or other issues.
- We have a growing body of literature on emulation efforts that have shown that it is often possible to estimate similar treatment effects with observational data—at least in certain disease areas
- Studies have also shown that despite mimicking the inclusion/exclusion criteria of trials that the distribution of patient characteristics in the observational data may differ from those of the RCT.
- Additional multivariate analysis using the balanced cohorts may help with residual confounding.
- As data and statistical methods continue to improve the reliability of causal inferences drawn from observational data should continue to improve.

Nilay Shah

Mayo Clinic

Assessing the Value of Real-World Data for Emulating Clinical Trials

Nilay Shah
Yale University-Mayo Clinic CERSI
February 16th, 2021



Disclosures

FDA U01FD005938

NHLBI R21HL140205

Team:

Joshua Wallach
Yihong Deng
Alyssa Berkowitz
Peter Noseworthy
Rozalina McCoy
Sanket Dhruva
Eric Polley
Xiaoxi Yao
Timothy Lyon
Jeph Herrin
William Crown
Joseph Ross

FDA Real World Evidence Team

CERSI
Yale-Mayo Clinic

Overview

1. Data Sources
2. Population representativeness
3. Real world practice vs. clinical trial design
4. Approaches to address selection
5. Trials discussed: PRONOUNCE, GRADE, CABANA

Key consideration: emulate trials to predict population and results prior to publication of results

CERSI
Yale-Mayo Clinic

Overview of the PRONOUNCE Trial

Randomly allocated 900 participants with advanced prostate cancer and cardiovascular disease to one of two drugs:

- Degarelix
- Leuprolide

Primary outcome: Time from randomization to first confirmed occurrence of the composite Major Adverse Event endpoint

- i.e., death due to any cause, myocardial infarction (fatal, non-fatal), or stroke (fatal, non-fatal) [Time frame: up to 336 days]

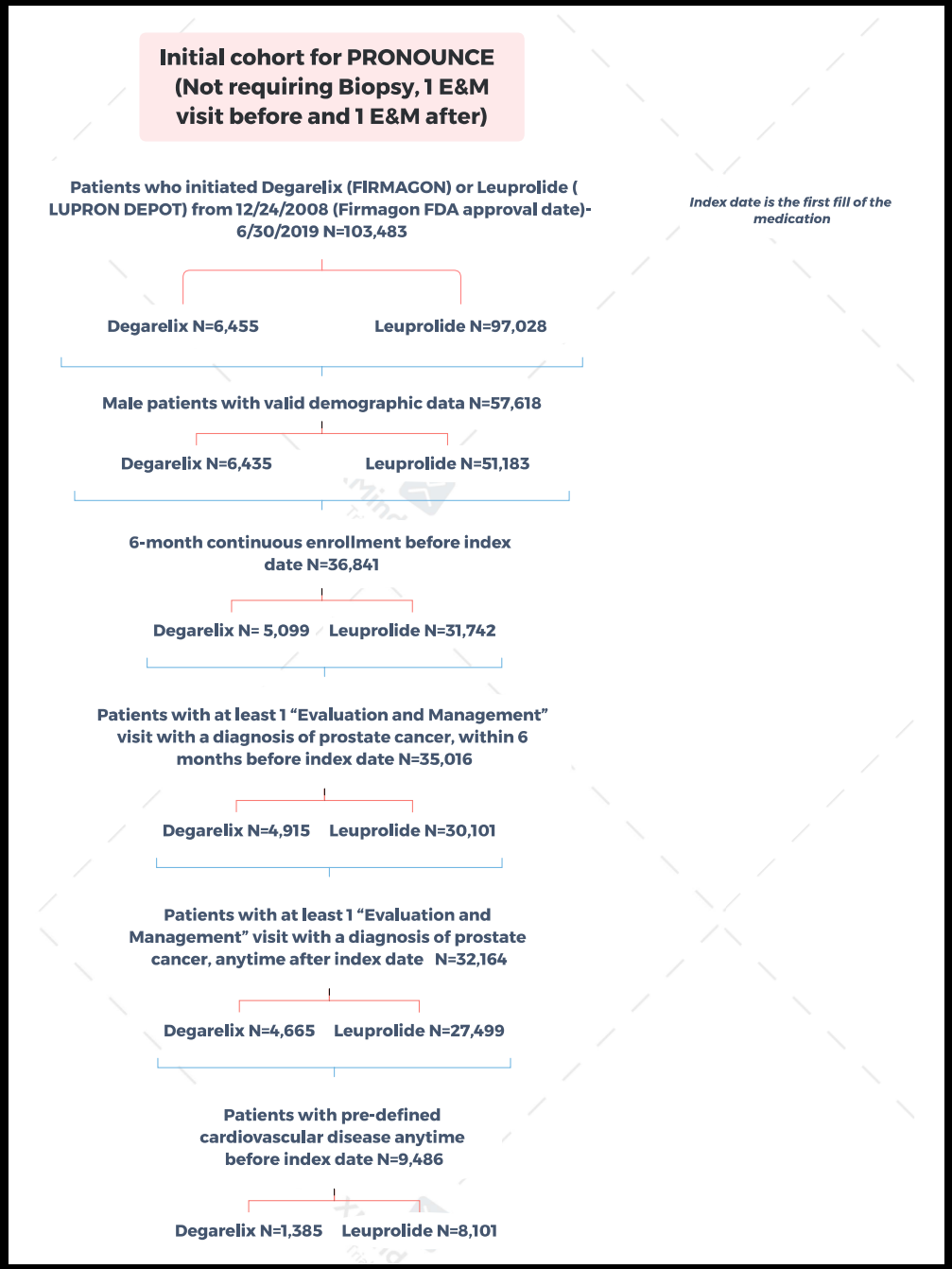
The trial is currently **active ("not recruiting")**, and the **results are not available** on ClinicalTrials.gov

CERSI
Yale-Mayo Clinic

PRONOUNCE Replication Cohort by drug

- **Cohort generated not requiring a prostate biopsy**
  - 7,928 patients were eligible for PRONOUNCE after applying the inclusion and exclusion criteria, removing potential crossover
    - Degarelix, n = 1,250
    - Leuprolide, n = 6,678

CERSI
Yale-Mayo Clinic



CERSI
Yale-Mayo Clinic

Updated analyses

- We identified cross-over within 336 days of follow-up:
  - Leuprolide arm: 19/6678 patients had a degarelix fill
  - Degarelix arm: **810/1250** patients had a leuprolide fill [This makes clinical sense, due to the longer dosing interval for leuprolide]

- Three analyses
  - Primary: Intention to treat
  - Secondary: Censoring patients when they switch; dropping any patients that crossed over between drugs

CERSI
Yale-Mayo Clinic

Statistical Analysis

- Propensity score matching to balance the difference in baseline characteristics between patients who received degarelix versus those who received leuprolide
  - One-to-one nearest neighborhood caliper matching
  - Caliper equal to 0.2 of the standard deviation of the logit of the propensity score
- Standardized differences used to assess the balance of covariates after matching and a standardized difference within 0.1 was considered acceptable
  - Covariates with standardized differences above 0.1 were adjusted for in the regression models.
- Cox proportional hazards regression was used to compare patients receiving degarelix versus those who received leuprolide for the primary and secondary outcomes in the propensity matched cohort, with robust sandwich estimates to account for the clustering within matched sets.

CERSI
Yale-Mayo Clinic

PRONOUNCE – Outcomes

Methods and Outcomes	No. Events	Person-years	Rate per 100 person-years	No. Events	Person-years	Rate per 100 person-years	Hazard Ratio (95% CI)	p-value
1. ITT	**Degarelix (N=1248)**			**Leuprolide (N=1248)**				
MACE	97	960.85	10.10	84	965.15	8.70	1.16 (0.87, 1.56)	0.31
Death	73	972.07	7.51	46	981.96	4.68	1.61 (1.11, 2.33)	0.01
Stroke	15	966.77	1.55	21	975.51	2.15	0.72 (0.37, 1.40)	0.34
MI	20	965.83	2.07	24	971.60	2.47	0.84 (0.47, 1.52)	0.57
2. Censor switch	**Degarelix (N=1248)**			**Leuprolide (N=1248)**				
MACE	55	476.13	11.55	84	964.90	8.71	1.54 (1.10, 2.17)	0.01
Death	38	481.37	7.89	46	981.72	4.69	2.11 (1.37, 3.24)	0.001
Stroke	9	478.72	1.88	21	975.27	2.15	0.98 (0.46, 2.08)	0.96
MI	13	478.79	2.72	24	971.35	2.47	1.14 (0.58, 2.24)	0.71
3. Dropped crossover	**Degarelix (N=440)**			**Leuprolide (N=440)**				
MACE	50	316.09	15.82	25	346.92	7.21	2.33 (1.41, 3.83)	0.001
Death	38	320.59	11.85	14	351.69	3.98	3.19 (1.71, 5.97)	0.00
Stroke	8	318.05	2.52	7	348.53	2.01	1.36 (0.48, 3.82)	0.56
	9	318.64	2.82	5	3.50	1.43	1.98 (0.63, 6.22)	0.24

CERSI
Yale-Mayo Clinic

Overview of the GRADE Trial

(Glycemia Reduction Approaches in Diabetes: A Comparative Effectiveness Study)

- 36-center pragmatic, unmasked, parallel treatment group RCT
- Randomly allocated 5,047 adults with type 2 diabetes, duration <10 years, treated with metformin and HbA1c 6.8-8.5% to one of 4 drugs:
  - Glimepiride (sulfonylurea)
  - Sitagliptin (DPP-4 inhibitor)
  - Liraglutide (GLP-1 receptor agonist)
  - Glargine (basal analog insulin)
- Currently **Active (not recruiting)** and the **results are not available**
  - Conceived in 2008 → recruitment 7/2013 - 9/2017 → anticipated complete follow-up and data collection by 7/2021

CERSI
Yale-Mayo Clinic

https://clinicaltrials.gov/ct2/show/NCT01794143
Diabetes Care. 2013 Aug;36(8):2254-61
Diabetes Care. 2019 Nov;42(11):2098-2107

Primary Outcome

- Time to **primary metabolic failure** of the assigned treatment
  - Time to an initial HbA1c ≥7.0%, subsequently confirmed at the next visit (at 3 months if HbA1c is 7-8.9%, or 3-6 weeks if HbA1c is ≥9.0%), while being treated at maximum tolerable doses of metformin and the assigned treatment.
- Time Frame: Quarterly for 4 to 7 years

CERSI
Yale-Mayo Clinic

https://clinicaltrials.gov/ct2/show/NCT01794143
Diabetes Care. 2013 Aug;36(8):2254-61
Diabetes Care. 2019 Nov;42(11):2098-2107[123]

Potential Outcomes

Primary Outcome

- Time to **primary metabolic failure** of the assigned treatment, defined by the time to HbA1c ≥7.0%.
- Anticipated deviation from GRADE:
  - We will not require a confirmatory HbA1c due to limited availability of laboratory results data within OLDW and variation in real-world HbA1c testing parameters.
  - Anticipate random and non-random variation in timing of available HbA1c results

CERSI
Yale-Mayo Clinic

Baseline Characteristics Before Matching

	Glargine (N=251)	Glimepiride (N=4329)	Liraglutide (N=696)	Sitagliptin (N=3007)	Total (N=8283)	Largest SMD
Age						0.40
Mean (SD)	60.2 (12.6)	63.0 (11.1)	54.9 (9.8)	62.0 (11.2)	61.8 (11.3)	
Median (IQR)	62.0 (51.0, 69.0)	65.0 (56.0, 71.0)	54.0 (48.0, 62.0)	64.0 (54.0, 70.0)	63.0 (54.0, 70.0)	
Age group, years						0.45
30-44	31 (12.4%)	270 (6.2%)	98 (14.1%)	237 (7.9%)	636 (7.7%)	
45-54	47 (18.7%)	726 (16.8%)	252 (36.2%)	540 (18.0%)	1565 (18.9%)	
55-64	66 (26.3%)	1126 (26.0%)	224 (32.2%)	786 (26.1%)	2202 (26.6%)	
65-74	81 (32.3%)	1628 (37.6%)	110 (15.8%)	1103 (36.7%)	2922 (35.3%)	
≥75	26 (10.4%)	579 (13.4%)	12 (1.7%)	341 (11.3%)	958 (11.6%)	
Gender						0.15
Female	133 (53.0%)	1986 (45.9%)	415 (59.6%)	1512 (50.3%)	4046 (48.8%)	
Male	118 (47.0%)	2343 (54.1%)	281 (40.4%)	1495 (49.7%)	4237 (51.2%)	
Race/Ethnicity						0.19
White	160 (63.7%)	2853 (65.9%)	493 (70.8%)	1816 (60.4%)	5322 (64.3%)	
Black	38 (15.1%)	554 (12.8%)	94 (13.5%)	387 (12.9%)	1073 (13.0%)	
Hispanic	29 (11.6%)	505 (11.7%)	75 (10.8%)	409 (13.6%)	1018 (12.3%)	
Asian	11 (4.4%)	243 (5.6%)	18 (2.6%)	251 (8.3%)	523 (6.3%)	
Other, unknown, missing	13 (5.2%)	174 (4.0%)	16 (2.3%)	144 (4.8%)	347 (4.2%)	
Annual Household Income						0.30
<$40,000	73 (29.1%)	1081 (25.0%)	118 (17.0%)	652 (21.7%)	1924 (23.2%)	
$40,000 - $74,999	55 (21.9%)	1192 (27.5%)	169 (24.3%)	772 (25.7%)	2188 (26.4%)	
$75,000 – $124,999	73 (29.1%)	1213 (28.0%)	231 (33.2%)	803 (26.7%)	2320 (28.0%)	
$125,000 – $199,999	16 (6.4%)	429 (9.9%)	106 (15.2%)	407 (13.5%)	958 (11.6%)	
≥200,000	8 (3.2%)	165 (3.8%)	45 (6.5%)	196 (6.5%)	414 (5.0%)	
Unknown/missing	26 (10.4%)	249 (5.8%)	27 (3.9%)	177 (5.9%)	479 (5.8%)	

CERSI
Yale-Mayo Clinic

GRADE Replication – Analyses

- Creating the Study Cohort
  - All patients in OLDW meeting GRADE eligibility criteria
  - Inverse probability of treatment weighting (IPTW)
    - Glargine arm was excluded due to very small sample size
    - After dropping glargine, groups were still not balanced due to liraglutide
      - High cost of the drug, which has limited its uptake particularly among older patients
      - Liraglutide is more often prescribed as a 3rd line agent or in the setting of markedly elevated HbA1c
    - Performed ATT weighting and matching in an effort to balance the treatment arms

CERSI
Yale-Mayo Clinic

Methods for Balancing Cohorts

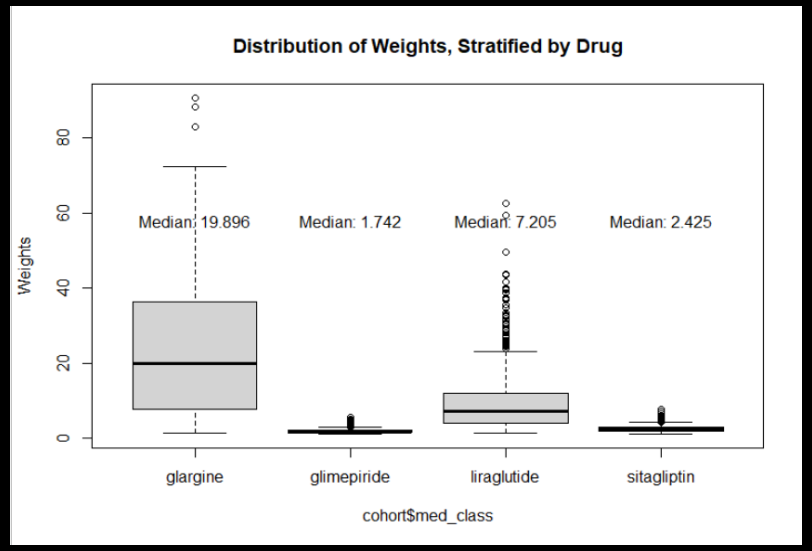
- Propensity score estimation
  - XGBoost multinomial model
  - Lasso multinomial regression model
  - **Generalized boosted logistic models (one for each treatment versus the others)**

- Weighting methods
  - Average Treatment Effect (ATE)
  - Average Treatment Effects on the Treated (ATT)
  - **Stabilized weights**

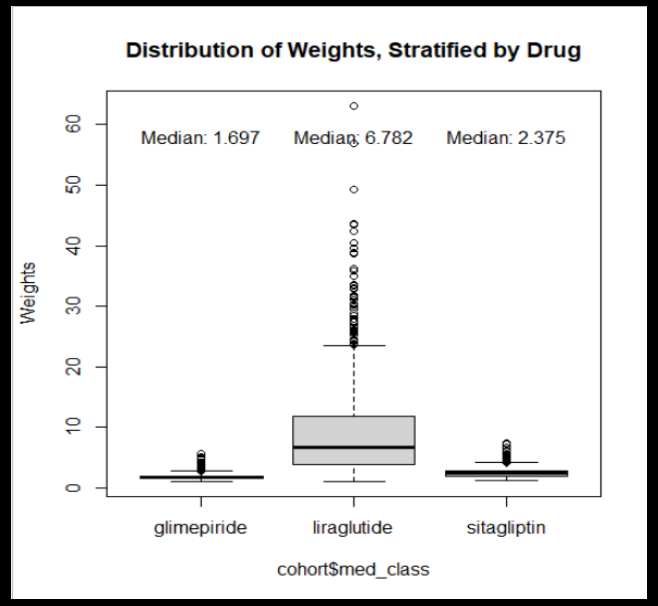
CERSI
Yale-Mayo Clinic

Weight Distributions (generalized boosted models)

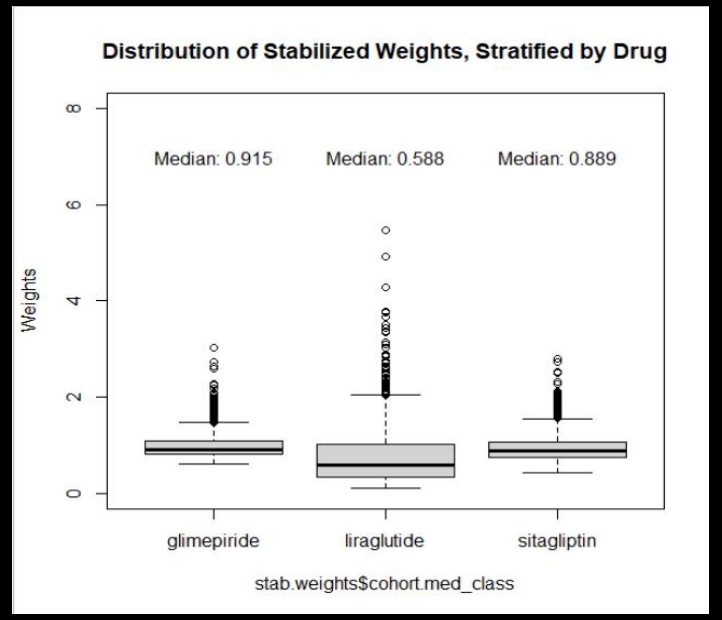
ATE



ATE-Drop Glargine



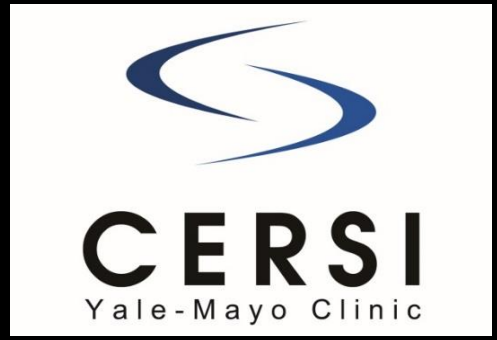
Stabilized weights -Drop Glargine



CERSI
Yale-Mayo Clinic

Generalizability of the CABANA Trial

Assessing Outcomes With Catheter Ablation for Atrial Fibrillation in Routine Practice

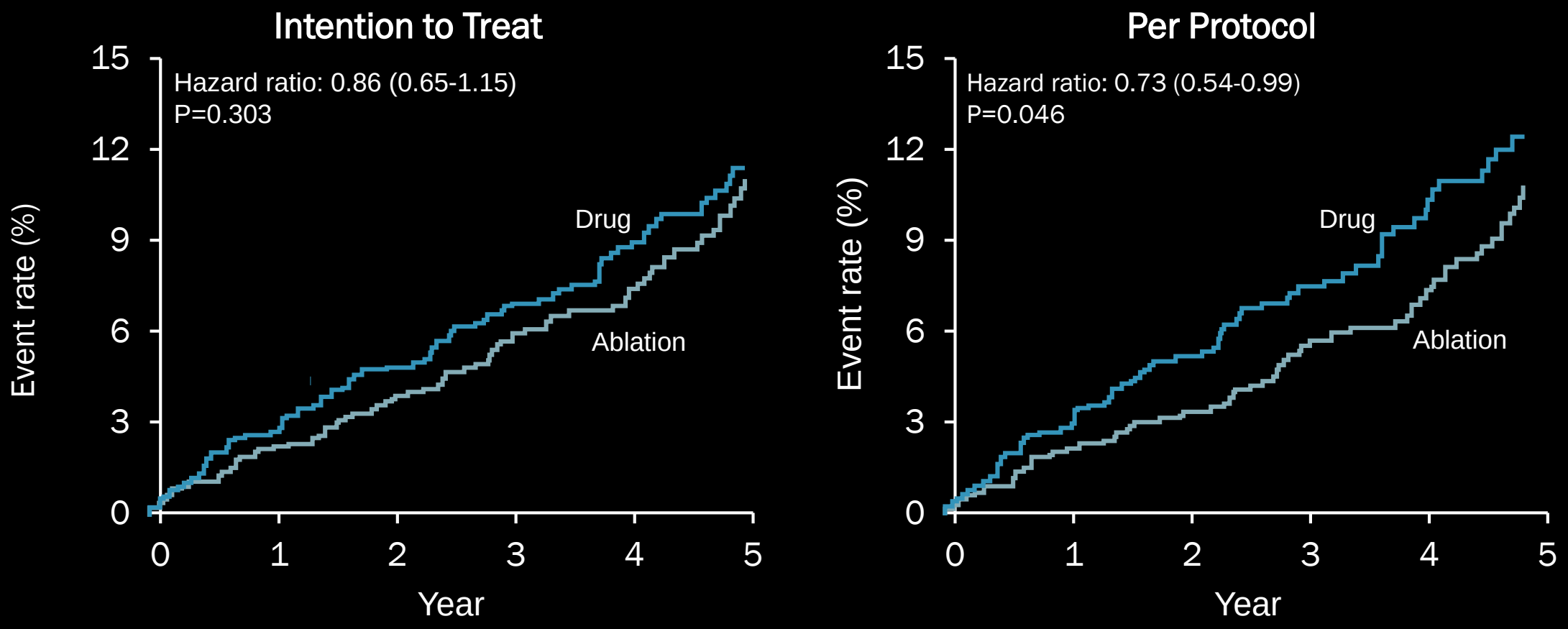


Background: Paired RCT-observational study

- CABANA is an important trial in EP
  - Compared ablation vs. med for AF cardiovascular risk reduction
  - Randomized 2,204 patients
- We initiated a complementary NIH-funded study
  - Conducted in parallel and completed prior to CABANA data lock

CERSI
Yale-Mayo Clinic

Primary CABANA Findings: Impact of Crossover

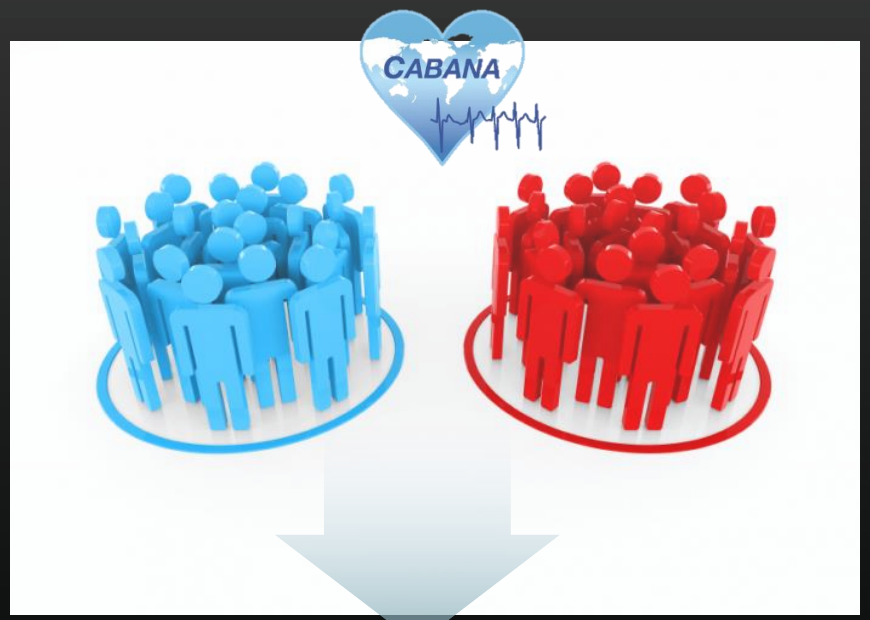


Treatment received: HR 0.67 (0.50, 089), P=0.006

What is the actual benefit of ablation?

CERSI
Yale-Mayo Clinic

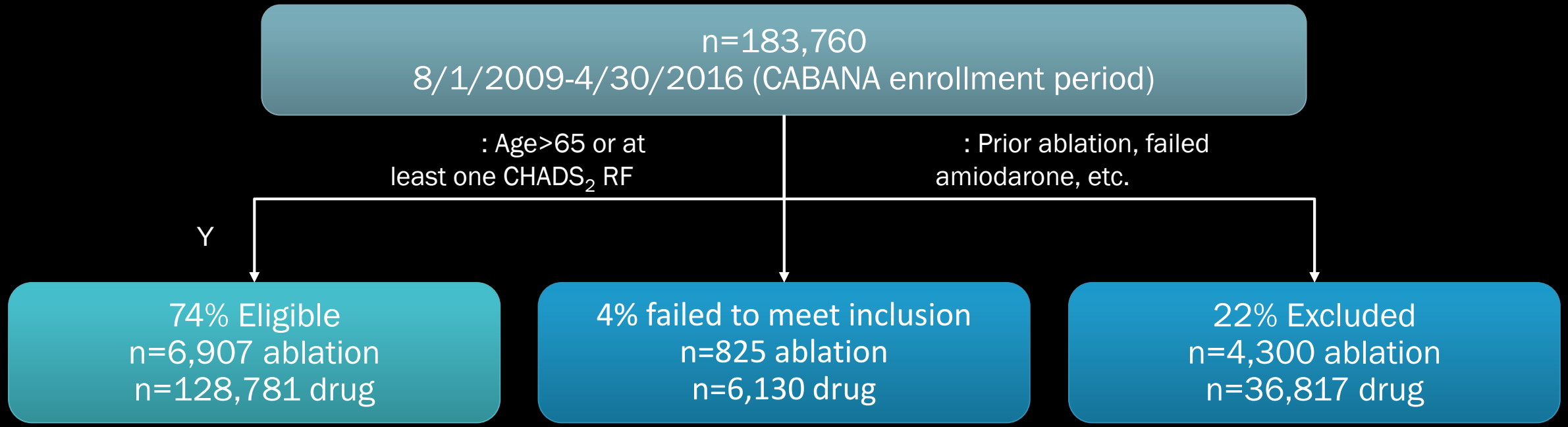
Paired RCT and Observational Data



What is the impact of ablation on cardiovascular outcomes?

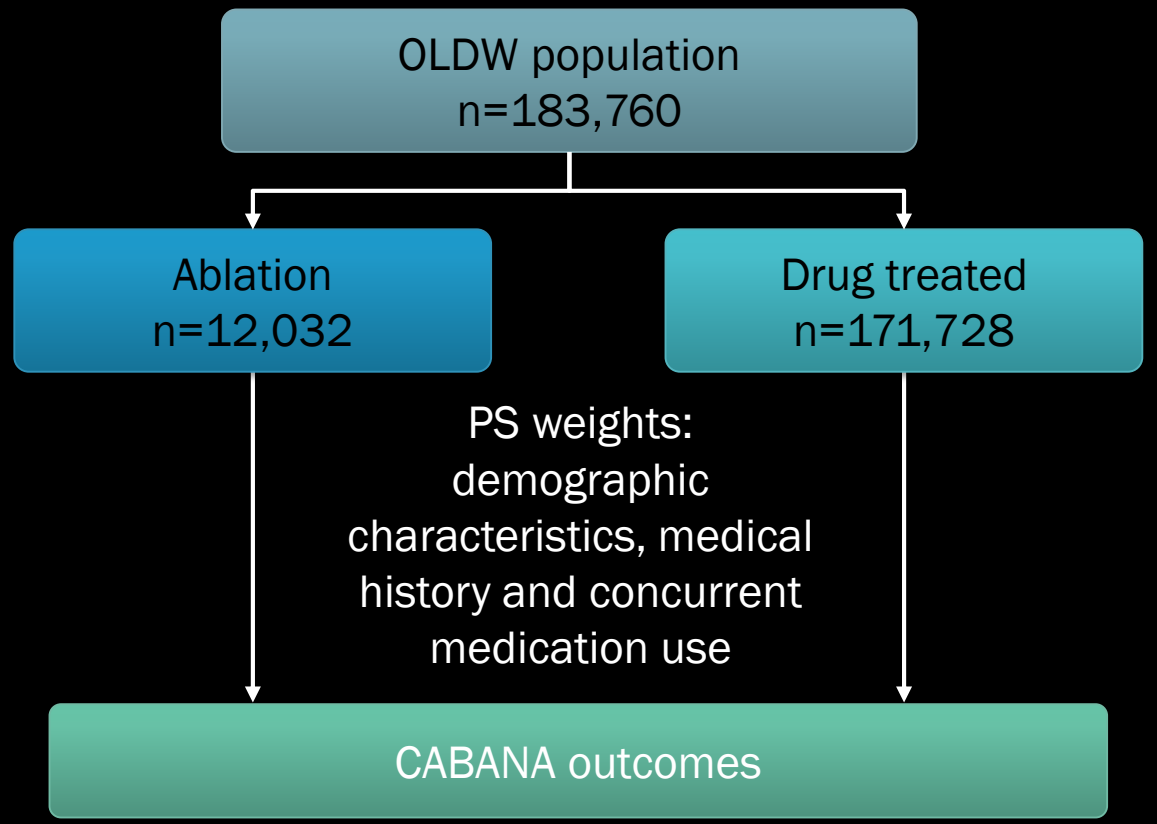
1. Do trial participants represent patients in everyday practice?
2. Can observational data help interpret the trial findings?
3. What is the treatment effect in excluded populations?

Q1: Do Trial Participants Represent Patients in Everyday Practice?



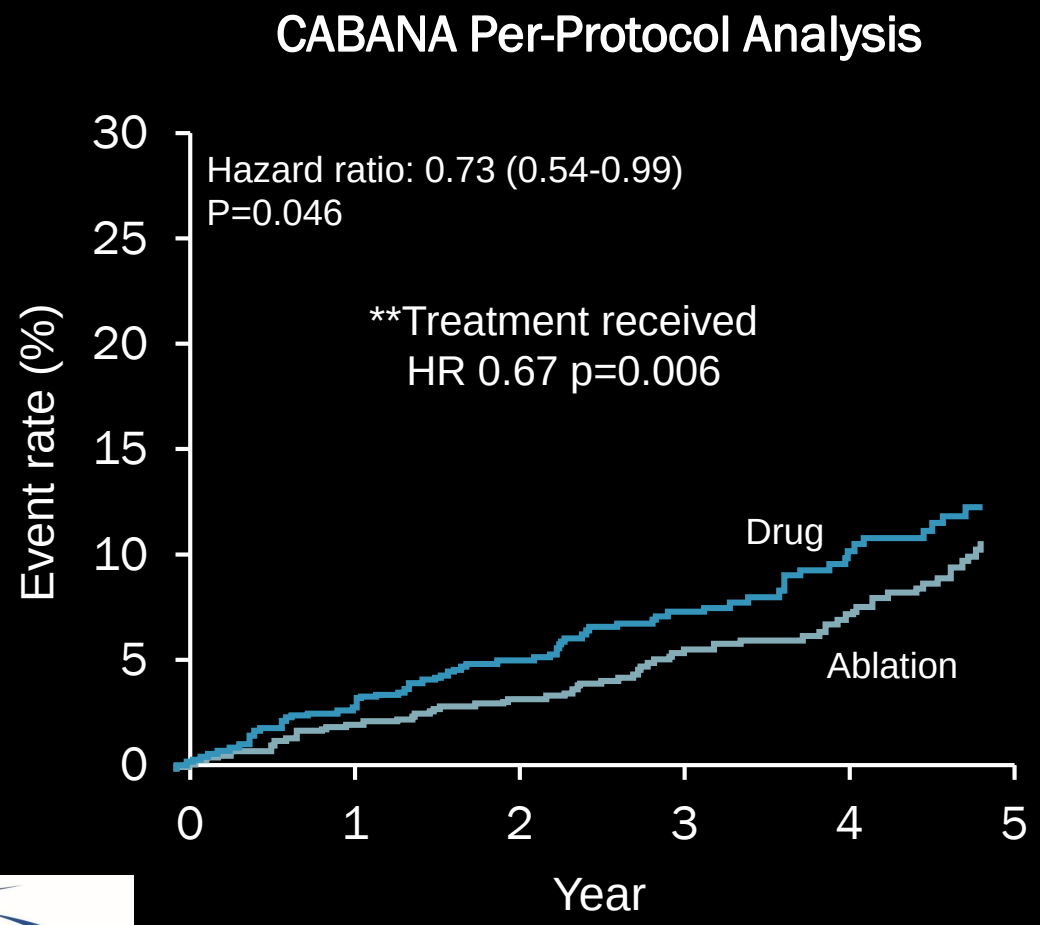
Q2: Can observational data help interpret the trial findings?

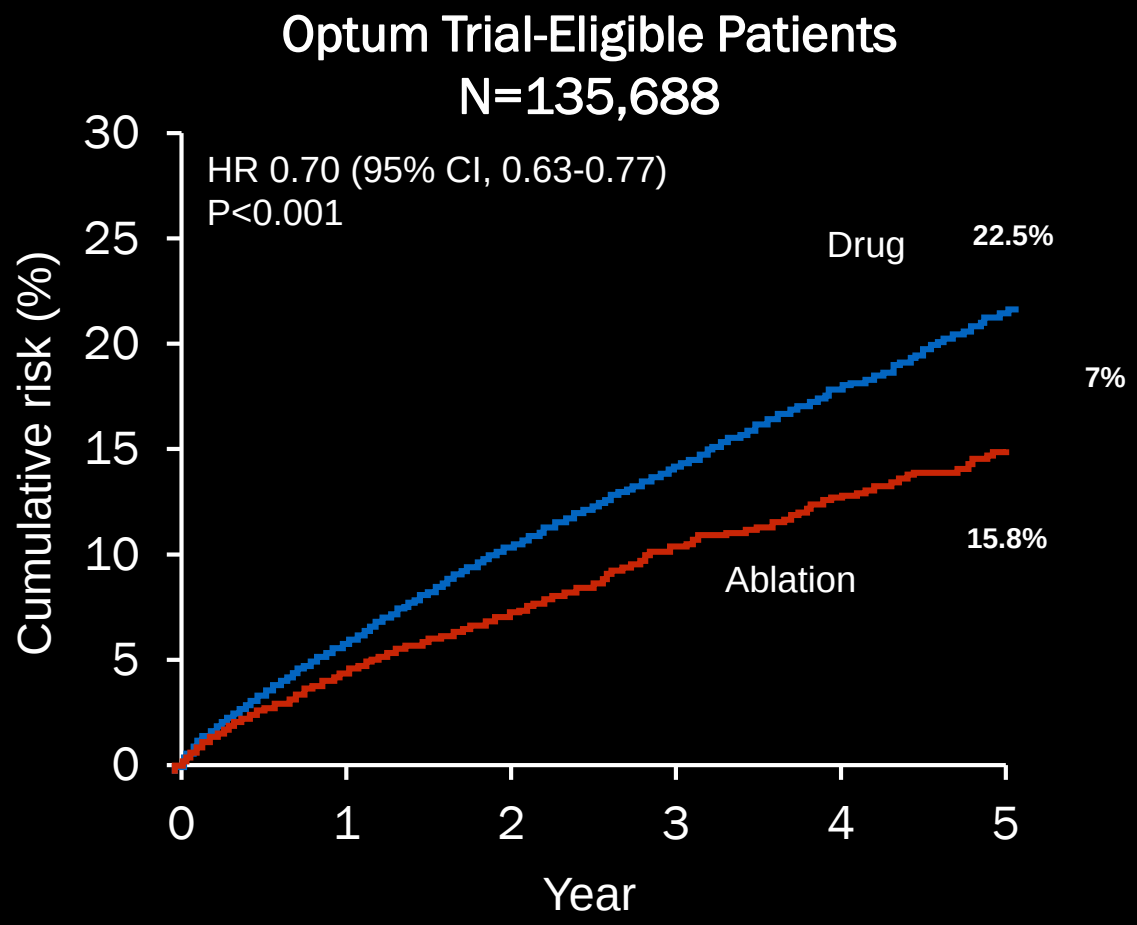
- PS overlap weighting to balance patients on 90 baseline characteristics
- Cox proportional hazards regression
- Primary CABANA outcome:
  - composite of mortality, stroke, major bleeding, and cardiac arrest



CERSI
Yale-Mayo Clinic

Q2: Can observational data help interpret the trial findings?

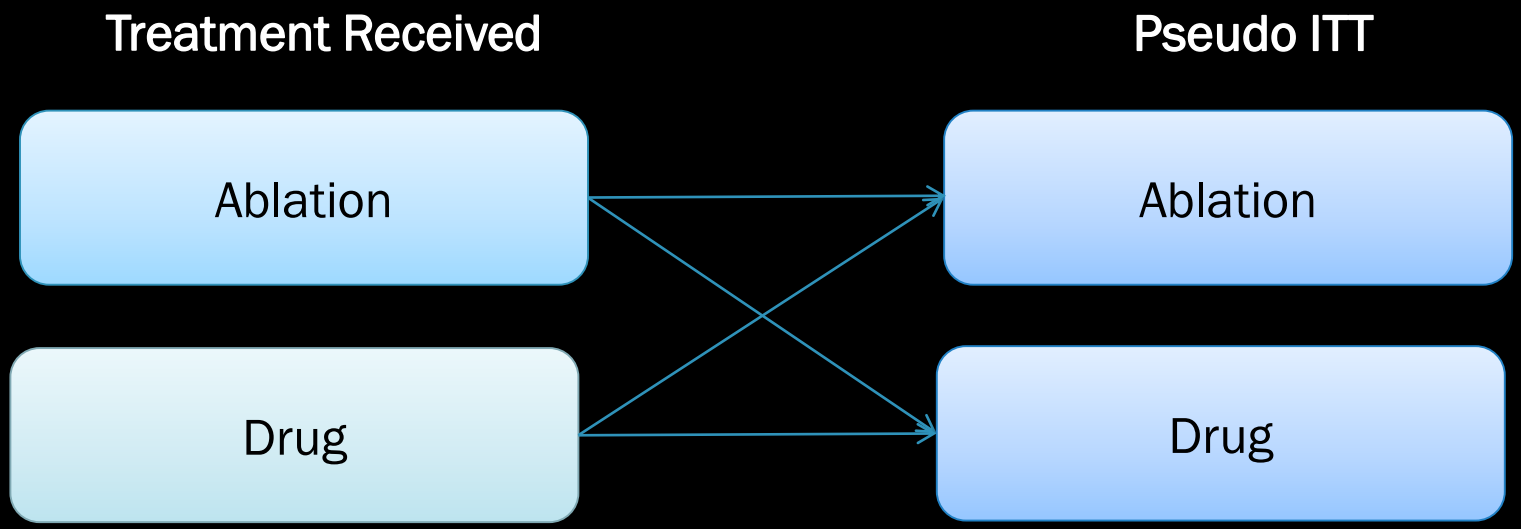




Larger Absolute Risk/Absolute Risk Reduction in Practice vs RCT

CERSI
Yale-Mayo Clinic

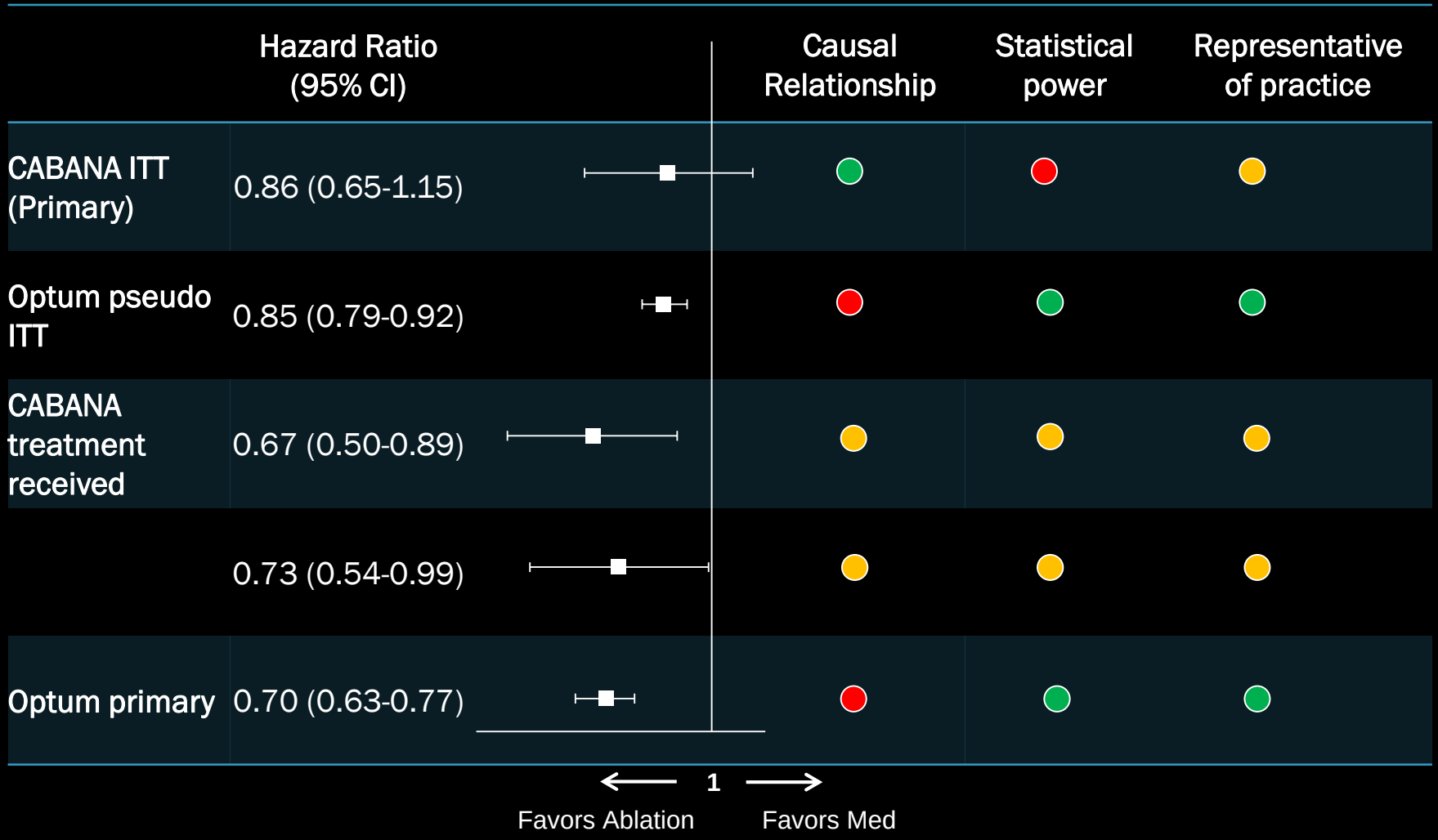
OptumLabs Pseudo Intent-to-Treat Analysis



Simulate crossover to mirror RCT ITT results

CERSI
Yale-Mayo Clinic

Complimentary Evidence from RCT and Observational Data

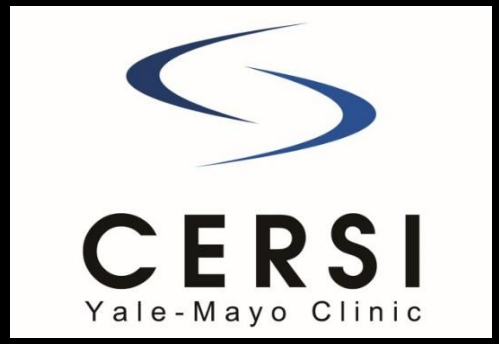


Key Consideration in Use Of Real-World Data for Emulating Clinical Trials

1. Data Sources → information on medication fills and clinical data: ideal to integrate claims and electronic health record data
2. Population representativeness → even when emulating trials in real-world data, this may be different
3. Real world practice vs. clinical trial design
4. Approaches to address selection → at a minimum, needs to conduct sensitivity analyses with multiple approaches; especially a key consideration for treatments that are rarely prescribed
5. Real world results and clinical trial results will not always align – what does that mean?

CERSI
Yale-Mayo Clinic

Questions?



Break — 15 Minutes

We will be back momentarily.

Session 3 will begin at 4:15 pm (U.S. Eastern).

Duke-Margolis is hiring!

Are you interested in real-world evidence, payment for medical products, or antimicrobial resistance?

We have multiple openings. Please visit https://healthpolicy.duke.edu/careers to learn more about opportunities.

Session 3: Reactions to Replication Results

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Panelists

- **Sebastian Schneeweiss**, Harvard Medical School
- **William Crown**, Brandeis University
- **Nilay Shah**, Mayo Clinic
- **Joseph Ross**, Yale University
- **Miguel Hernán**, Harvard T.H. Chan School of Public Health
- **Robert Temple**, U.S. Food and Drug Administration

Day 1 Adjournment

Evaluating RWE from Observational Studies in Regulatory Decision-Making: Lessons Learned from Trial Replication Analyses

February 16 & 17, 2021

FDA U.S. FOOD & DRUG ADMINISTRATION | Duke MARGOLIS CENTER for Health Policy

Welcome & Overview | Day 2

Mark McClellan
Duke-Robert J. Margolis, MD, Center for Health Policy

Meeting Agenda

Day One

- Session 1: Principles for Using Rigorous Observational Study Designs to Replicate the Results of Clinical Trials
- Session 2: Session 2: Presentations from Trial Replication Projects
- Session 3: Reactions to Replication Results

Day Two

- Session 4: Key Themes Emerging from Replication Efforts
- Session 5: Observational Studies: Opportunities, Limitations, and Next Steps

Virtual Meeting Reminders

- Visit the Duke-Margolis website (https://healthpolicy.duke.edu/events) for meeting materials, including the agenda, speaker biographies, and discussion topics.

- Questions for our panelists? Feel free to submit questions via Zoom's Q&A function.

- Join the conversation @Duke-Margolis #TrialReplication

Session 4: Key Themes Emerging from Replication Efforts

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Lucinda Orsini

COMPASS Pathways

www.ispor.org

ISPOR

Improving healthcare decisions

RWE and RCT Replication

Lucinda S. Orsini, DPM, MPH
VP – Value and Outcomes Research, COMPASS Pathways
Former Associate Chief Science Officer – ISPOR

Making RWE Useful Requires

- Quality Production
  - Careful data collection and/or curation
  - Appropriate analytic methods
  - Good procedural practices for transparent study process
  - Replicability/reproducibility
- Responsible Consumption
  - Informed interpretation
  - Fit-for-purpose application

How important is it that RWE 'replicates' RWE?

What we don't mean:

We don't want to imply that all RWE studies need to calibrate against an RCT – that would defeat the purpose of RWE

It is obvious that RWE complements and expands the evidence generated by RCTs and often expands the line of inquiry

- Different populations
- Different treatment patterns
- Different endpoints
- Different comparators

Key Consideration in Use Of Real-World Data for Emulating Clinical Trials

1. Data Sources → information on medication fills and clinical data: ideal to integrate claims and electronic health record data
2. Population representativeness → even when emulating trials in real-world data, this may be different
3. Real world practice vs. clinical trial design
4. Approaches to address selection → at a minimum, needs to conduct sensitivity analyses with multiple approaches; especially a key consideration for treatments that are rarely prescribed
5. Need to determine approaches for validated outcomes in real world data. There is a potential role for NLP in the future.
6. Real world results and clinical trial results will not always align – what does that mean?

How important is it that RWE 'replicates' RWE?

What we don't mean:

We don't want to imply that all RWE studies need to calibrate against an RCT – that would defeat the purpose of RWE

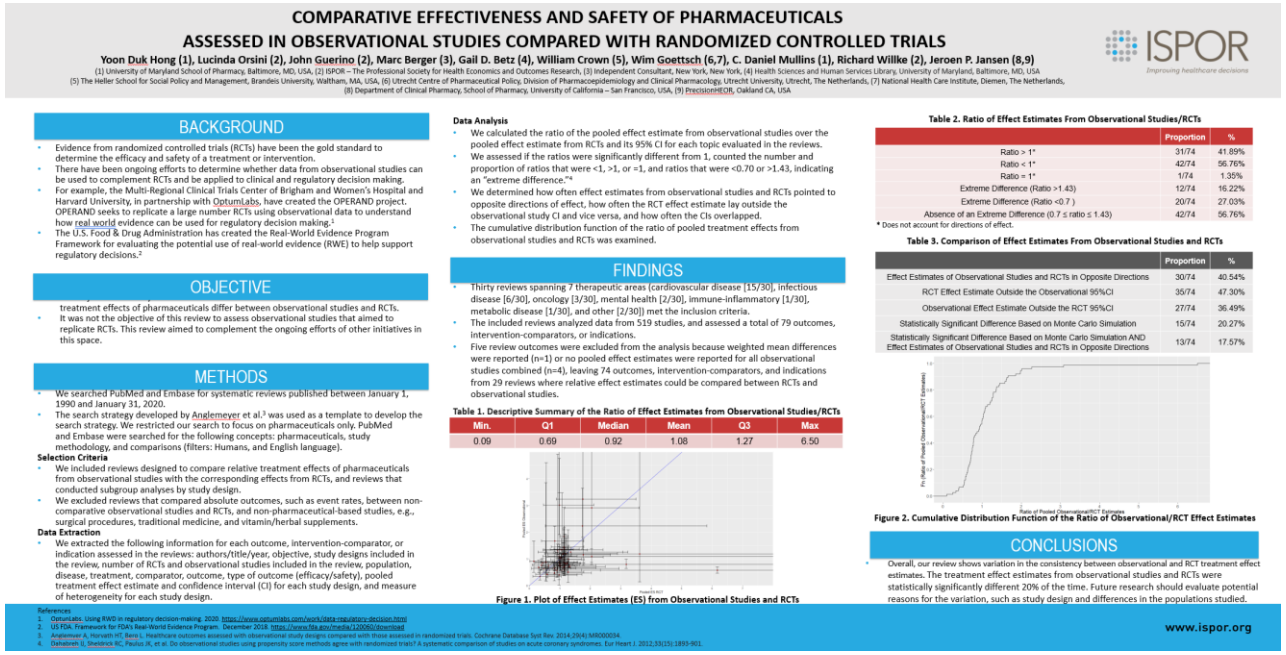
It is obvious that RWE complements and expands the evidence generated by RCTs and often expands the line of inquiry

- Different populations
- Different treatment patterns
- Different endpoints
- Different comparators

Key Consideration in Use Of Real-World Data for Emulating Clinical Trials

1. Data Sources → information on medication fills and clinical data: ideal to integrate claims and electronic health record data
2. Population representativeness → even when emulating trials in real-world data, this may be different
3. Real world practice vs. clinical trial design
4. Approaches to address selection → at a minimum, needs to conduct sensitivity analyses with multiple approaches; especially a key consideration for treatments that are rarely prescribed
5. Need to determine approaches for validated outcomes in real world data. There is a potential role for NLP in the future.
6. Real world results and clinical trial results will not always align – what does that mean?

Review of Reviews comparing RCT and observational studies



COMPARATIVE EFFECTIVENESS AND SAFETY OF PHARMACEUTICALS
ASSESSED IN OBSERVATIONAL STUDIES COMPARED WITH RANDOMIZED CONTROLLED TRIALS

Yoon Duk Hong (1), Lucinda Orsini (2), John Guerino (2), Marc Berger (3), Gail D. Betz (4), William Crown (5), Wim Goettsch (6,7), C. Daniel Mullins (1), Richard Willke (2), Jeroen P. Jansen (8,9)

(1) University of Maryland School of Pharmacy, Baltimore, MD, USA, (2) ISPOR – The Professional Society for Health Economics and Outcomes Research, (3) Independent Consultant, New York, New York, (4) Health Sciences and Human Services Library, University of Maryland, Baltimore, MD, USA (5) The Heller School for Social Policy and Management, Brandeis University, Waltham, MA, USA, (6) Utrecht Centre of Pharmaceutical Policy, Division of Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands, (7) National Health Care Institute, Diemen, The Netherlands, (8) Department of Clinical Pharmacy, School of Pharmacy, University of California – San Francisco, USA, (9) PrecisionHEOR, Oakland CA, USA

BACKGROUND

- Evidence from randomized controlled trials (RCTs) have been the gold standard to determine the efficacy and safety of a treatment or intervention.
- There have been ongoing efforts to determine whether data from observational studies can be used to complement RCTs and be applied to clinical and regulatory decision making.
- For example, the Multi-Regional Clinical Trials Center of Brigham and Women's Hospital and Harvard University, in partnership with OptumLabs, have created the OPERAND project. OPERAND seeks to replicate a large number RCTs using observational data to understand how real world evidence can be used for regulatory decision making.[1]
- The U.S. Food & Drug Administration has created the Real-World Evidence Program Framework for evaluating the potential use of real-world evidence (RWE) to help support regulatory decisions.[2]

OBJECTIVE

- treatment effects of pharmaceuticals differ between observational studies and RCTs.
- It was not the objective of this review to assess observational studies that aimed to replicate RCTs. This review aimed to complement the ongoing efforts of other initiatives in this space.

METHODS

- We searched PubMed and Embase for systematic reviews published between January 1, 1990 and January 31, 2020.
- The search strategy developed by Anglemeyer et al.[3] was used as a template to develop the search strategy. We restricted our search to focus on pharmaceuticals only. PubMed and Embase were searched for the following concepts: pharmaceuticals, study methodology, and comparisons (filters: Humans, and English language).

Selection Criteria

- We included reviews designed to compare relative treatment effects of pharmaceuticals from observational studies with the corresponding effects from RCTs, and reviews that conducted subgroup analyses by study design.
- We excluded reviews that compared absolute outcomes, such as event rates, between non-comparative observational studies and RCTs, and non-pharmaceutical-based studies, e.g., surgical procedures, traditional medicine, and vitamin/herbal supplements.

Data Extraction

- We extracted the following information for each outcome, intervention-comparator, or indication assessed in the reviews: authors/title/year, objective, study designs included in the review, number of RCTs and observational studies included in the review, population, disease, treatment, comparator, outcome, type of outcome (efficacy/safety), pooled treatment effect estimate and confidence interval (CI) for each study design, and measure of heterogeneity for each study design.

Data Analysis

- We calculated the ratio of the pooled effect estimate from observational studies over the pooled effect estimate from RCTs and its 95% CI for each topic evaluated in the reviews.
- We assessed if the ratios were significantly different from 1, counted the number and proportion of ratios that were <1, >1, or =1, and ratios that were <0.70 or >1.43, indicating an "extreme difference."*
- We determined how often effect estimates from observational studies and RCTs pointed to opposite directions of effect, how often the RCT effect estimate lay outside the observational study CI and vice versa, and how often the CIs overlapped.
- The cumulative distribution function of the ratio of pooled treatment effects from observational studies and RCTs was examined.

FINDINGS

- Thirty reviews spanning 7 therapeutic areas (cardiovascular disease [15/30], infectious disease [6/30], oncology [3/30], mental health [2/30], immune-inflammatory [1/30], metabolic disease [1/30], and other [2/30]) met the inclusion criteria.
- The included reviews analyzed data from 519 studies, and assessed a total of 79 outcomes, intervention-comparators, or indications.
- Five review outcomes were excluded from the analysis because weighted mean differences were reported (n=1) or no pooled effect estimates were reported for all observational studies combined (n=4), leaving 74 outcomes, intervention-comparators, and indications from 29 reviews where relative effect estimates could be compared between RCTs and observational studies.

Table 1. Descriptive Summary of the Ratio of Effect Estimates from Observational Studies/RCTs

Min.	Q1	Median	Mean	Q3	Max
0.09	0.69	0.92	1.08	1.27	6.50

Figure 1. Plot of Effect Estimates (ES) from Observational Studies and RCTs

Table 2. Ratio of Effect Estimates From Observational Studies/RCTs

	Proportion	%
Ratio > 1*	31/74	41.89%
Ratio < 1*	42/74	56.76%
Ratio = 1*	1/74	1.35%
Extreme Difference (Ratio >1.43)	12/74	16.22%
Extreme Difference (Ratio <0.7)	20/74	27.03%
Absence of an Extreme Difference (0.7 ≤ ratio ≤ 1.43)	42/74	56.76%

* Does not account for directions of effect.

Table 3. Comparison of Effect Estimates From Observational Studies and RCTs

	Proportion	%
Effect Estimates of Observational Studies and RCTs in Opposite Directions	30/74	40.54%
RCT Effect Estimate Outside the Observational 95%CI	35/74	47.30%
Observational Effect Estimate Outside the RCT 95%CI	27/74	36.49%
Statistically Significant Difference Based on Monte Carlo Simulation	15/74	20.27%
Statistically Significant Difference Based on Monte Carlo Simulation AND Effect Estimates of Observational Studies and RCTs in Opposite Directions	13/74	17.57%

Figure 2. Cumulative Distribution Function of the Ratio of Observational/RCT Effect Estimates

CONCLUSIONS

- Overall, our review shows variation in the consistency between observational and RCT treatment effect estimates. The treatment effect estimates from observational studies and RCTs were statistically significantly different 20% of the time. Future research should evaluate potential reasons for the variation, such as study design and differences in the populations studied.

References

1. OptumLabs. Using RWD in regulatory decision-making. 2020. https://www.optumlabs.com/work/data-regulatory-decision.html
2. US FDA. Framework for FDA's Real-World Evidence Program. December 2018. https://www.fda.gov/media/120060/download
3. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. Cochrane Database Syst Rev. 2014;29(4):MR000034.
4. Dahabreh IJ, Sheldrick RC, Paulus JK, et al. Do observational studies using propensity score methods agree with randomized trials? A systematic comparison of studies on acute coronary syndromes. Eur Heart J. 2012;33(15):1893-901.

www.ispor.org

- Overall, our review shows variation in the consistency between observational and RCT treatment effect estimates.
- The treatment effect estimates from observational studies and RCTs were statistically significantly different 20% of the time.
- Future research should evaluate potential reasons for the variation, such as study design and differences in the populations studied

COMPARATIVE EFFECTIVENESS AND SAFETY OF PHARMACEUTICALS ASSESSED IN OBSERVATIONAL STUDIES COMPARED WITH RANDOMIZED CONTROLLED TRIALS,
Poster Presented at ISPOR Virtual Europe 2020
https://www.ispor.org/heor-resources/presentations-database/presentation/euro2020-3282/107871

RWE Transparency Initiative

Received: 22 April 2020 | Revised: 12 June 2020 | Accepted: 23 June 2020
DOI: 10.1002/pds.5079

COMMENTARY

WILEY

Improving transparency to build trust in real-world secondary data studies for hypothesis test recommendations and a road m evidence transparency initiative

Lucinda S. Orsini[1] | Brigitta Monz[2] | C. D
Gregory Daniel[5] | Hans-Georg Eichler[6] |
Marc Berger[1] | Nirosha M. Lederer[5] | Pal
Shirley V. Wang[9] | William Crown[10] | V

VALUE HEALTH. 2020; 23(9):1128–1136

ScienceDirect

Contents lists available at sciencedirect.com
Journal homepage: www.elsevier.com/locate/jval

ISPOR Report

Improving Transparency to Build Trust in Real-World Secondary Data Studies for Hypothesis Testing—Why, What, and How: Recommendations and a Road Map from the Real-World Evidence Transparency Initiative

Lucinda S. Orsini, DPM, MPH,* Marc Berger, MD, William Crown, PhD, Gregory Daniel, PhD, MPH, Hans-Georg Eichler, MD, Wim Goettsch, PhD, Jennifer Graff, PharmD, John Guerino, MHS, Pall Jonsson, PhD, Nirosha Mahendraratnam Lederer, PhD, Brigitta Monz, MD, MPH, MA, C. Daniel Mullins, PhD, Sebastian Schneeweiss, MD, ScD, David Van Brunt, PhD, Shirley V. Wang, PhD, ScM, Richard J. Willke, PhD

Modeling good study 'hygiene'

Transparency

CT.gov registration:

- Complete protocol of each emulation

1. **RCT Details**
2. **Person responsible for implementation of replication in Aetion**
3. **Data Source(s)**
4. **Study Design Diagram**
5. **Cohort Identification**
   - 5.1 Inclusion/exclusion criteria for cohort entry
   - 5.2 Flowchart of the study cohort assembly
6. **Variables**
   - 6.1 Exposure-related variables:
   - 6.2 Preliminary Covariates:
   - 6.3 Outcome variables and study follow-up:
7. **Initial Feasibility Analysis**
   - Aetion report name:
   - Date conducted:
8. **Initial Power Assessment**
9. **Balance Assessment after PS matching**
   - Aetion report name:
   - Date conducted:
10. **Final Power Assessment**
    - Aetion report name:
    - Date conducted:
11. **Study Confidence and Concerns**
12. **Register study protocol on clinicalTrials.gov**
13. **Comparative Analyses**
    - Aetion report name:
    - Date conducted:
14. **Requested Results**
15. **References**

Comparative analysis starts after registration

20

OPERAND

Transparency

- Both teams registered their study protocols on the EU-PAS registry before they were given access to any data.
  - http://www.encepp.eu/encepp_studies/indexRegister.shtml

RWE – Bad Image Continues

This article has been retracted: N Engl J Med. DOI: 10.1056/NEJMc2021225

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19

Mandeep R. Mehra, M.D
SreyRam Kuy, M.D., M.H.S., Timoth

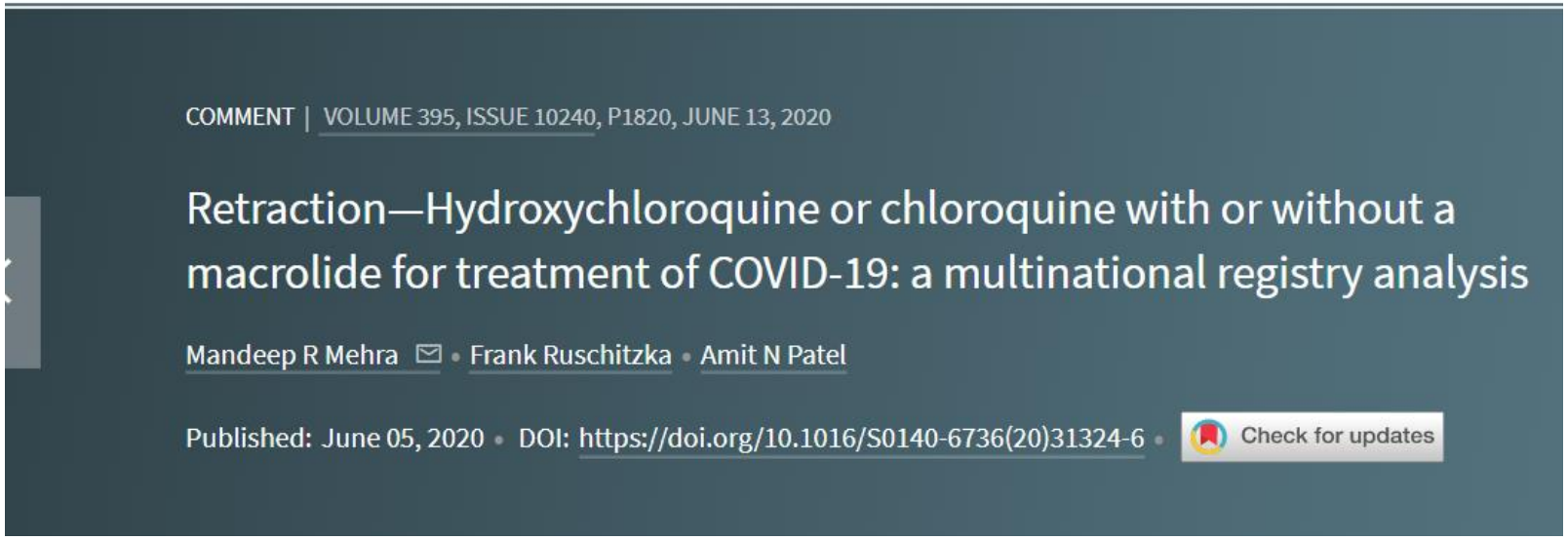
THE LANCET

The NEW ENGLAND JOURNAL of MEDICINE

SOUNDING BOARD

The Magic of Randomization versus the Myth of Real-World Evidence

Rory Collins, F.R.S., Louise Bowman, M.D., F.R.C.P., Martin Landray, Ph.D., F.R.C.P., and Richard Peto, F.R.S.



Josie Briggs

Patient-Centered Outcomes Research Institute

Josie Briggs

JASN EIC

PCORI- Senior Advisor

Duke-Margolis RWE Symposium

Feb 17, 2021

A few comments from the journal editor perspective

- Guilty as charged
- Applaud a transformative set of new standards for observational research on efficacy of interventions

- Continued worries:
  - True match of initiator and non-initiator
  - Outcome measures – do they capture what matters
  - Ascertainment bias especially for outcome measures

- Help needed:
  - Data-sharing expectations – what is practical?
  - Pre-registration ?
  - Guards against publication bias

A few comments from the public sector funder perspective

- Value of pre- emulation studies
  - Refine questions – for whom? How delivered? When? Where? Etc.
  - Clarify equipoise
- Value of negative studies
  - Important to know what doesn’t really work
  - Less is often more
- Worry- We still need the gold standard trials

Michele Jonsson-Funk

UNC Gillings School of Public Health

Michele Jonsson Funk, PhD
Associate Professor of Epidemiology
Director, Center for Pharmacoepidemiology
Gillings School of Global Public Health
University of North Carolina at Chapel Hill

Duke-Margolis RWE Symposium

17 February 2021

When is $HR_{RCT} \neq HR_{RWE}$ *and* HR_{RWE} is unbiased?

- Random error (not systematic error) in one or the other
- Answering different questions
  - Different estimands (tx effect in treated vs. total popn)
- HTE and different population distributions
  - If background rate of the outcome differs and true effect is not null, HTE guaranteed on absolute or relative scale
- Treatment itself is different
  - RCT: 2 doses 28d (-3 / +7) days apart
  - RWE: 2 doses up to 8 weeks apart
- Outcome differs (FU duration, frequency of competing risks)
- Adherence differs
  - Same Rx, same population. In RCT, Rx provided for free. In RWE, Rx requires that patient pay co-pay (or out of pocket) which not all can afford. Due to cost differences between tx of interest and alternatives, tx will appear less effective.

A few take away messages

- Recognizing when the data are not sufficient to support the analysis is key
  - Deep understanding of data sources and context in which they were generated is essential to assess whether data are fit-for-purpose
- Concerns about loss of sample size should not justify inclusion of individuals who are not appropriate for the target trial of interest (prevalent users)
- Fancy statistical analyses cannot make up for fundamental errors in study design (eg use of crystal balls, time travel)
- Key features of design that limit bias in RWE differ from those for RCTs
  - Thus, expertise needed to critically review RWE differs
- Bias due to unmeasured confounding less problematic than other sources but cannot be ignored

Worry about unmeasured confounding is common.

- Systematic review of observational, head-to-head cohort studies of drugs or biologics
- Published in high-impact medical and epidemiology journals
- 3 years: 2017, 2018, and 2019
- 83 publications identified
  - 43 in major medical journals
  - 40 in epidemiology journals

- 89% (74/83) listed “residual,” “unmeasured,” or “uncontrolled” confounding as a limitation

Major Medical Journals (n=43)

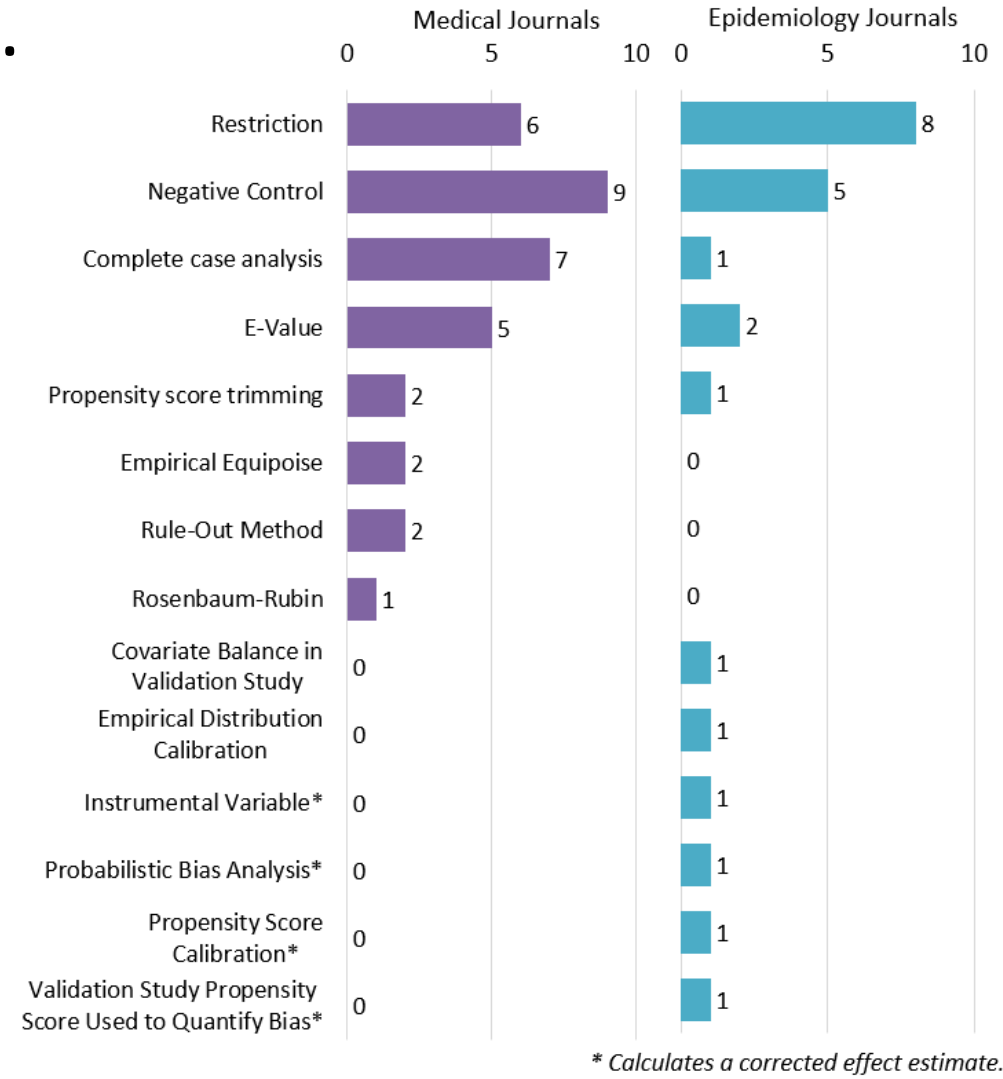
- *JAMA*
- *New England Journal of Medicine*
- *British Medical Journal*
- *The Lancet*
- *Annals of Internal Medicine*

Major Epidemiology Journals (n=40)

- *American Journal of Epidemiology*
- *Annals of Epidemiology*
- *Epidemiology*
- *European Journal of Epidemiology*
- *International Journal of Epidemiology*
- *Journal of Clinical Epidemiology*
- *Journal of Environmental and Community Health*
- *Pharmacoepidemiology and Drug Safety*

Explicitly assessing the threat is not.

- 32 (43%) of the 74 articles did not report sensitivity analysis to assess the potential bias due to unmeasured / uncontrolled confounding.
- Of those that did, few used methods that formally produced a corrected estimate of the treatment effect.



** Calculates a corrected effect estimate.*

Robert Ball

U.S. Food and Drug Administration

Break — 15 Minutes

We will be back momentarily.

Session 5 will begin at 2:40 pm (U.S. Eastern).

Duke-Margolis is hiring!

Are you interested in real-world evidence, payment for medical products, or antimicrobial resistance?

We have multiple openings. Please visit https://healthpolicy.duke.edu/careers to learn more about opportunities.

Session 5: Observational Studies: Opportunities, Limitations, and Next Steps

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Nancy Dreyer

IQVIA

Frank Harrell

Vanderbilt University

Rob Reynolds

GlaxoSmithKline

John Concato

U.S. Food and Drug Administration

Closing Remarks & Meeting Adjournment

Moderator: Mark McClellan, Duke-Robert J. Margolis, MD, Center for Health Policy

Thank You!

Contact Us

healthpolicy.duke.edu

Subscribe to our monthly newsletter at
dukemargolis@duke.edu

1201 Pennsylvania Avenue, NW, Suite 500
Washington, DC 20004

DC office: 202-621-2800
Durham office: 919-419-2504

Follow Us

DukeMargolis

@DukeMargolis

@DukeMargolis

Duke Margolis