Lessons Learned from Trial Replication Analyses: Findings from the DUPLICATE Demonstration Project

May 10, 2022

U.S. FOOD & DRUG ADMINISTRATION

Duke MARGOLIS CENTER for Health Policy

Welcome and Introductions

Rachele Hendricks-Sturrup

Research Director, Real World Evidence

Duke-Margolis Center for Health Policy

Statement of Independence

The Robert J. Margolis, MD, Center for Health Policy is part of Duke University, and as such it honors the tradition of academic independence on the part of its faculty and scholars. Neither Duke nor the Margolis Center take partisan positions, but the individual members are free to speak their minds and express their opinions regarding important issues.

For more details on relevant institutional policies, please refer to the Duke Faculty Handbook, including the Code of Conduct and other policies and procedures. In addition, regarding positions on legislation and advocacy, Duke University policies are available at http://publicaffairs.duke.edu/government.

Virtual Meeting Reminders

- Attendees are encouraged to contribute throughout the meeting with questions in the Zoom Q&A function.
- This meeting is being recorded, and the recording and slide deck will be posted on the Duke-Margolis event page in the weeks following the meeting.

Meeting Agenda

12:00 pm - Welcome and Introductions

12:05 pm - Opening Remarks from FDA

12:15 pm - Overview of Technical Findings: The RCT-DUPLICATE Demonstration Project

1:15 pm - Break

1:20 pm - Stakeholder Reactions to Result Findings, and Implications

2:25 pm - Closing Remarks

2:30 pm - Adjournment

FDA Opening Remarks

John Concato

U.S. Food and Drug Administration

U.S. FOOD & DRUG ADMINISTRATION

Duke-Margolis Public Webinar

Lessons Learned from RCT-DUPLICATE

10 May 2022

John Concato, MD, MS, MPH
Associate Director for Real-World Evidence Analytics
Office of Medical Policy
Center for Drug Evaluation and Research
U.S. Food and Drug Administration

Disclaimer

- **Views and opinions expressed are those of the presenter and should not be attributed to the Food and Drug Administration**

- **No conflicts of interest exist related to this presentation**

21st Century Cures Act (2016) – RWE

- **FDA *established* a program to evaluate the potential use of real-world evidence (RWE) to:**
  - **Support a new indication for a drug approved under section 505(c)**
  - **Satisfy post-approval study requirements**
- **Draft framework *issued* in December 2018:**
  - **Describe sources of RWE, challenges, pilot opportunities, etc.**
- **Draft guidance for industry *issued* in Sep, Oct, Nov, & Dec 2021**
- **Standard for substantial evidence remains unchanged; *commitments met* for Prescription Drug User Fee Act (PDUFA) VI**

Background: 'Real-World' Definitions (FDA 2018)

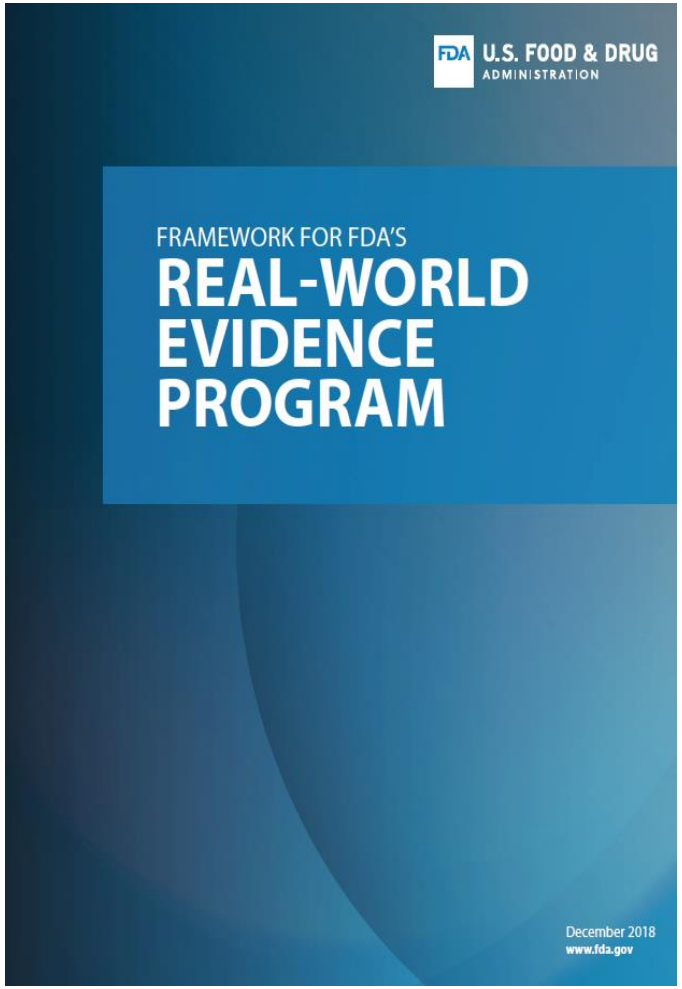
Real World Data (RWD) are data relating to patient health status and/or delivery of health care routinely collected from a variety of sources

- electronic health records (EHRs)
- medical claims data
- product and disease registries
- patient-generated data, including from in-home settings
- other sources that can inform on health status, such as "wearable" devices

Real World Evidence (RWE) is clinical evidence regarding the usage and potential benefits/risks of a medical product derived from analysis of RWD

Generated using different study designs, including but not limited to randomized trials (e.g., pragmatic clinical trials), externally controlled trials, or observational studies

FDA RWE Framework (2018)



- **Applies to Center for Drug Evaluation and Research (CDER) and Center for Biologics Evaluation and Research (CBER)**
- **Multifaceted program to implement RWE:**
  - **internal processes**
  - **external stakeholder engagement**
  - **guidance development**
  - **demonstration projects**

https://www.fda.gov/media/120060/download

FDA 'RWE Guidance' (Sep-Dec 2021)

Guidance for Industry

DRAFT GUIDANCE

Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products

Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products

Data Standards for Drug and Biological Product Submissions Containing Real-World Data

Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products

Ongoing RWE Demonstration Projects - Examples

- ‘OneSource’ project
- Linking RCTs w/ RWD
- ‘ICAREdata’ project

- RCT-DUPLICATE trial emulations
- Statistics for RCT designs w/ hybrid controls

- Evaluating confounded treatment effects
- Targeted learning framework for causal effect estimation

Opinions Regarding Study Design

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

“[…] because of the potential biases in observational studies, such studies cannot generally be trusted […] the replacement of randomized trials with nonrandomized observational analyses is a false solution to the serious problem of ensuring that patients receive treatments that are both safe and effective.” (Collins, *New Engl J Med* 2020;382:674)

‘Misunderstanding randomized controlled trials’

“We argue that any special status for RCTs is unwarranted. Which method is likely to yield a good causal inference depends on what we are trying to discover as well as on what is already known.” (Deaton & Cartwright, *Soc Sci Med*, 2018;210:2)

Real-World Evidence — Where Are We Now?

John Concato, M.D., M.P.H., and Jacqueline Corrigan-Curay, J.D., M.D.

Randomized, Interventional Study		Nonrandomized, Interventional Study	Nonrandomized, Noninterventional Study
Traditional randomized trial using RWD in planning	**Trial in clinical practice settings, with pragmatic elements**	**Externally controlled trial**	**Observational study**
RWD used to assess enrollment criteria and trial feasibility RWD used to support selection of trial sites	Selected outcomes identified using, e.g., health records data, claims data, or data from digital health technologies RCT conducted using, e.g., electronic case report forms for health records data or claims data	Single-group trial with external control group derived from RWD	Cohort study Case–control study Case–crossover study

Generation of RWE

Increasing reliance on RWD

Reliance on RWD in Representative Types of Study Design.

RCT denotes randomized, controlled trial; RWD real-world data; and RWE real-world evidence.

N ENGL J MED 386;18 NEJM.ORG MAY 5, 2022

RCTs and Observational Studies

RCT-DUPLICATE as an evaluation of observational studies:

- **Project title: *R*andomized *C*ontrolled *T*rials *D*uplicated *U*sing *P*rospective *L*ongitudinal *I*nsurance *C*laims: *A*pplying *T*echniques of *E*pidemiology**
- **Note: seeking to emulate a hypothetical randomized trial when designing a non-interventional study is a fundamentally different task**
- **Excerpt from project website: "If principled nonrandomized study approaches based on healthcare databases can consistently match the results of published trials and predict the results of ongoing trials, then we gain confidence in the validity of future real-world data analyses that may be performed in the absence of randomized trial evidence"**

RCT-DUPLICATE – Selected Process Steps

- Prospective engagement with FDA during protocol development and initial feasibility and power calculations
- FDA review of final definitions of cohort identification, exposure, outcome, and covariates
- While still blind to differential outcome, final power analyses and covariate balance checks are completed; joint go/no-go decision
- Study protocol registered on ClinicalTrials.gov
- Analyze outcome data and calculate effect measures
- Document findings
- Apply prespecified measures of agreement
- Audit trail visible to FDA throughout the process; FDA sub-team can engage in additional post-hoc sensitivity analyses for training purposes

CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov

Overview of Technical Findings: The RCT-DUPLICATE Demonstration Project

12:15-1:15pm

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

Results from the RCT DUPLICATE[1] initiative

Sebastian Schneeweiss, MD, ScD
Professor of Medicine and Epidemiology

Shirley V Wang, PhD
Associate Professor of Medicine

Jessica Franklin, PhD
Associate Professor of Medicine**

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

[1] Randomized Controlled Trials Duplicated Using Prospective Longitudinal Insurance Claims: Applying Techniques of Epidemiology
** Formerly with the Brigham and Women's Hospital, now at Optum Inc.

This study was funded by FDA under contracts HHSF223201710186C and HHSF223201710146C

Disclosures

Dr. Schneeweiss

- PI, Sentinel Innovation Center (FDA)
- Co-Chair, Mass General Brigham 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, UCB
- Consulting fees from Aetion, Inc. (incl. equity)

Dr. Wang

- Supported by grants from FDA Sentinel, NHLBI, NIA, NICHD

Dr. Franklin

- Current employment by Optum, Inc.

Table of Contents

Key publications on our rationale and methodology

When and How Can Real World Data Analyses Substitute for Randomized Controlled Trials?

Jessica M. Franklin[1] and Sebastian Schneeweiss[1]

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]

Emulation Differences vs. Biases When Calibrating Real-World Evidence Findings Against Randomized Controlled Trials

Jessica M. Franklin[1,*], Robert J. Glynn[1], Samy Suissa[2] and Sebastian Schneeweiss[1]

Circulation

ORIGINAL RESEARCH ARTICLE

Emulating Randomized Clinical Trials With Nonrandomized Real-World Evidence Studies

First Results From the RCT DUPLICATE Initiative

Jessica M. Franklin, PhD
Elisabetta Patorno, MD, DrPH
Rishi J. Desai, MS, PhD
Robert J. Glynn, PhD, ScD
David Martin, MD, MPH
Kenneth Quinto, MD, MPH
Ajinkya Pawar, PhD
Lily G. Bessette, BS
Hemin Lee, MD, MPH
Elizabeth M. Garry, PhD
Nileesa Gautam, BS
Sebastian Schneeweiss, MD, ScD

Executive Summary

Knowledge gap: When can RWE come to causal conclusions on treatment effects?

Objective: To emulate 30 RCTs using clinical practice data and compare findings

Approach: Using claims data, we applied a pre-defined, transparent process to emulate RCTs with RWE and compare treatment effect estimates

Findings:

- 50% of the selected RCTs could be emulated closely regarding design, analysis
- Closely emulated RCTs found comparable treatment effects to the RWE studies
- RCT and RWE findings were more likely to diverge when there were substantive emulation challenges; perhaps answering different target questions or due to bias

Implications: Rigorous epidemiologic methods combined with fit-for-purpose data and target trial approaches can produce valid inference from RWE studies

RCT-DUPLICATE: A methods demonstration project

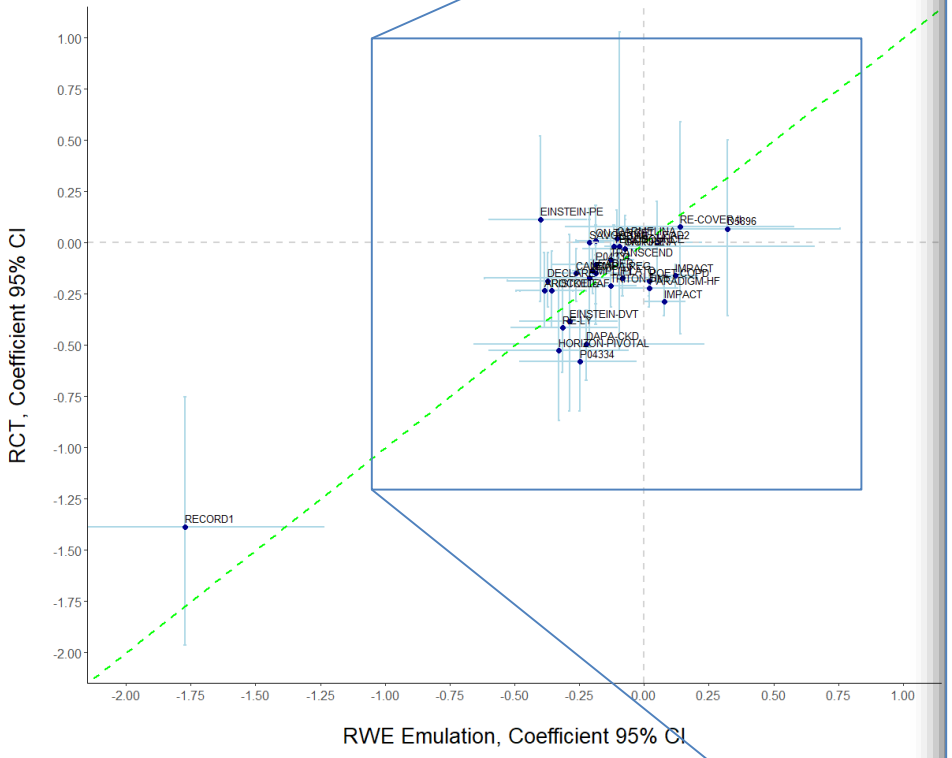
Objective: To understand and improve the validity of RWE studies to support regulatory decision making

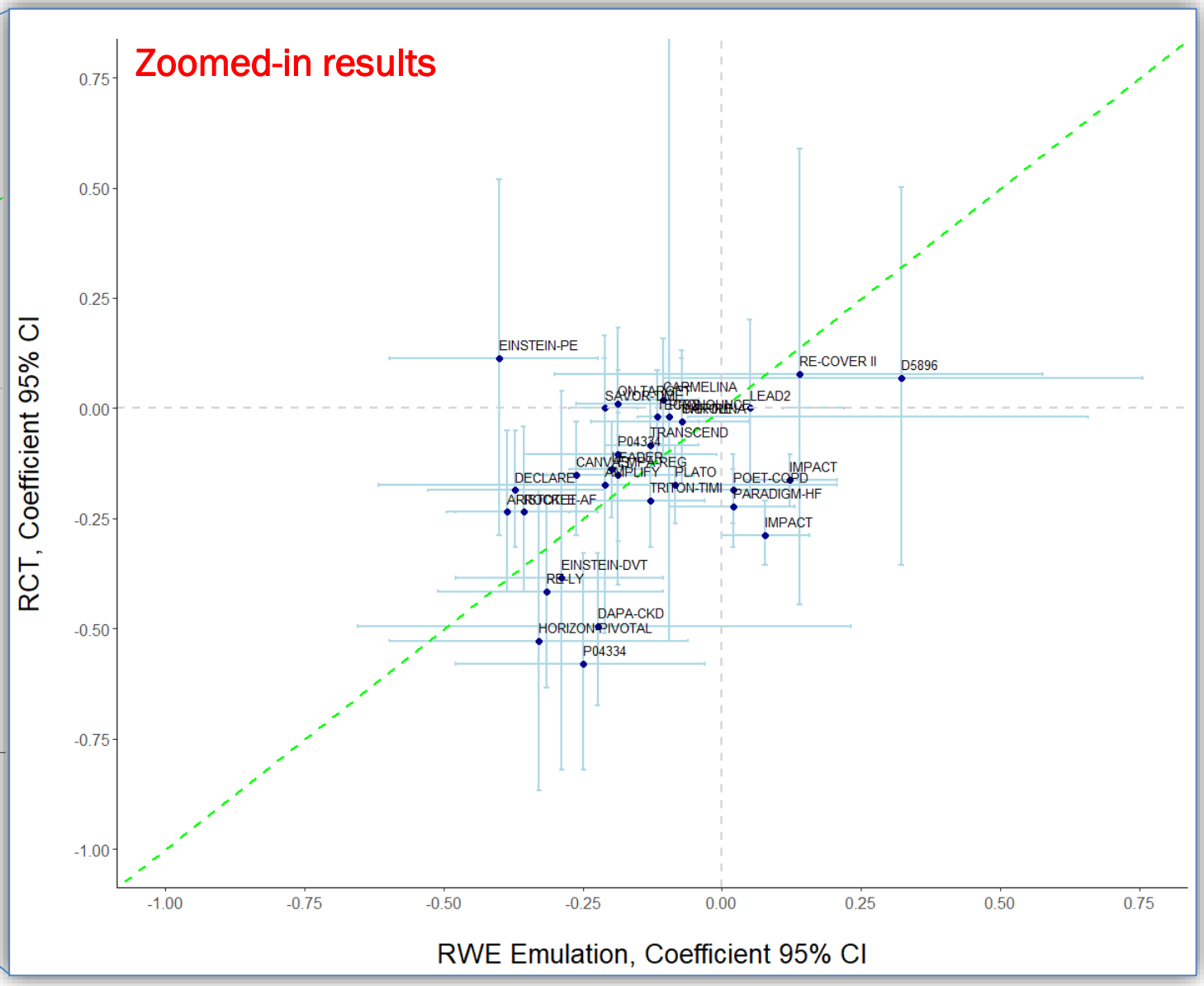
1	2	3
Emulate **30 RCTs** and predict **7 RCTs** considered by FDA	Test a **process** with FDA to evaluate RWE studies	Factors that predict replication success, causal estimates
Learnings: Had we replaced an RCT with a single RWE study would we have come to the same decision?	**Learnings:** How to conduct transparent, reproducible RWE studies and enable regulators to re-analyze data?	**Learnings:** Identify factors that predictably increase validity of RWE studies.

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

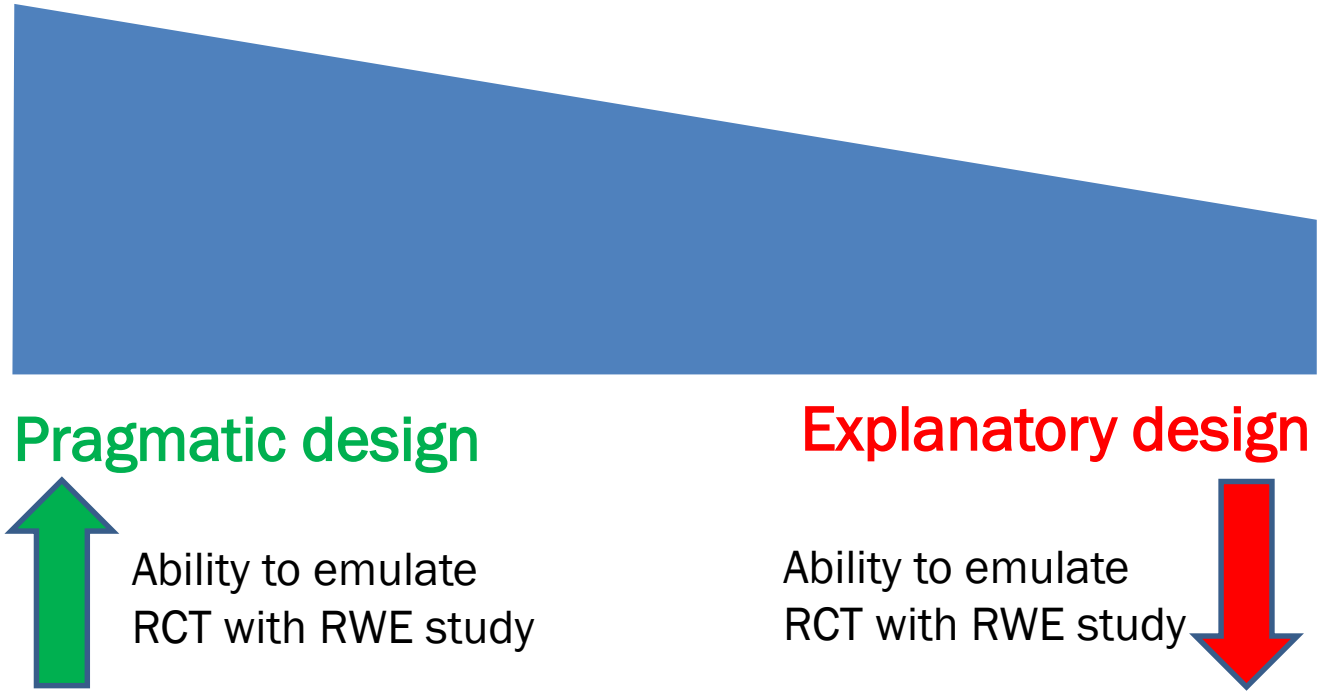
32 RCT-RWE emulation results

Pearson's overall = 0.80; 0.63-0.90





When and How Can Real-World Data Analyses Align With Randomized Controlled Trials?



Thorpe et al 2009, Journal of Clinical Epidemiology

Sox et al 2017, JAMA

Emulation differences vs Bias: *Are we asking a different question?*

Emulation Differences

Differences between RCT and RWE

Population

- Inclusion-exclusion
- Run-in periods with subject selection

Treatment strategy

- Loading dose, step-up therapy, allowable co-medication
- Placebo

Outcome ascertainment

- Measurement definition
- Primary vs. secondary data collection

Follow up

- Time-varying hazard
- Measures to maximize adherence

Bias

Differences between RWE treatment arms

Confounding

- Un- or mis-measured outcome predictors

Outcome ascertainment

- Differential surveillance
- Misclassification

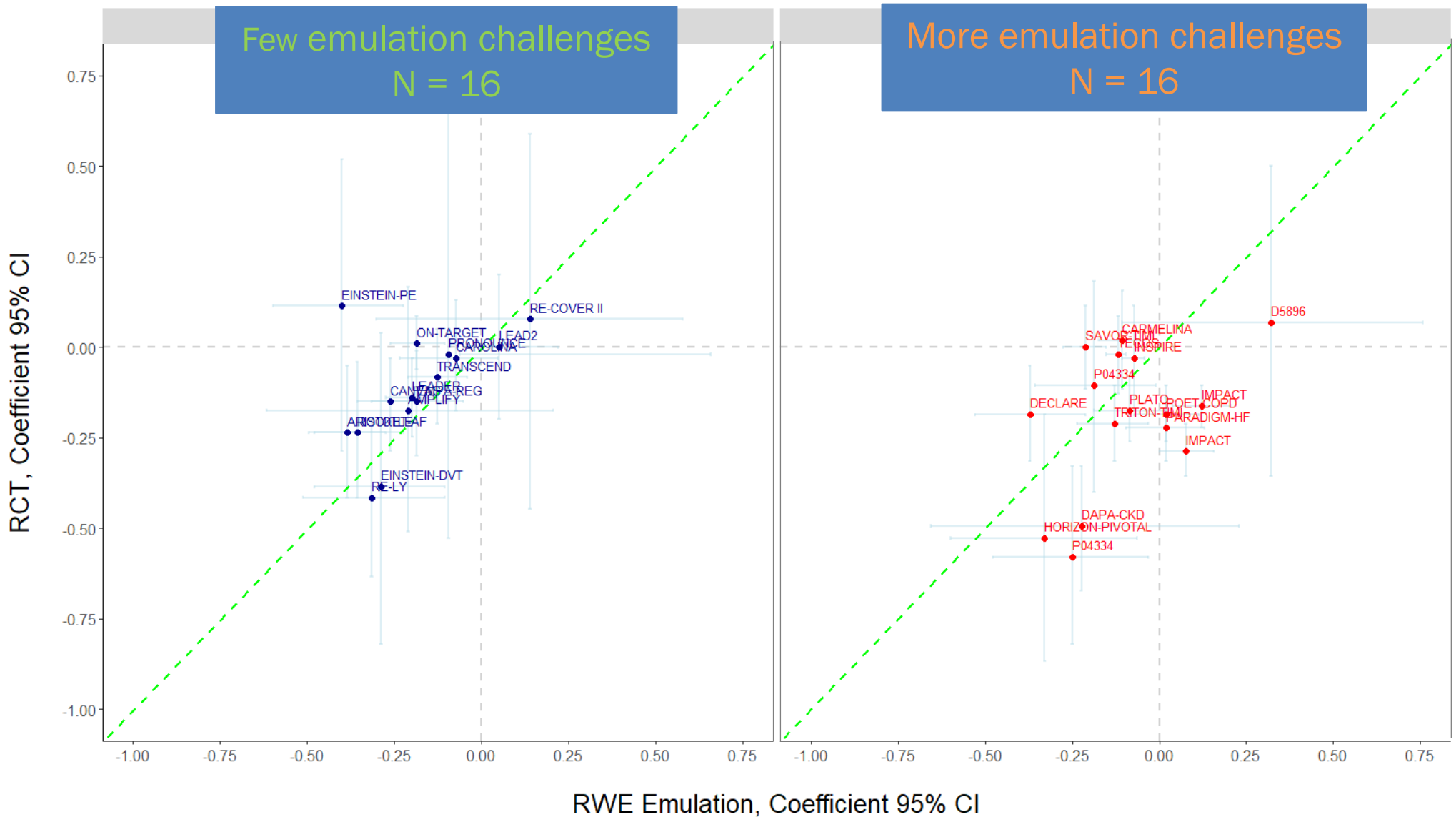
Follow up

- Differential duration
- Informative censoring

Emulation Differences vs. Biases When Calibrating Real-World Evidence Findings Against Randomized Controlled Trials

Jessica M. Franklin[1,*], Robert J. Glynn[1], Samy Suissa[2] and Sebastian Schneeweiss[1]

Few emulation challenges vs more emulation challenges



Total # trials = 32, includes 30 emulations of completed trials and 2 predictions of ongoing trials.

Trials 1 – 11*

	Trial name	RCT	RWE[1] Adjusted	RWE[1] Unadjusted	Std. Diff.	Agreement	Close emulation?
Diabetes	LEADER	0.87 (0.78, 0.97)	0.82 (0.76, 0.87)	0.57 (0.54, 0.61)	0.90	RA/EA/SD	Y
Diabetes	DECLARE	0.83 (0.73, 0.95)	0.69 (0.59, 0.81)	0.47 (0.41, 0.53)	1.76	RA/–/SD	N
Diabetes	EMPA-REG	0.86 (0.74, 0.99)	0.83 (0.73, 0.95)	0.63 (0.57, 0.70)	0.35	RA/EA/SD	Y
Diabetes	CANVAS	0.86 (0.75, 0.97)	0.77 (0.70, 0.85)	0.58 (0.54, 0.62)	1.34	RA/EA/SD	Y
Diabetes	CARMELINA	1.02 (0.89, 1.17)	0.90 (0.84, 0.96)	0.90 (0.86, 0.95)	1.61	–/EA/SD	N
Diabetes	TECOS	0.98 (0.88, 1.09)	0.89 (0.86, 0.91)	0.81 (0.79, 0.84)	1.71	–/EA/SD	N
Diabetes	SAVOR-TIMI	1.00 (0.89, 1.12)	0.81 (0.76, 0.86)	0.65 (0.62, 0.69)	3.16	–/–/–	N
Diabetes	LEAD2	0.00 (-0.20, 0.20)	0.05 (0.11, 0.22)	0.01 (0.11, 0.13)	-0.37	RA/EA/SD	Y
Antiplatelet	TRITON-TIMI	0.81 (0.73, 0.90)	0.88 (0.79, 0.97)	0.70 (0.65, 0.76)	-1.11	RA/EA/SD	N
Antiplatelet	PLATO	0.84 (0.77, 0.92)	0.92 (0.83, 1.02)	0.84 (0.78, 0.91)	-1.31	–/EA/SD	N
Antiplatelet	ISAR-REACT5	1.36 (1.09, 1.70)	n/a[2]	n/a[2]	n/a[2]	n/a[2]	N

1) Pooled estimate across databases
2) Chi-square test indicated that results were heterogeneous by database, therefore results were not pooled

* Close emulation refers to trials where there were few emulation challenges
RA = regulatory agreement (point est and CI on same side of null); *EA* = estimate agreement (RWE estimate in CI of RCT); *SD* = std diff agreement (≤≤2)

Trials 12 – 22*

	Trial name	RCT	RWE[1] Adjusted	RWE[1] Unadjusted	Std. Diff.	Agreement	Close emulation?
Atrial Fibrillation	ARISTOTLE	0.79 (0.66, 0.95)	0.68 (0.61, 0.76)	0.66 (0.62, 0.71)	1.36	RA/EA/SD	Y
	RE-LY	0.66 (0.53, 0.82)	0.73 (0.60, 0.90)	0.67 (0.58, 0.78)	-0.66	RA/EA/SD	Y
	ROCKET-AF	0.79 (0.66, 0.96)	0.70 (0.62, 0.80)	0.76 (0.69, 0.84)	1.00	RA/EA/SD	Y
VTE	EINSTEIN-DVT	0.68 (0.44, 1.04)	0.75 (0.62, 0.90)	0.85 (0.76, 0.95)	-0.42	–/EA/SD	Y
	EINSTEIN-PE	1.12 (0.75, 1.68)	0.67 (0.55, 0.80)	0.73 (0.64, 0.83)	2.28	–/–/–	Y
	RE-COVER II	1.08 (0.64, 1.80)	1.15 (0.74, 1.78)	1.48 (1.09, 2.00)	-0.18	RA/EA/SD	Y
	AMPLIFY	0.84 (0.60, 1.18)	0.81 (0.54, 1.23)	0.64 (0.50, 0.82)	0.13	RA/EA/SD	Y
	RECORD1	0.25 (0.14, 0.47)	0.17 (0.10, 0.29)	0.25 (0.18, 0.34)	0.63	RA/EA/SD	Y
Hypertension	TRANSCEND	0.92 (0.81, 1.05)	0.88 (0.81, 0.96)	0.80 (0.74, 0.85)	0.55	–/EA/SD	Y
	ON-TARGET	1.01 (0.94, 1.09)	0.83 (0.77, 0.90)	0.68 (0.64, 0.72)	3.46	–/–/–	Y

1) Pooled estimate across databases
2) Chi-square test indicated that results were heterogeneous by database, therefore results were not pooled

* Close emulation refers to trials where there were few emulation challenges
RA = regulatory agreement (point est and CI on same side of null); *EA* = estimate agreement (RWE estimate in CI of RCT); *SD* = std diff agreement (≤≤2)

Trials 23 – 30*

	Trial name	RCT	RWE[1] Adjusted	RWE[1] Unadjusted	Std. Diff.	Agreement	Close emulation?
Osteoporosis	HORIZON-PIVOTAL	0.59 (0.42, 0.83)	0.72 (0.55, 0.94)	1.08 (0.86, 1.35)	-0.90	RA/EA/SD	N
Osteoporosis	VERO	0.44 (0.29, 0.68)	n/a[2]	n/a[2]	n/a[2]	n/a[2]	N
Chronic Kidney	DAPA-CKD	0.61 (0.51, 0.72)	0.80 (0.52, 1.26)	0.41 (0.29, 0.58)	-1.10	–/–/SD	N
Heart Failure	PARADIGM-HF	0.80 (0.73, 0.87)	1.02 (0.91, 1.14)	0.95 (0.90, 1.02)	-3.42	–/–/–	N
Asthma	P04334	0.56 (0.44, 0.72)	0.78 (0.62, 0.97)	0.87 (0.76, 0.99)	-1.95	RA/–/SD	N
Asthma	D5896	1.07 (0.70, 1.65)	1.38 (0.90, 2.13)	1.41 (1.00, 1.98)	-0.81	RA/EA/SD	N
COPD	IMPACT	0.85 (0.80, 0.90)	1.13 (1.04, 1.23)	1.22 (1.15, 1.30)	-5.46	–/–/–	N
COPD	POET-COPD	0.83 (0.77, 0.90)	1.02 (0.93, 1.12)	1.05 (0.99, 1.12)	-3.27	–/–/–	N
COPD	INSPIRE	0.97 (0.84, 1.12)	0.93 (0.90, 0.96)	0.83 (0.81, 0.85)	0.56	RA/EA/SD	N

1) Pooled estimate across databases
2) Chi-square test indicated that results were heterogeneous by database, therefore results were not pooled

* Close emulation refers to trials where there were few emulation challenges
RA = regulatory agreement (point est and CI on same side of null); *EA* = estimate agreement (RWE estimate in CI of RCT); *SD* = std diff agreement (≤≤2)

Prediction of ongoing Phase IV RCTs (2 of 7)

	Trial name	RCT	RWE[1] Adjusted	RWE[1] Unadjusted	Std. Diff.	Agreement	Close emulation?
Diabetes	CAROLINA[2]	**0.98** (0.84, 1.14)	**0.91** (0.79, 1.05)	**0.92** (0.83-1.01)	0.70	RA/EA/SD	Y
Prostate cancer	PRONOUNCE[3]	**1.28** (0.59, 2.79)	**1.35** (0.94, 1.93)	**1.70** (1.30, 2.21)	-0.12	RA/EA/SD	Y

1) Pooled estimate across databases
2) *Patorno E, Schneeweiss S, Gopalakrishnan C, Martin D, Franklin JM. Using real-world data to predict findings of an ongoing phase IV cardiovascular outcome trial: cardiovascular safety of linagliptin versus glimepiride. Diabetes Care. 2019;42:2204-10*
3) *Merola D, Schneeweiss S, Sreedhara S, Zabotka LE, Quinto K, Concato J, Wang SV. Using real-world data to predict results of an ongoing phase IV oncology trial: comparative safety of degarelix vs. leuprolide in advanced prostate cancer. Manuscript in preparation.*

* Close emulation refers to trials where there were few emulation challenges

RA = regulatory agreement (point est and CI on same side of null); *EA* = estimate agreement (RWE estimate in CI of RCT); *SD* = std diff agreement (≤≤2)

Outline

Emulation Challenges

Example

Lessons Learned

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Emulation Challenges

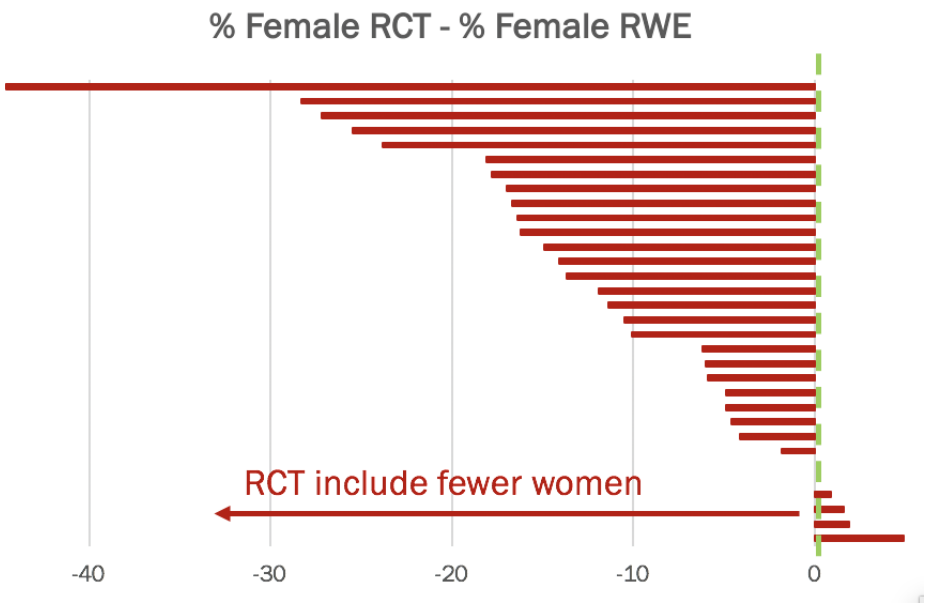
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

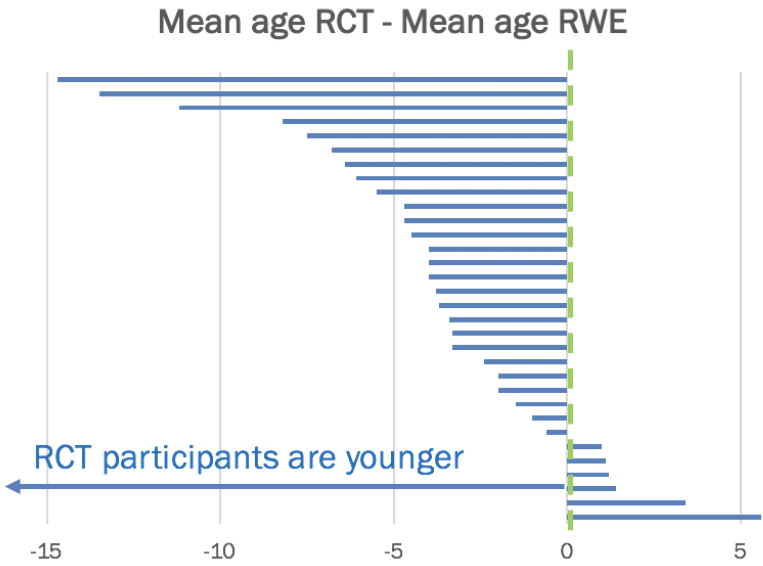
- Well emulated
- Sufficiently emulated
- Difficult to emulate

PARADIGM-HF (Phase 3)
Inclusion
Age>=18, HFrEF, HF hospitalization within 12 months Stable on ACEis/ARBs and beta-blocker therapies
Exclusion
• Allergy, intolerance, and contraindication to any of the study drugs • History of angioedema • Treatment with both ACEis AND ARBs • Acute decompensated HF • Symptomatic hypotension • Low eGFR/renal dysfunction • Hyperkalemia • ACS, Stroke, TIA, CABG, PCI, Other CV Procedures, Carotid Angioplasty within 3 months • Coronary/carotid artery disease or PCI within 6 mo. after visit 1 • CRT device within 3 months prior to visit 1 or intent to implant • History of heart transplant, on transplant list, or with LVAD • History of severe pulmonary disease • Peripartum- or chemotherapy- induced cardiomyopathy • Untreated ventricular arrhythmia with syncopal episodes • Symptomatic bradycardia or 2nd & 3rd degree AV block • Hemodynamically significant mitral and/or aortic valve disease • Active IBD, Duodenal/gastric ulcers • Hepatic disease • Cholestyramine or colestipol resins • Presence of any disease with a life expectancy of <5 years • *Ivabradine use*

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness





Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Good	Trial had active comparator
Moderate	Placebo emulated by drug expected to be unrelated to the outcome AND cohort characteristics well balanced, OR active comparator had to be modified for feasibility reasons
Poor	Placebo emulated by drug expected to be unrelated to the outcome AND expectation of residual confounding from characteristics poorly measured in claims (e.g. SES)

Diabetes

Placebo control emulation

(2nd line diabetes drug – **DPP4i vs sulfonylureas**, 3P MACE)

Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
CARMELINA	Linagliptin vs placebo	3P MACE	**1.02** (0.89, 1.17)	**0.90** (0.84, 0.96)	1.61	NI	*	EA	SD
TECOS	Sitagliptin vs placebo	3P MACE + angina	**0.98** (0.88, 1.09)	**0.89** (0.86, 0.91)	1.71	NI	*	EA	SD
SAVOR-TIMI	Saxagliptin vs placebo	3P MACE	**1.00** (0.89, 1.12)	**0.81** (0.76, 0.86)	3.16¶	NI	*	–	–
CAROLINA	Linagliptin vs glimepiride	3P MACE	**0.98** (0.84, 1.14)	**0.91** (0.79, 1.05)	0.70	NI	RA	EA	SD

Good

Moderate

Poor

Comparator emulation

- Expensive new drug vs older cheap drug
- Difficult to capture SES differences?

Diabetes

Placebo control emulation

(2nd line diabetes drug – **GLP1, SGLT2i vs DPP4i**, 3P MACE)

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

Good

Moderate

Poor

Comparator emulation

- Comparing expensive newer drugs
- Closer therapeutic alternatives

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Assessed with high specificity
Lower specificity or high missingness

Outcome emulation

	RCT						RWE					
	Exposed			Comparator			Exposed			Comparator		
	Events	N	Rate*	Events	N	Rate*	Events	N	Rate*	Events	N	Rate*
ROCKET-AF	188	6,958	1.7	241	7,004	2.2	419	51,318	1.5	518	51,318	2.4
PARADIGM-HF	914	4,187	21.8	1,117	4,212	26.5	645	3,033	46.4	636	3,033	44.6
LEAD2	n/a	482	1.0	n/a	242	1.0	n/a	373	1.0	n/a	373	0.9

Exposed rates

Comparator rates

Assessed with high specificity

Lower specificity or high missingness

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

In-hospital start of medication

Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--

MarketScan	1.20 (0.95, 1.51)
Optum	0.73 (0.52, 1.01)
P for homogeneity	0.01

Good

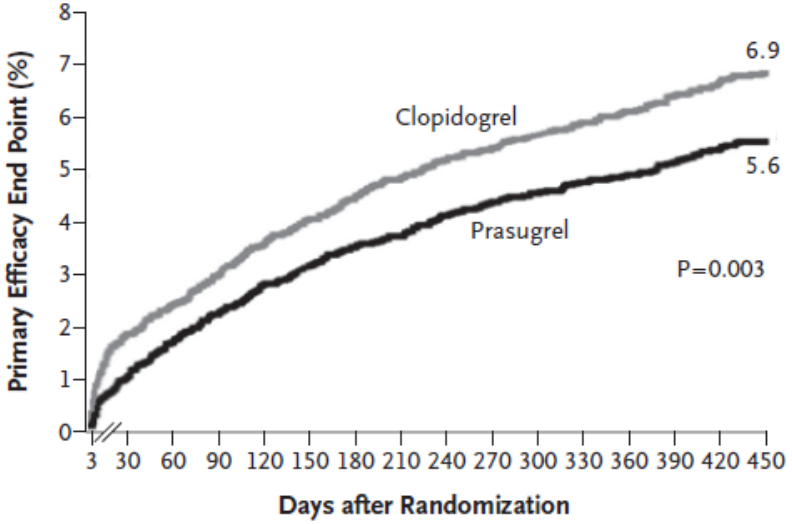
Moderate

Poor

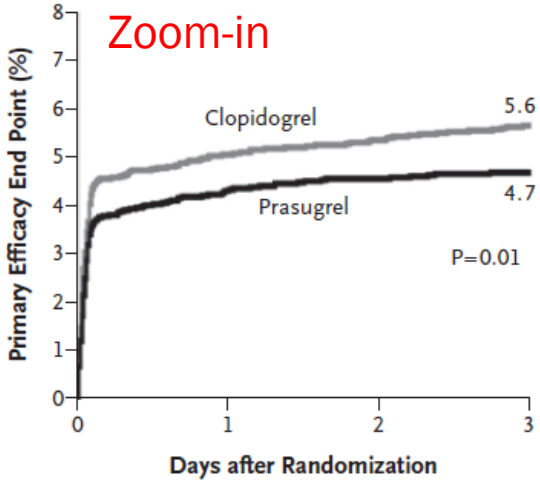
In-hospital start of medication

Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--



Wiviott et al, NEJM 2007



Take-home points:

- RCT shows early and immediate effect – starting FU while in hospital
- RWE study question targets patients who survive until discharge and fill 1st Rx
- Cannot capture early effect without linked hospital + outpatient Rx data

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Dose-titration during follow up

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--

"We compared ticagrelor (180-mg loading dose, 90 mg twice daily thereafter) and clopidogrel (300-to-600-mg loading dose, 75 mg daily thereafter)"
Wallentin, NEJM 2009

"The first 3 weeks patients will receive rivaroxaban 15 mg twice-daily followed by rivaroxaban 20 mg once-daily." (EINSTEIN protocol)

Good

Moderate

Poor

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	**EA**	**SD**
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-

Emulation Challenges

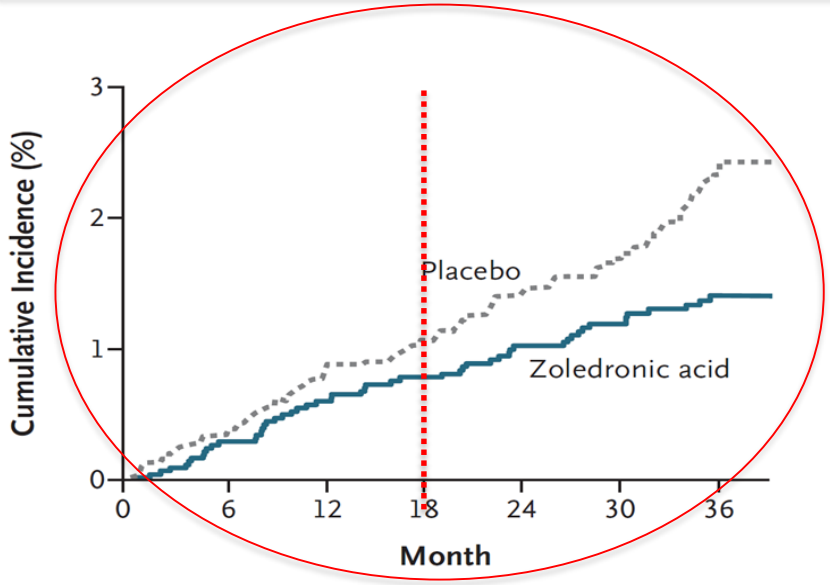
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Delayed treatment effects

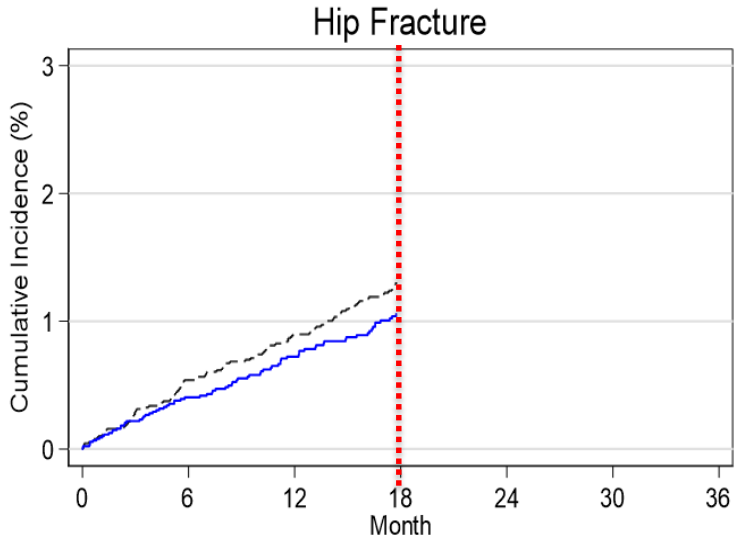
Osteoporosis

HORIZON-PIVOTAL (osteoporosis, hip fracture)

RCT	RWE
HR_{36mo} = 0.59 (0.42, 0.83)	HR_{36mo} = ??
HR_{18mo} = 0.75	HR_{18mo} = 0.75 (0.58, 0.97)



No. at Risk							
Zoledronic acid	3875	3807	3674	3553	3494	3387	3161
Placebo	3861	3806	3694	3577	3499	3397	3144

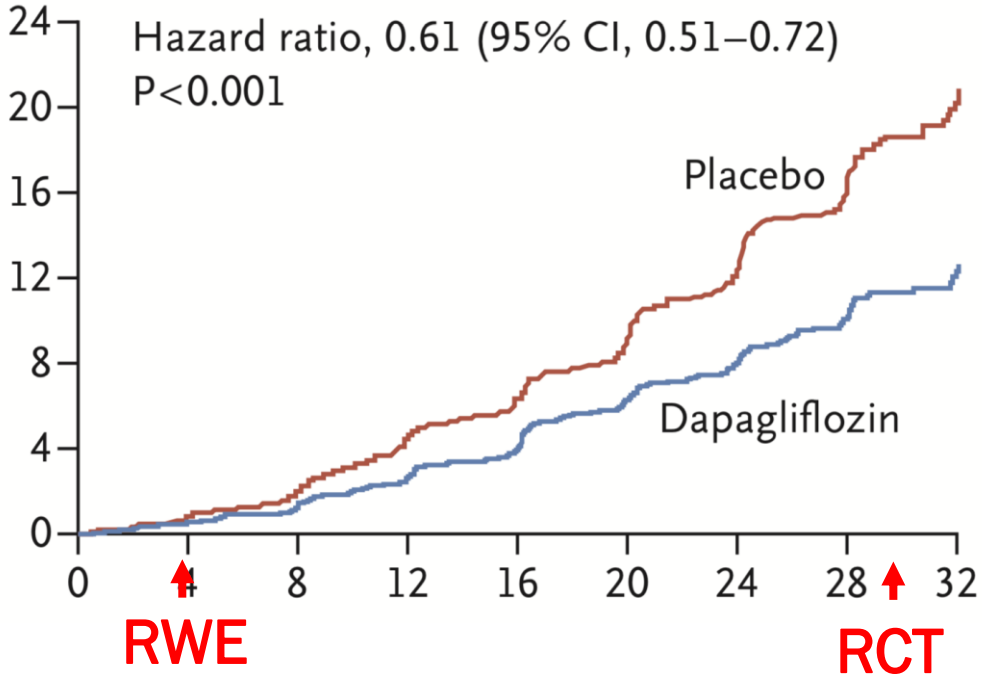


Number at risk							
Raloxifene	9003	7753	6768	0	0	0	0
Zoledronic acid	9003	7766	6743	0	0	0	0

Delayed effect with long follow up

CKD

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
DAPA-CKD	Dapaglifloxin vs placebo	Sustained decline in eGFR, ESRD, death	**0.61** (0.51, 0.72)	**0.80** (0.52, 1.26)	-1.10	Sup	-	-	**SD**



Median Follow Up

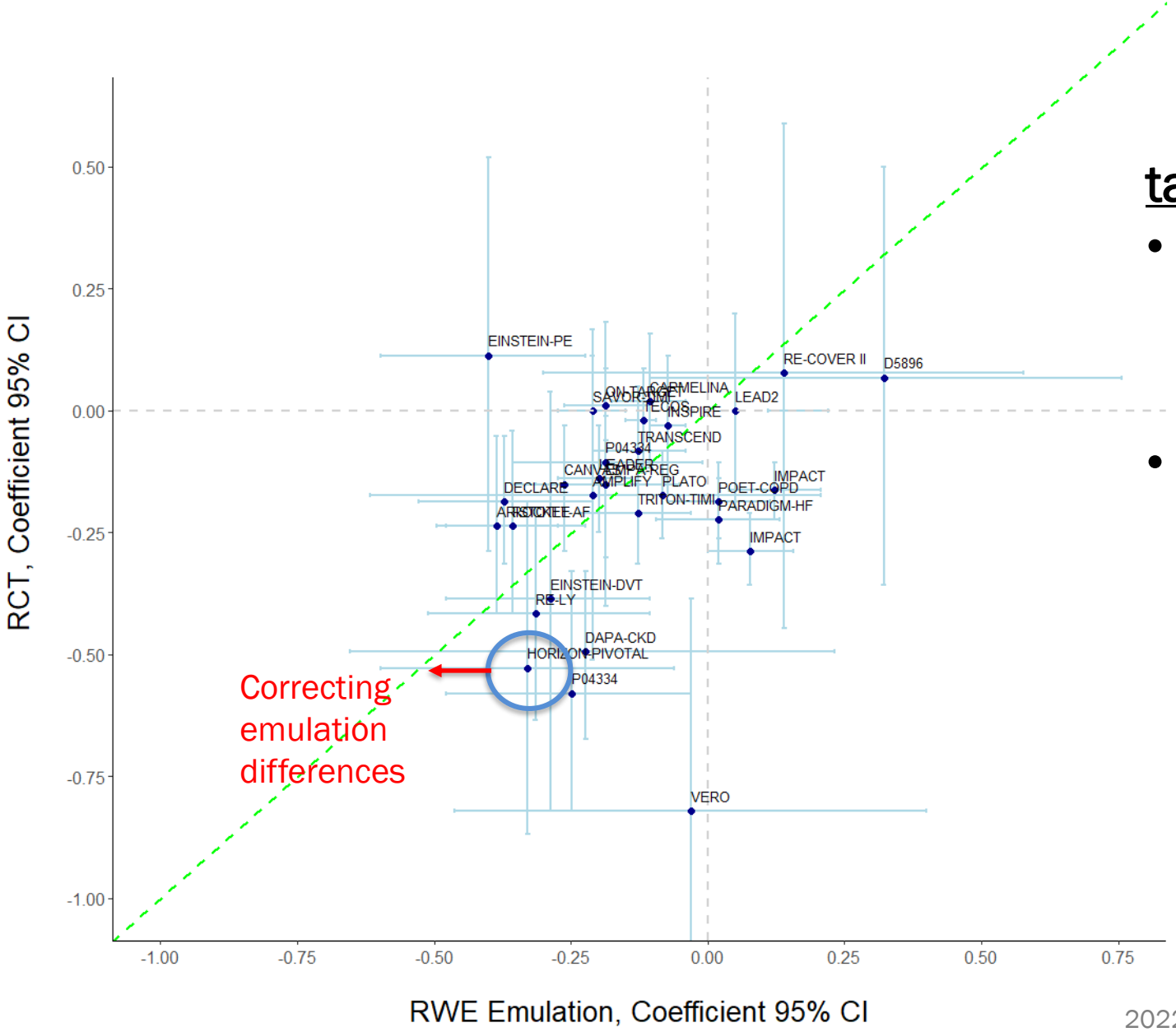
Median Follow Up

Good

Moderate

Poor

RCT by RWE findings



take-home points:

- Challenging to replicate trial findings for outcomes with long induction or time-varying hazard
- Patients in clinical practice may not experience benefit observed in explanatory trial

Emulation Challenges

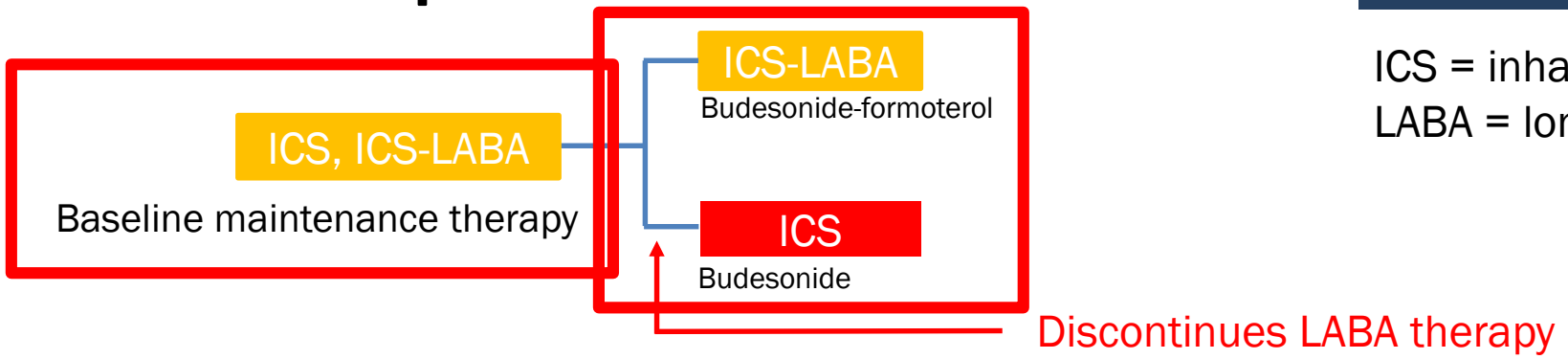
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

On placebo
Both treatment groups
On 1 treatment arm

Discontinuation of maintenance therapy → short term ↑ exacerbation

D5896
Treatment: ICS-LABA vs ICS
Outcome: Serious asthma related events

ICS = inhaled corticosteroid
LABA = long-acting beta agonist



Assumptions Scenario 1 (our RWE emulation):

- Truth is upper bound of non-inferiority limit (1.32)
- 50% of patients were on LABA at baseline
- No effect of discontinuation

Baseline	Randomized: ICS-LABA	Randomized: ICS
No LABA use	29	22
LABA use	29	22
	58	44

RR = 58/44 = **1.32**

Assumptions Scenario 2:

- Truth is upper bound of non-inferiority limit (1.32
- 50% of patients were on LABA at baseline
- *Discontinuation increases risk of outcome by 50%*

Baseline	Randomized: ICS-LABA	Randomized: ICS
No LABA use	29	22
LABA use	29	22+11
	58	55

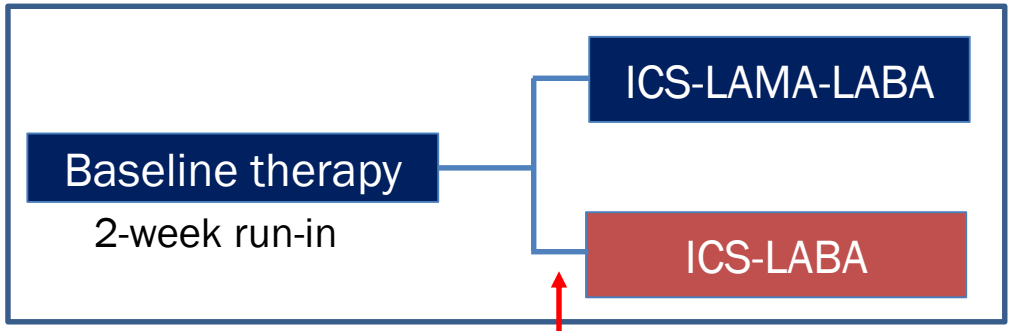
RR = 58/55 = **1.05**

D5896	1.07 (0.70, 1.65)
Pooled RWD	1.38 (0.90, 2.13)

Suissa S et al. *Chest* 2013 143(5). May 2013" 1208-1213

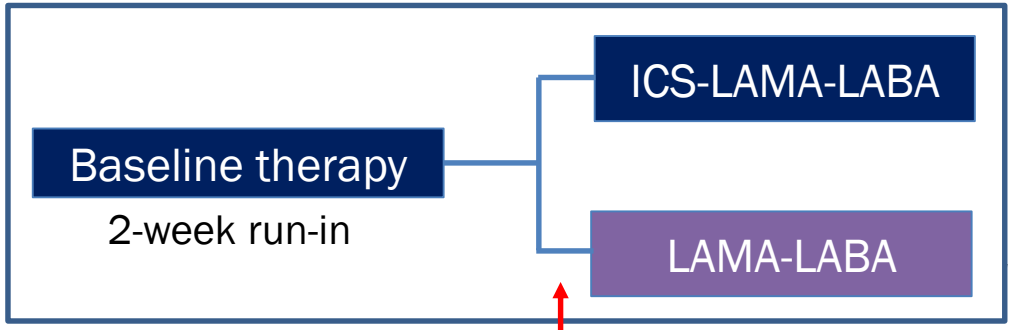
RCT: D/c of maintenance therapy → short term increase in exacerbation
RWE: Confounding by stage of therapy/disease severity

IMPACT trial and COPD exacerbation event



Discontinues LAMA therapy

RCT	0.85 (0.80, 0.90)
RWE	1.13 (1.04, 1.23)



Discontinues ICS therapy

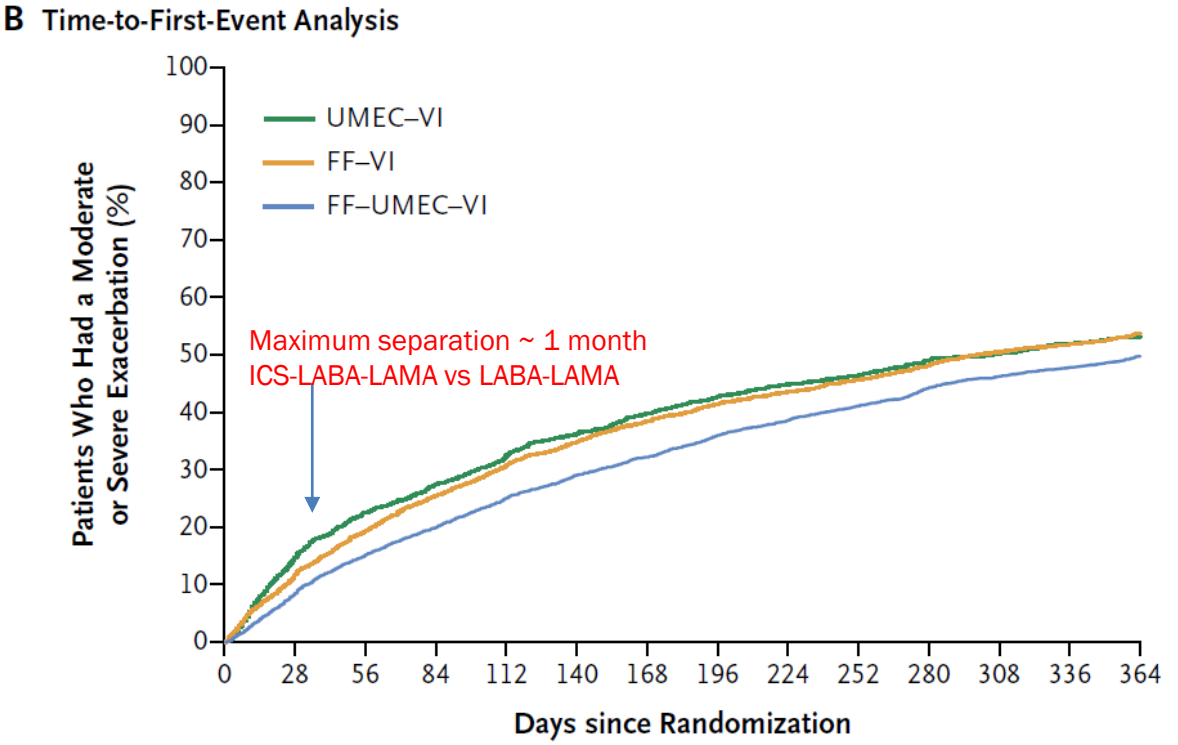
RCT	0.75 (0.70, 0.81)
RWE	1.08 (1.00, 1.17)

38% on triple therapy during baseline!

Randomization

RCT: D/c of maintenance therapy → short term increase in exacerbation

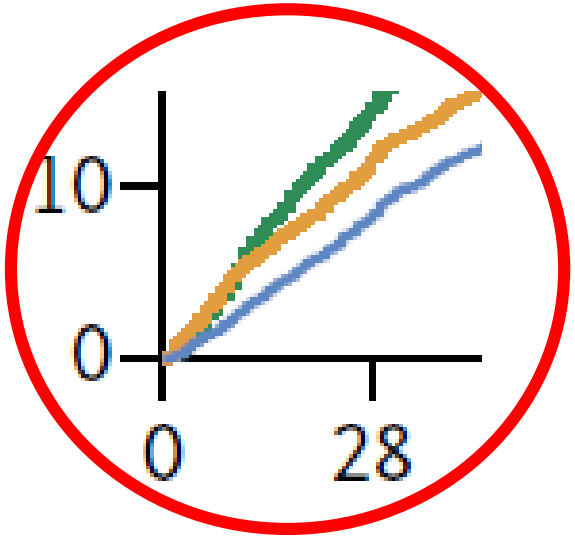
IMPACT trial and COPD exacerbation event



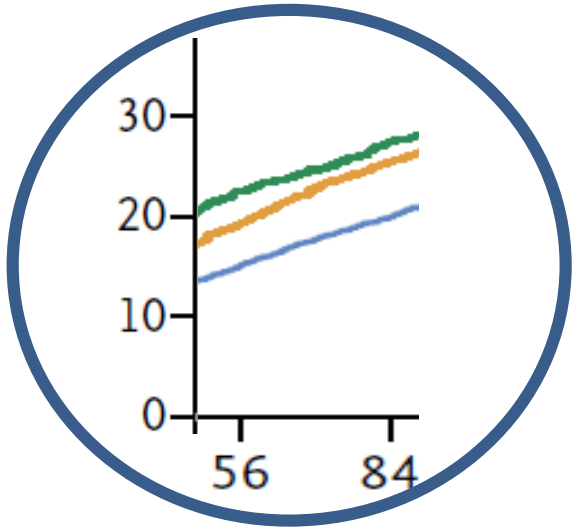
No. at Risk

UMEC–VI	2070	1721	1516	1406	1301	1201	1123	1059	1001	971	917	884	851	642
FF–VI	4134	3554	3133	2838	2620	2410	2250	2120	2004	1823	1823	1729	1671	1228
FF–UMEC–VI	4151	3758	3408	3186	2954	2752	2614	2457	2324	2216	2085	1988	1919	1419

First 30-day interval
= treatment effect



Second 30-day interval
= no treatment effect



Suissa S, Ariel A. Triple therapy trials in COPD: a precision medicine opportunity.
Eur Respir J 2018; 52: 1801848 [https://doi.org/10.1183/13993003.01848-2018].

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Robustness of findings across multiple data sources

For 2 out of 32 trials, we observed results that diverged by database and could not be pooled.

ISAR-REACT5

Ticagrelor vs prasugrel on 3PMACE

	Result
RCT	**1.36** (1.09, 1.70)
MarketScan	**1.20** (0.95, 1.51)
Optum	**0.73** (0.52, 1.01)
Pooled	n/a

p for homogeneity <0.03

VERO

Teriparatide vs risedronate on vertebral fracture

	Result
RCT	**0.44** (0.29, 0.68)
MarketScan	**1.33** (0.80, 2.20)
Optum	**0.43** (0.19, 0.96)
Pooled	n/a

Take-home point:

- Important to replicate in multiple databases

"substantial evidence of effectiveness…2 adequate and well controlled investigations"
FDA Guidance for Industry

Robustness to alternative design/analysis specifications

For 2 out of 32 trials, colleagues independently asked similar questions using the same data sources

Take-home point:

- Important to investigate robustness of evidence to reasonable alternative choices

Desai et al.

PARADIGM	**0.80** (0.73, 0.87)
RCT-DUPLICATE	**0.97** (0.87-1.08)
Initiators of ACE/ARB vs sacubitril/valsartan	0.92 (0.84, 1.00)
Switchers from ACE to ARB vs sacubitril/valsartan	0.79 (0.74, 0.85)
Combined	**0.84** (0.80, 0.89)

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

JAMA Internal Medicine | Original Investigation

Use of Health Care Databases to Support Supplemental Indications of Approved Medications

Michael Fralick, MD; Aaron S. Kesselheim, MD, JD, MPH; Jerry Avorn, MD; Sebastian Schneeweiss, MD, ScD

Original research

Effectiveness of angiotensin-neprilysin inhibitor treatment versus renin-angiotensin system blockade in older adults with heart failure in clinical care

Rishi J Desai ,[1] Elisabetta Patorno,[1] Muthiah Vaduganathan,[2] Mufaddal Mahesri,[1] Kristyn Chin,[1] Raisa Levin,[1] Scott D Solomon,[2] Sebastian Schneeweiss[1]

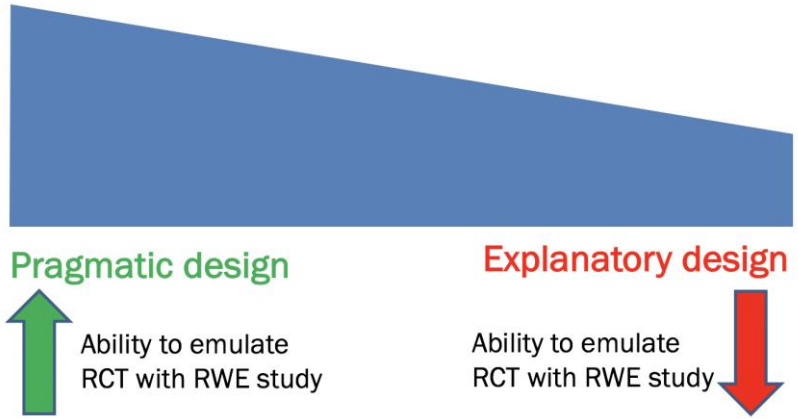
Investigating subtle differences in exposure, outcome, inclusion-exclusion criteria, covariates, follow-up

Outline

Emulation Challenges

Example

Lessons Learned



Challenges with emulation of trial design expected to shift the target question for RWE study vs RCT

a) Start of follow up in hospital (hospital Rx data not available in claims, but may be available in linked data)
b) Run-in that selects responders to one treatment arm
c) Mixing effect of randomization and discontinuation of baseline maintenance therapy
d) Delayed effect over long follow up
e) Differences in population distribution coupled with effect modification
f) Inadequate emulation of the exposure or outcome

Few emulation challenges = None of { a, b, c, d } AND comparator and outcome emulation are at least moderate, with >1 classified as good

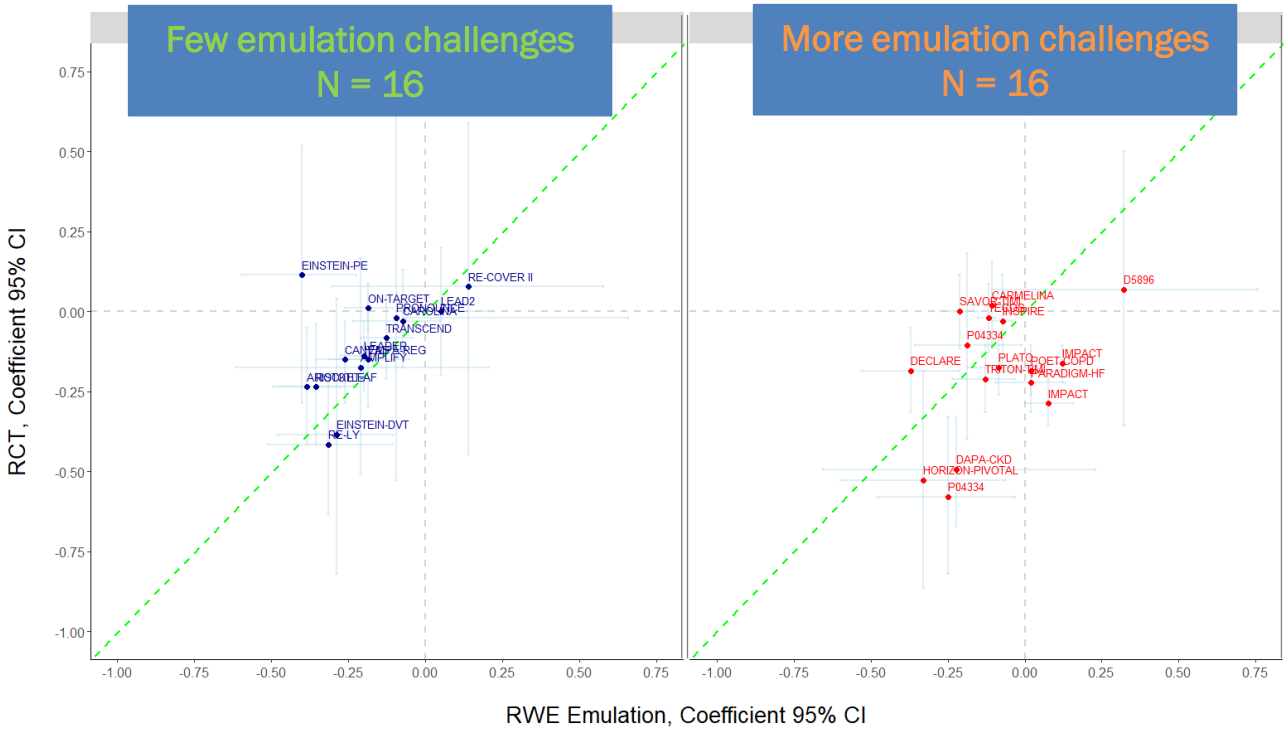
More emulation challenges = a OR b OR c OR d OR poor comparator emulation OR neither comparator and outcome emulation are classified as good

Few emulation challenges vs more emulation challenges

Pearson's overall = 0.80; 0.63-0.90

	Few emulation challenges N = 16	More emulation challenges N = 16
Pearson's	0.93 (0.79, 0.97)	0.46 (-0.05, 0.78)
ICC, 95% CI	0.89 (0.68, 0.96)	0.41 (-0.03, 0.73)
RA	12 (75%)	6 (38%)
EA	14 (88%)	7 (44%)
SD	14 (88%)	10 (63%)

ICC = intraclass correlation coefficient; CI = confidence interval; RA = regulatory agreement; EA = estimate agreement; SD = standardized difference agreement



Take-home points:

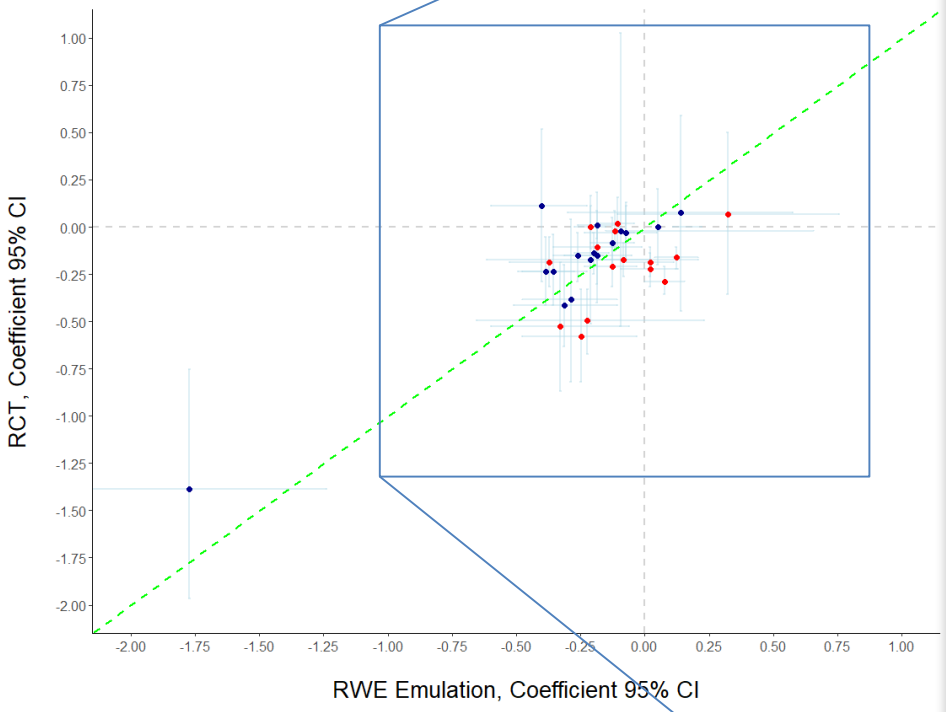
Recall: For this methods project, the goal was to emulate published RCTs as closely as possible:

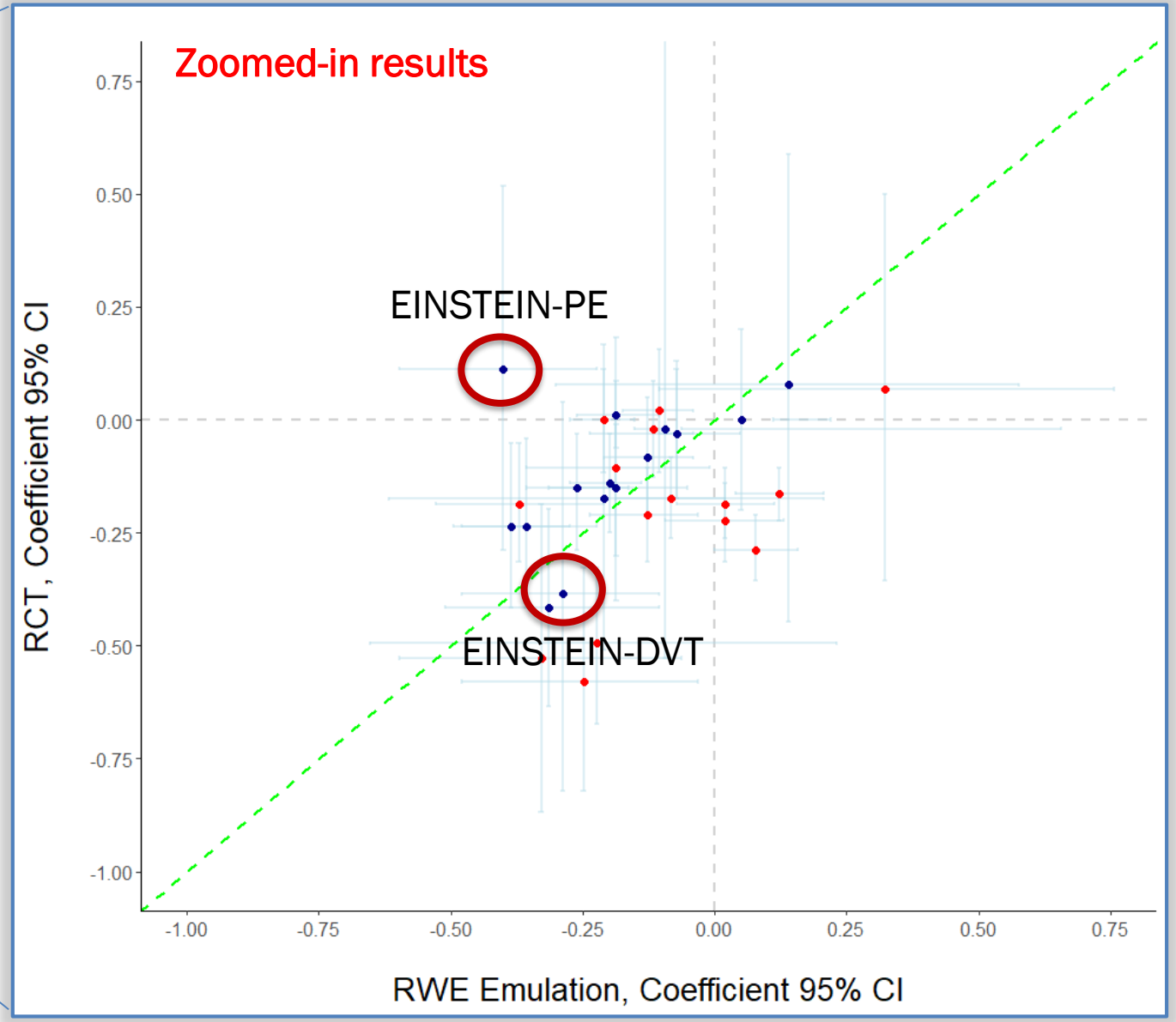
- Few emulation challenges → closer agreement in effect estimates
- More emulation challenges → less agreement in RCT/RWE effect estimates: diverge on target question/pop^{n}? Different answers may be correct.

Total # trials = 32, includes 30 emulations of completed trials and 2 predictions of ongoing trials.

32 RCT-RWE emulation results

Pearson's overall = 0.80; 0.63-0.90





- Few emulation challenges
- More emulation challenges

Chance or other factors?

VTE

Trial name	Comparator	Endpoint	RCT	RWE	Stand. Diff.	Test	Agreement			Indication
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	-	-	-	PE

It remains unclear why EINSTEIN-DVT and EINSTEIN-PE trials have separating results. Meta-analysis of 6 trials* finds no heterogeneity of effects in patients presenting with DVT or PE.

**Dentali F, et al. Intern Emerg Med. 2015*

Good

Moderate

Poor

Take-home points

1. RWE studies come to the same conclusions as RCTs when we are able to emulate well, i.e. target the same question

2. There is more nuance to evaluation of replicability of trial results with RWE than can be found in binary agreement metrics.
   - Residual bias, random error
   - Efficacy vs effectiveness
   - Single trial as reference standard

3. In evaluating when and how RWE studies complement RCTs, we should think about the target trial design that would match the need/question of end users (ideal vs pragmatic)

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

Discussion

Future: calibrating our RWE tool kit

- Repository of well-documented studies that illustrate the agreement between RCTs and RWE when the RWE study is designed to emulate the RCT as closely as possible.

- 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 categorizations like:
        - Population
        - Length of 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

Summary of take-home messages (in more detail)

- Divergence in RCT and RWE findings can come from many sources, most likely emulation differences and confounding
- Published trial is an imperfect benchmark to calibrate against if trial didn't ask the same question targeted by RWE study
- Single trial was used as reference standard for causal effects and sometimes there is substantial variation between trials
- Challenging to emulate:
  - Explanatory design features (e.g. run-in that selects responders)
  - Findings for outcomes with long induction or time varying hazard
    - Patients in clinical practice may not experience benefit observed in explanatory trials
  - If circumstances of use of comparator is different from the exposure of interest in clinical practice (e.g. imbalanced cost, prior-authorization)
- Important to:
  - Explore and understand robustness of evidence to reasonable alternative choices
  - Evaluate replicability of RWE in multiple data sources

Harvard study team:

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

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

FDA colleagues:

Drs. Martin, Quinto, Concato, Corrigan-Curay, Paraoan, Bradley, and Li

Aetion team:

Drs. Garry, Rassen, and Isaman, Gibbs, Gilpin

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

Back up slides

Emulation Challenges

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Emulation Challenges

Diabetes

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

DUPLICATE Diabetes Trial Emulations

Diabetes

Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
LEADER	Liraglutide vs placebo	3P MACE	**0.87** (0.78, 0.97)	**0.82** (0.76, 0.87)	0.90	NI	**RA**	**EA**	**SD**
DECLARE	Dapagliflozin vs placebo	HHF + CV death	**0.83** (0.73, 0.95)	**0.69** (0.59, 0.81)	1.76	NI	**RA**	**–**	**SD**
EMPA-REG	Empagliflozin vs placebo	3P MACE	**0.86** (0.74, 0.99)	**0.83** (0.73, 0.95)	0.35	NI	**RA**	**EA**	**SD**
CANVAS	Canagliflozin vs placebo	3P MACE	**0.86** (0.75, 0.97)	**0.77** (0.70, 0.85)	1.34	NI	**RA**	**EA**	**SD**
CARMELINA	Linagliptin vs placebo	3P MACE	**1.02** (0.89, 1.17)	**0.90** (0.84, 0.96)	1.61	NI	*	**EA**	**SD**
TECOS	Sitagliptin vs placebo	3P MACE + angina	**0.98** (0.88, 1.09)	**0.89** (0.86, 0.91)	1.71	NI	*	**EA**	**SD**
SAVOR-TIMI	Saxagliptin vs placebo	3P MACE	**1.00** (0.89, 1.12)	**0.81** (0.76, 0.86)	3.16¶	NI	*	**–**	**–**
CAROLINA	Linagliptin vs glimepiride	3P MACE	**0.98** (0.84, 1.14)	**0.91** (0.79, 1.05)	0.70	NI	**RA**	**EA**	**SD**
LEAD2	Liraglutide vs glimepiride	Change in hbA1c	**0.0** (-0.2, 0.2)	**0.05** (-0.11, 0.22)		NI	**RA**	**EA**	**SD**

Better placebo proxy

Worse placebo proxy

Missing HbA1c

Good | Moderate | Poor

Diabetes

Placebo control

(2nd line diabetes drug – **DPP4i vs sulfonylureas**, 3P MACE)

Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
CARMELINA	Linagliptin vs placebo	3P MACE	**1.02** (0.89, 1.17)	**0.90** (0.84, 0.96)	1.61	NI	*	EA	SD
TECOS	Sitagliptin vs placebo	3P MACE + angina	**0.98** (0.88, 1.09)	**0.89** (0.86, 0.91)	1.71	NI	*	EA	SD
SAVOR-TIMI	Saxagliptin vs placebo	3P MACE	**1.00** (0.89, 1.12)	**0.81** (0.76, 0.86)	3.16¶	NI	*	–	–
CAROLINA	Linagliptin vs glimepiride	3P MACE	**0.98** (0.84, 1.14)	**0.91** (0.79, 1.05)	0.70	NI	RA	EA	SD

Good

Moderate

Poor

Comparator emulation

- Expensive new drug vs older cheap drug
- Difficult to capture SES differences?

Placebo control

(2nd line diabetes drug – GLP1, SGLT2i vs DPP4i, 3P MACE)

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

Good

Moderate

Poor

Comparator emulation

- Comparing expensive newer drugs
- Closer therapeutic alternatives

LEAD2 – hbA1c outcome, multiple imputation

STEP 1 – Missing data exploration

1. Assumption that **HbA1c values at the end of follow-up** were missing at random (MAR)
2. Examined number and proportion of missing values for each censoring reason
3. Compared baseline characteristics of patients with and without missing values

STEP 2 – MI model specification

Selection of independent variables:

1. Auxiliary variables:
   - Variables that were imbalanced between patients with (n. 479) vs without (n. 267) an HbA1 lab result during follow-up
   - Variables associated with difference in HbA1c from baseline to follow up
2. Analytical variable: Variable from analytical model (**exposure**)

List of auxiliary variables
Age
Gender
Race
Hypertension
Hyperlipidemia
Cerebrovascular diseases
Diabetic Neuropathy
Mood Disorders
Use of statins or other lipid lowering drugs
Business type
Frailty Score

STEP 3 – MI analysis specification:

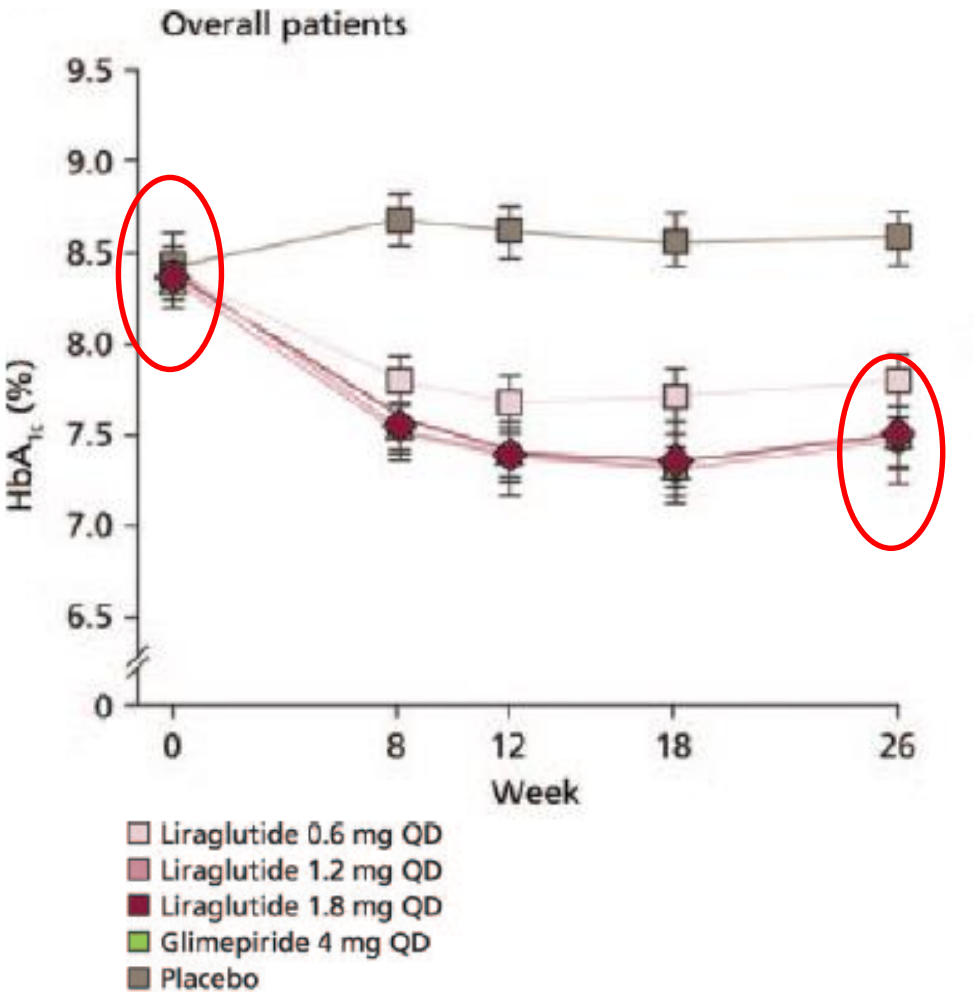
1. Imputation phase (**n. of imputations = 15**)
2. Analysis phase – **ANALYTICAL MODEL: Outcome (HbA1c difference) = exposure (liraglutide = 1; glimepiride = 0)**
3. Pooling phase

Results - HbA_{1c} levels over time

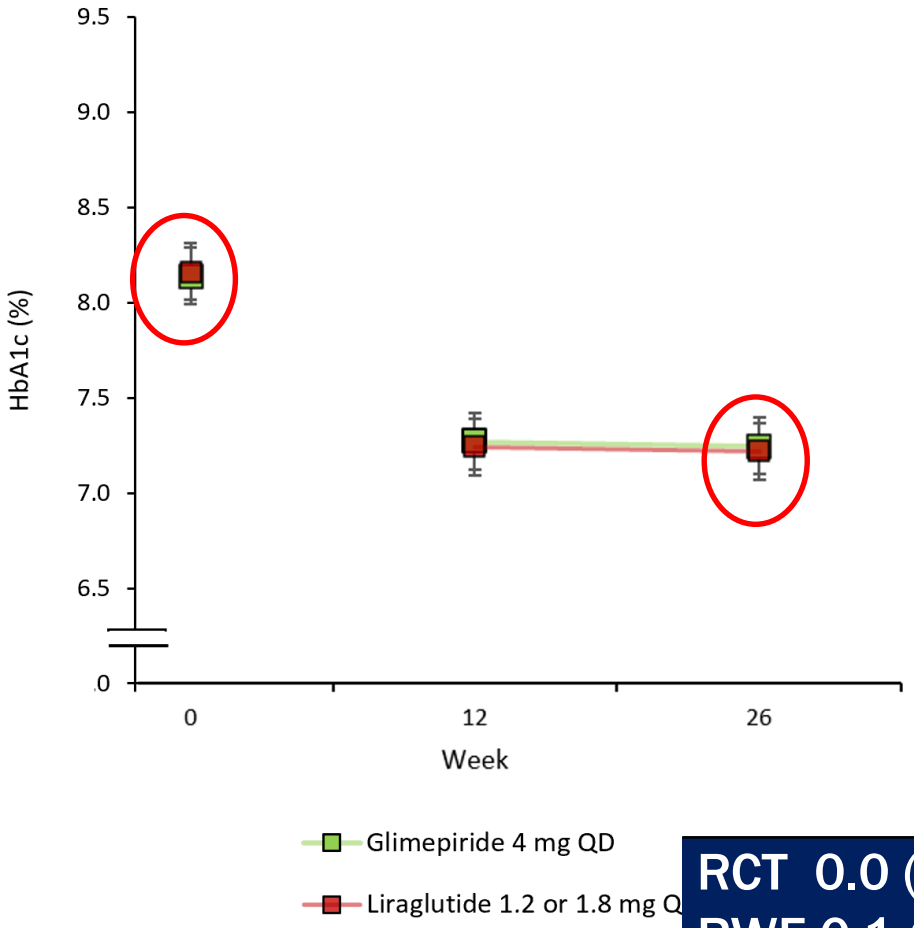
HbA_{1c}	LEAD2	RWD
Baseline	8.4	8.2
6 months	7.5	7.3

Diabetes

LEAD2 Trial



DUPLICATE - RWE Lead-2 Trial



RCT 0.0 (-0.2, 0.2)
RWE 0.1 (-0.1, 0.2)

Emulation Challenges

Antiplatelet

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

In-hospital start of medication

Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--

MarketScan	1.20 (0.95, 1.51)
Optum	0.73 (0.52, 1.01)
P for homogeneity	0.01

Good

Moderate

Poor

Variation in co-medication in multi-national trials

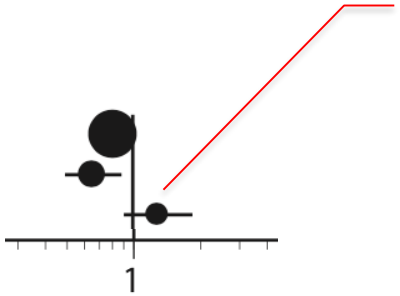
Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--

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)



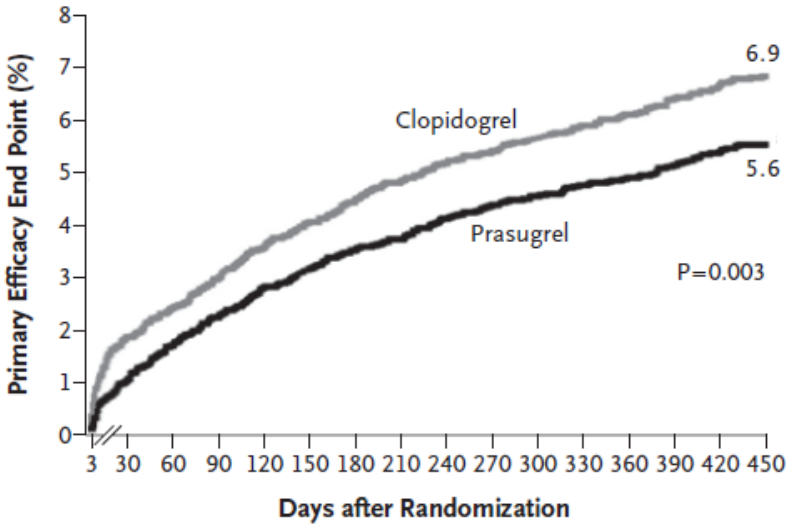
Note: RCT-DUPLICATE used U.S. data sources only

** Mahaffey KW et al. Circ 2011*

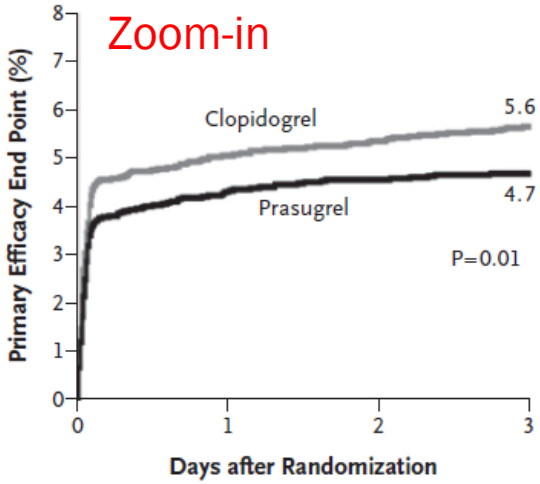
In-hospital start of medication

Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--



Wiviott et al, NEJM 2007



take-home points:

- RCT shows early and immediate effect – starting FU while in hospital
- RWE study question targets patients who survive until discharge and fill 1st Rx
- Cannot capture early effect without linked hospital + outpatient Rx data

In-hospital start of medication

Antiplatelet

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRITON-TIMI	Prasugrel vs clopidogrel	3P MACE	**0.81** (0.73, 0.90)	**0.88** (0.79, 0.97)	-1.11	Sup	**RA**	**EA**	**SD**
PLATO	Ticagrelor vs clopidogrel	3P MACE	**0.84** (0.77, 0.92)	**0.92** (0.83, 1.02)	-1.31	Sup	**–**	**EA**	**SD**
ISAR-REACT5	Ticagrelor vs prasugrel	3P MACE	**1.36** (1.09, 1.70)	--	--	Sup	--	--	--

Criticisms of ISAR-REACT5

- 83% had only telephone follow up, 7% written correspondence
- ITT analysis, ~1/3 or more discontinued therapy
- Compared different treatment strategies, not therapies
- Surprisingly low prasugrel outcome rate compared to prior trials

1 year event rates from trials*

	Prasugrel	Ticagrelor	Clopidogrel
TRITON-TIMI	9.9		12.1
PLATO		9.8	11.7
ISAR-REACT5	6.8	9.3	

*TRITON-TIMI and PLATO outcome: *cardiovascular death*, stroke, MI
ISAR-REACT outcome: *all-cause death*, stroke, MI

Ostrowska, M et al. Folia Cardiologica 2019
Kow, C et al. European Heart Journal 2021

Emulation Challenges

Atrial fibrillation

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

DOAC treatment for Afib

Atrial fibrillation

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
ARISTOTLE	Apixaban vs warfarin	Stroke/ systemic embolism	**0.79** (0.66, 0.95)	**0.68 (0.61, 0.76)**	1.36	NI	**RA**	**EA**	**SD**
RE-LY	Dabigatran vs warfarin	Stroke/ systemic embolism	**0.66** (0.53, 0.82)	**0.73 (0.60, 0.90)**	-0.66	NI	**RA**	**EA**	**SD**
ROCKET-AF	Rivaroxaban vs warfarin	Stroke/ systemic embolism	**0.79** (0.66, 0.96)	**0.70 (0.62, 0.80)**	1.00	NI	**RA**	**EA**	**SD**

"Patients taking VKAs at the time of randomization stopped their VKA drug on the day of randomization and began the assigned drug when the INR fell <2.0 (if randomized to dabigatran) or <3.0 (if randomized to warfarin)."

- RE-LY rationale and design, Ezekowitz, et al, American Heart Journal 2009

take-home points:

- Trials included treatment naïve initiators and prevalent users who "switched" versus "continued" after a washout (similar estimates in both strata)
- RWE studies included patients who were initiators after 6-month washout

Good | Moderate | Poor

Emulation Challenges

VTE

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Summary:
5 DOAC – VTE trials for different indications

VTE

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
RE-COVER II	Dabigatran vs Warfarin	VTE, VTE related death	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	RA	EA	SD	VTE
AMPLIFY	Apixaban vs Enoxaparin/ Warfarin	VTE, VTE related death	**0.84** (0.60, 1.18)	**0.76** (0.53, 1.09)	0.40	NI	RA	EA	SD	VTE
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-	PE
RECORD1	Rivaroxaban vs Enoxaparin	VTE/ all cause death	**0.25** (0.14, 0.47)	**0.17** (0.10, 0.29)	0.91	NI	RA	EA	SD	Hip

Good

Moderate

Poor

Summary:
5 DOAC – VTE trials for different indications

VTE

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
RE-COVER II	Dabigatran vs Warfarin	VTE, VTE related death	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	RA	EA	SD	VTE
AMPLIFY	Apixaban vs Enoxaparin/ Warfarin	VTE, VTE related death	**0.84** (0.60, 1.18)	**0.76** (0.53, 1.09)	0.40	NI	RA	EA	SD	VTE
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-	PE
RECORD1	Rivaroxaban vs Enoxaparin	VTE/ all cause death	**0.25** (0.14, 0.47)	**0.17** (0.10, 0.29)	0.91	NI	RA	EA	SD	Hip

Good

Moderate

Poor

Summary:
5 DOAC – VTE trials for different indications

VTE

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
RE-COVER II	Dabigatran vs Warfarin	VTE, VTE related death	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	RA	EA	SD	VTE
AMPLIFY	Apixaban vs Enoxaparin/ Warfarin	VTE, VTE related death	**0.84** (0.60, 1.18)	**0.76** (0.53, 1.09)	0.40	NI	RA	EA	SD	VTE
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-	PE
RECORD1	Rivaroxaban vs Enoxaparin	VTE/ all cause death	**0.25** (0.14, 0.47)	**0.17** (0.10, 0.29)	0.91	NI	RA	EA	SD	Hip

Good

Moderate

Poor

VTE

Summary:
5 DOAC – VTE trials for different indications

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
RE-COVER II	Dabigatran vs Warfarin	VTE, VTE related death	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	RA	EA	SD	VTE
AMPLIFY	Apixaban vs Enoxaparin/ Warfarin	VTE, VTE related death	RCT	RWE	0.40	NI	RA	EA	SD	VTE
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-	PE
RECORD1	Rivaroxaban vs Enoxaparin	VTE/ all cause death	**0.25** (0.14, 0.47)	**0.17** (0.10, 0.29)	0.91	NI	RA	EA	SD	Hip

Good

Moderate

Poor

It remains unclear why EINSTEIN-DVT and EINSTEIN-PE trials have separating results. Meta-analysis of 6 trials* finds no heterogeneity of effects in patients presenting with DVT or PE.

Dentali F, et al. Intern Emerg Med. 2015

Study or Subgroup	NOACs Events	NOACs Total	VKA Events	VKA Total	Weight	Risk Ratio M-H, Random, 95% CI	Year
4.1.1 PE							
EINSTEIN-PE	50	2419	44	2413	13.1%	1.13 [0.76, 1.69]	2012
HOKUSAI	47	1650	65	1669	15.5%	0.73 [0.51, 1.06]	2013
AMPLIFY	21	900	23	886	6.2%	0.90 [0.50, 1.61]	2013
RECOVER-I&II	23	795	25	807	6.8%	0.93 [0.53, 1.63]	2014
Subtotal (95% CI)		**5764**		**5775**	**41.6%**	**0.90 [0.72, 1.13]**	
Total events	141		157				
Heterogeneity: Tau² = 0.00; Chi² = 2.50, df = 3 (P = 0.48); I² = 0%							
Test for overall effect: Z = 0.91 (P = 0.37)							
4.1.2 DVT							
EINSTEIN-DVT	36	1731	51	1718	11.9%	0.70 [0.46, 1.07]	2010
HOKUSAI	83	2468	81	2453	23.3%	1.02 [0.75, 1.38]	2013
AMPLIFY	38	1698	47	1736	11.8%	0.83 [0.54, 1.26]	2013
RECOVER-I&II	45	1758	37	1747	11.4%	1.21 [0.79, 1.86]	2014
Subtotal (95% CI)		**7655**		**7654**	**58.4%**	**0.93 [0.75, 1.16]**	
Total events	202		216				
Heterogeneity: Tau² = 0.01; Chi² = 3.81, df = 3 (P = 0.28); I² = 21%							
Test for overall effect: Z = 0.65 (P = 0.52)							
Total (95% CI)		**13419**		**13429**	**100.0%**	**0.92 [0.80, 1.07]**	
Total events	343		373				
Heterogeneity: Tau² = 0.00; Chi² = 6.37, df = 7 (P = 0.50); I² = 0%							
Test for overall effect: Z = 1.11 (P = 0.27)							
Test for subgroup differences: Chi² = 0.04, df = 1 (P = 0.84), I² = 0%							

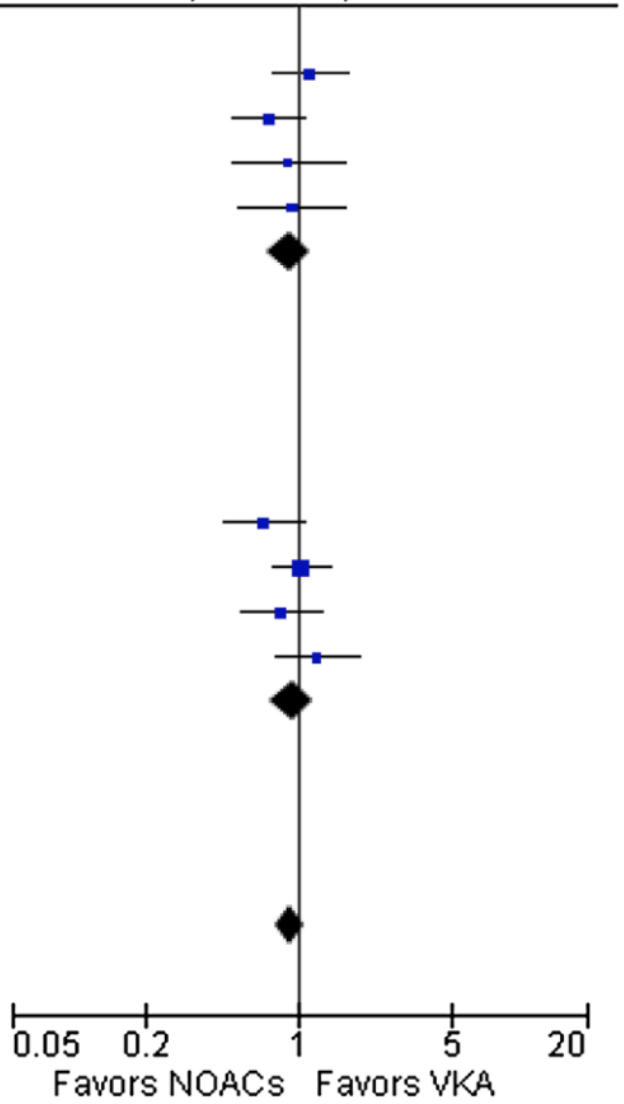


Fig. 1 Forest plot of the primary efficacy outcome (recurrent VTE or death related to VTE) in patients receiving non-vitamin K oral anticoagulants (NOACs) or vitamin K antagonists (VKA)

Summary:
5 DOAC – VTE trials for different indications

VTE

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
RE-COVER II	Dabigatran vs Warfarin	VTE, VTE related death	**1.08** (0.64, 1.80)	**1.10** (0.76, 1.60)	-0.06	NI	RA	EA	SD	VTE
AMPLIFY	Apixaban vs Enoxaparin/ Warfarin	VTE, VTE related death	**0.84** (0.60, 1.18)	**0.76** (0.53, 1.09)	0.40	NI	RA	EA	SD	VTE
EINSTEIN-DVT	Rivaroxaban vs Enoxaparin/VKA	VTE	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	-0.42	NI	*	EA	SD	DVT
EINSTEIN-PE	Rivaroxaban vs Enoxaparin/VKA	VTE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	2.21	NI	*	-	-	PE
RECORD1	Rivaroxaban vs Enoxaparin	VTE/ all cause death	**0.25** (0.14, 0.47)	**0.17** (0.10, 0.29)	0.91	NI	RA	EA	SD	Hip

Good

Moderate

Poor

take-home points:

- Single trial was used as reference standard for causal effects.
- Sometimes trials don't agree.

Emulation Challenges

Hypertension

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Antihypertensives: Telmisartan

Hypertension

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
TRANSCEND	Telmisartan vs placebo	3P MACE + HHF	**0.92** (0.81, 1.05)	**0.88** (0.81, 0.96)	0.55	Sup	*	**EA**	**SD**
ON-TARGET	Telmisartan vs Ramipril	3P MACE + HHF	**1.01** (0.94, 1.09)	**0.83** (0.77, 0.90)	3.46	NI	–	–	–

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)

Investigating subtle differences in exposure, outcome, inclusion-exclusion criteria, covariates, follow-up

Good

Moderate

Poor

take-home point:

- Important to investigate robustness of evidence to reasonable alternative choices

Emulation Challenges

Osteoporosis

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Delayed effect with long follow up

Osteoporosis

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
HORIZON-PIVOTAL	Zoledronic acid vs placebo	Hip Fracture	**0.59** (0.42, 0.83)	**0.72** (0.55, 0.94)	0.90	Sup	**RA**	**EA**	**SD**
VERO	Teriparatide vs risedronate	Vertebral fracture	**0.44** (0.29, 0.68)	-	-2.54	Sup	–	–	–

MarketScan	1.33 (0.80, 2.20)
Optum	0.43 (0.19, 0.96)
P for homogeneity	0.02

Good

Moderate

Poor

Delayed treatment effects

HORIZON-PIVOTAL (osteoporosis, hip fracture)

RCT

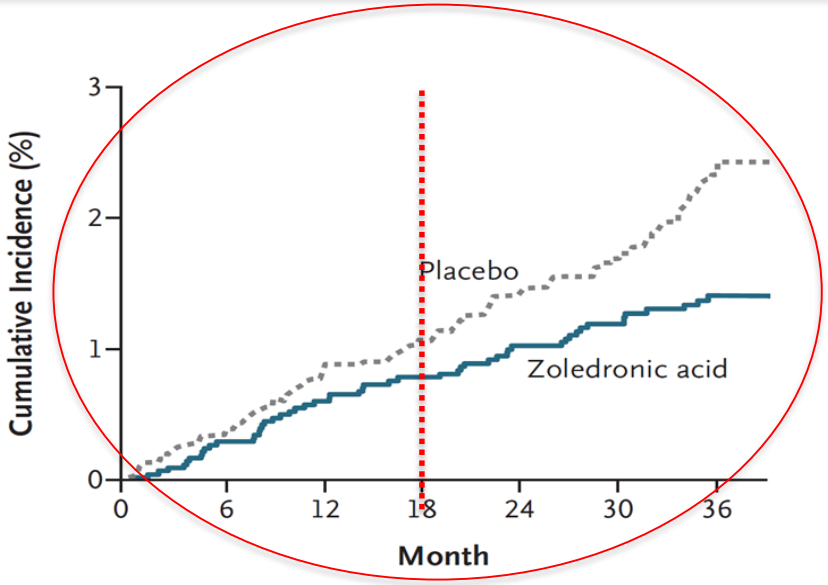
HR_{36mo} = 0.59 (0.42, 0.83)

HR_{18mo} = 0.75

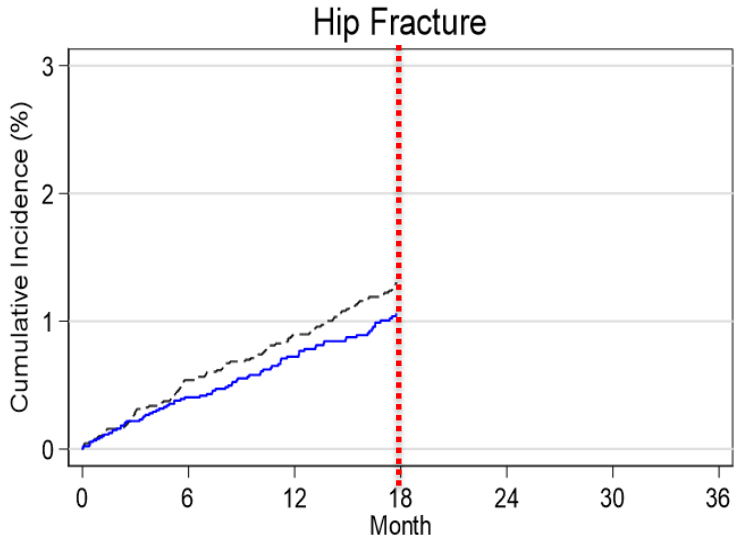
RWE

HR_{36mo} = ??

HR_{18mo} = 0.75 (0.58, 0.97)

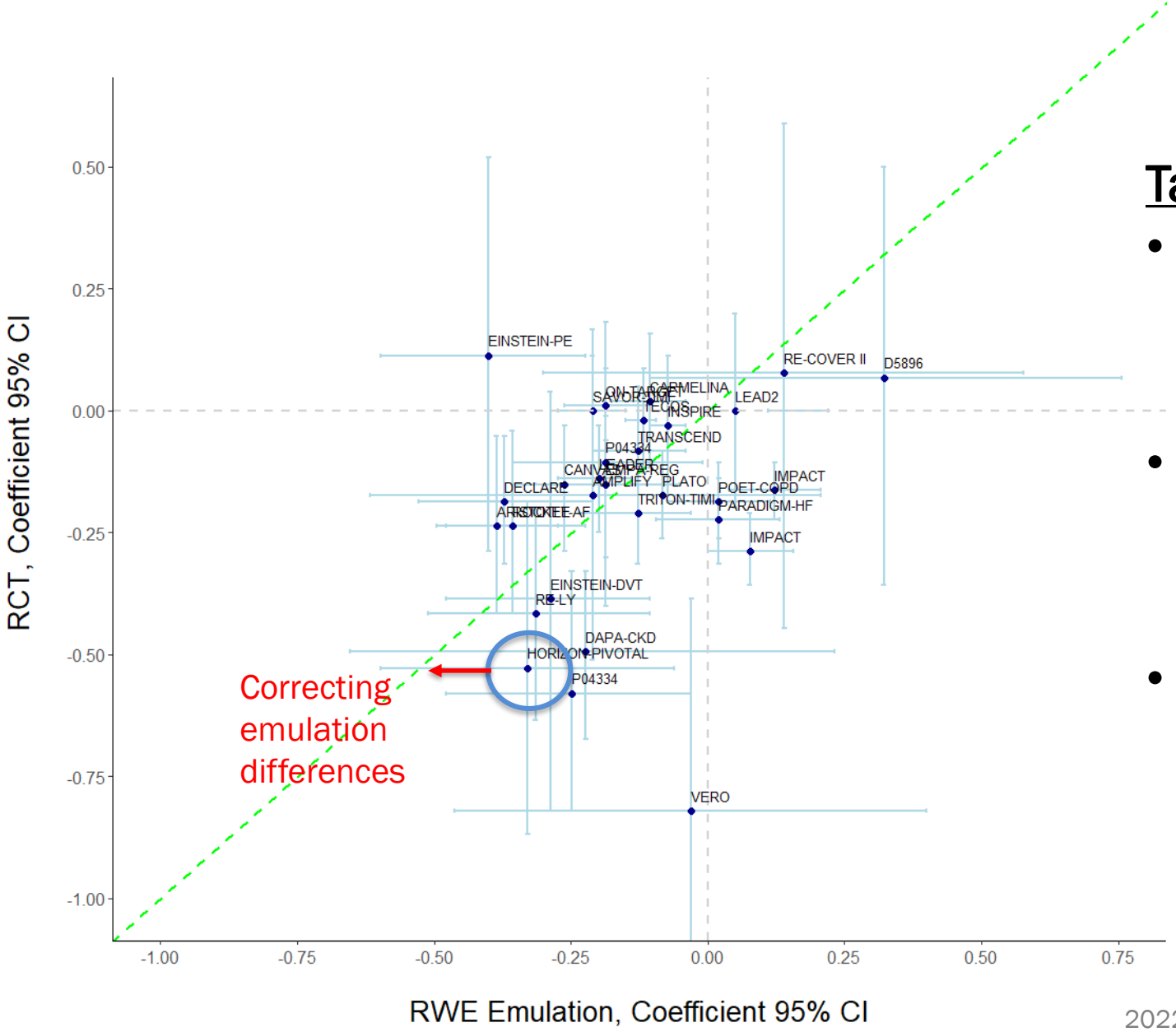


No. at Risk							
Zoledronic acid	3875	3807	3674	3553	3494	3387	3161
Placebo	3861	3806	3694	3577	3499	3397	3144



Number at risk							
Raloxifene	9003	7753	6768	0	0	0	0
Zoledronic acid	9003	7766	6743	0	0	0	0

RCT by RWE findings



Take-home points:

- Challenging to replicate trial findings for outcomes with long induction or time varying hazard
- Patients in clinical practice may not experience benefit observed in explanatory trial
- Important to replicate in multiple databases

Emulation Challenges

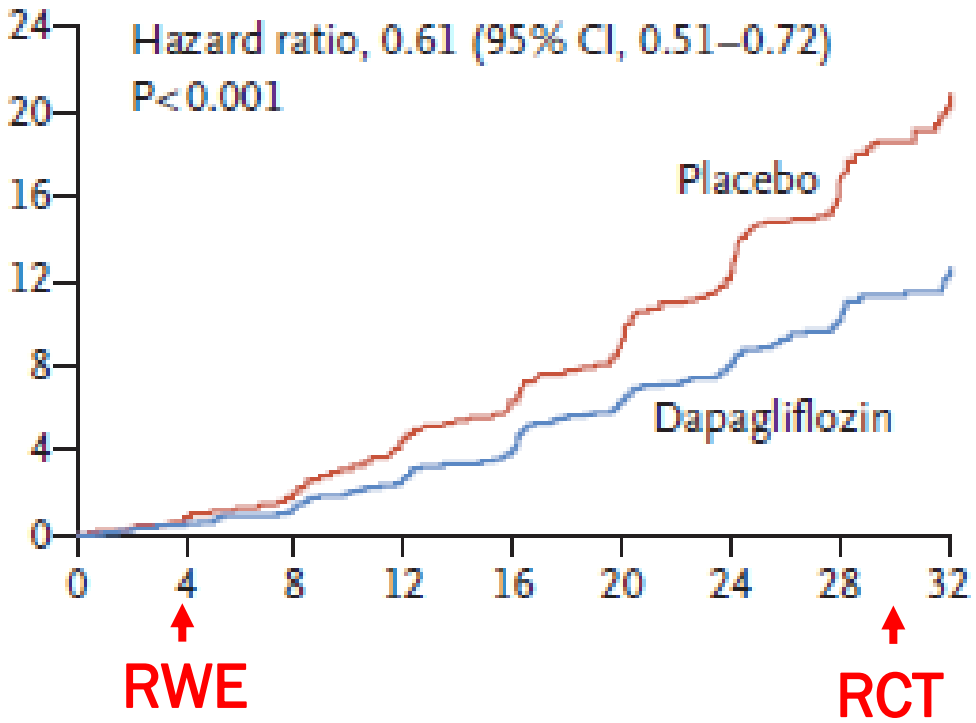
CKD

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Delayed effect with long follow up

CKD

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement		
DAPA-CKD	Dapaglifloxin vs placebo	Sustained decline in eGFR, ESRD, death	**0.61** (0.51, 0.72)	**0.80** (0.52, 1.26)	-1.10	Sup	-	-	**SD**



Median Follow Up

Median Follow Up

Good

Moderate

Poor

Emulation Challenges

Heart Failure

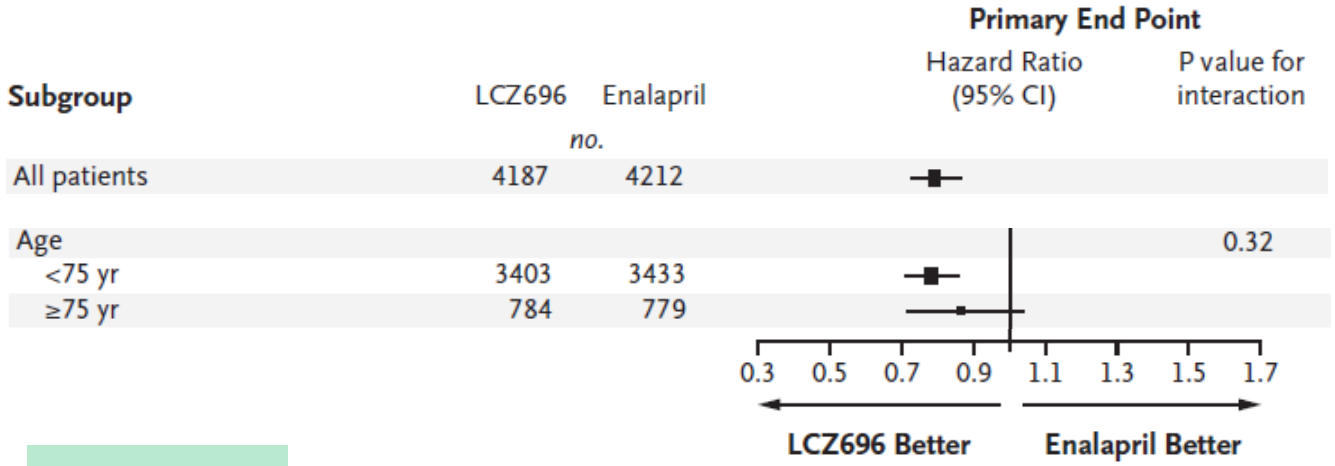
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Heart failure: Sacubitril/ Valsartan (Entresto)

Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
PARADIGM-HF	Sacubitril/valsartan vs Enalapril	HHF/CV death	**0.80** (0.73, 0.87)	**0.97** (0.87-1.08)	-3.42	Sup	–	–	–

Treatment effect reduced in those 75+?

RCT by age:



RWE by age:

0.89 (0.77-1.02)

1.04 (0.89-1.23)

Good

Moderate

Poor

Heart failure: Sacubitril/ Valsartan (Entresto)

Trial name	Comparator emulation	Endpoint emulation+	RCT result	RWE results	Stand. Diff.	Test	Agreement		
PARADIGM-HF	Sacubitril/valsartan vs Enalapril	HHF/CV death	**0.80** (0.73, 0.87)	**0.97** (0.87-1.08)	-3.42	Sup	–	–	–

Original research

Effectiveness of angiotensin-neprilysin inhibitor treatment versus renin-angiotensin system blockade in older adults with heart failure in clinical care

Rishi J Desai ,[1] Elisabetta Patorno,[1] Muthiah Vaduganathan,[2] Mufaddal Mahesri,[1] Kristyn Chin,[1] Raisa Levin,[1] Scott D Solomon,[2] Sebastian Schneeweiss[1]

0.92 (0.84, 1.00) Initiators of ACE/ARB vs sacubitril/valsartan
0.79 (0.74, 0.85) Switchers from ACE to ARB vs sacubitril/valsartan
0.84 (0.80, 0.89) Combined

Investigating subtle differences in exposure, outcome, inclusion-exclusion criteria, covariates, follow-up

Good

Moderate

Poor

take-home point:

- Important to explore and understand robustness of evidence to reasonable alternative choices

Emulation Challenges

Asthma/COPD

- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation of maintenance therapy at randomization
- Robustness

Run-in on placebo vs run-in on medication under investigation

Immediate transient effect of discontinuation on outcome

Run-in selects responders to 1 Tx arm + Discontinuation of maintenance therapy → short term ↑ exacerbation

Asthma/COPD

Trial name	Comparator	Endpoint	RCT result	RWE results	Stand. Diff.	Test	Agreement			Indication
P4334	Mometasone furoate/formoterol vs formoterol	Exacerbation	0.56 (0.44, 0.72)	0.78 (0.62-0.97)	-1.95	Sup	**RA**	–	**SD**	asthma
D5896	Budesonide + formoterol vs budesonide	Asthma related death, intubation, hosp	1.07 (0.70, 1.65)	1.38 (0.90-2.13)	-0.81	NI	**RA**	**EA**	**SD**	asthma
IMPACT	Fluticasone, umeclidinium, vilanterol vs fluticasone, vilanterol	Exacerbation	0.85 (0.80, 0.90)	1.13 (1.04-1.23)	-5.46	Sup	–	–	–	COPD
POET-COPD	Tiotropium vs salmeterol	Exacerbation	0.83 (0.77, 0.90)	1.02 (0.93-1.12)	-3.27	Sup	–	–	–	COPD
INSPIRE	Advair vs tiotropium	Exacerbation	0.97 (0.84, 1.12)	0.93 (0.90-0.96)	0.56	Sup	**RA**	**EA**	**SD**	COPD

Good

Moderate

Poor

Asthma/COPD

Run-in selects responders to 1 Tx arm + Discontinuation of maintenance therapy → short term ↑ exacerbation

Conceptual Question: What is the risk of exacerbation with Drug A vs Drug B?

RCT

Operational Question: In *responders to run-in therapy* (Drug A) who are randomized to Drug A vs Drug B ***AND*** to *simultaneously discontinue* baseline maintenance medication, what is the risk of exacerbation?

RWD

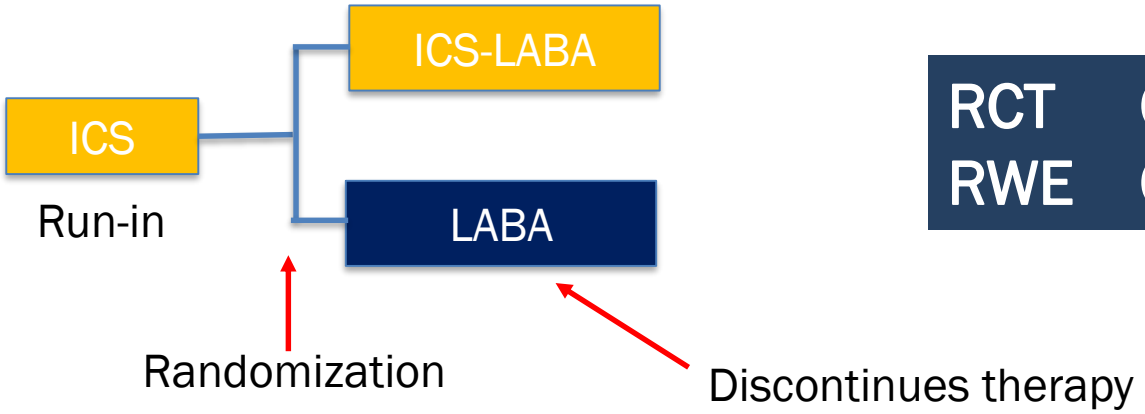
Operational Question: In patients who initiate Drug A vs Drug B (with a washout of X days) ***AND*** have similar baseline therapy, what is the risk of exacerbation?

take-home point:

Some explanatory design features in trials cannot be emulated with clinical practice data → RCT and RWE study will target different questions

Run-in selects responders to 1 Tx arm

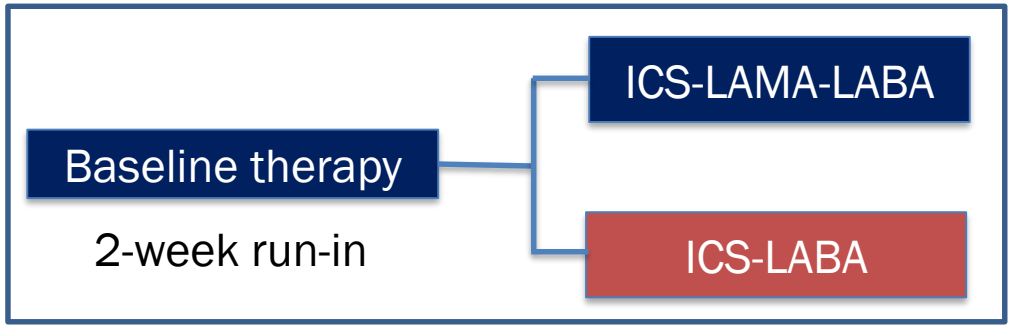
P4334, asthma exacerbation



RCT	0.56 (0.44, 0.72)
RWE	0.89 (0.72, 1.10)

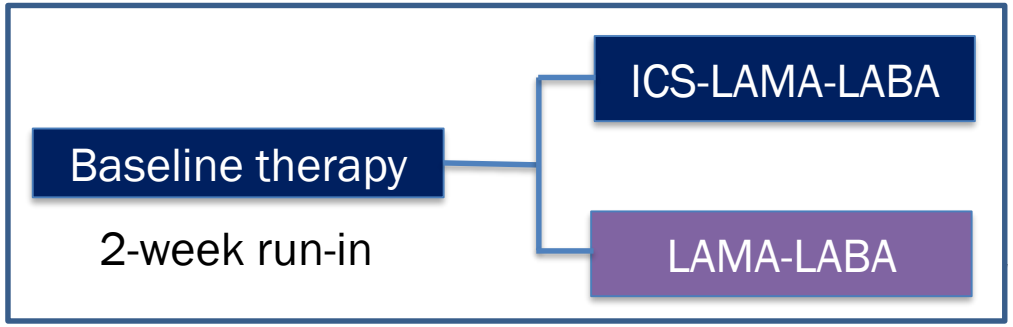
RCT: Discontinuation of maintenance therapy → short term ↑ exacerbation
RWE: Confounding by stage of therapy/disease severity

IMPACT, COPD exacerbation



Discontinues therapy

RCT	0.85 (0.80, 0.90)
RWE	1.13 (1.04, 1.23)



38% on triple therapy

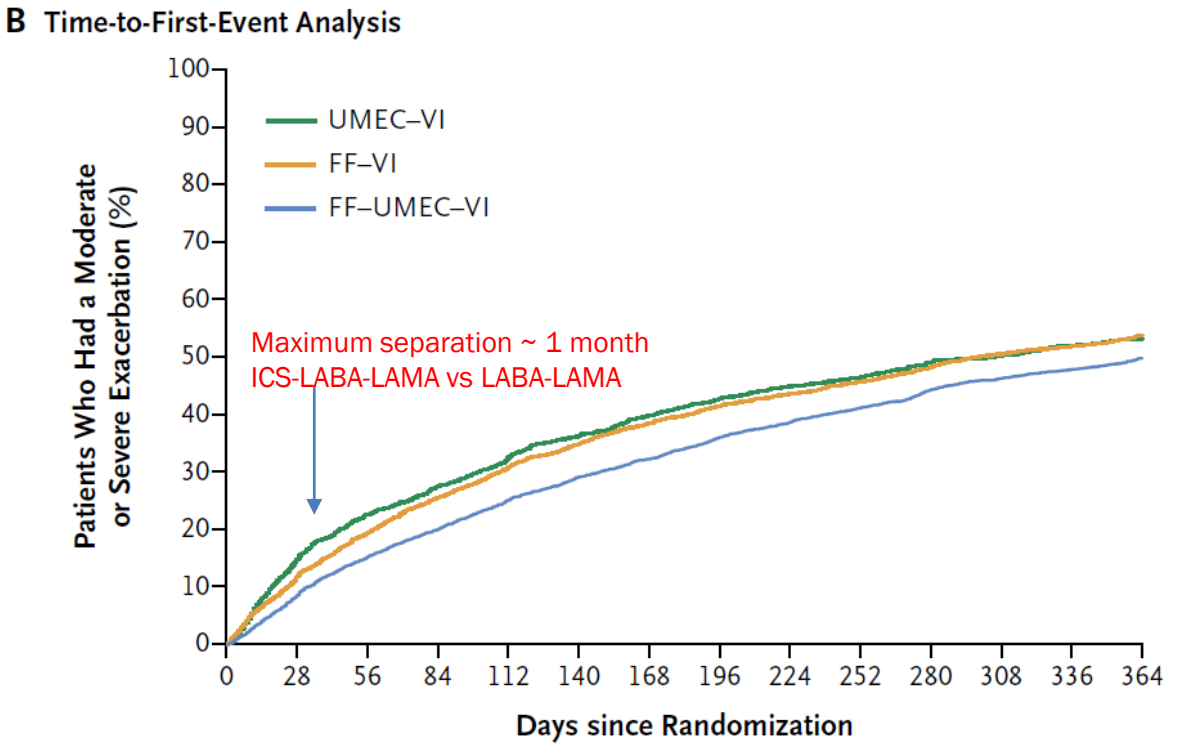
Discontinues therapy

RCT	0.75 (0.70, 0.81)
RWE	1.08 (1.00, 1.17)

Randomization

Discontinuation of maintenance therapy → short term ↑ exacerbation

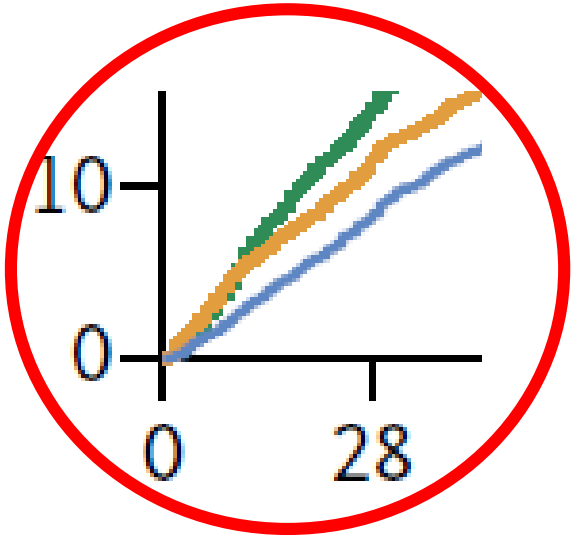
IMPACT trial and COPD exacerbation event



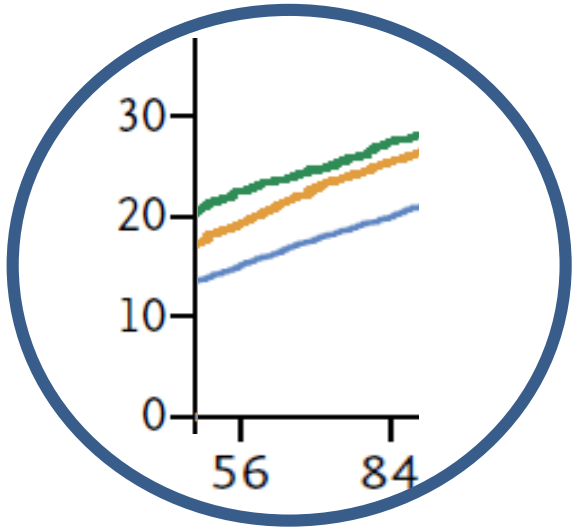
No. at Risk

UMEC–VI	2070	1721	1516	1406	1301	1201	1123	1059	1001	971	917	884	851	642
FF–VI	4134	3554	3133	2838	2620	2410	2250	2120	2004	1823	1823	1729	1671	1228
FF–UMEC–VI	4151	3758	3408	3186	2954	2752	2614	2457	2324	2216	2085	1988	1919	1419

First 30-day interval



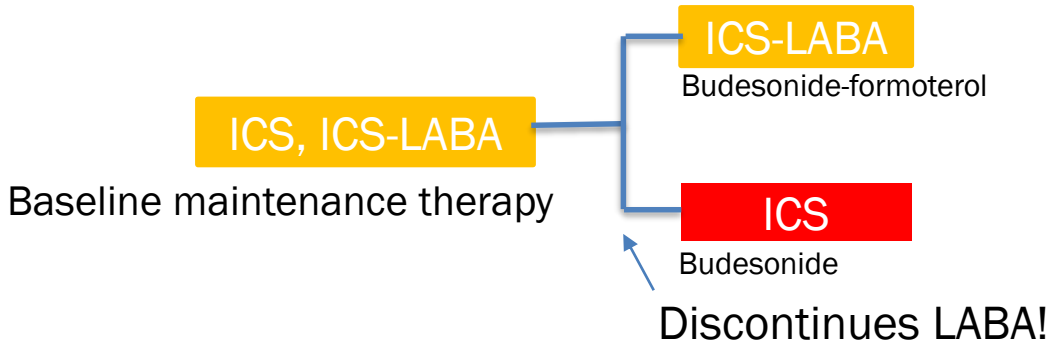
Second 30-day interval



Suissa S, Ariel A. Triple therapy trials in COPD: a precision medicine opportunity. Eur Respir J 2018; 52: 1801848 [https://doi.org/10.1183/13993003.01848-2018].

Non-inferiority bound translates to RR 1.32 (upper limit 95% CI <2)

D5896
Treatment: ICS-LABA vs ICS
Outcome: Serious asthma related events



Assumptions Scenario 1:

- Truth is upper bound of non-inferiority limit (1.32)
- 50% of patients were on LABA at baseline
- No effect of discontinuation

Baseline	Randomized: ICS-LABA	Randomized: ICS
No LABA use	29	22
LABA use	29	22
	58	44

RR = 58/44 = 1.32

Assumptions Scenario 2:

- Truth is upper bound of non-inferiority limit (1.32
- 50% of patients were on LABA at baseline
- *Discontinuation increases risk of outcome by 50%*

Baseline	Randomized: ICS-LABA	Randomized: ICS
No LABA use	29	22
LABA use	29	22+11
	58	55

RR = 58/55 = 1.05

D5896	1.07 (0.70, 1.65)
Pooled RWD	1.38 (0.90, 2.13)

Suissa S et al. *Chest* 2013 143(5). May 2013" 1208-1213

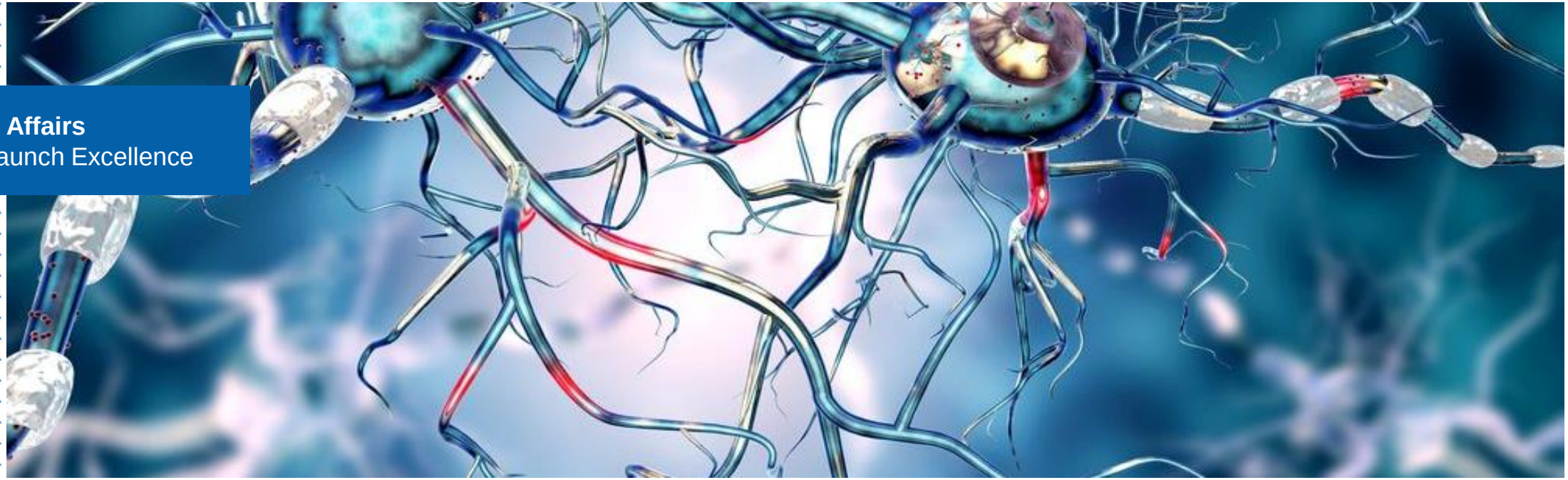
Break

1:05 pm – 1:20 pm

Stakeholder Reactions to Result Findings, and Implications

1:20 pm – 2:25 pm

Global Medical Affairs
Evidence and Launch Excellence



Lessons Learned from Trial Replication Analyses: Findings from the DUPLICATE Demonstration Project

Melvin (Skip) Olson, ScD
Head Integrated Evidence Strategy and Innovation
Novartis
May 2022

NOVARTIS | Reimagining Medicine

Reactions to the Lessons Learned

Why RWE?

- Many uses
- Many stakeholders

Emulation framework

- Builds toolbox for defining “fit-for-purpose”
- Focus on the real issues, not duplication per se

Next steps

- Do we need MOTS?
- Refined emulation framework

A huge round of applause to the RCT DUPLICATE team

Remarks on the RCT DUPLICATE project

Issa J. Dahabreh

CAUSALab
A Center to Learn What Works
HARVARD
T.H. CHAN
SCHOOL OF PUBLIC HEALTH

Target trial emulation view

Core methods of a highly pragmatic **target trial** can be **emulated** by observational analysis (except for randomization)

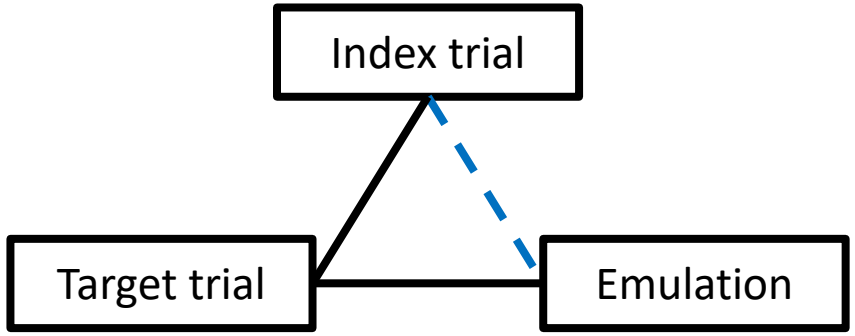
Emulation estimates have a causal interpretation (under assumptions)

Hernán & Robins; American Journal of Epidemiology 2016
Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available

Benchmarking view

Target trial methods chosen to be similar to those in a highly pragmatic **index trial**

Close **emulation** of **target trial** methods $\Longrightarrow$ agreement of **emulation** and **index trial** results (under assumptions)

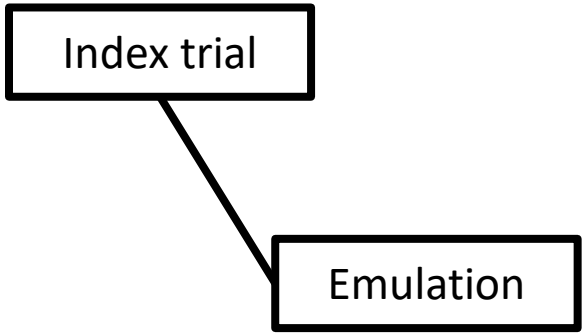


Dahabreh, Robins, Hernán; Epidemiology 2020
Benchmarking Observational Methods by Comparing Randomized Trials and Their Emulations

RCT DUPLICATE view

Emulate **index trials**

Focus on agreement of results between **index trial** and **emulation**



Index trials in RCT DUPLICATE are not pragmatic

Index trial

Index trials sometimes differ from any **target trial** that can be emulated by observational analysis of claims data

Examples:

- Placebo control
- Run-in periods

Target trial

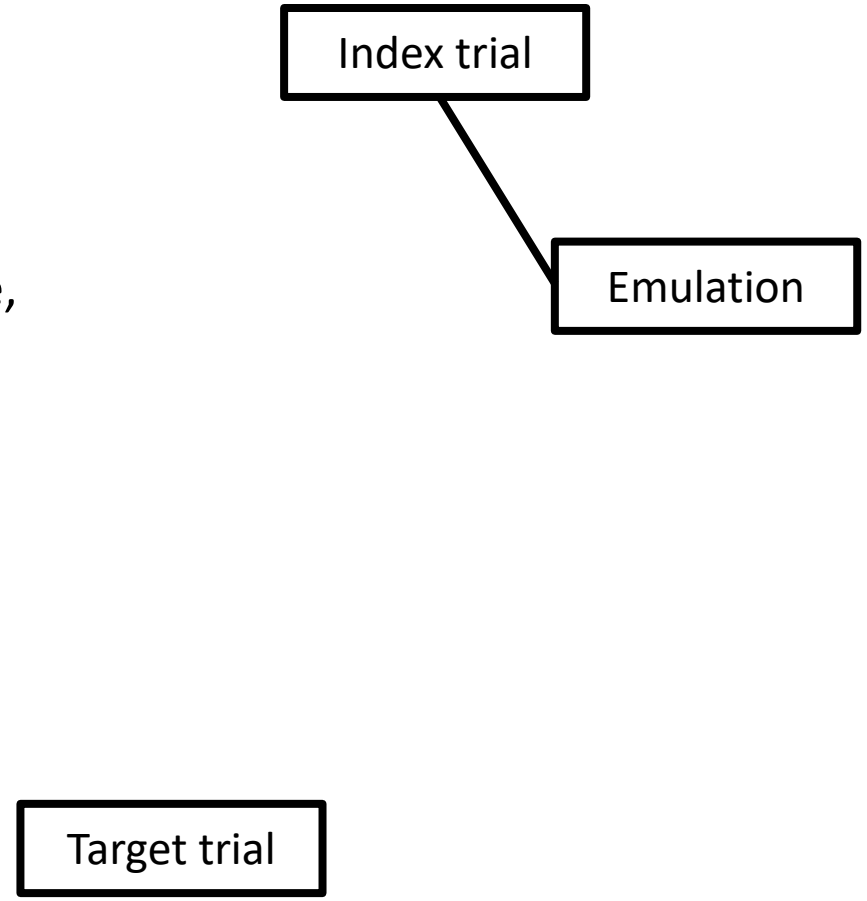
Methods chosen to improve results agreement

Examples:

- Placebo emulated by initiation of an active agent
- Censoring on non-adherence, no adjustment for selection bias

Not “in the spirit” of target trial emulation

Emphasis on predictive performance



Causal – predictive spectrum of comparisons

	Black box prediction of index trial results using observational analyses	Emulation of target trials with observational analyses
ATTITUDE:	Study of studies, meta-regression	Causal inference
OBJECT OF STUDY:	Response surface of results, pattern recognition	Causal estimands
SUCCESSS:	Good predictive performance	Close emulation of methods $\Longrightarrow$ conclusion (estimate) agreement

Practical comparisons live in the *uncomfortable middle* of the spectrum.

Adjournment

Lessons Learned from Trial Replication Analyses: Findings from the DUPLICATE Demonstration Project

May 10, 2022

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