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

Results from the RCT DUPLICATE[1] initiative

Shirley V Wang, PhD, Sebastian Schneeweiss, MD, ScD, and the RCT-DUPLICATE Initiative

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

[1] **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

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

Key publications on our rationale, methodology and results

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]

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

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]

Original Investigation

April 25, 2023

Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses

Results of 32 Clinical Trials

Shirley V. Wang, PhD, ScM[1]; Sebastian Schneeweiss, MD, ScD[1]; and the RCT-DUPLICATE Initiative

» Author Affiliations

JAMA. 2023;329(16):1376-1385. doi:10.1001/jama.2023.4221

RCT-DUPLICATE: A demonstration project

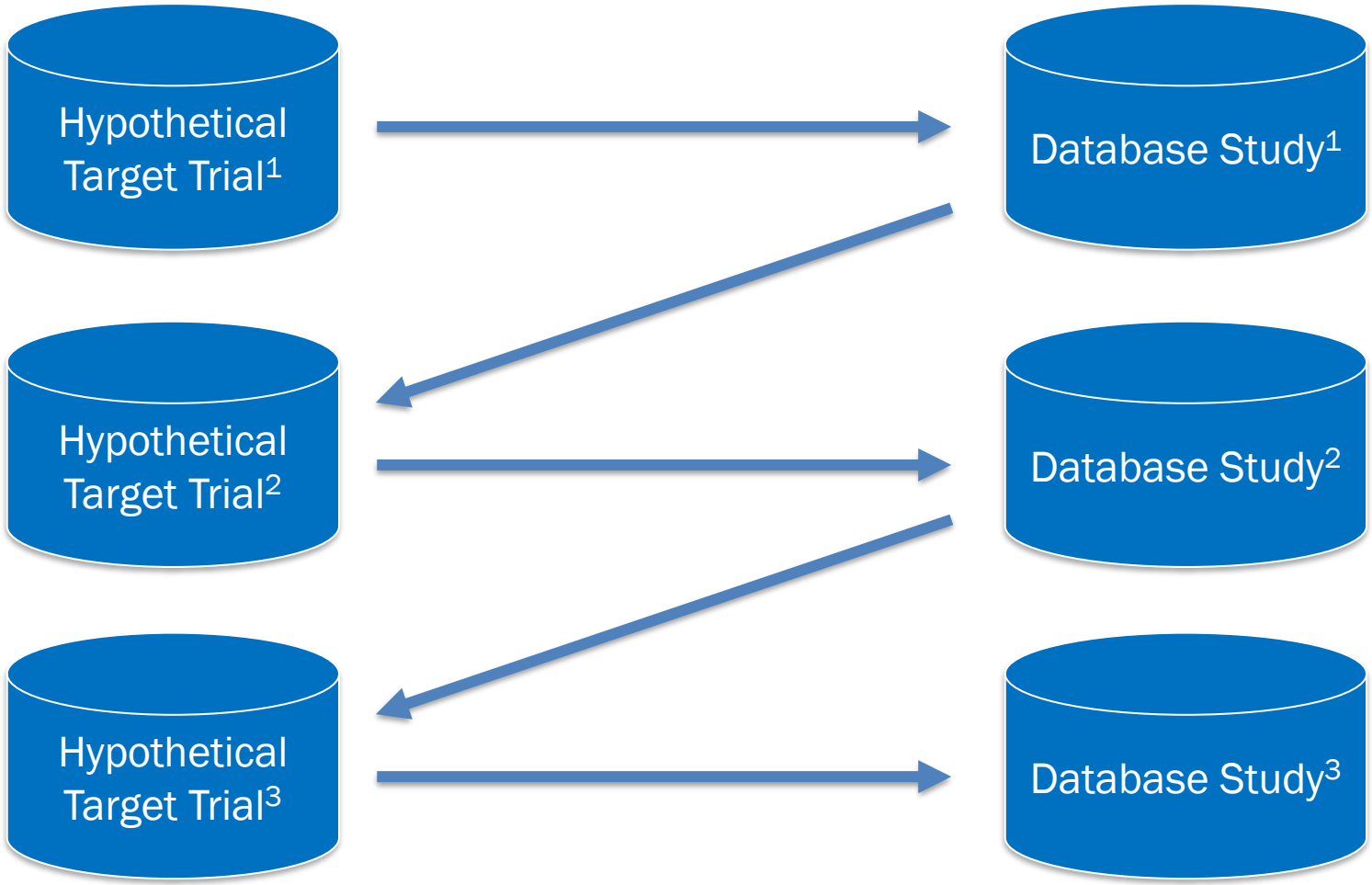
A family of studies aimed to understand and improve the validity of RWE studies for regulatory decision making

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 there been a similarly designed RWE study instead of an RCT** 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

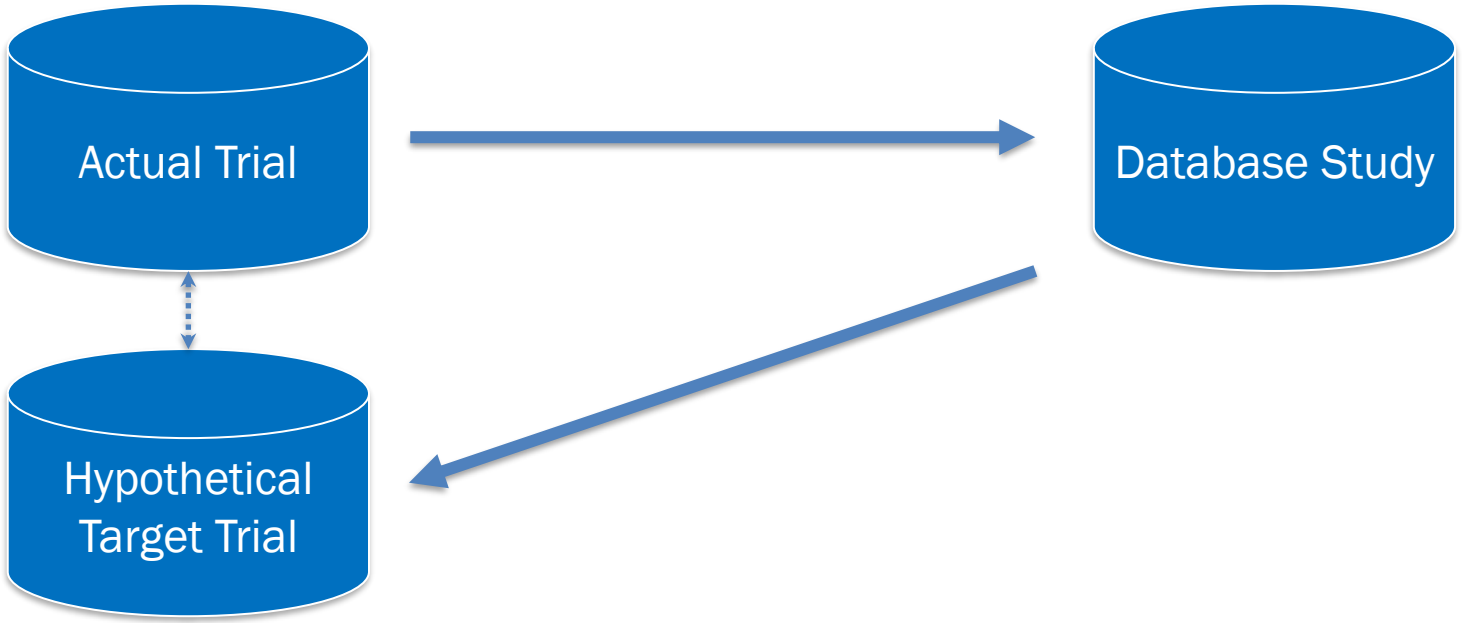
Designing a database study to mimic a hypothetical trial

Iterate until data and design are fit-for-purpose for relevant question



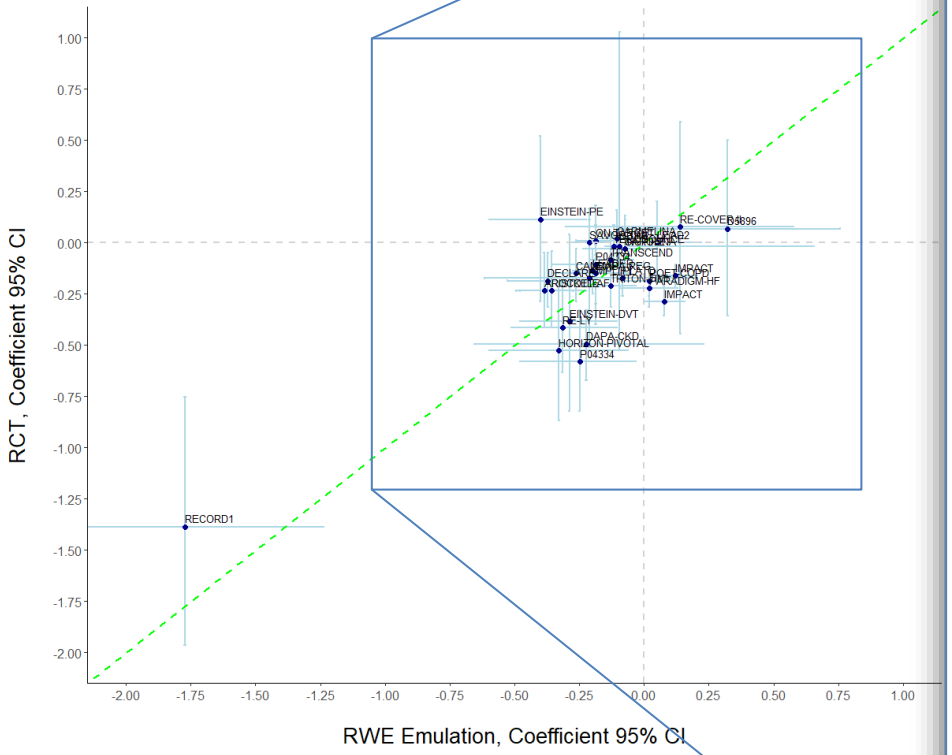
Emulation of actual RCTs to use RCT results as reference standard

Hypothetical target trial $\overset{?}{\approx}$ *Actual published trial*

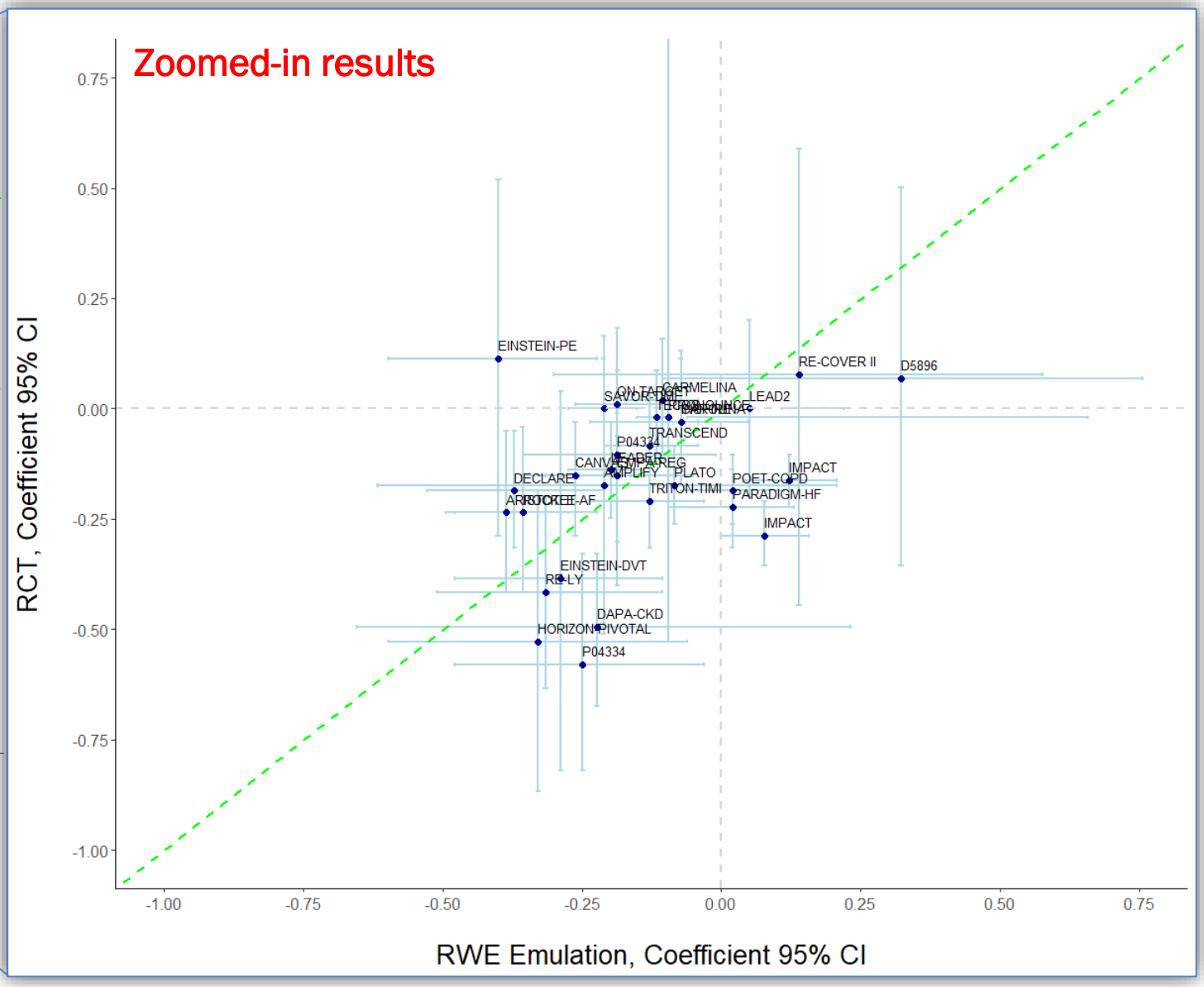


32 RCT-RWE emulation results

Pearson's overall = 0.80; 0.62-0.90



Zoomed-in results



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]

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
	DECLARE	0.83 (0.73, 0.95)	0.69 (0.59, 0.81)	0.47 (0.41, 0.53)	1.76	RA/–/SD	N
	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
	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
	CARMELINA	1.02 (0.89, 1.17)	0.90 (0.84, 0.96)	0.90 (0.86, 0.95)	1.61	–/EA/SD	N
	TECOS	0.98 (0.88, 1.09)	0.89 (0.86, 0.91)	0.81 (0.79, 0.84)	1.71	–/EA/SD	N
	SAVOR-TIMI	1.00 (0.89, 1.12)	0.81 (0.76, 0.86)	0.65 (0.62, 0.69)	3.16	–/–/–	N
	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
	PLATO	0.84 (0.77, 0.92)	0.92 (0.83, 1.02)	0.84 (0.78, 0.91)	-1.31	–/EA/SD	N
	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]		Std. Diff.	Agreement	Close emulation?	
			Adjusted	Unadjusted				
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 Differences

- 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 Differences

- 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

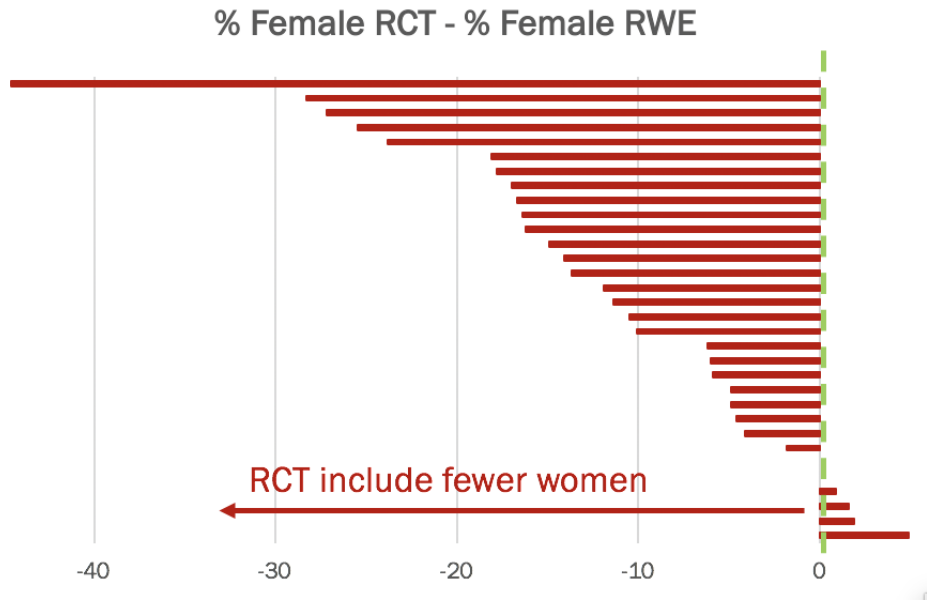
Age>=18, HFrEF, HF hospitalization within 12 months
Stable on ACEis/ARBs and beta-blocker therapies

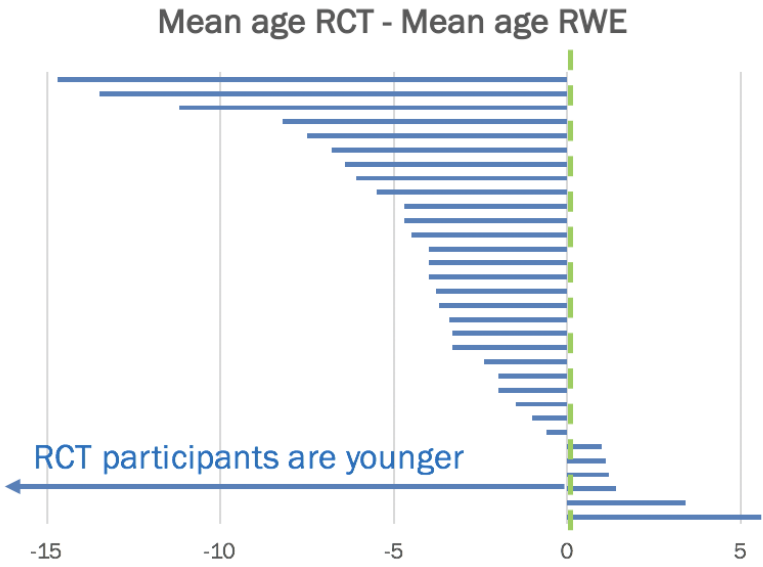
Exclusion

- 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 Differences

- 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 Differences

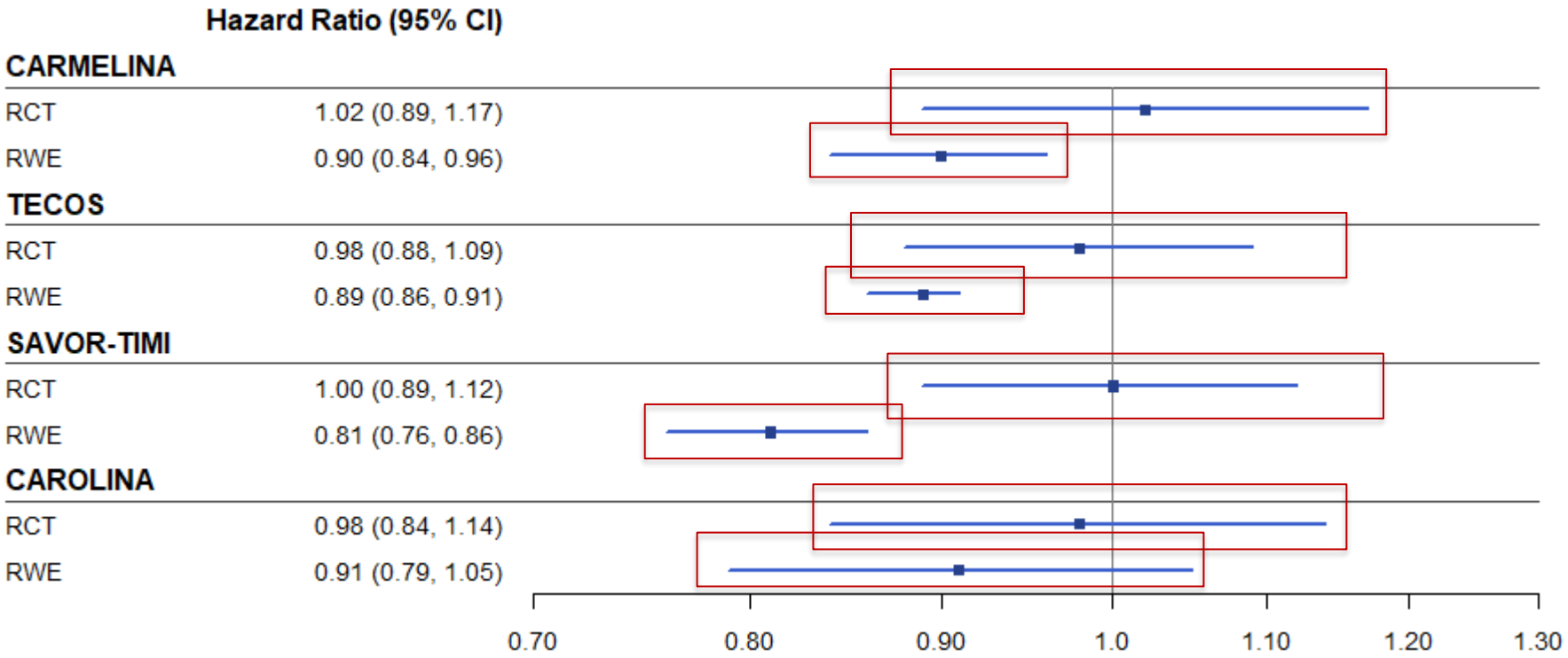
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation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

Placebo control emulation

RCT: DPP4i vs placebo and risk of 3P MACE

RWE: DPP4i vs 2nd gen sulfonylurea and risk of 3P MACE

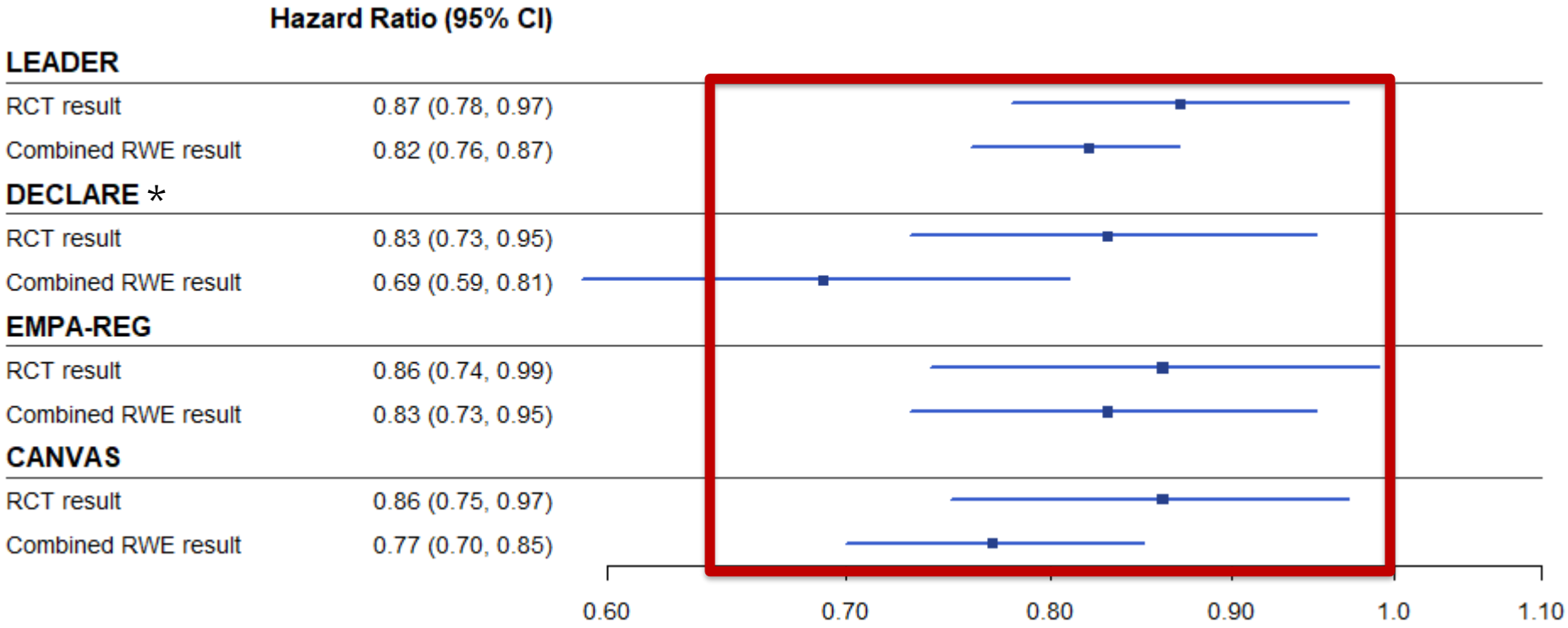


- Expensive new drug vs older cheap drug
- Difficult to capture SES differences?

Placebo control emulation

RCT: GLP1, SGLT2i vs placebo and risk of 3P MACE*

RWE: GLP1, SGLT2i vs DPP4i and risk of 3P MACE*



- Comparing expensive newer drugs
- Closer therapeutic alternatives

* DECLARE was HHF and death

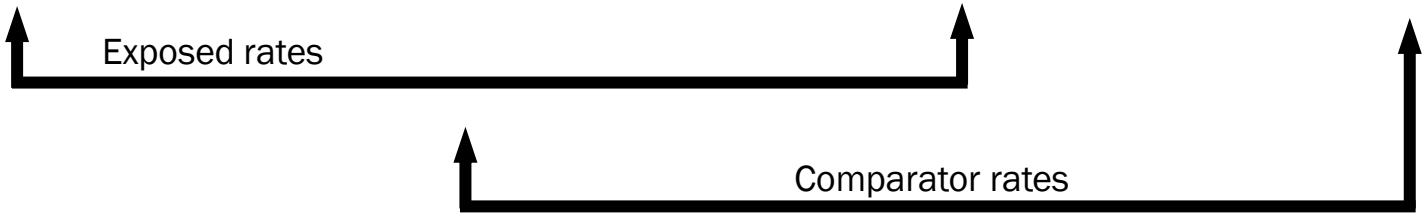
Emulation Differences

- 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



Assessed with high specificity

Lower specificity or high missingness

Emulation Differences

- 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

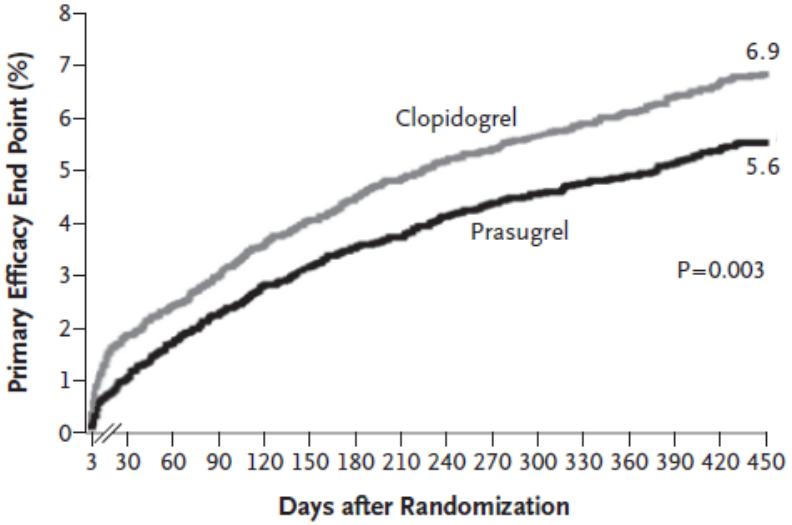
Moderate

Poor

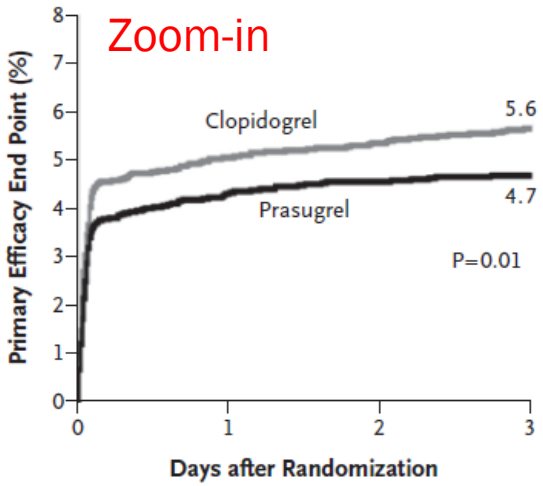
In-hospital start of medication

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 Differences

- 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 Differences

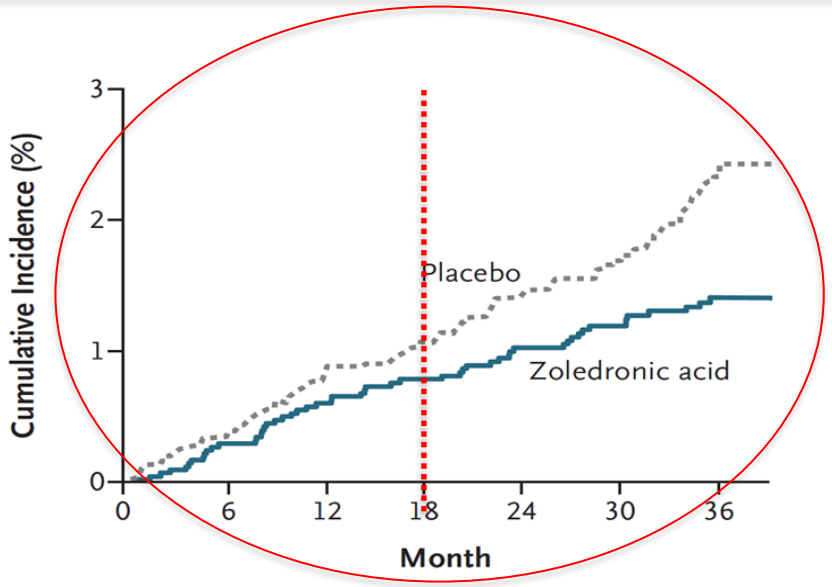
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation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

Delayed treatment effects

Osteoporosis

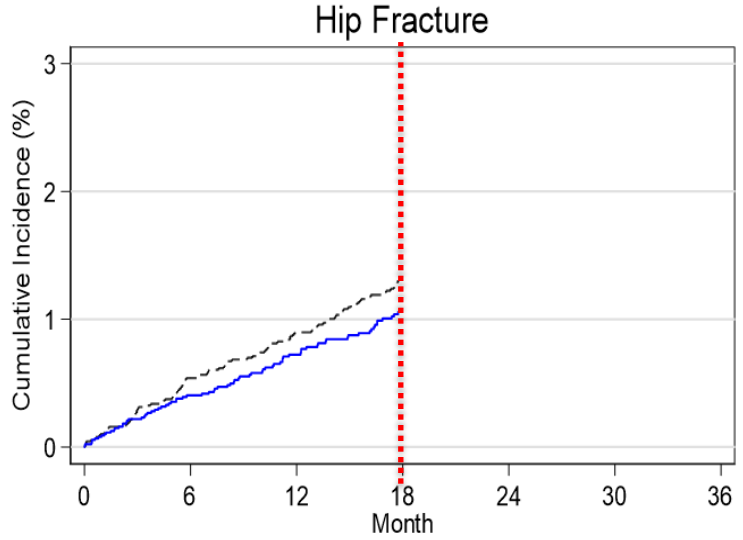
HORIZON-PIVOTAL (osteoporosis, hip fracture)

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



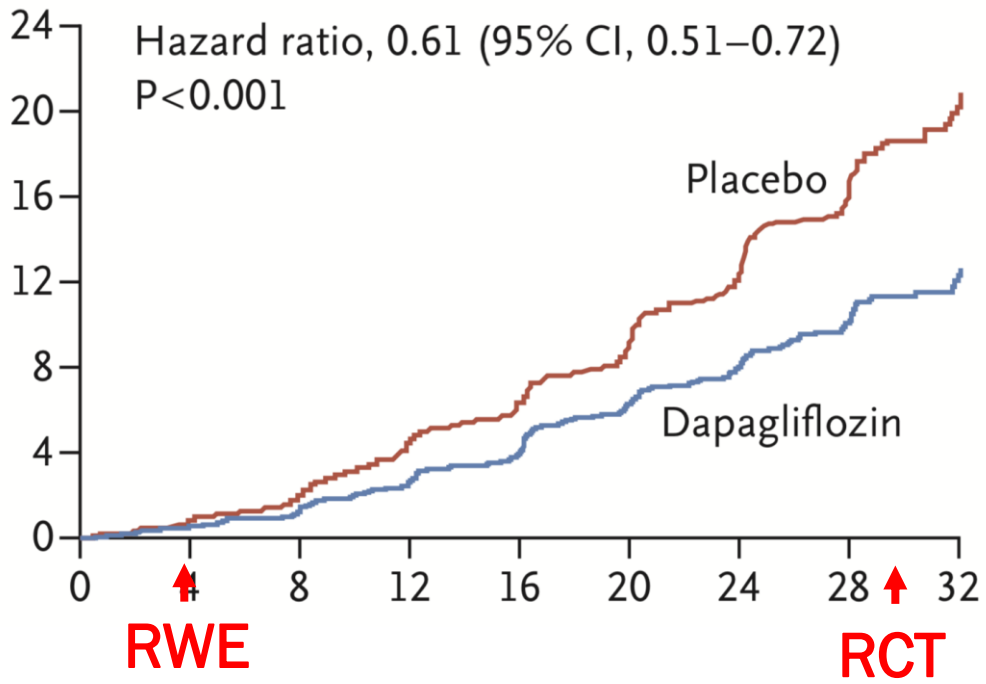
Number at risk

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	Dapagliflozin 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

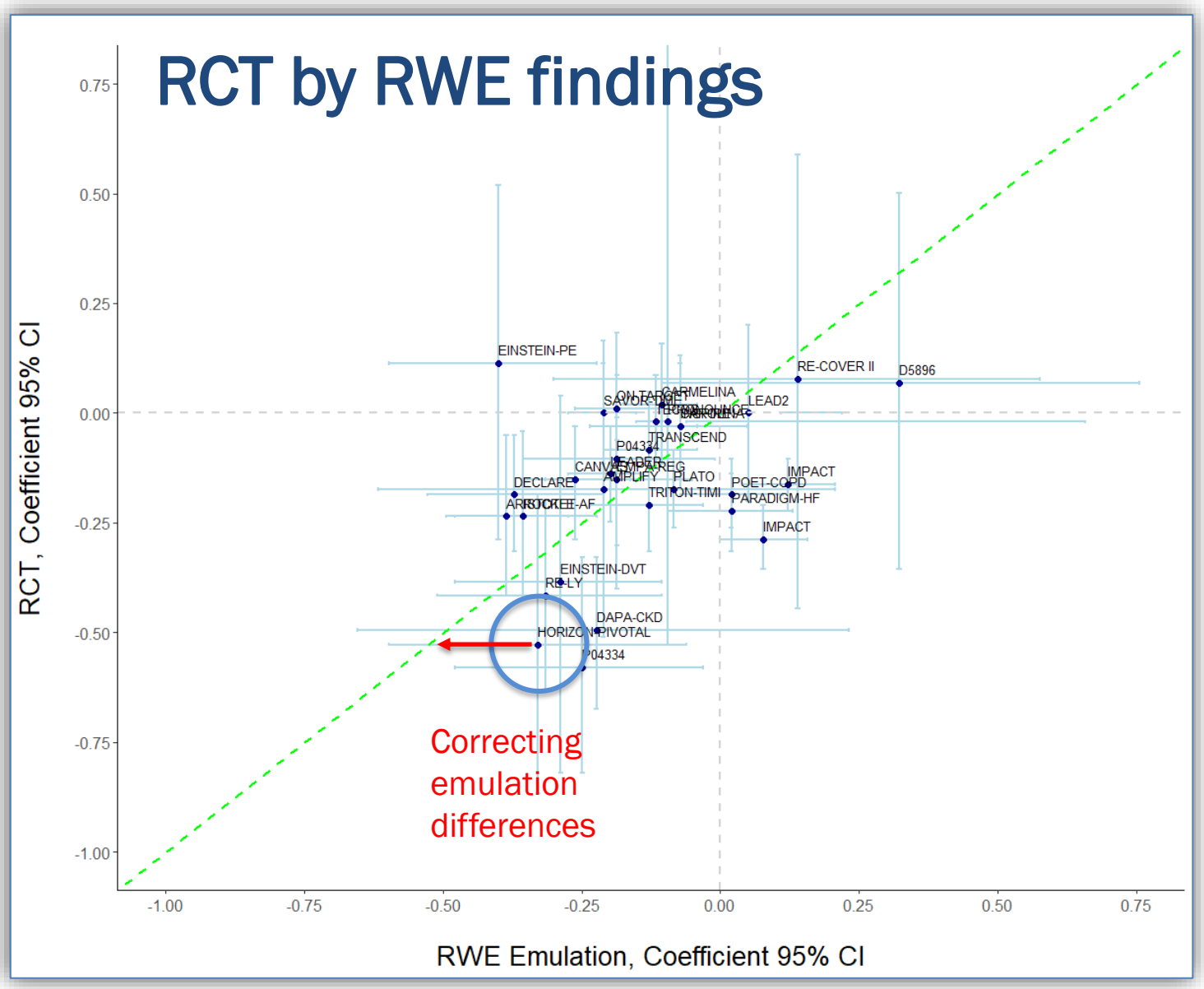
Median Follow Up

Good

Moderate

Poor

RCT by RWE findings



- *Emulation difference:* time varying effect over long follow up in RCT + low adherence in clinical practice

- Correction for design difference → closer calibration

Take home points:

- Challenging to replicate trial findings when effect is delayed
- Patients in clinical practice may not experience full benefit seen in explanatory trial

Emulation Differences

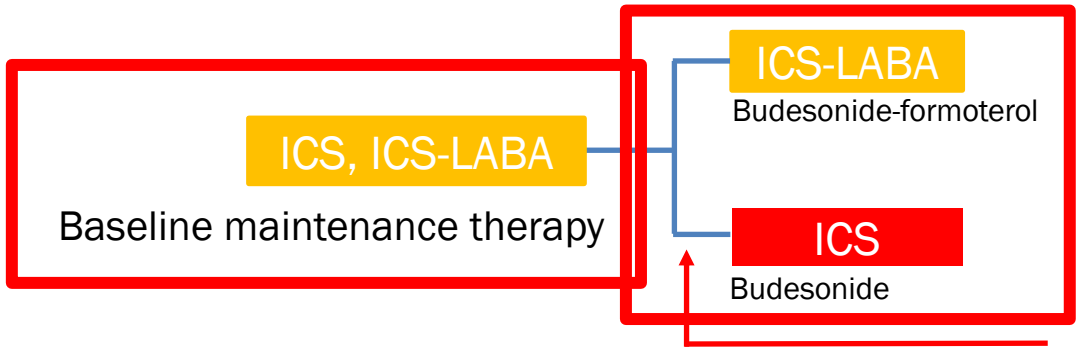
- Inclusion-exclusion emulation
- Population distribution
- Comparator emulation (good, moderate, poor)
- Outcome emulation (good, moderate)
- Placebo control
- In-hospital start of medication
- Dose titration protocol during follow-up
- Delayed effect with a long follow-up window
- Run-in window
- Discontinuation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

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

		Randomized	
		ICS-LABA	ICS
Baseline	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%

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

RR = 58/55 = 1.05

D5896	1.07 (0.70, 1.65)

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

Emulation Differences

- 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, Elisabetta Patorno, Muthiah Vaduganathan, Mufaddal Mahesri, Kristyn Chin, Raisa Levin, Scott D Solomon, Sebastian Schneeweiss

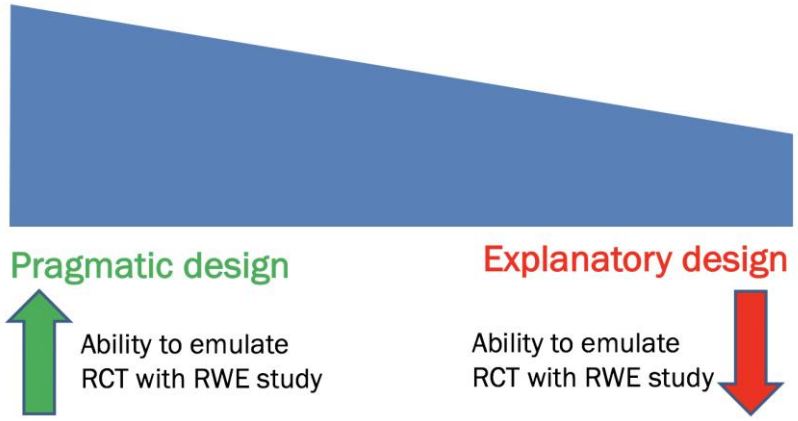
Investigating subtle differences in exposure, outcome, inclusion-exclusion criteria, covariates, follow-up

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

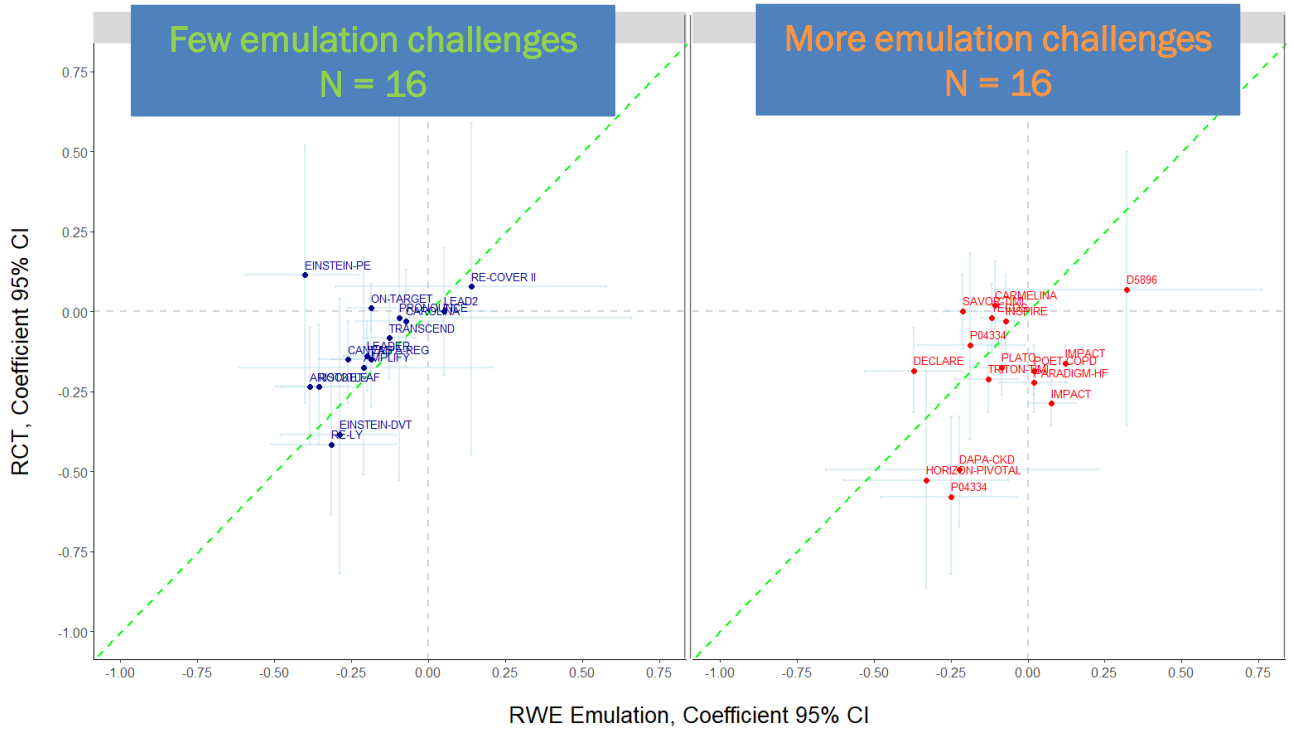
More emulation challenges = a OR b OR c OR d OR poor comparator emulation OR neither comparator and outcome emulation are classified as good

Few emulation challenges vs more emulation challenges

Pearson's overall = 0.80; 0.62-0.90

	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*	16 (100%)	9 (56%)
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:

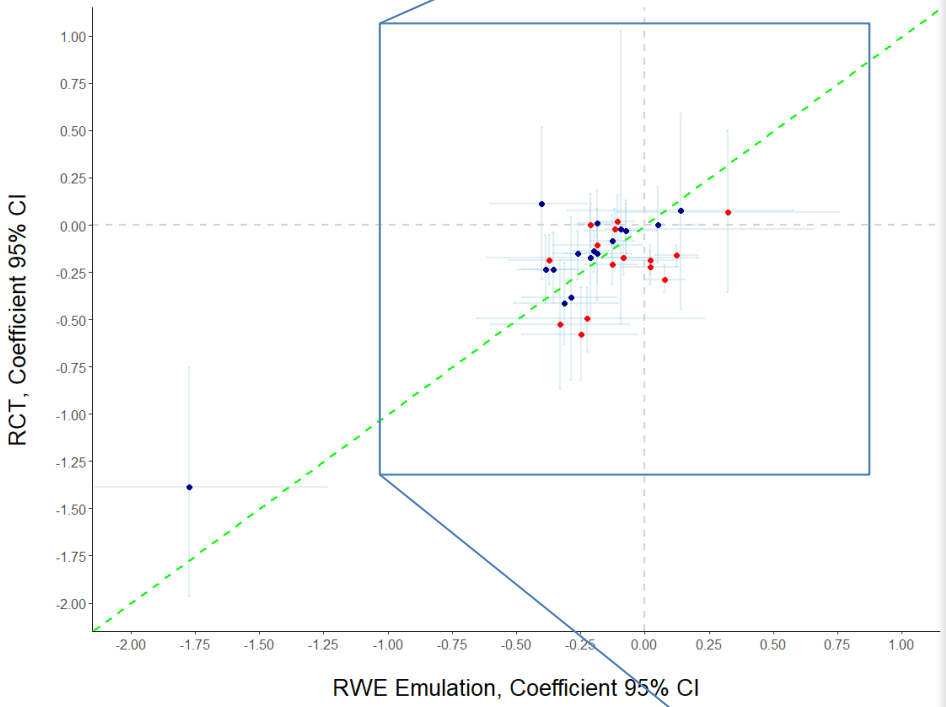
- Few emulation challenges → closer agreement in effect estimates
- More emulation challenges → less agreement in RCT/RWE effect estimates: diverge on target question/pop^{n}? Different answers may be correct.

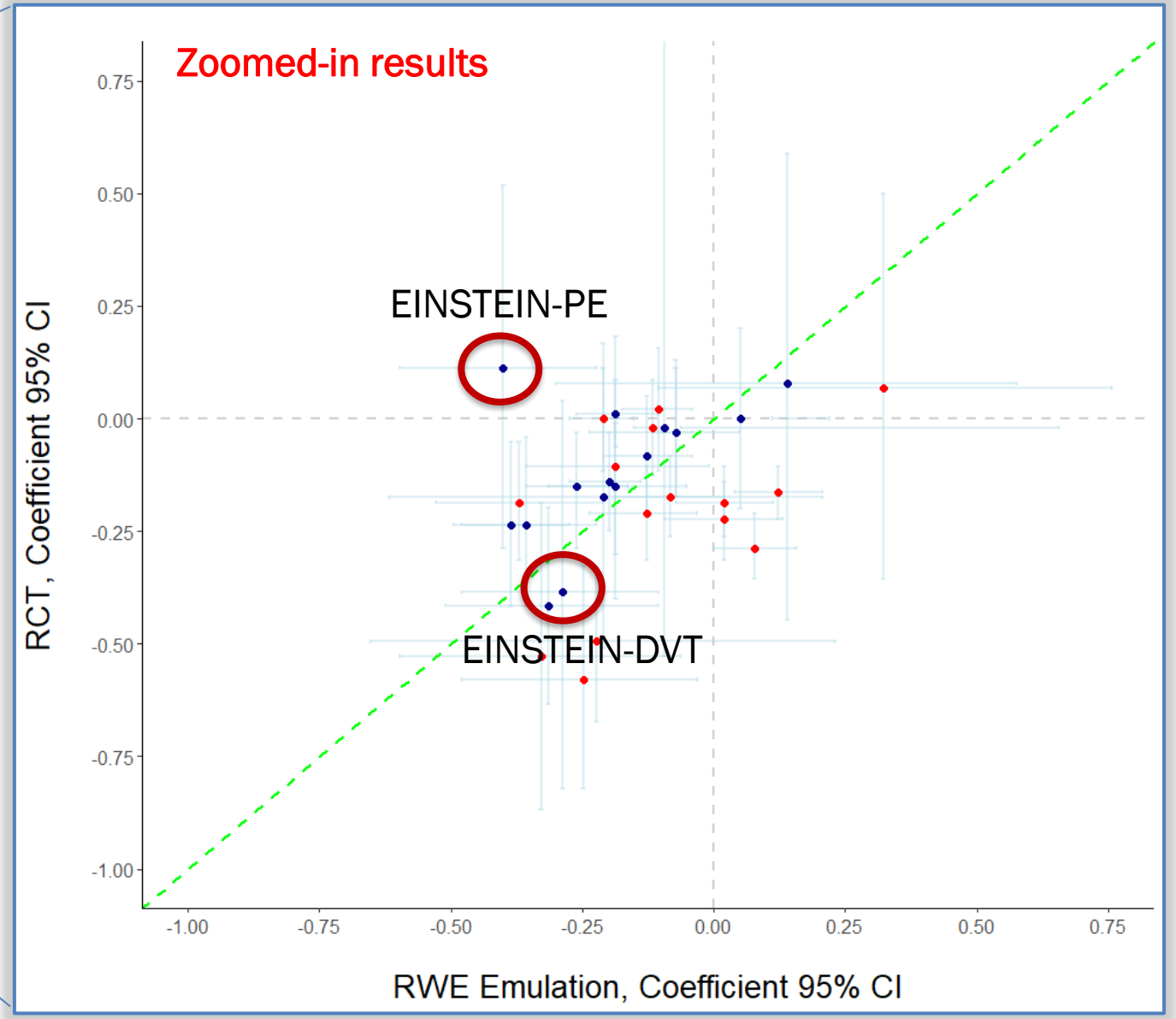
Total # trials = 32, includes 30 emulations of completed trials and 2 predictions of ongoing trials.

* Proportions represent full or partial regulatory agreement. Full agreement only was 75% vs 38%)

32 RCT-RWE emulation results

Pearson's overall = 0.80; 0.63-0.90





- Few emulation challenges
- More emulation challenges

Chance? (or other factors)

Trial name	RCT result	RWE results	Agreement		
			Statistical significance	Estimate	Standardized difference
EINSTEIN-DVT	**0.68** (0.44, 1.04)	**0.75** (0.63, 0.89)	Partial	Yes	Yes
EINSTEIN-PE	**1.12** (0.75, 1.68)	**0.68** (0.58, 0.81)	No	No	No

- Both RCTs met non-inferiority criteria
- P-value for homogeneity 0.09

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

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

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