Generating reliable insights from RWE for decision-making: highlights from RCT-DUPLICATE

Shirley V Wang Ph.D., Sebastian Schneeweiss, M.D., Sc.D.
Division of Pharmacoepidemiology and Pharmacoeconomics
Department of Medicine, Brigham and Women's Hospital,
Harvard Medical School, Boston

Disclosures

- The work to be discussed today was funded by:
  - FDA (contracts HHSF223201710186C and HHSF223201710146C)
- I am also principal investigator on other grants from FDA and NIH (NHLBI, NIA, NICHD)

Full Summary

https://healthpolicy.duke.edu/events/findings-duplicate-demonstration-project

RCT-DUPLICATE: A demonstration project

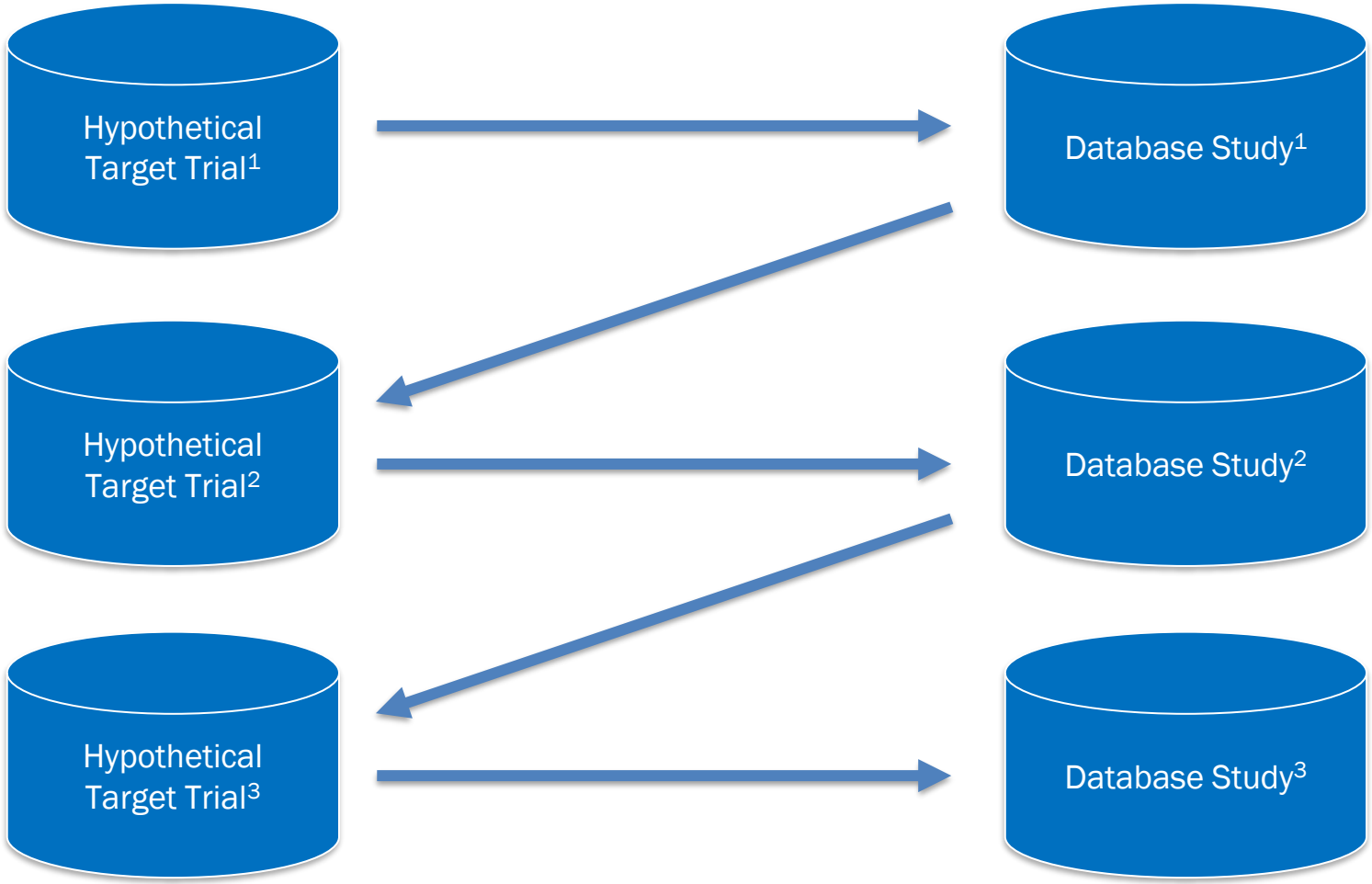
A family of studies aimed to understand and improve the validity of RWE studies for regulatory decision making

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 we replaced an RCT with a single similarly- designed 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

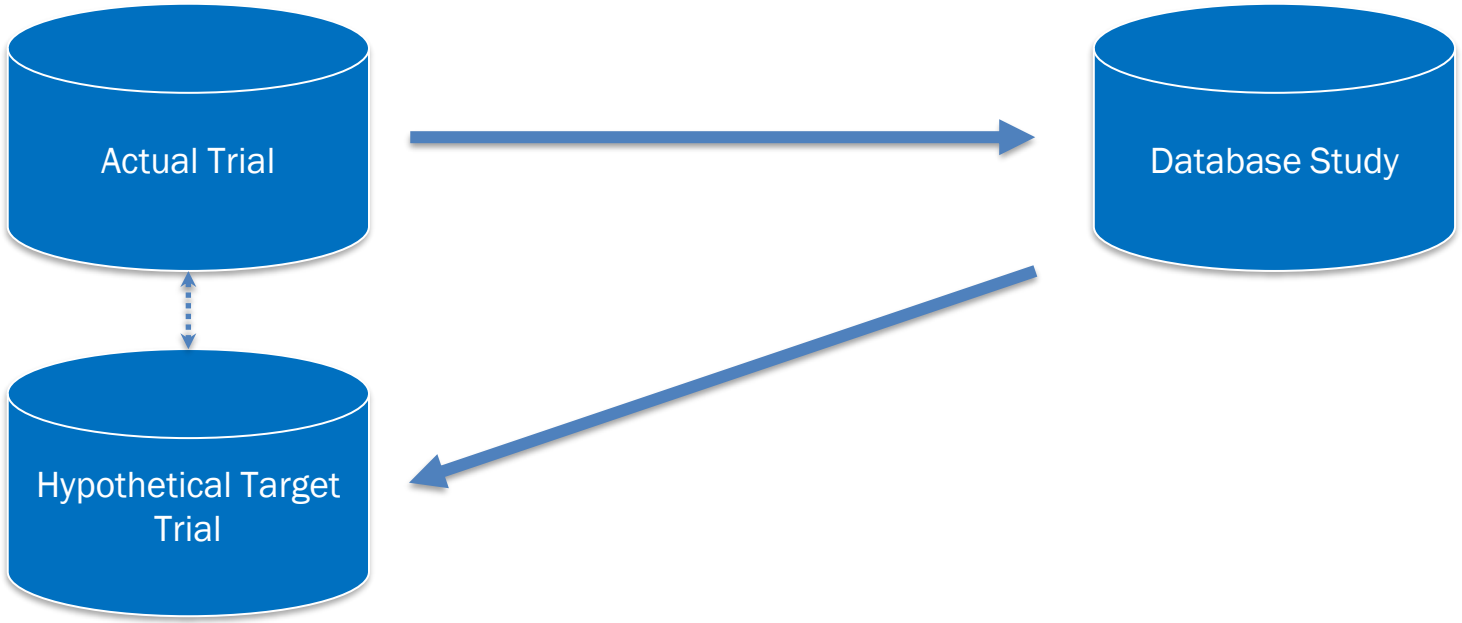
Designing a database study to mimic a hypothetical trial

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



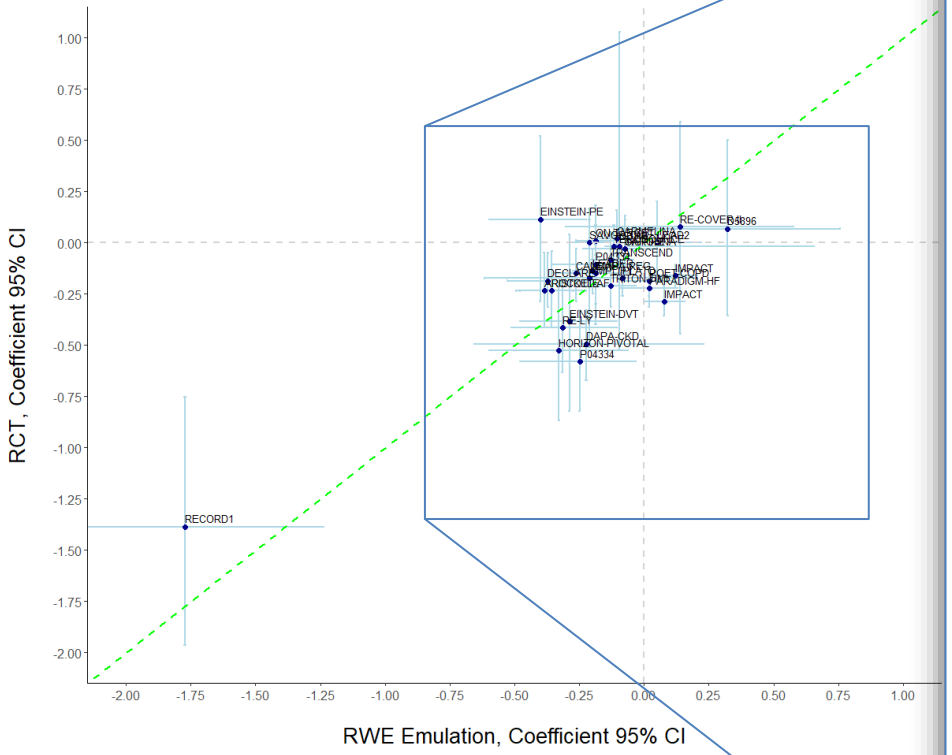
Emulation of actual RCTs as reference standard

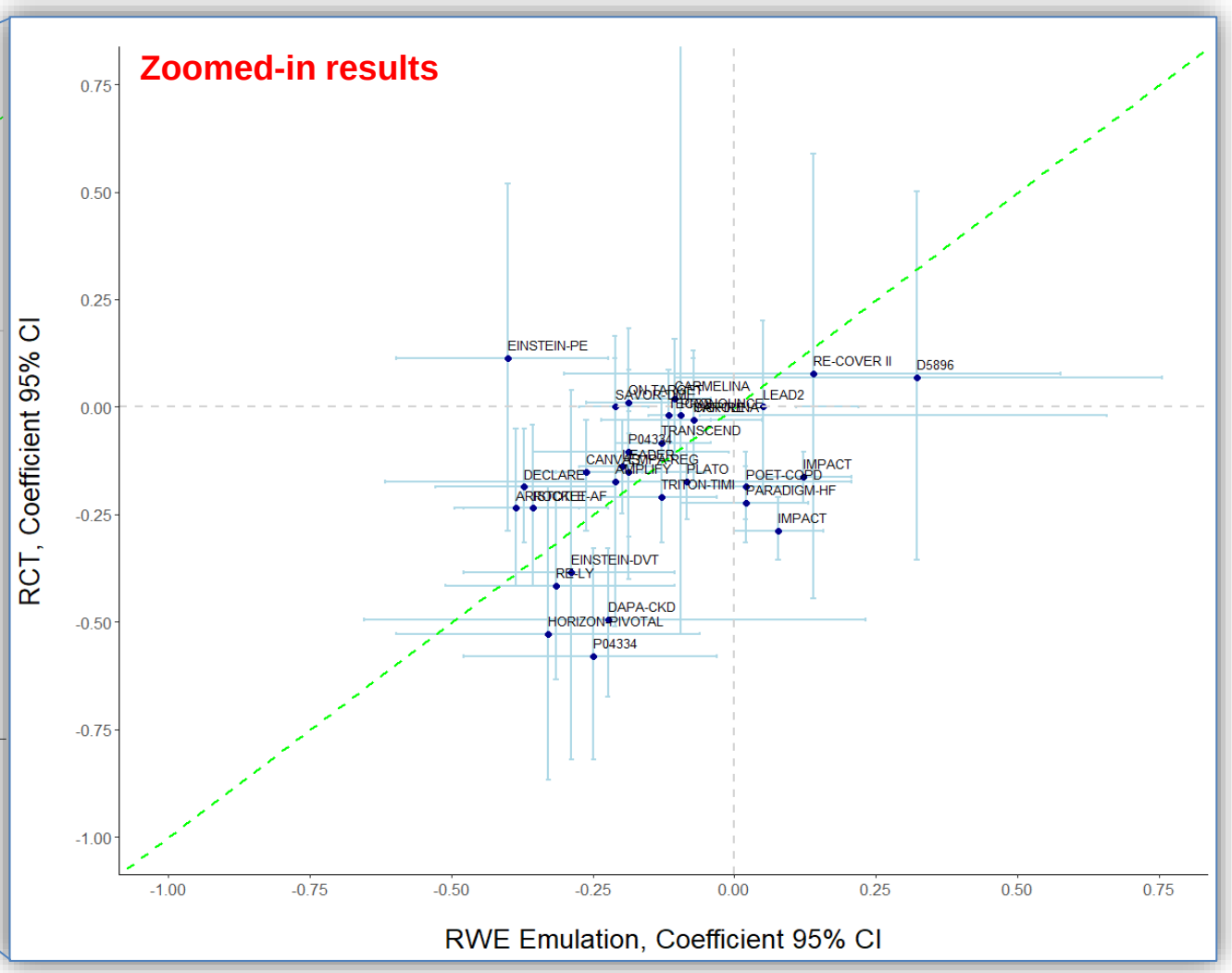
Hypothetical target trial ≈ Actual published trial



Calibration RCT vs RWE

Pearson's overall = 0.80; 0.62-0.90





Bias vs Emulation Differences

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

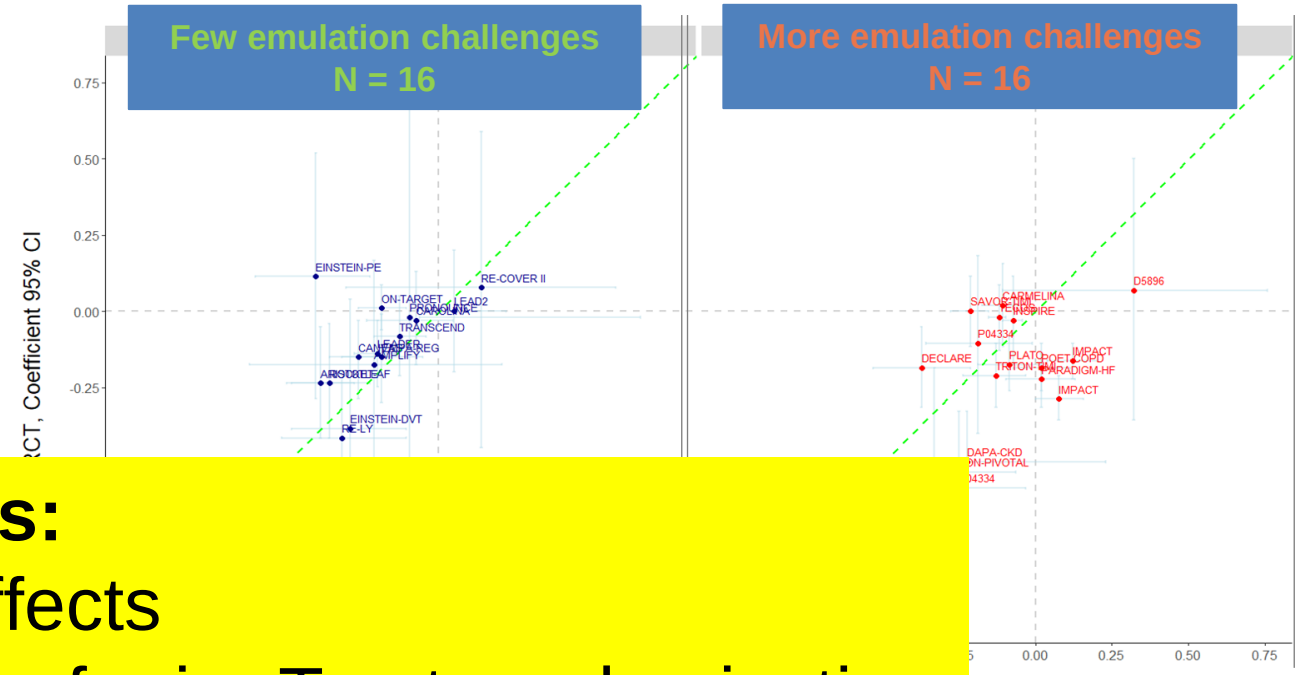
More emulation challenges = a OR b OR c OR d OR poor comparator emulation OR neither comparator and outcome emulation are classified as good

RCT-DUPLICATE findings of 32 RCT emulations

Pearson's overall = 0.80; 0.63-0.90

	Few emulation challenges N = 16	More emulation challenges N = 16
Pearson's	0.93 (0.80, 0.98)	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%)	
SD	14 (88%)	

ICC = intraclass correlation coefficie
regulatory agreement; EA = estimate
agreement



Two case studies:

1. Time varying effects
2. Discontinuation of prior Tx at randomization
3. Chance or other factors

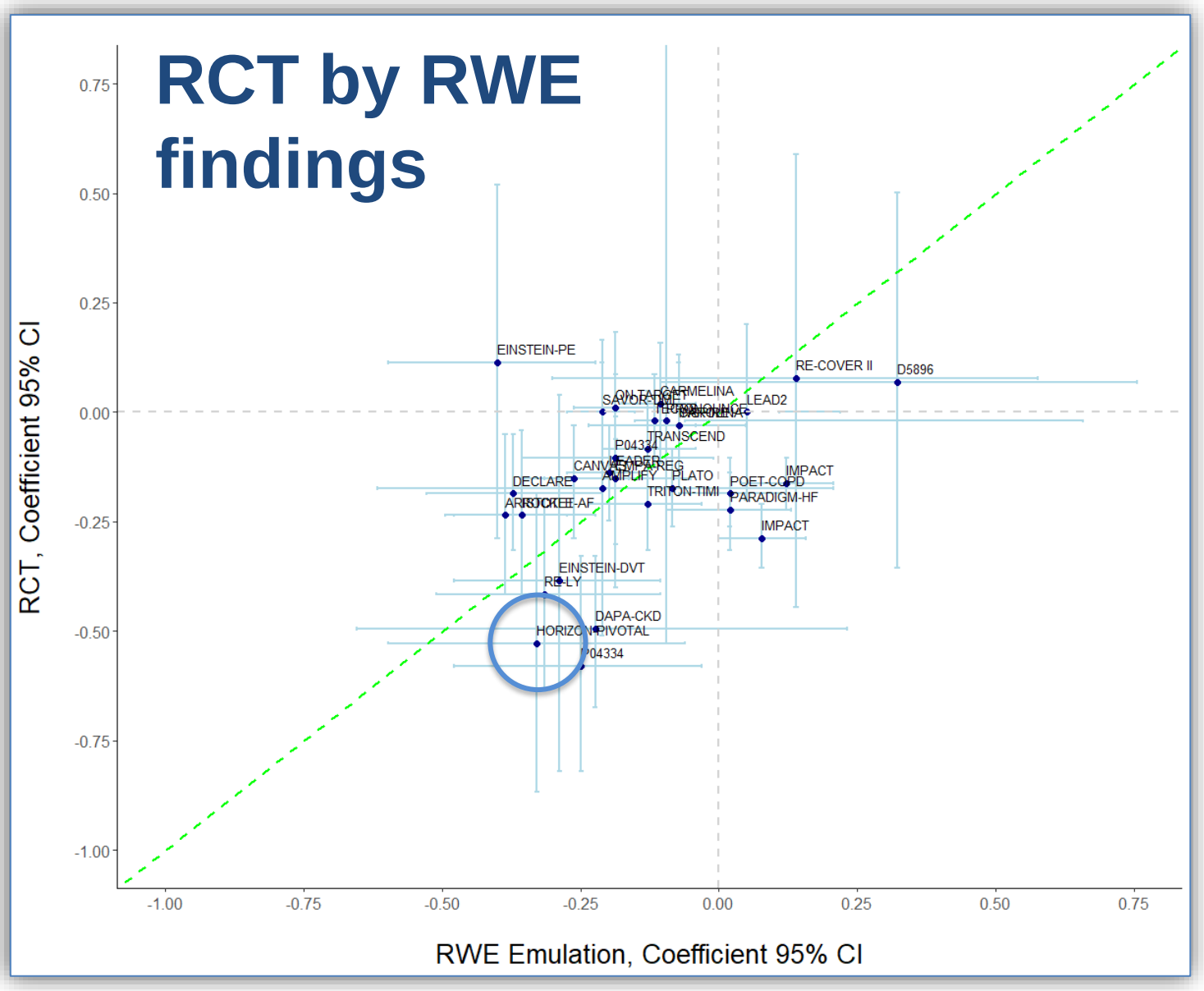
Take-home points:

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.

1. Time varying treatment effects

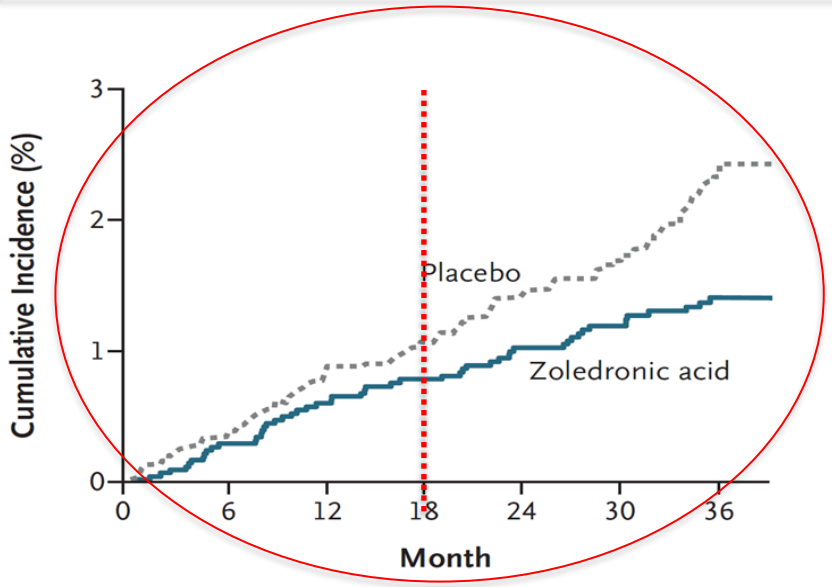


HORIZON-PIVOTAL
RCT: zoledronic acid vs **placebo**
RWE: zoledronic acid vs ***raloxifene***
Outcome: hip fracture

1. Time varying treatment effects

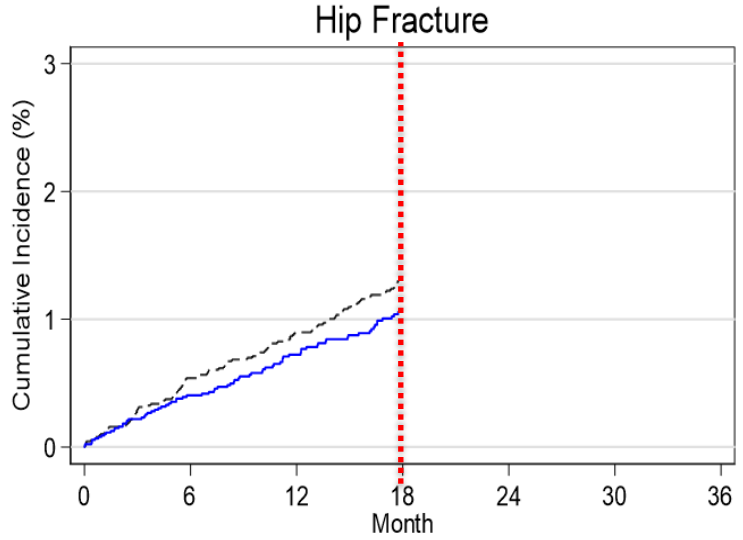
HORIZON-PIVOTAL (osteoporosis, hip fracture)

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

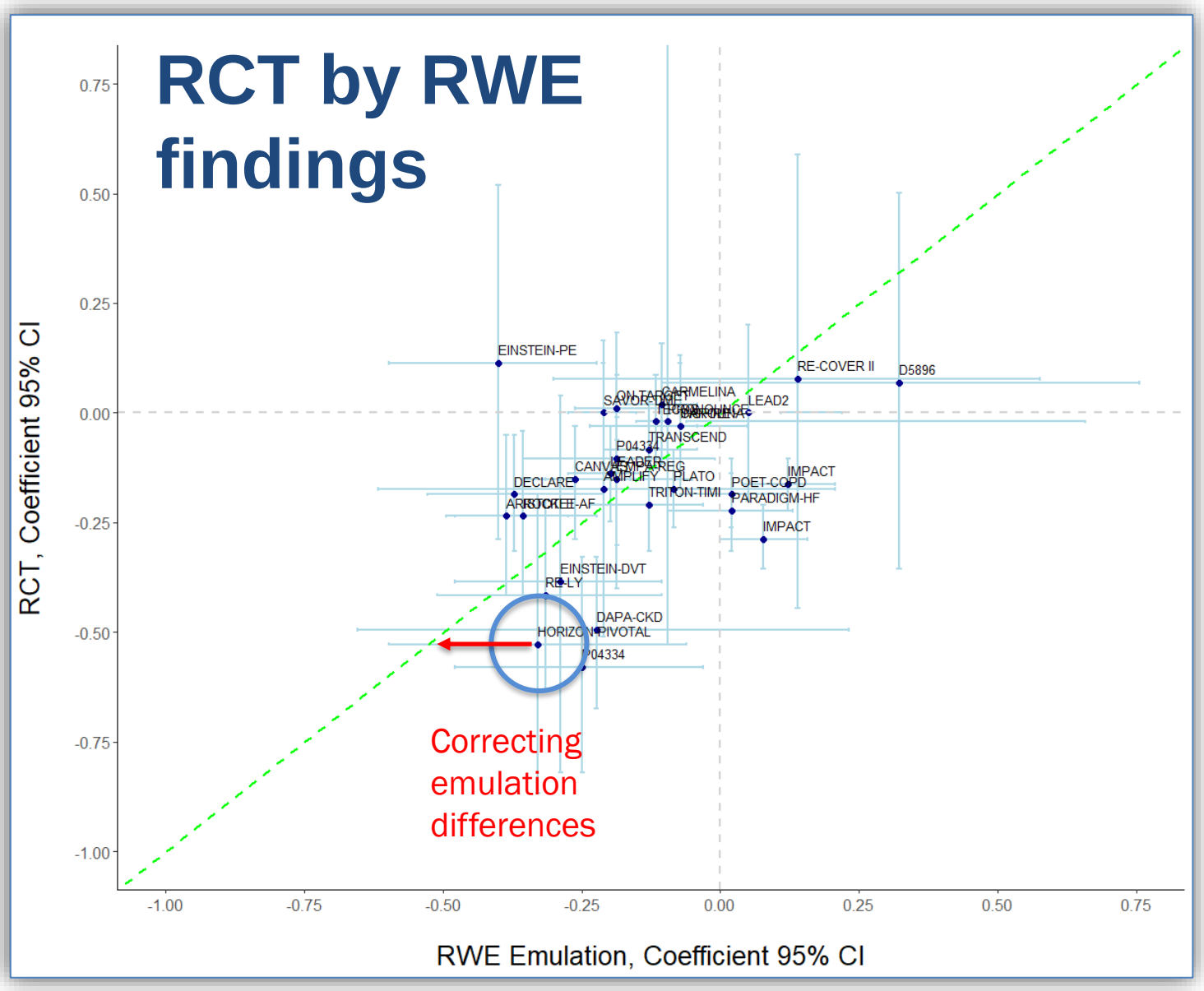
Zoledronic acid	3875	3807	3674	3553	3494	3387	3161
Placebo	3861	3806	3694	3577	3499	3397	3144



Number at risk

Raloxifene	9003	7753	6768	0	0	0	0
Zoledronic acid	9003	7766	6743	0	0	0	0

1. Time varying treatment effects



- Short time on treatment + time varying effect was emulation difference affecting Horizon Pivotal and other trials
- Correction for difference → closer calibration

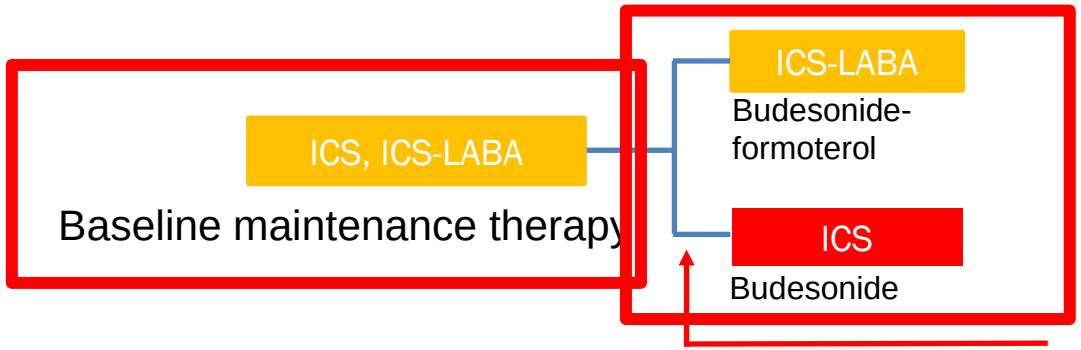
Take home points:

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

2. 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%

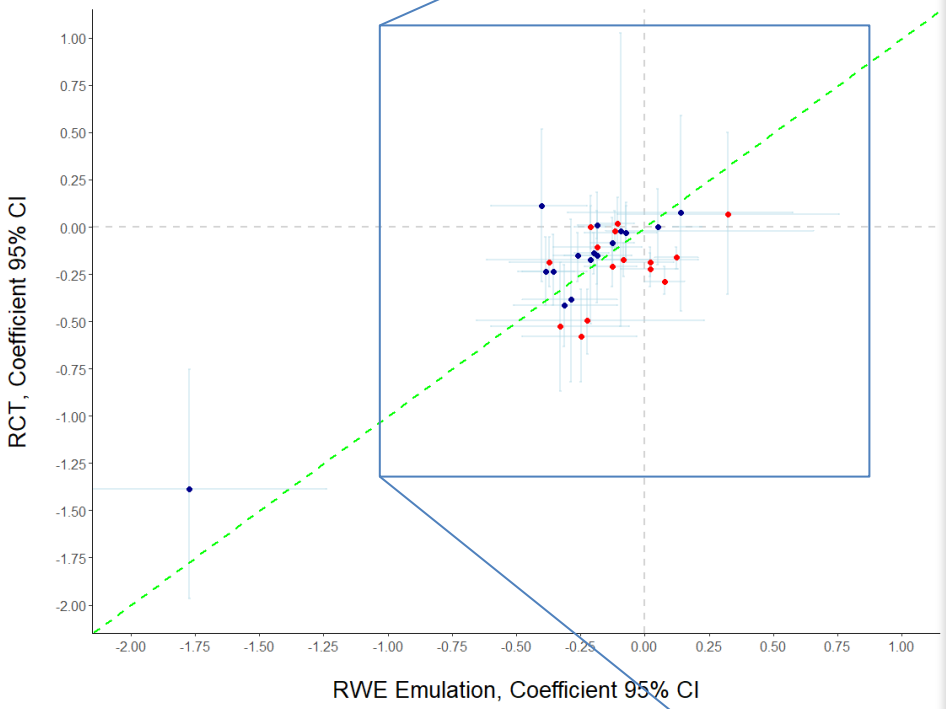
		Randomized	
		ICS-LABA	ICS
Baseline	No LABA use	29	22
	LABA use	29	22+11
		58	55

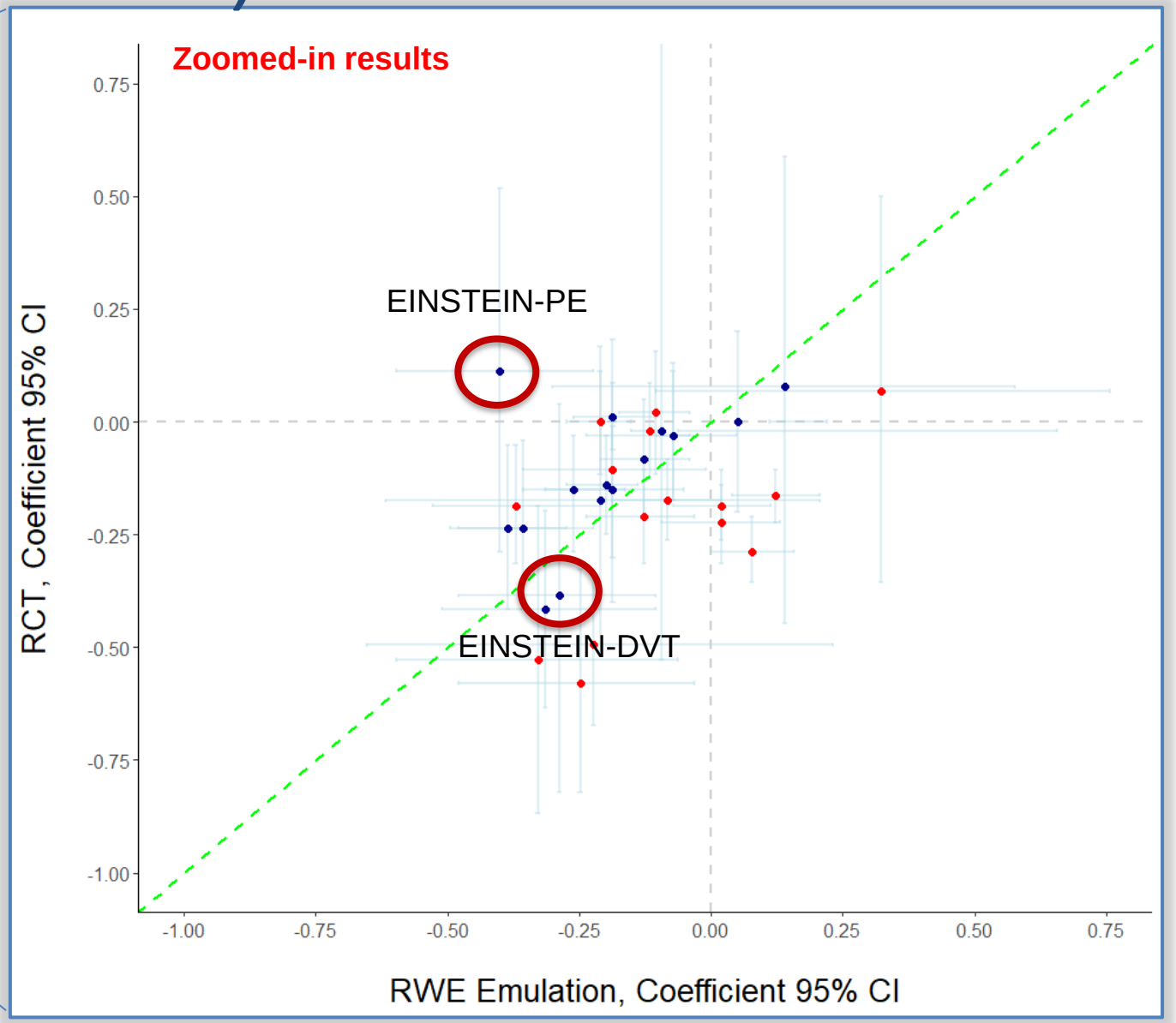
RR = 58/55 = 1.05

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

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

3. Chance? (or other factors)





- Few emulation challenges
- More emulation challenges

3. Chance? (or other factors)

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

- Both 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*

Good

Moderate

Poor

Take-home points

1. Evaluation of replicability of trial results with RWE studies requires nuance
    - Residual bias, random error
    - Efficacy vs effectiveness
    - Single trial as reference standard
2. Think about the target trial that would match the question for end users when evaluating when and how RWE studies complement RCTs (ideal vs pragmatic)

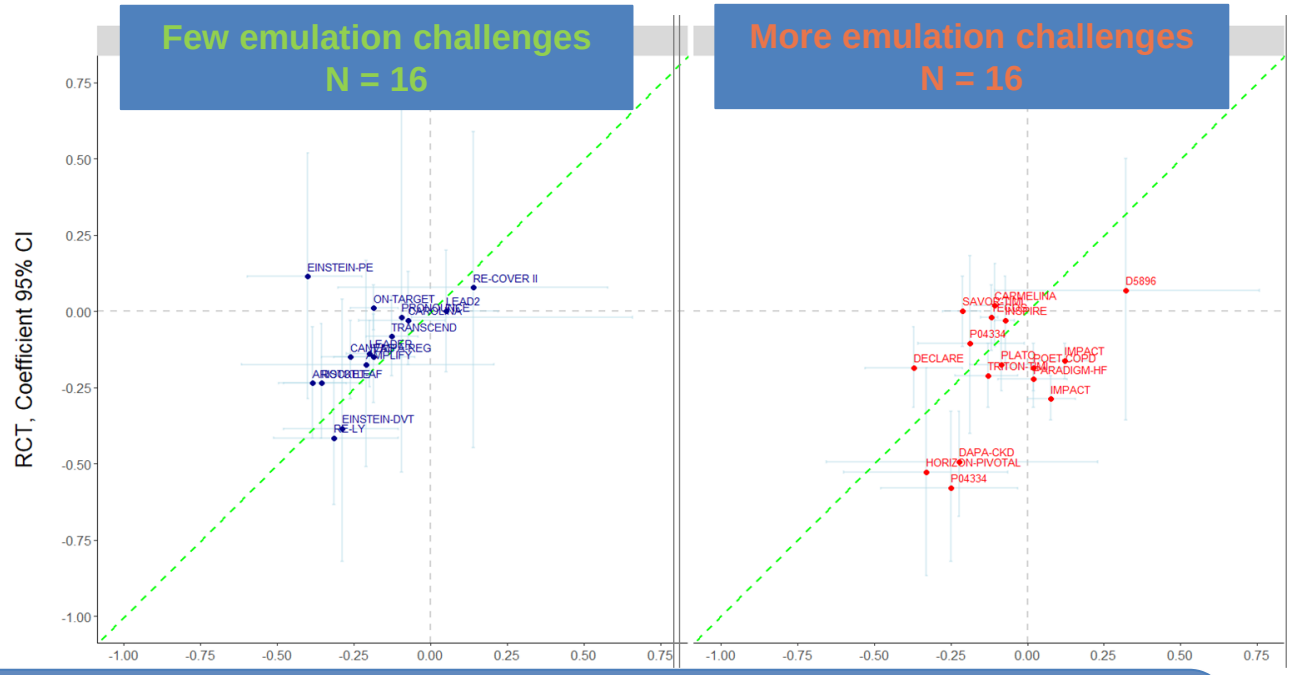
With data that are fit-for-purpose and proper design and analysis, non-randomized real-world evidence studies come to similar conclusions about a drug's treatment effect as randomized trials

RCT-DUPLICATE findings of 32 RCT emulations

Pearson's overall = 0.80; 0.63-0.90

	Few emulation challenges N = 16	More emulation challenges N = 16
Pearson's	0.93 (0.80, 0.98)	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



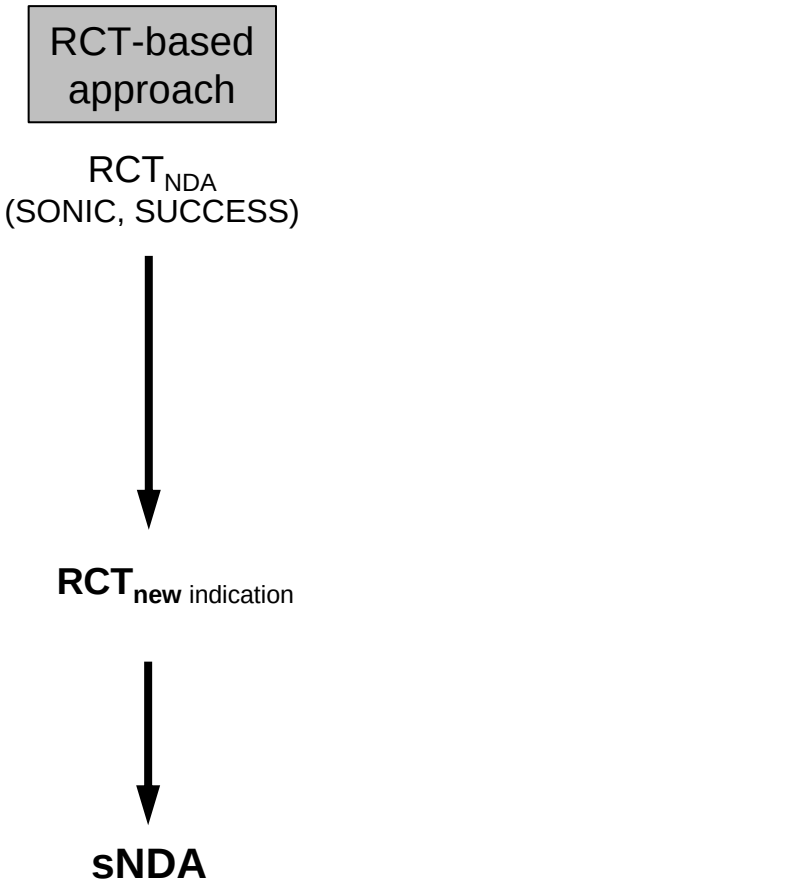
RWE studies come to the same conclusions

if they emulate an RCT design well and data are fit-for-purpose

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

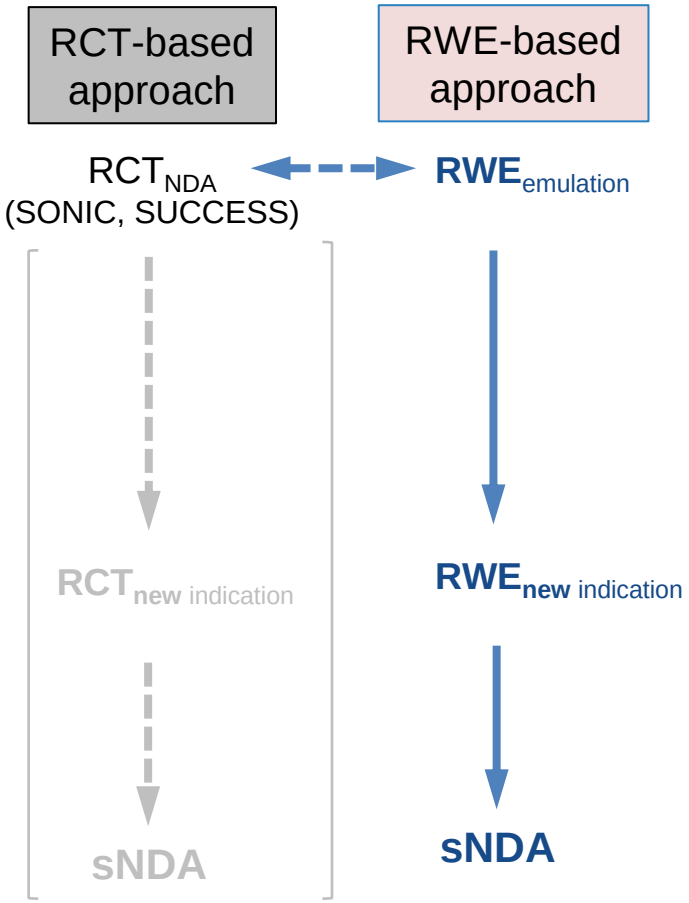
A 2-stage process to increased confidence in RWE

Typically process towards a supplemental NDA (sNDA)



A 2-stage process to increased confidence in RWE

From 2 randomized trials to a study on the effectiveness of a new combination therapy not studied in RCTs

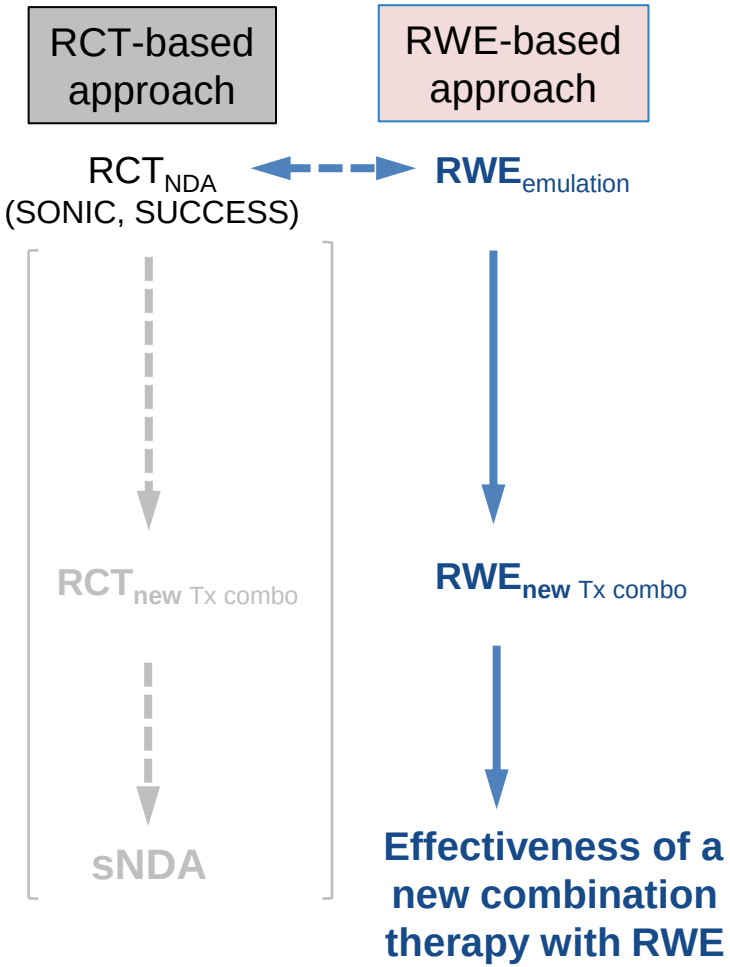


Stage 1: RWE emulation of completed RCTs to confirm the validity of the RWE approach (data and analysis)

Stage 2: RWE study of interest with boosted confidence from successful Stage 1 RCT emulations

A 2-stage process to increased confidence in RWE

From 2 randomized trials to a study on the effectiveness of a new combination therapy not studied in RCTs



Stage 1: RWE emulation of completed RCTs (SONIC, SUCCESS) to confirm the validity of the RWE approach (data and analysis)

Stage 2: RWE study of interest with boosted confidence from successful Stage 1 RCT emulations

SUCCESS emulation

Emulation of a randomized controlled trial in ulcerative colitis with US and French claims data: Infliximab with thiopurines compared to infliximab monotherapy

Julien Kirchgesner[1,2] | Rishi J. Desai[1] | Maria C. Schneeweiss[1] | Laurent Beaugerie[2] | Seoyoung C. Kim[1,3] | Sebastian Schneeweiss[1]

SONIC emulation

Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease

Julien Kirchgesner[1,2,*], Rishi J. Desai[1], Laurent Beaugerie[2], Seoyoung C. Kim[1,3] and Sebastian Schneeweiss[1]

RWE study

Decreased risk of treatment failure with vedolizumab and thiopurines combined compared with vedolizumab monotherapy in Crohn's disease

Julien Kirchgesner [1,2,3] Rishi J Desai,[3] Maria C Schneeweiss,[3] Laurent Beaugerie,[1,2] Sebastian Schneeweiss,[3] Seoyoung C Kim[3,4]

1. Kirchgesner J, Desai RJ, Schneeweiss MC, Beaugerie L, Kim SC, Schneeweiss S. Emulation of a randomized controlled trial in ulcerative colitis with US and French claims data: Infliximab with thiopurines compared to infliximab monotherapy. Pharmacoepidemiol Drug Saf. 2022 Feb;31(2):167-175.
2. Kirchgesner J, Desai RJ, Beaugerie L, Kim SC, Schneeweiss S. Calibrating Real-World Evidence Studies Against Randomized Trials: Treatment Effectiveness of Infliximab in Crohn's Disease. Clin Pharmacol Ther. 2022 Jan;111(1):179-186.
3. Kirchgesner J, Desai RJ, Schneeweiss MC, Beaugerie L, Schneeweiss S, Kim SC. Decreased risk of treatment failure with vedolizumab and thiopurines combined compared with vedolizumab monotherapy in Crohn's disease. Gut. 2022 Apr 6:gutjnl-2022-327002. doi: 10.1136/gutjnl-2022-327002. Epub ahead of print.

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, Tesfaye, Umarje, York, Zabotka, Zakoul

FDA colleagues:

Drs. Martin, Quinto, Concato, Corrigan-Curay, Paraoan

Aetion team:

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

Expert advisor panel:*

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

swang1@bwh.harvard.edu
@Shirley_Wang_V

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