External Control Arm analyses with calibration and hybrid designs

Sebastian Schneeweiss, MD, ScD
Professor of Medicine and Epidemiology, Harvard Medical School, Boston

Chief, Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine
Brigham and Women's Hospital

Nov 2024

Funding

- In parts by FDA (HHSF223201710186C and HHSF…46C)
- In parts by the FDA Sentinel Innovation Center (75F40119F19002)
- In parts by the NIH: NHLBI (R01-HL141505)
  NIAMS (R01-AR080194)
- In parts by the Burroughs Wellcome Fund
- Additional funding came from PCORI

Disclosures

- PI, Sentinel Innovation Center (FDA)
- Co-Chair, Mass General Brigham Center for Integrated Healthcare Data Research
- PI of research grants awarded to BWH by Bayer, UCB, Boehringer Ingelheim, Takeda
- Equity in Aetion, Inc.

EVERY decision maker in healthcare wants the best evidence possible

YET, most operate under constraints that make them willing to compromise

Approved drugs using RWE as “supporting evidence” (FDA)

Drug	Indication	Approval	Data
Carbaglu (carglumicacid)	NAGS deficiency	2010	Retrospective, non-random, unblinded case series of 23 patients compared to historical control group
Blincyto (Blinatumomab)	Acute Lymphoblastic Leukemia	2014	Single-arm trial, Reference group weighted analysis of patient level data on chart review of 694 patients at EU and US study sites
Omegaven (fish oil triglycerides)	Parenteral nutrition	2018	Two single-arm trials, matched to historical control arm from associated cholestasis hospital record
Ibrance (palbociclib)	Male breast cancer	2019	Data from electronic health records and post-marketing reports of the real-world use of IBRANCE in male patients
Voxzogo (vosoritide)	Achondroplasia	2021	Observational, retrospective AchNH registry served as external patients 2+ years control to two small supportive Phase II studies
Orencia (abatacept)	Prophylaxis of acute graft versus host disease	2021	Registry-based clinical study using real world data from the Center for International Blood & Bone Marrow Transplant Research

Concato, Currigan-Curay NEJM 2022

Prograf: RWE as “substantial evidence” (FDA)

Drug	Indication	Approval	Data
Carbaglu (carglumicacid)	NAGS deficiency	2010	Retrospective, non-random, unblinded case series of 23 patients compared to historical control group
Blincyto (Blinatumomab)	Acute Lymphoblastic Leukemia	2014	Single-arm trial, Reference group weighted analysis of patient level data on chart review of 694 patients at EU and US study sites
Omegaven (fish oil triglycerides)	Parenteral nutrition	2018	Two single-arm trials, matched to historical control arm from associated cholestasis hospital record
Ibrance (palbociclib)	Male breast cancer	2019	Data from electronic health records and post-marketing reports of the real-world use of IBRANCE in male patients
Voxzogo (vosoritide)	Achondroplasia	2021	Observational, retrospective AchNH registry served as external patients 2+ years control to two small supportive Phase II studies
Orencia (abatacept)	Prophylaxis of acute graft versus host disease	2021	Registry-based clinical study using real world data from the Center for International Blood & Bone Marrow Transplant Research
Prograf (Tacrolimus)	Prevention of organ rejection after allogenic lung transplantation	2021	Non-interventional treatment arm vs. historical controls using registry of transplant recipients

Concato, Currigan-Curay NEJM 2022

FDA self-described RWE use case for effectiveness claims

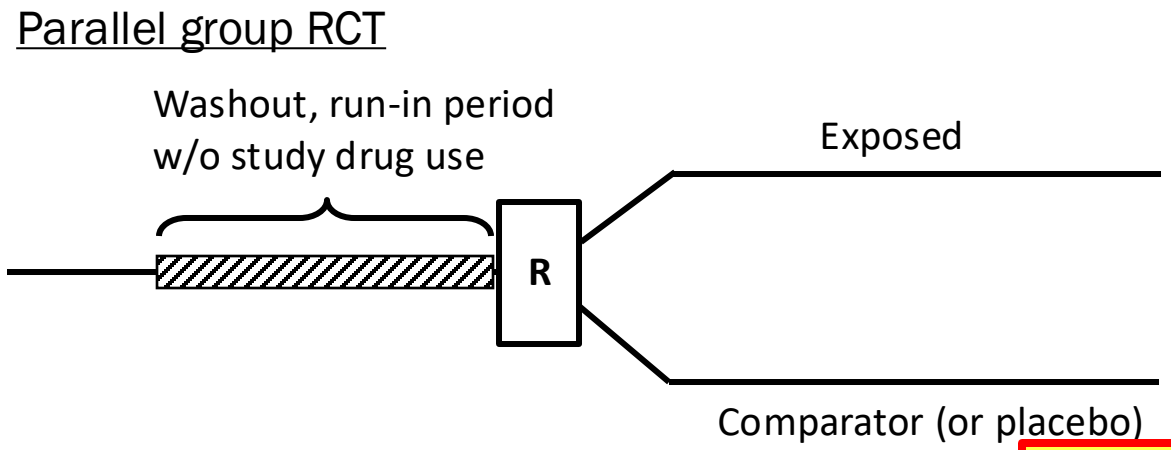
When can real-world data generate real-world evidence?

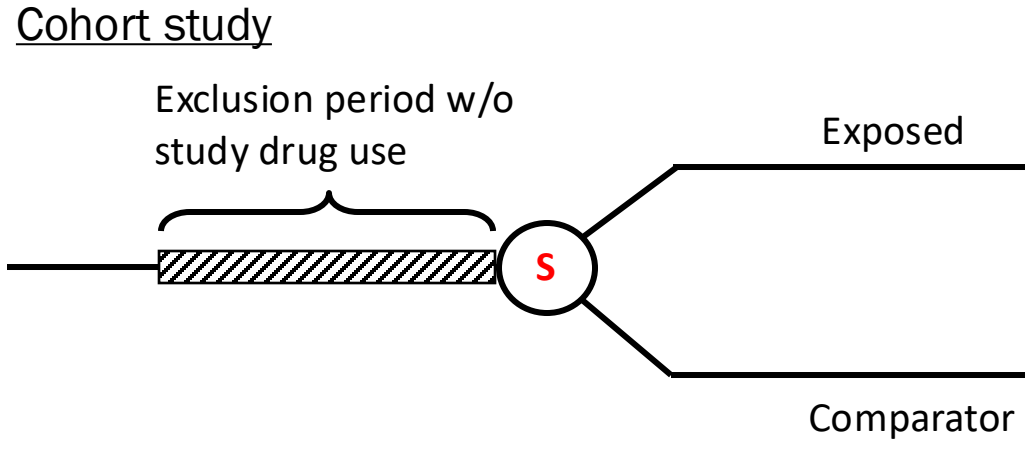
Motiur Rahman[1] | Gerald Dal Pan[2] | Peter Stein[3] | Mark Levenson[4] | Stefanie Kraus[5] | Aloka Chakravarty[6] | Donna R. Rivera[7] | Richard Forshee[8] | John Concato[1,9]

- Purposes of Using RWD/RWE as Part of the Submission (select all that apply)
  - To support safety and/or effectiveness for a product not previously approved by FDA
  - To support labeling changes for an approved product, including
    - Add or modify an indication
    - Change dose, dose regimen, or route of administration
    - Expand the labeled indication of the product to a new population
    - Add comparative effectiveness information
    - Add or modify safety information
    - Other labeling change—specify
  - To support or satisfy a PMR/PMC

Pharmacoepi Drug Safety 2023

Causal study designs: Contemplate the target trial





Key difference to RCT #1:
No baseline randomization

Key difference to RCT #2:
Secondary data

FDA submission using an external control arm:

Blincyto (blinatumomab) for treatment of adult relapsed/refractory acute lymphoblastic leukemia (ALL)

GOAL

In patients with ALL obtain a relative measurement of complete remission and overall survival between blinatumomab and standard of care.

SINGLE ARM TRIAL

189 patients with Ph-negative, B-precursor R/R ALL enrolled in the single-arm trial exposed to blinatumomab.

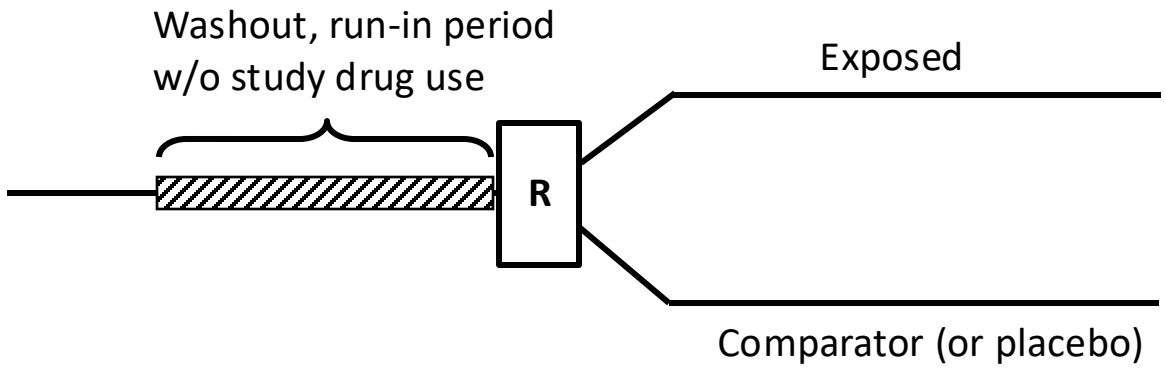
EXTERNAL CONTROL ARM

Individual patients from six national study groups and and five large treatment centers with Ph-negative, B-precursor R/R ALL treated with standard of care.

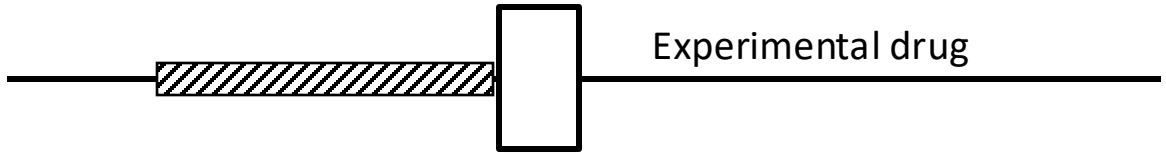
Sources: FDA's NDA/BLA statistical review
Gökbuget N, Kelsh M, Chia V, et al. Blinatumomab vs historical standard therapy of adult relapsed/refractory acute lymphoblastic leukemia. Blood Cancer J. 2016

Causal study designs: Contemplate the target trial

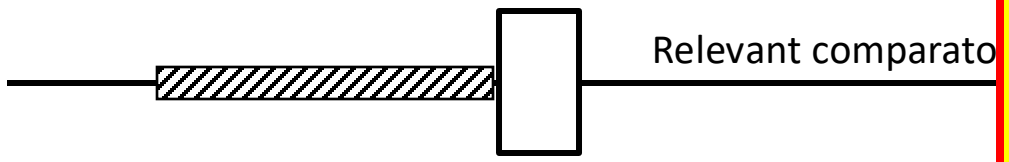
Parallel group RCT



Single arm trial (SAT): Always primary data



External ctrl arm (ECA): Usually secondary data



Key difference to RCT #3: Differential data source

Ability to measure key covariates and outcomes

VARIABLE	EXPERIMENTAL ARM	EXTERNAL CONTROL ARM
Age	✓	✓
Gender	✓	✓
Duration between initial diagnosis and salvage therapy	✓	✓
Region	✓	✓
Prior allogeneic stem cell transplant	✓	✓
Prior number of salvage therapies	✓	✓
Primary refractory and in first salvage therapy	✓	✓
Refractory to last salvage therapy	✓	✓

Source: FDA's NDA/BLA statistical review; Gokbuget et al.

OUTCOME	EXPERIMENTAL ARM	EXTERNAL CONTROL ARM
Complete remission(CR)	✓	✓
Overall Survival (OS)	✓	✓

"Per medical reviewer, the 2 stratification factors (age and prior lines of treatment) are the two most important factors that could be related to the CR outcome." **- FDA Statistical Review**

RWD to build external control groups

FDA U.S. FOOD & DRUG ADMINISTRATION

ODAC Briefing Document

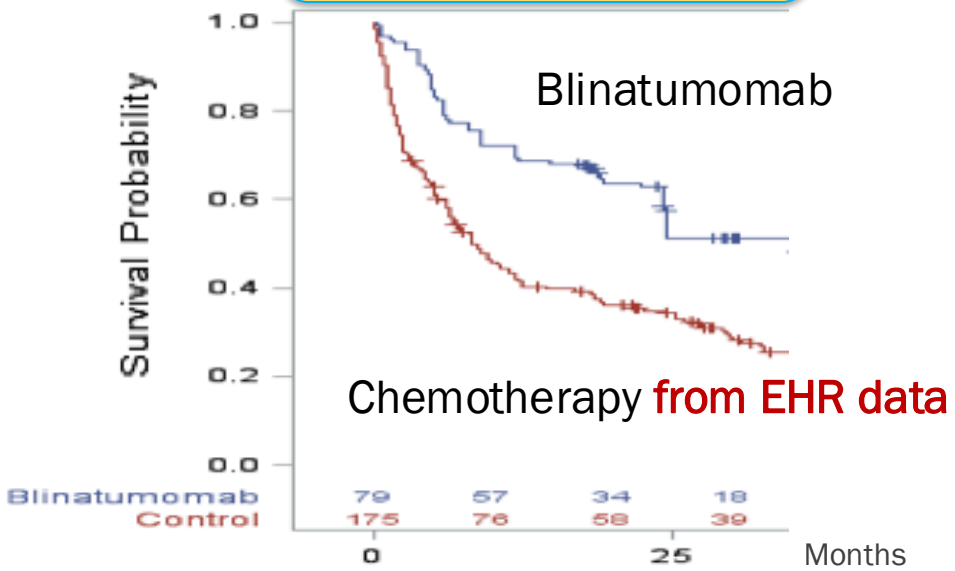
PS-IPTW weighted studies
MT103-203 (TRT) and
20120148 (CTRL)

Blood Cancer Journal

Blinatumomab vs historical standard therapy of adult relapsed/ refractory acute lymphoblastic leukemia

N Gökbuget[1], M Kelsh[2], V Chia[2], A Advani[3], R Bassan[4], H Dombret[5], M Doubek[6], AK Fielding[7], S Giebel[8], V Haddad[9], D Hoelzer[1], C Holland[10], N Ifrah[11], A Katz[2], T Maniar[12], G Martinelli[13], M Morgades[14], S O'Brien[15], J-M Ribera[14], JM Rowe[16], A Stein[17], M Topp[18], M Wadleigh[19] and H Kantarjian[15]

Overall survival
HR = 0.61 (0.40-0.94)



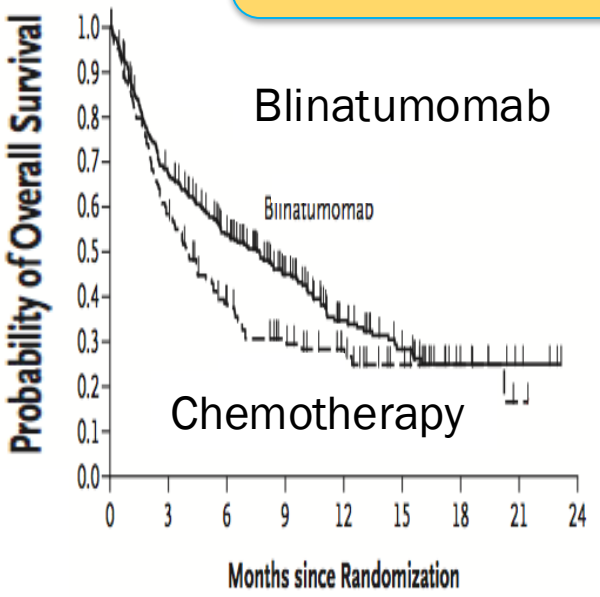
ORIGINAL ARTICLE

Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia

TOWER

Hagop Kantarjian, M... Gökbuget, M.D.,

Overall survival
HR = 0.71 (0.55-0.93)

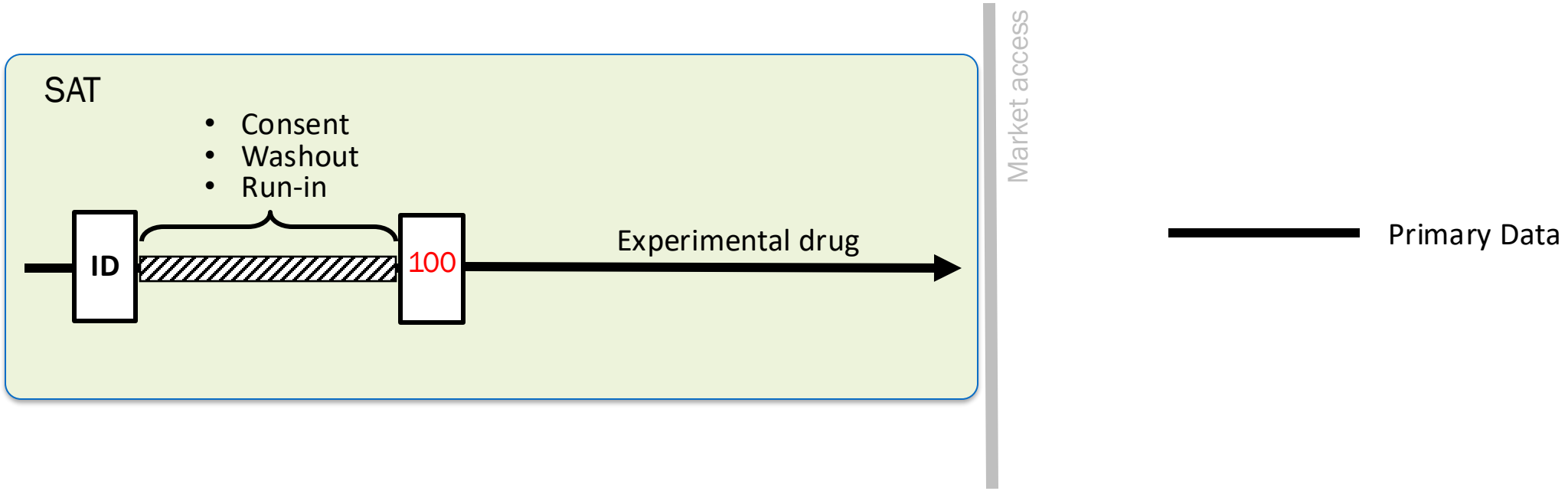


Why was this a successful use case for an ECA?

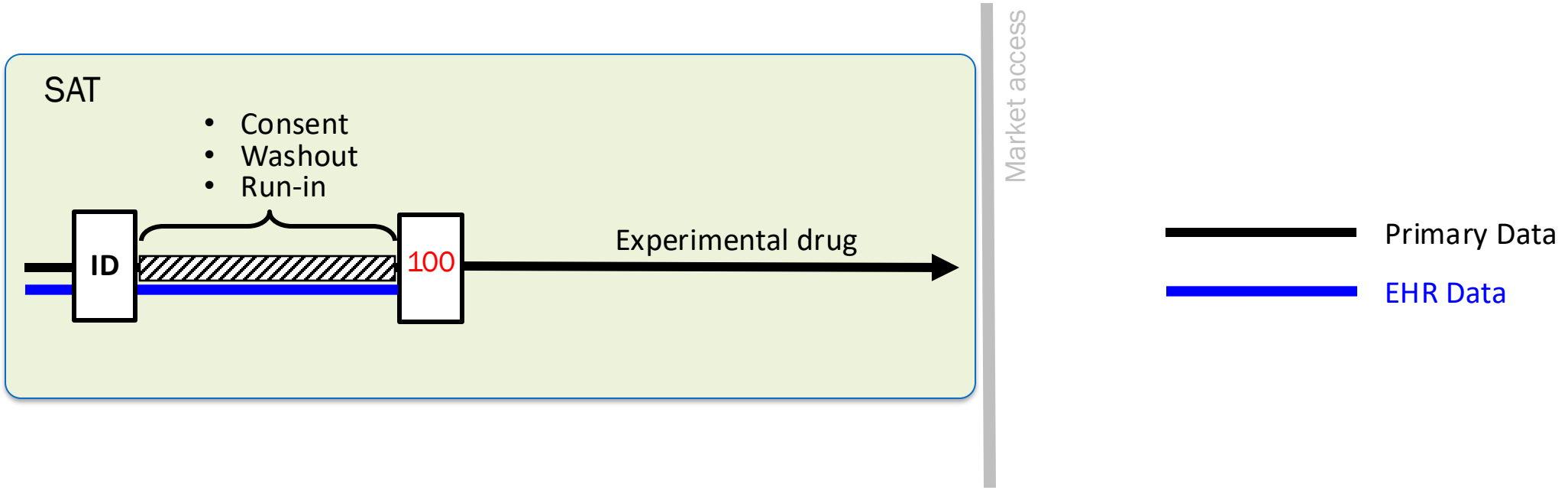
Well-defined natural history	✓ ALL is well understood
Objective endpoint	✓ **Well-defined outcomes** by standard criteria for CR as bone marrow blasts < 5%; overall survival
Patient comparability	✓ Drew a very large sample of patients (~2,400) and was able to **closely mimic trial eligibility criteria**
Good confounder measurement	✓ **Key confounders measured**: Early engagement with FDA resulted in confirmation of confounder adjustment
Large effect size	✓ ECA demonstrated a **substantial benefit** in CR: (SAT CR **42%** [34-49%]; ECA CR **24%** [20-27%])

Sources: FDA's Demonstrating Substantial Evidence of Effectiveness for Human Drug and Biological Products Guidance for Industry *and* Guidance for Industry on Choice of Control Group and Related Issues in Clinical Trials. EMA's ICH Topic E. 10 Choice of Control Group in Clinical Trials.

External control arms (ECAs) for single arm trials



ECA calibration design



EHR data are often incomplete

Fitness of real-world data for clinical trial data collection: Results and lessons from a HARMONY Outcomes ancillary study

Bradley G Hammill[1,2], Jeffrey D Leimberger[1], Zachary Lampron[1], Sudha R Raman[2], Emily C O'Brien[1,2], Keele E Wurst[3], Sally Mountcastle[4], Marianne Cunnington[3], Salim Janmohamed[5] and Lesley H Curtis[1,2]

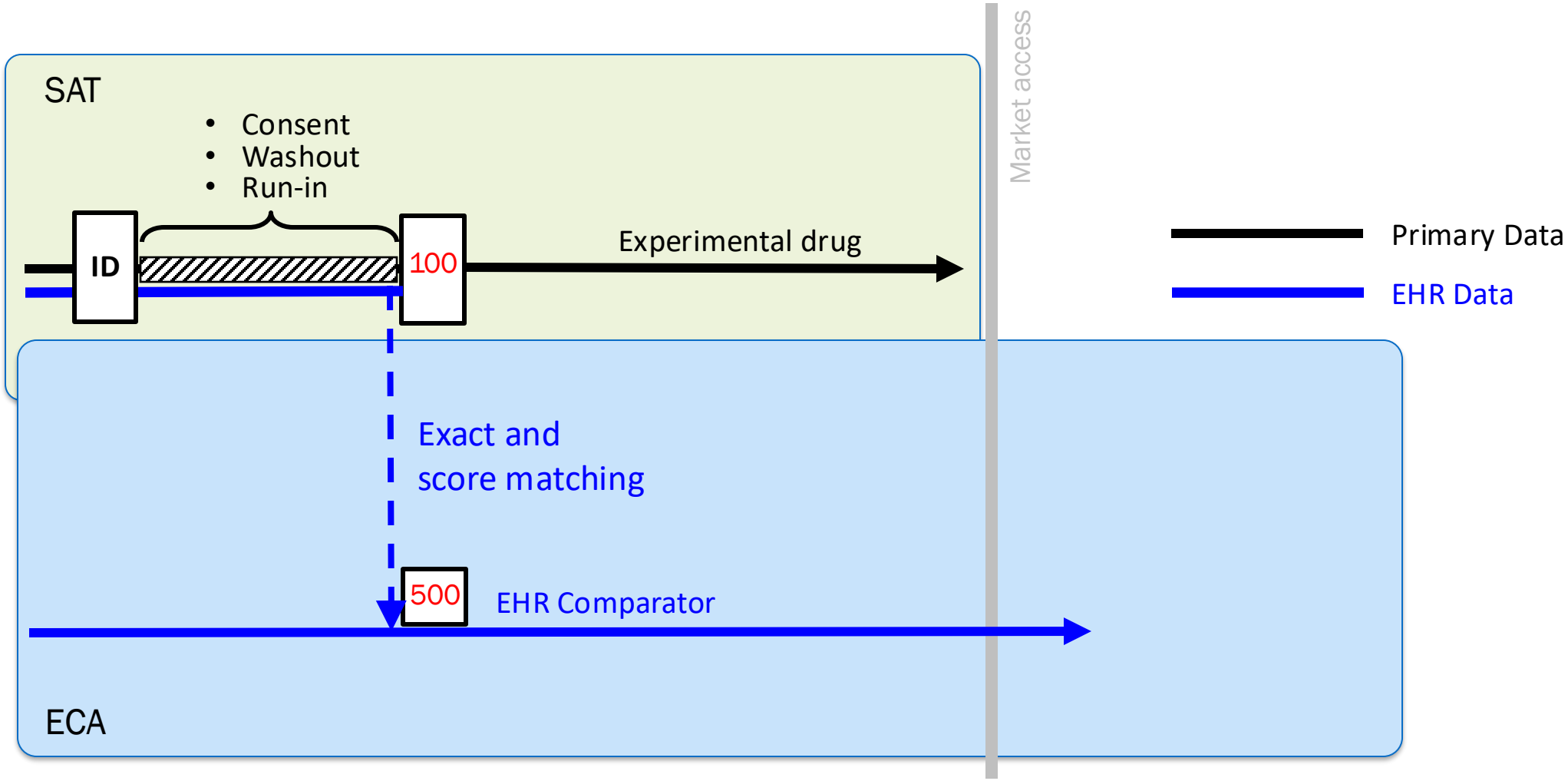
Measure	Sensitivity	Specificity
Demographics		
Female sex	100 (79–100)	100 (93–100)
White race	80 (67–90)	100 (77–100)
African American race	92 (62–100)	100 (94–100)
Hispanic ethnicity*	0 (0–84)	79 (67–88)
Medical history		
Myocardial infarction	24 (11–40)	94 (79–99)
Coronary artery disease	49 (36–63)	80 (44–98)
Stroke	44 (14–79)	100 (94–100)
Transient ischemic attack	0 (0–84)	100 (95–100)
Carotid artery disease	0 (0–71)	99 (92–100)
Heart failure	53 (27–79)	96 (87–100)
Valvular heart disease	67 (9–99)	94 (85–98)
Atrial fibrillation	56 (21–86)	97 (89–100)
Hypertension	55 (42–67)	100 (48–100)
Hyperlipidemia	49 (36–62)	50 (16–84)
Diabetic eye disease	50 (12–88)	95 (87–99)
Diabetic neuropathy	41 (24–61)	83 (67–93)
Medications		
ACE inhibitor	32 (18–48)	100 (88–100)
Angiotensin receptor blocker	33 (13–59)	96 (87–100)
P2Y12 inhibitor	52 (31–72)	100 (92–100)
Anti-hyperglycemic medication	46 (34–59)	NA
DPP-IV inhibitors	33 (8–70)	83 (72–92)

Table 2. Performance of baseline characteristics derived from EHR data compared to the trial database.

Many standard EHR systems are not capturing what we measure in RCTs

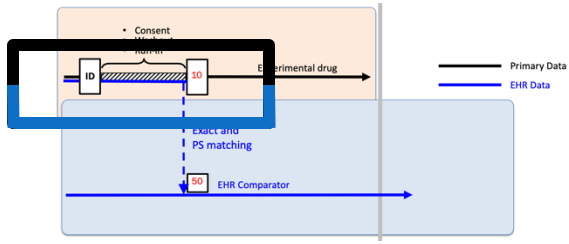
Note:
Sensitivity <1 means the EHR data source misses medications dispensed to patients.

ECA calibration design



SAT = single arm trial; ECA = external control arm; ICA = internal control arm;
EHR = electronic health record; RWE = Real-world evidence study

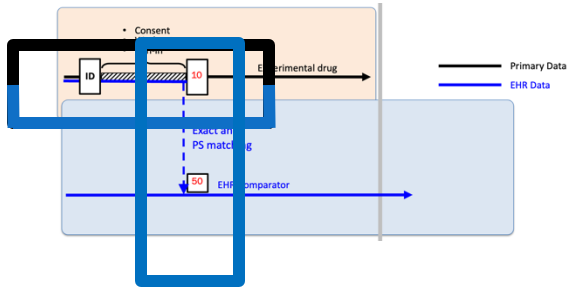
Measurement calibration in practice



	These are the **same patients** in the SAT	
Pre-exposure characteristics	Primary data from **SAT**	EHR data from **SAT**
N (hypothetical)	100	100
Severity marker A (mg) % missing	100 mg 1%	80 mg 30%
Severity marker B (%)	50%	30%
Severity marker C (%)	20%	15%
Comorbidity A (%)	30%	40%
Comorbidity B (%)	20%	10%
Medication A (%)	20%	15%
Medication B (%)	20%	10%

SAT = Single Arm Trial; ECA = External Control Arm
Note: This table only addresses the measurement issues and not unobserved confounding

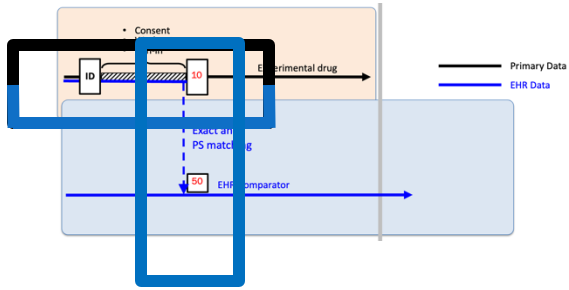
Measurement calibration in practice



Pre-exposure characteristics	Primary data from **SAT**	EHR data from **SAT**	EHR data from ***ECA***
	These are the **same patients** in the SAT		
N (hypothetical)	100	100	500
Severity marker A (mg) % missing	100 mg 1%	80 mg 30%	80 mg 30%
Severity marker B (%)	50%	30%	30%
Severity marker C (%)	20%	15%	15%
Comorbidity A (%)	30%	40%	40%
Comorbidity B (%)	20%	10%	10%
Medication A (%)	20%	15%	15%
Medication B (%)	20%	10%	10%
		These are **equal patients** in SAT and ECA, not same	

SAT = Single Arm Trial; ECA = External Control Arm
Note: This table only addresses the measurement issues and not unobserved confounding

Measurement calibration in practice



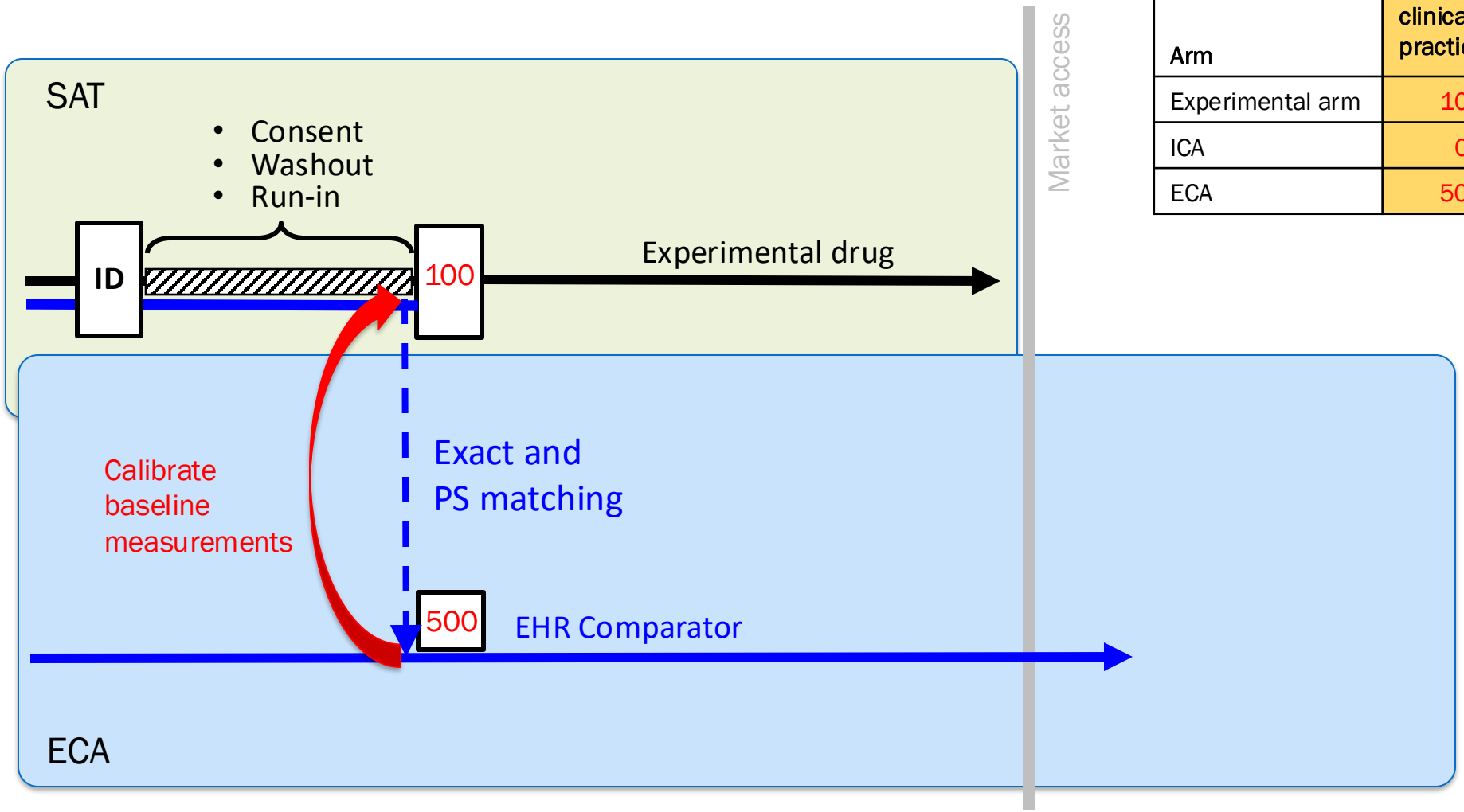
Pre-exposure characteristics	These are the **same patients** in the SAT: Primary data from **SAT**	EHR data from **SAT**	EHR data from ***ECA***
N (hypothetical)	100	100	500
Severity marker A (mg) % missing	100 mg 1%	80 mg 30%	80 mg 30%
Severity marker B (%)	50%	30%	30%
Severity marker C (%)	20%	15%	15%
Comorbidity A (%)	30%	40%	40%
Comorbidity B (%)	20%	10%	10%
Medication A (%)	20%	15%	15%
Medication B (%)	20%	10%	10%
		These are **equal patients** in SAT and ECA, not same	

SAT = Single Arm Trial; ECA = External Control Arm
Note: This table only addresses the measurement issues and not unobserved confounding

ECA calibration design

Table: Distribution of patients in each arm (hypothetical)

Arm	Calibration RCT extra-polation to clinical practice	Hybrid Efficient trial design	E2E E2E evidence development system
Experimental arm	100		
ICA	0		
ECA	500		



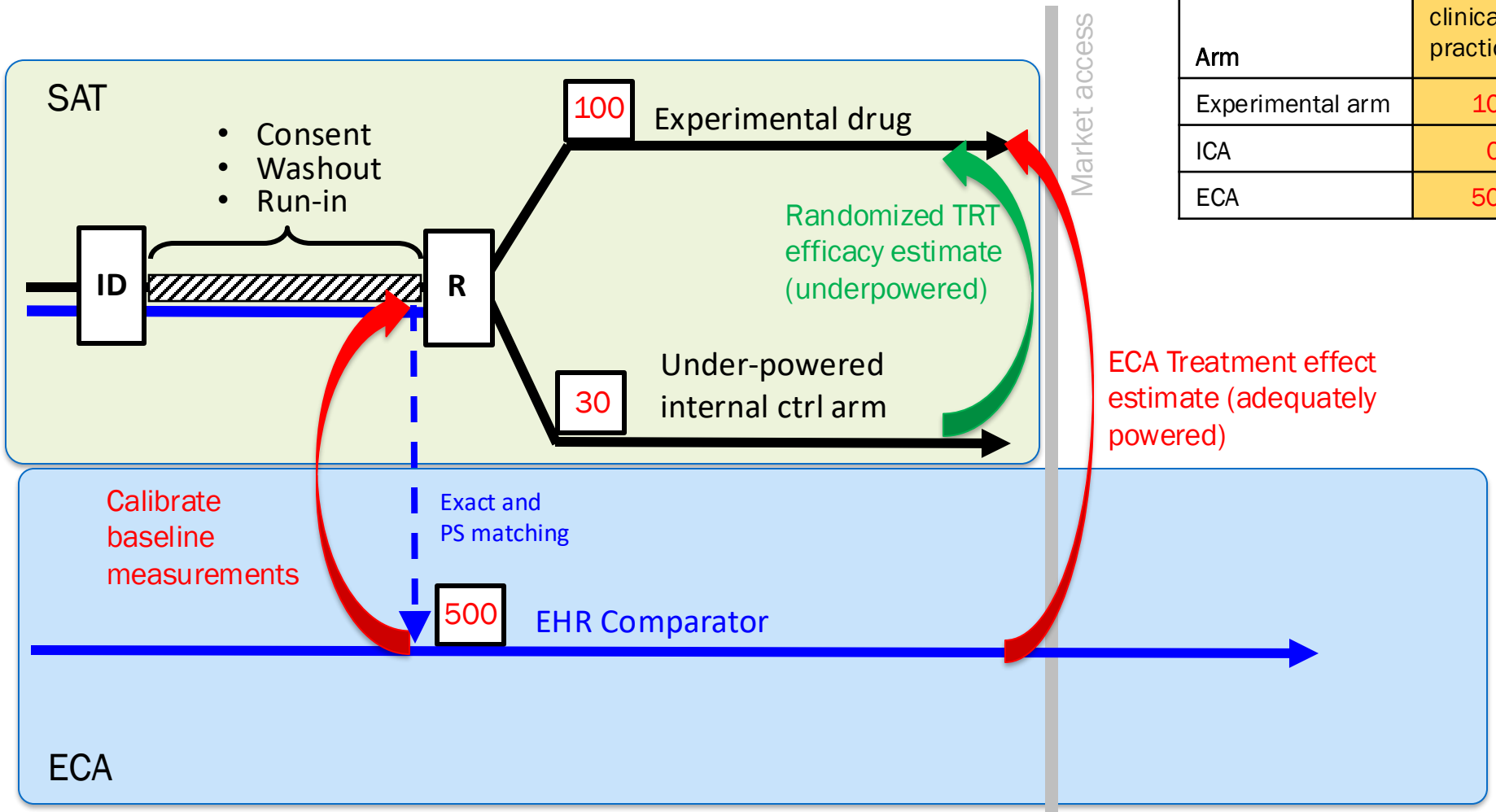
SAT = single arm trial; ECA = external control arm; ICA = internal control arm;
EHR = electronic health record; RWE = Real-world evidence study

Schneeweiss S. CPT 2024

ECA hybrid design

Table: Distribution of patients in each arm (hypothetical)

Arm	Calibration RCT extra-polation to clinical practice	**Hybrid** **Efficient trial design**	E2E E2E evidence development system
Experimental arm	100	100	
ICA	0	30	
ECA	500	500	



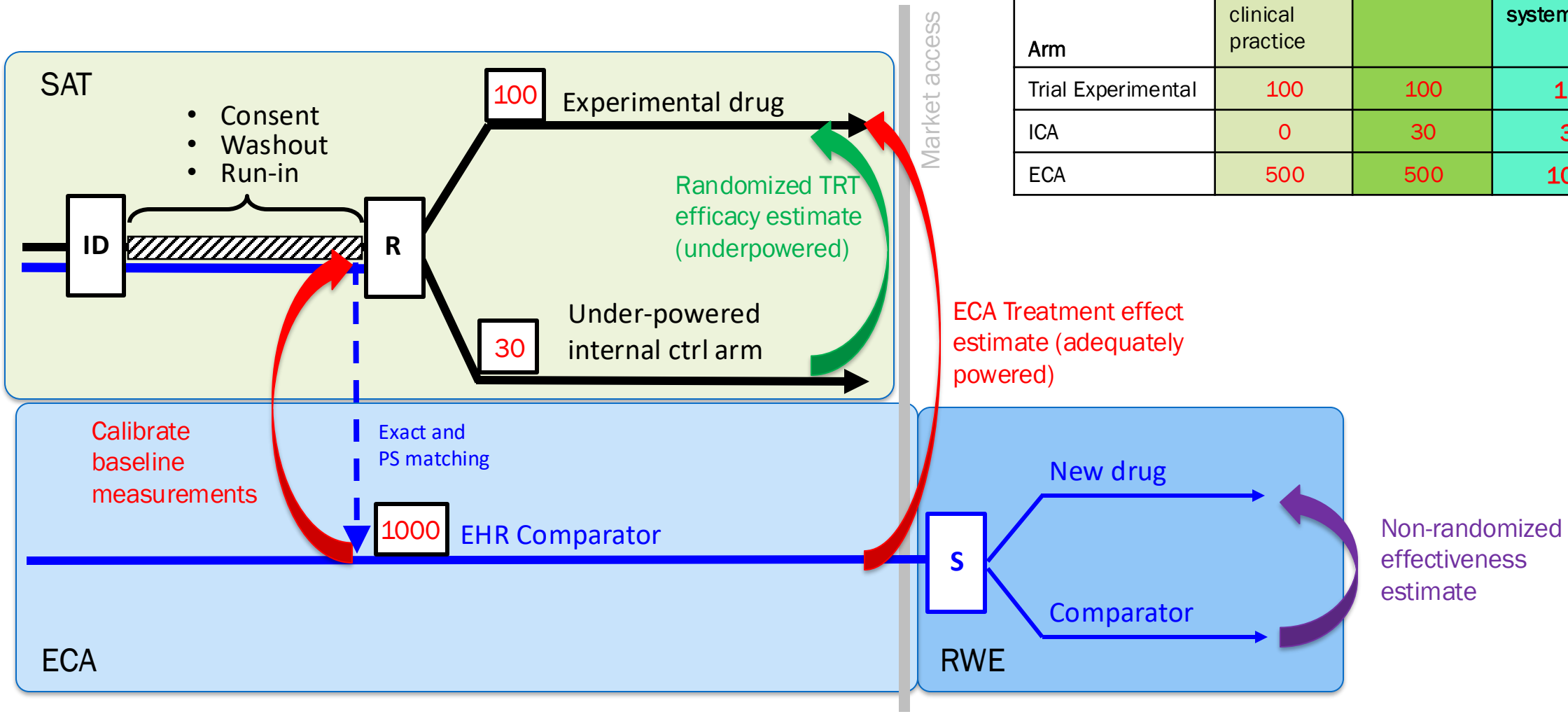
SAT = single arm trial; ECA = external control arm; ICA = internal control arm;
EHR = electronic health record; RWE = Real-world evidence study

Tan WK, et al. Contemp Clin Trials Commun 2022; Schneeweiss S. CPT 2024

Efficacy to Effectiveness (E2E) system

Table: Distribution of patients in each arm (hypothetical)

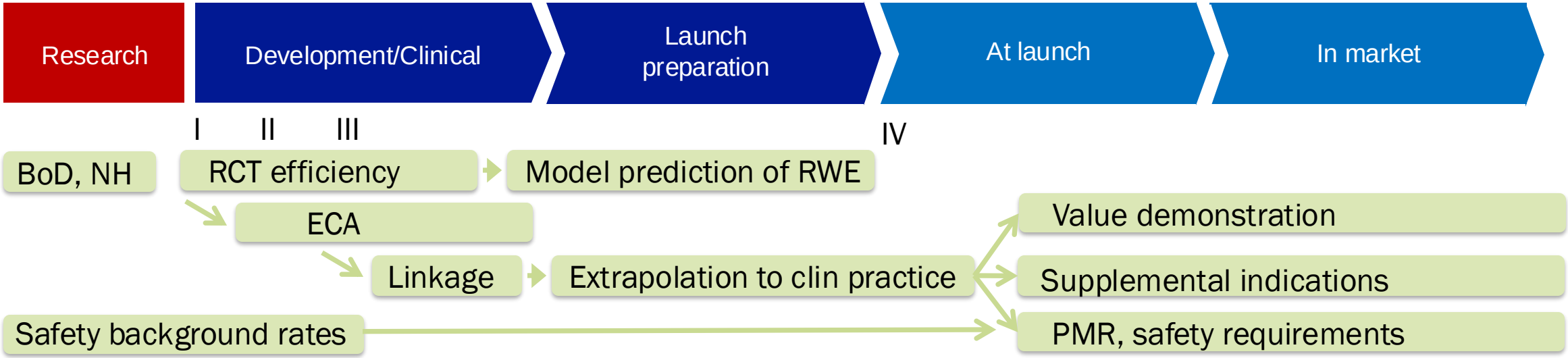
Arm	Calibration RCT extra-polation to clinical practice	Hybrid Efficient trial design	E2E **E2E evidence development system**
Trial Experimental	100	100	100
ICA	0	30	30
ECA	500	500	1000



SAT = single arm trial; ECA = external control arm; ICA = internal control arm;
EHR = electronic health record; RWE = Real-world evidence study

Schneeweiss S. CPT 2024

Building a RWE value chain

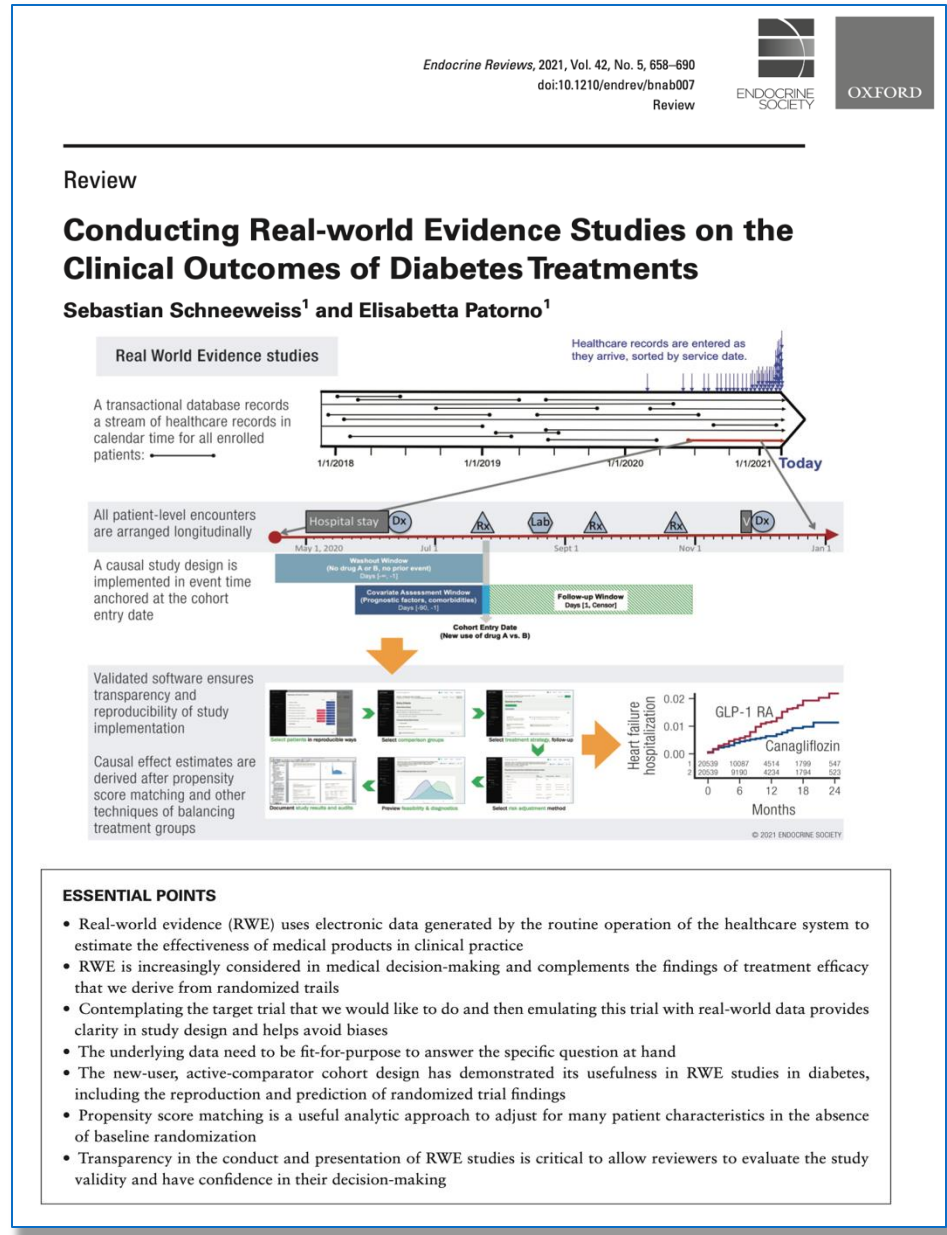


BoD = Burden of disease; NH = Natural history; ECA = External control arm, incl. calibration and hybrid designs; PMR = post marketing requirement

Schneeweiss S. CPT 2024

What can we do?

- Anchor RWE in **Integrated Evidence Plans**
  - start even before Phase 1
- Think about your ECA when you plan the SAT – ECA as afterthoughts are more likely to fail
- Base RWE communication in **Target Trial language**
- Keep in mind: **Regulatory openness** ≠ lower standards



Schneeweiss S, Patorno E. Conducting Real-world Evidence Studies on the Clinical Outcomes of Diabetes Treatments. Endocr Rev. 2023 Sep 28;42(5):658-690.