DATACK

PHUSE Single Day Events Japan 2024

Overview of the Target Trial Emulation

Datack Inc.
Ryozo Wakabayashi

Appropriately designed RWD studies can complement evidence from RCTs

Research

JAMA | Original Investigation

Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses
Results of 32 Clinical Trials

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

RCT DUPLICATE initiative

- Investigated whether database studies explicitly designed to emulate past and ongoing RCTs of medications can generate similar causal conclusions
- As a result, the conclusions were generally consistent with those of the RCTs.
  - Pearsoncorrelation r, 0.82
  - 75% statistical significance agreement
  - 66% estimate agreement
  - 75%standardized difference agreement)
- RWD studies appropriately designed to emulate RCTs can yield causal conclusions similar to those of the RCTs.

Table 1. Effect Estimates and Agreement Metrics

Study No.	Trial name	Effect estimates (95% CI)			Standardized difference[c]	Agreement		
		RCT	Database studyAdjusted[a,b]	Database study crude[a,b]		Statistical significance	Estimate	Standardized difference
1	LEADER	0.87 (0.78 to 0.97)	0.82 (0.76 to 0.87)	0.57 (0.54 to 0.61)	0.90	SA	EA	SD
2	DECLARE-TIMI58	0.83 (0.73 to 0.95)	0.69 (0.59 to 0.81)	0.47 (0.41 to 0.53)	1.76	SA		SD
3	EMPA-REG	0.86 (0.74 to 0.99)	0.83 (0.73 to 0.95)	0.63 (0.57 to 0.70)	0.35	SA	EA	SD
4	CANVAS	0.86 (0.75 to 0.97)	0.77 (0.70 to 0.85)	0.58 (0.54 to 0.62)	1.34	SA	EA	SD
5	CARMELINA	1.02 (0.89 to 1.17)	0.90 (0.84 to 0.96)	0.90 (0.86 to 0.95)	1.61	SAP	EA	SD
6	TECOS	0.98 (0.88 to 1.09)	0.89 (0.86 to 0.91)	0.81 (0.79 to 0.84)	1.71	SAP	EA	SD
7	SAVOR-TIMI	1.00 (0.89 to 1.12)	0.81 (0.76 to 0.86)	0.65 (0.62 to 0.69)	3.16	SAP		
8	LEAD-2	0 (−0.20 to 0.20)	0.05 (−0.11 to 0.22)	0.01 (−0.11 to 0.13)	−0.37	SA	EA	SD
9	TRITON-TIMI	0.81 (0.73 to 0.90)	0.88 (0.79 to 0.97)	0.70 (0.65 to 0.76)	−1.11	SA	EA	SD
10	PLATO	0.84 (0.77 to 0.92)	0.92 (0.83 to 1.02)	0.84 (0.78 to 0.91)	−1.31		EA	SD
11	ISAR-REACT 5	1.36 (1.09 to 1.70)	NA[d]	NA[d]	NA[d]	NA[d]		
12	ARISTOTLE	0.79 (0.66 to 0.95)	0.68 (0.61 to 0.76)	0.66 (0.62 to 0.71)	1.36	SA	EA	SD
13	RE-LY	0.66 (0.53 to 0.82)	0.73 (0.60 to 0.90)	0.67 (0.58 to 0.78)	−0.66	SA	EA	SD
14	ROCKET AF	0.79 (0.66 to 0.96)	0.70 (0.62 to 0.80)	0.76 (0.69 to 0.84)	1.00	SA	EA	SD
15	EINSTEIN DVT	0.68 (0.44 to 1.04)	0.75 (0.62 to 0.90)	0.85 (0.76 to 0.95)	−0.42	SAP	EA	SD
16	EINSTEIN PE	1.12 (0.75 to 1.68)	0.67 (0.55 to 0.80)	0.73 (0.64 to 0.83)	2.28	SAP		
17	RE-COVER II	1.08 (0.64 to 1.80)	1.15 (0.74 to 1.78)	1.48 (1.09 to 2.00)	−0.18	SA	EA	SD
18	AMPLIFY	0.84 (0.60 to 1.18)	0.81 (0.54 to 1.23)	0.64 (0.50 to 0.82)	0.13	SA	EA	SD
19	RECORD1	0.25 (0.14 to 0.47)	0.17 (0.10 to 0.29)	0.25 (0.18 to 0.34)	0.63	SA	EA	SD
20	TRANSCEND	0.92 (0.81 to 1.05)	0.88 (0.81 to 0.96)	0.80 (0.74 to 0.85)	0.55		EA	SD
21	ONTARGET	1.01 (0.94 to 1.09)	0.83 (0.77 to 0.90)	0.68 (0.64 to 0.72)	3.46	SAP		
22	HORIZON-PFT	0.59 (0.42 to 0.83)	0.72 (0.55 to 0.94)	1.08 (0.86 to 1.35)	−0.90	SA	EA	SD
23	VERO	0.44 (0.29 to 0.68)	NA[d]	NA[d]	NA[d]	NA[d]	NA[d]	NA[d]
24	DAPA-CKD	0.61 (0.51 to 0.72)	0.80 (0.52 to 1.26)	0.41 (0.29 to 0.58)	−1.10	SD		SD
25	PARADIGM-HF	0.80 (0.73 to 0.87)	1.02 (0.91 to 1.14)	0.95 (0.90 to 1.02)	−3.42			
26	P04334[e,f]	0.56 (0.44 to 0.72)	0.78 (0.62 to 0.97)	0.87 (0.76 to 0.99)	−1.95	SA	SD	SD
27	D5896	1.07 (0.70 to 1.65)	1.38 (0.90 to 2.13)	1.41 (1.00 to 1.98)	−0.81	SA	EA	SD
28	IMPACT[e,g]	0.85 (0.80 to 0.90)	1.13 (1.04 to 1.23)	1.22 (1.15 to 1.30)	−5.46			
29	POET-COPD	0.83 (0.77 to 0.90)	1.02 (0.93 to 1.12)	1.05 (0.99 to 1.12)	−3.27			
30	INSPIRE[h]	0.97 (0.84 to 1.12)	0.93 (0.90 to 0.96)	0.83 (0.81 to 0.85)	0.56	SA	EA	SD
31	CAROLINA[i]	0.98 (0.84 to 1.14)	0.91 (0.79 to 1.05)	0.92 (0.83-1.01)	0.70	SA	EA	SD
32	PRONOUNCE[i]	1.28 (0.59 to 2.79)	1.35 (0.94 to 1.93)	1.70 (1.30 to 2.21)	−0.12	SA	EA	SD

Wang SV, et al. Emulation of Randomized Clinical Trials With Nonrandomized Database Analyses: Results of 32 Clinical Trials. JAMA. 2023 Apr 25;329(16):1376-1385..

Limitations of clinical trials

Randomized Controlled Trial：RCT

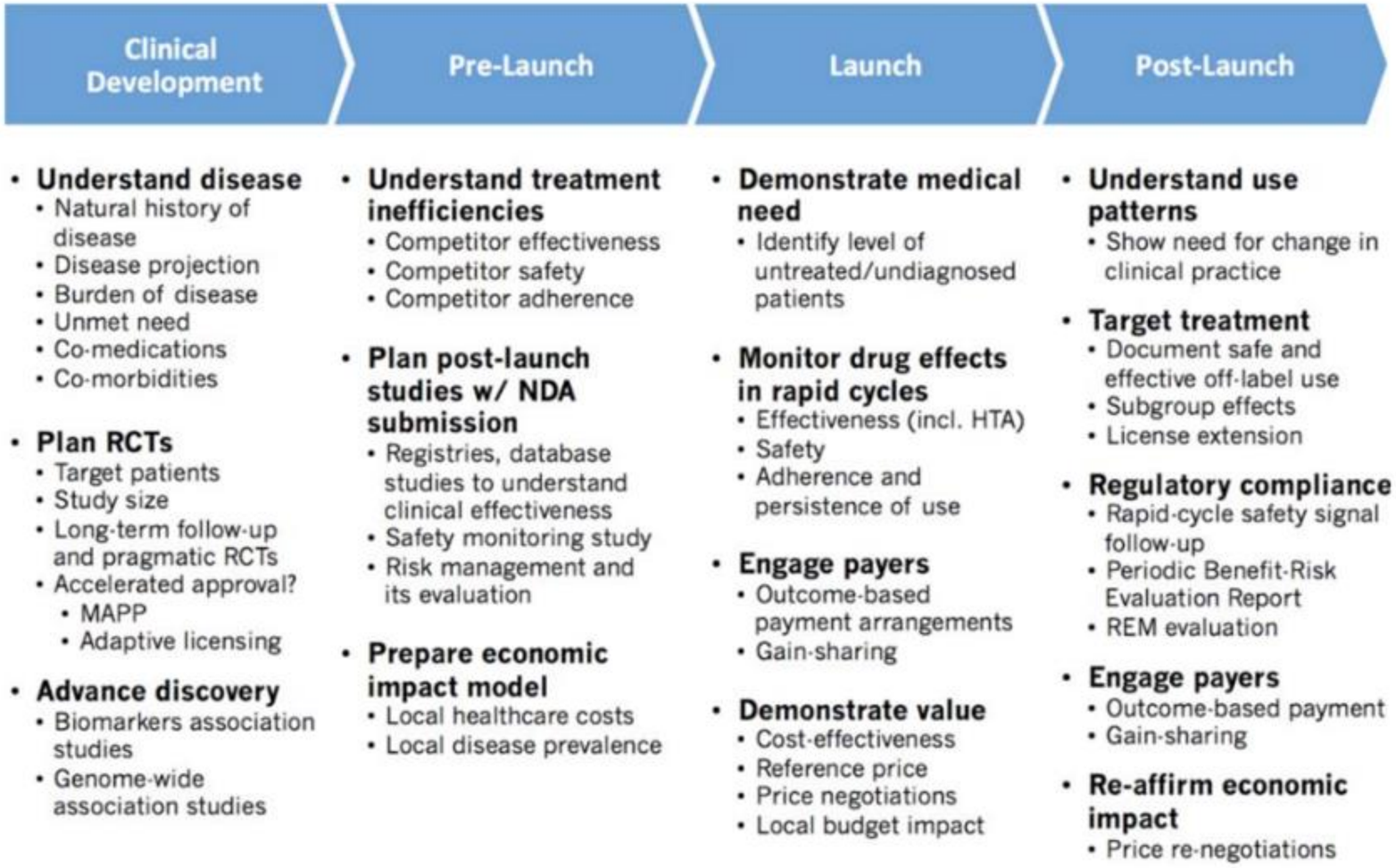
- The gold standard for evaluating the efficacy and safety of medications
- High internal validity
  - RCTs minimize biases related to the study and provide high internal validity for testing causal effects

Limitations of clinical trial

- Cost (financial and time) of implementation and ethical concerns
- 『5 toos』of RCT
  - Too few：The sample size of study subjects is limited
  - Too narrow：Specific patient groups
  - Too median-aged：Elderly patients and children are excluded
  - Too simple：simple treatment strategy
  - Too brief：The treatment period is short, and the long-term outcomes can not evaluate.
- As clinical data packages become more diverse, relying solely on randomized controlled trials is increasingly challenging

Utilization of RWD in the drug development process

RWD is expected to be used across various phases of the drug development process



Schneeweiss S. Improving therapeutic effectiveness and safety through big healthcare data. Clin Pharmacol Ther. 2016 Mar;99(3):262-5.

Utilization of RWD in the regulatory decision making process

Regulatory authorities worldwide have issued guidance on using real-world data in drug development.

- 21st Century Cures Act of 2016、FDA's Real-WorldEvidenceProgram
- ICH E8 outlines the use of external controls based on diverse data sources and highlights key considerations for their application

Efforts in Japan

- Numerous regulatory measures have been implemented to promote the use of real-world data in the drug approval process
- Real-world data has been utilized in decision-making contexts for regulatory approval of several drugs

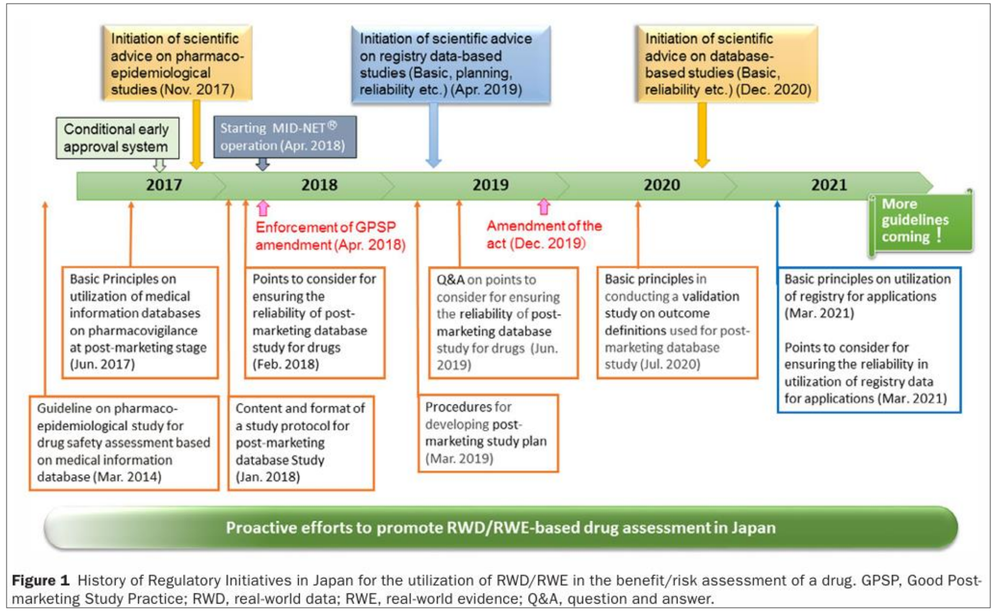


Figure 1 History of Regulatory Initiatives in Japan for the utilization of RWD/RWE in the benefit/risk assessment of a drug. GPSP, Good Post-marketing Study Practice; RWD, real-world data; RWE, real-world evidence; Q&A, question and answer.

Table 3 List of RWD-based studies on drug safety assessment conducted by the PMDA

Research purpose (target events)	Target drugs	Data source	Summary of results and regulatory actions (decisions)
Prescription survey	Triptan	NDB	・A certain number of patients were suspected of overusing triptans ・Used as a reference for revising the package insert with more precautions
Blood coagulability	DAAs (https://doi.org/10.1007/s43441-020-00247-8)	MID-NET®	・Improvement of the liver function by DAAs might be related to the fluctuation in blood coagulability in patients receiving both DAA and warfarin ・Used as a reference for revising the package insert with more precautions
Prescription survey	Valsartan	NDB	・Identified approximate number of patients, periods and doses of prescriptions ・Confirmed that no new additional safety measures were required
Thrombocytopenia	G-CSF (https://doi.org/10.1002/cpt.2263)	MID-NET®	・Increased risk of thrombocytopenia by pegfilgrastim ・Used as the major evidence for revising the package insert with more precautions
Renal dysfunction	DAAs	MID-NET®	・Observed different risks of renal dysfunction among DAAs ・Confirmed that the current warning on the package insert was appropriate and no new additional safety measures were required
Retinal detachment	Fluoroquinolone antibiotics	NDB	・Observed no increased risk of retinal detachment by fluoroquinolone ・Confirmed that no new additional safety measures were required
Cardiovascular events	Anti-hyperuricemia drugs	NDB	・Observed no increased cardiovascular risk by anti-hyperuricemia drugs ・Confirmed that no new additional safety measures were required

DAAs, direct-acting antivirals against hepatitis C; NDB, Japanese National Claims Database; PMDA, Pharmaceuticals and Medical Devices Agency; RWD, real-world data.

Nishioka K, et al. Evolving Acceptance and Use of RWE for Regulatory Decision Making on the Benefit/Risk Assessment of a Drug in Japan. Clin Pharmacol Ther. 2022 Jan;111(1):35-43. doi: 10.1002/cpt.2410. Epub 2021 Sep 18. PMID: 34528701; PMCID: PMC9290512.

Challenges of RWD study

Comparability between groups

- confounding, eligibility criteria when using RWD as an external control, and endpoint assessment

Common biases in RWD studies

- prevalent user bias / immortal time bias / confounding by indication / protopathic bias / etc
- The lack of randomization is not the only issue

Examples of RWD studies that reached misleading conclusions due to inappropriate study design

- Postmenopausal hormone therapy and coronary heart disease（example of prevalent user bias）
  - In an RWD study comparing current users vs. never users, hormone therapy was associated with a lower risk of coronary heart disease
  - However, in an RCT comparing initiators vs. non-initiators, hormone therapy was associated with a higher risk of CHD
- Statin and cancer（example of selection bias arising from the use of post-baseline information）
  - RWD studies suggested that statin use reduces cancer risk
  - However, RCTs found no association between statin use and cancer incidence

RWD studies using secondary data are highly flexible in analysis
⇒ increasing the risk of various biases

Grodstein F, et al. Postmenopausal estrogen and progestin use and the risk of cardiovascular disease. N Engl J Med. 1996 Aug 15;335(7):453-61.
Manson JE,et al; Women's Health Initiative Investigators. Estrogen plus progestin and the risk of coronary heart disease. N Engl J Med. 2003 Aug 7;349(6):523-34.
Dickerman BA,et al. Avoidable flaws in observational analyses: an application to statins and cancer. Nat Med. 2019 Oct;25(10):1601-1606.

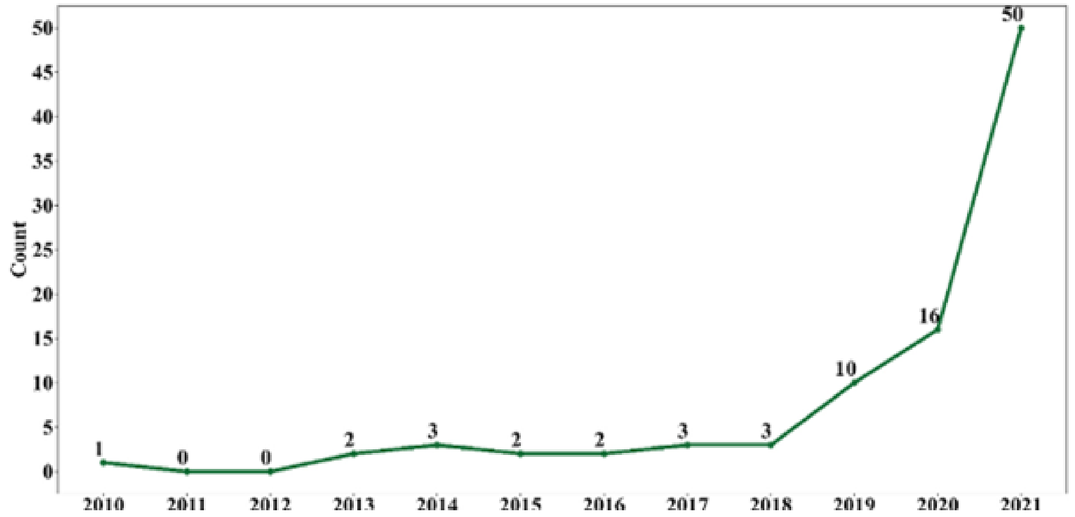
Target Trial Emulation

Concetp of the Target Trial Emulation

- Emulate an ideal randomized controlled trial (target trial) using observational data
- If the emulation is successful, the analysis of the observational data yields the same effect estimates as target trial

Target Trial

- The target trial refers to an ideal (hypothetical) randomized controlled trial designed to clarify causal effects
  - It does not refer to previously reported RCTs, although emulation of such RCTs (RCT replication) is possible.
  - The framework primarily assumes emulation of pragmatic trials, as placebo-controlled or blinded designs cannot be emulated



The publication trend of studies using Target Trial framework.

Zuo H, Yu L, Campbell SM, Yamamoto SS, Yuan Y. The implementation of target trial emulation for causal inference: a scoping review. J Clin Epidemiol. 2023 Oct;162:29-37.

Process of the Target Trial Emulation

Process of the Target Trial Emulation

- Step 1. Specify Target Trial protocol
- Step 2. Emulate Target Trial protocol using RWD

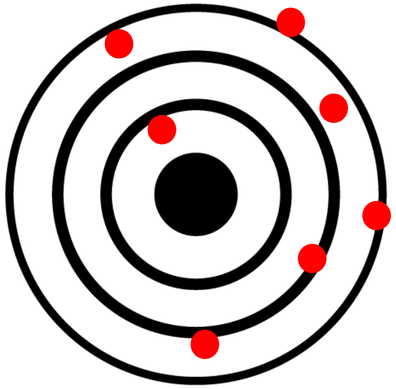
Table 1. A Summary of the Protocol of a Target Trial to Estimate the Effect of Postmenopausal Hormone Therapy on the 5-Year Risk of Breast Cancer

Protocol Component	Description		Emulation in RWD (Operational definition)
Eligibility criteria	Postmenopausal women within 5 years of menopause between the years 2005 and 2010 and with no history of cancer and no use of hormone therapy in the past 2 years.	→	XXXX
Treatment strategies	Refrain from taking hormone therapy during the follow-up. Initiate estrogen plus progestin hormone therapy at baseline and remain on it during the follow-up unless you are diagnosed with deep vein thrombosis, pulmonary embolism, myocardial infarction, or cancer.	→	XXXX
Assignment procedures	Participants will be randomly assigned to either strategy at baseline and will be aware of the strategy to which they have been assigned.	→	XXXX
Follow-up period	Starts at randomization and ends at diagnosis of breast cancer, death, loss to follow-up, or 5 years after baseline, whichever occurs first.	→	XXXX
Outcome	Breast cancer diagnosed by an oncologist within 5 years of baseline.	→	XXXX
Causal contrasts of interest	Intention-to-treat effect, per-protocol effect	→	XXXX
Analysis plan	Intention-to-treat effect estimated via comparison of 5-year cancer risks among individuals assigned to each treatment strategy. Per-protocol effect estimation requires adjustments for pre- and postbaseline prognostic factors associated with adherence to the strategies of interest. All analyses will be adjusted for pre- and postbaseline prognostic factors associated with loss to follow-up (57). This analysis plan implies that the investigators prespecify and collect data on the adjustment factors.	→	XXXX

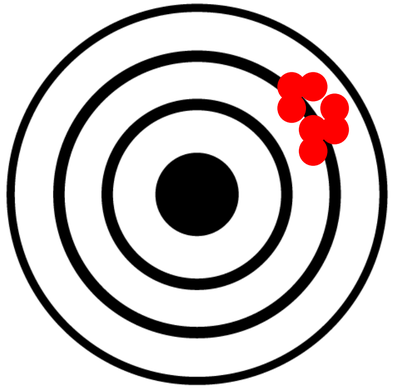
Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol. 2016 Apr 15;183(8):758-64.

What is the Validity ?

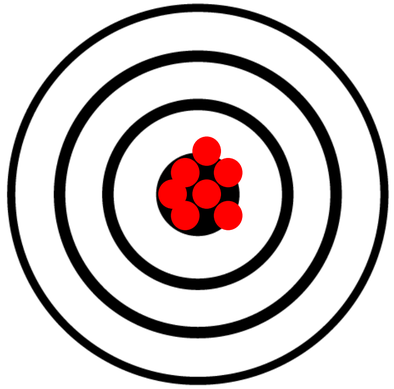
Not Reliable
Not Valid



Reliable
Not Valid



Reliable
Valid



- In general, “validity” refers to the distance from the ideal target (gold standard)
- In research design of RWD study, “validity of the study” indicates the degree of deviation from the ideal target trial
- Designing a valid study requires first clearly defining the target trial and then emulating it as closely as possible

Process of the Target Trial Emulation（Example）

Component	Target trial	Emulation in Swedish Renal Registry
Eligibility	Individuals (both sexes) aged 18 years or older with new-onset CKD G4/5 (defined as an eGFR <30 ml/min/1.73m^2 using the CKD-EPI equation) who were adherent users of RASi and had no history of kidney transplantation or dialysis.	Same as target trial. A medication possession ratio >80% in the 2 years prior to inclusion was used as a proxy for adherence.
Treatment strategies	Stop RASi within 6 months and remain off treatment vs. continue RASi.	In our main analysis, we compared the treatment strategies stopping RASi within 6 months and remaining off treatment after first observed eGFR <30 ml/min/1.73m^2 or continue RASi during the entire follow-up.
Treatment assignment	Eligible individuals are randomly assigned to one of the two strategies and are aware of the treatment strategy they are assigned to (i.e., no blinding).	Randomisation is emulated via cloning of individuals and assigning each replicate to a treatment strategy.
Follow-up	For each individual follow-up starts at the time of assignment to a strategy (i.e., baseline is the moment when eGFR first drops <30 ml/min/1.73m^2) and ends at the occurrence of death, major cardiovascular event, kidney replacement therapy or 5 years, whichever comes first.	Same as target trial.
Primary end point	All-cause mortality. Secondary endpoints include major adverse cardiovascular events and initiation of kidney replacement therapy.	Same as target trial. All-cause mortality is identified from the Swedish Death Registry. Cardiovascular hospitalizations are identified through ICD-10 codes in the National Patient Registry: myocardial infarction: I21, I22; cerebrovascular event: G45, G464, G463, I63, I64, I693, I694, I698. Initiation of KRT is ascertained from the Swedish Renal Registry.
Causal contrast	Intention-to-treat effect. Per protocol effect.	Per protocol effect: effect of adhering to the strategies as specified under "Treatment strategies" during follow-up.
Statistical analysis	Intention-to-treat analysis. Per protocol analysis: Individuals are artificially censored when they deviate from their assigned strategy as follows: *Stop within 6 months and remain off treatment:* Censored at the beginning of the 6th month if not stopped. Censored if RASi treatment is restarted after discontinuation at any moment during follow-up. *Continue:* Censored when individual stopped during follow-up. Note that inverse probability weighting is required also in a randomized trial to validly estimate the per-protocol effect. IP weights are estimated as a function of time-fixed variables (age, sex, calendar year) and the following baseline variables and time-varying variables: eGFR, systolic and diastolic blood pressure, comorbidities (ischemic heart disease, myocardial infarction, arrhythmia, heart failure, peripheral vascular disease, cerebrovascular disease, diabetes, chronic pulmonary disease, cancer), medication use (beta blockers, calcium channel blockers, diuretic, statins, antiplatelet) and hospitalizations (total number of hospitalizations in previous year, AKI hospitalization in previous year, hyperkalemia hospitalization). Standardized, weighted survival curves under each strategy.	Same as per protocol analysis. We created an expanded dataset including 2 replicates for each included individual and assigned one replicate to each treatment strategy. We adjusted for the following baseline and time-varying variables and assumed that adjustment for these variables was sufficient to adjust for informative censoring: age, sex, calendar year, eGFR, systolic and diastolic blood pressure, comorbidities (ischemic heart disease, myocardial infarction, arrhythmia, heart failure, peripheral vascular disease, cerebrovascular disease, diabetes, chronic pulmonary disease, cancer), medication use (beta blockers, calcium channel blockers, diuretic, statins, antiplatelet) and hospitalizations (total number of hospitalizations in previous year, AKI hospitalization in previous year, hyperkalemia hospitalization). Sensitivity analyses were additionally adjusted for ACR and potassium values.

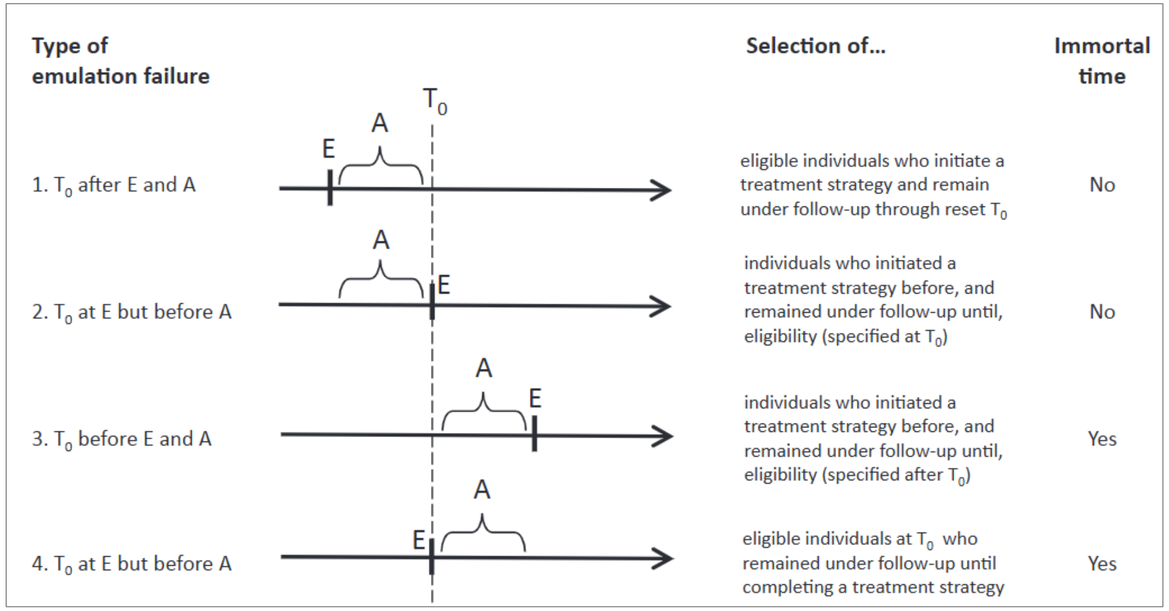
Fu EL, et al. Stopping Renin-Angiotensin System Inhibitors in Patients with Advanced CKD and Risk of Adverse Outcomes: A Nationwide Study. J Am Soc Nephrol. 2021 Feb;32(2):424-435. doi: 10.1681/ASN.2020050682. Epub 2020 Dec 28.

Reason to specify the target trial

Why to specify the target trial？

- The lack of randomization is not the only issue in RWD studies
- Design flaws, such as the improper setting of “time zero”, are common and can easily lead to biases
  - Immortal time bias / Prevalent user bias (selection bias) /Measurement error
- Clearly defining the Target Trial helps avoid methodological pitfalls and ensures a valid study design
  - Deviations from the Target Trial are a source of bias in RWD studies

Biases associated with the setting of time zero



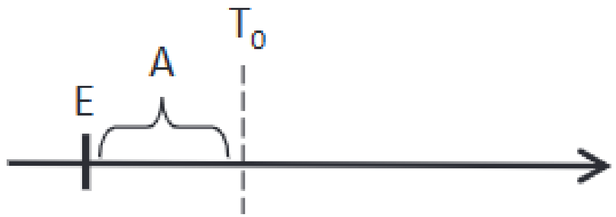
- Bias may occur when the following three time points are misaligned
  - Time zero
  - Eligibility
  - Treatment assignment
- In RWD,
  - There may be multiple time points at which eligibility criteria are met
  - There may be difficult to define “Time zero” of follow-up

Hernán MA, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov;79:70-75.

Common Biases

Prevalent user bias

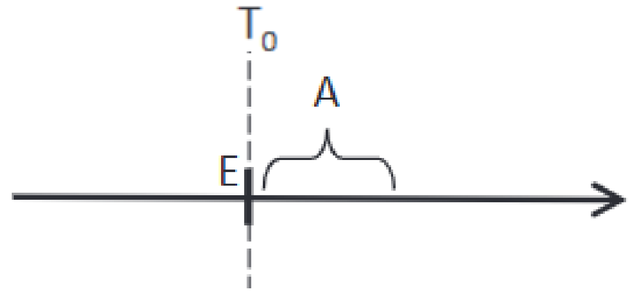
1. T_0 after E and A



- Comparison
  - Prevalent users (who had initiated therapy before baseline) vs Never users
- Bias
  - Failure to identify the early adverse effect of the drug
  - the group of prevalent users might have been relatively depleted of susceptible (**Selection bias**)

Immortal time bias

4. T_0 at E but before A



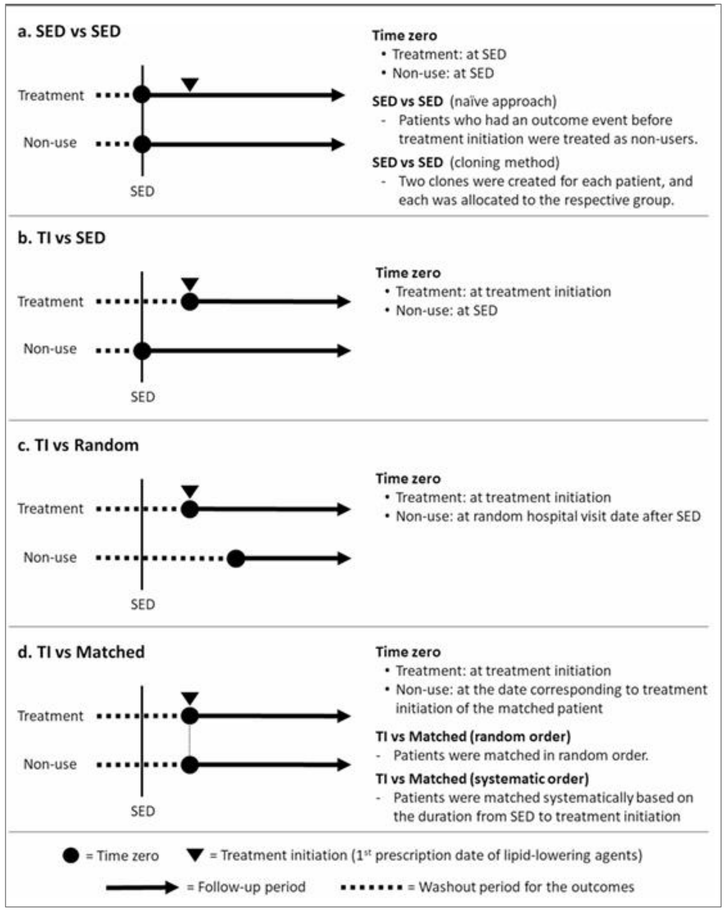
- Comparison
  - Initiating treatment **after time zero** vs. not initiating treatment
- Bias
  - Patients in a treatment group inevitably remain event-free until they receive the study treatment (Immortal time)
  - As a result, overestimate the incidence rate in the non-use group

Impact of the biases associated with the setting of time zero

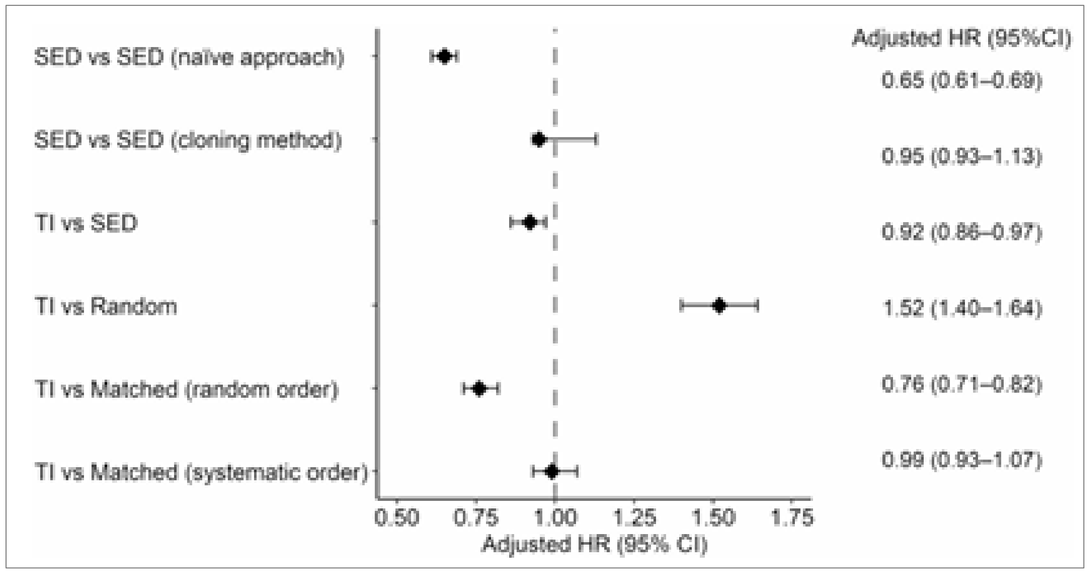
ORIGINAL RESEARCH ARTICLE

Impact of "time zero" of Follow-Up Settings in a Comparative Effectiveness Study Using Real-World Data with a Non-user Comparator: Comparison of Six Different Settings

Ryozo Wakabayashi[1,2] · Takahiro Hirano[1,2] · Thomas Laurent[1,2] · Yoshiki Kuwatsuru[3] · Ryohei Kuwatsuru[1,3]



The treatment effects, based on the adjusted hazard ratio for the outcome, varied among the six time-zero setting methods, even though the same dataset was used for the same research question.



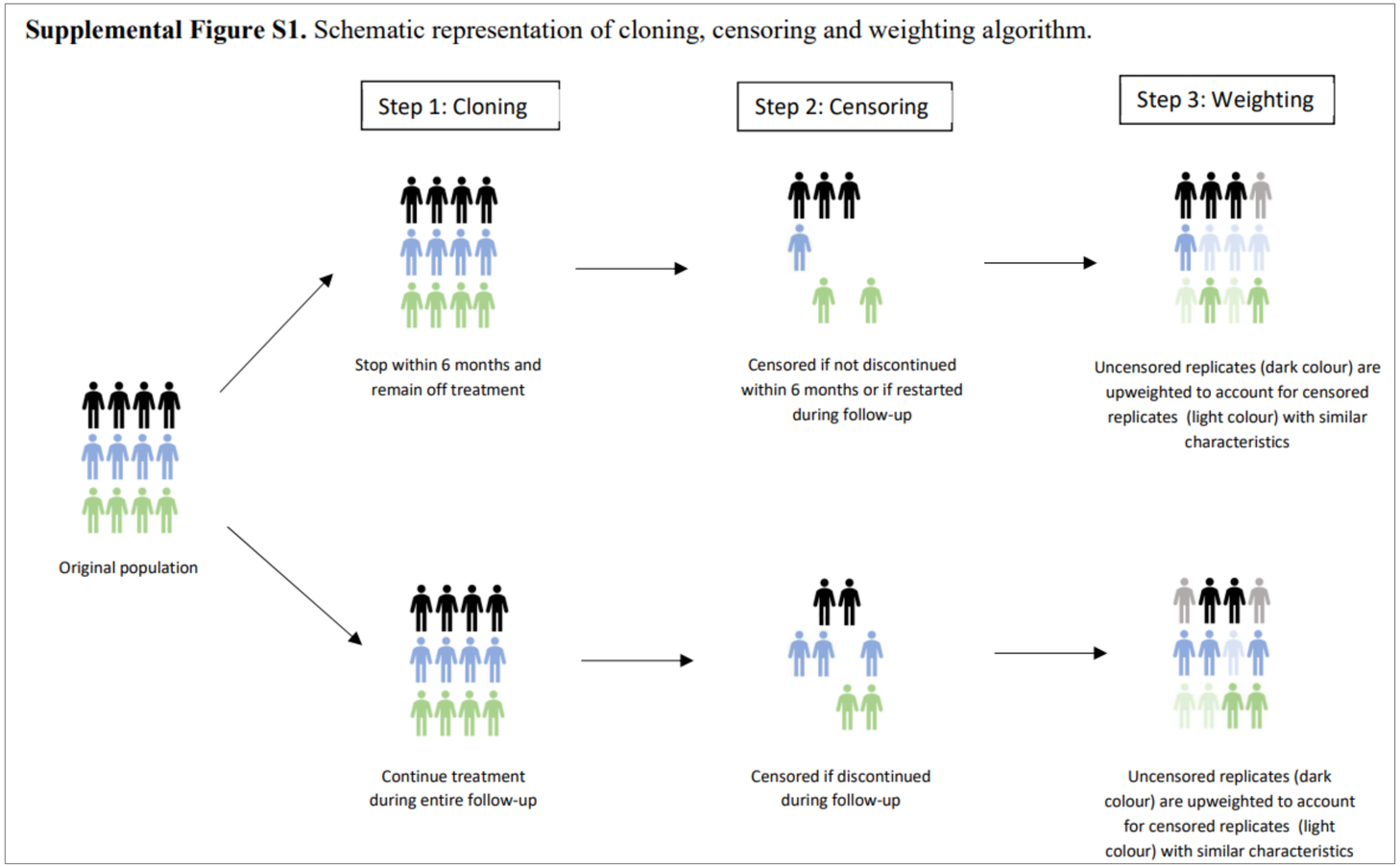
Wakabayashi R, et al. Impact of "time zero" of Follow-Up Settings in a Comparative Effectiveness Study Using Real-World Data with a Non-user Comparator: Comparison of Six Different Settings. Drugs Real World Outcomes. 2023 Mar;10(1):107-117

Addressing time zero-related challenges in Target Trial Emulation

Time zero-related issues	Example	Approach
Time zero is hard to define (meet the eligibility criteria at multiple times)	P: Postmenopausal women with an intact uterus I : Hormone therapy C: No hormone therapy O: Breast cancer	1. Chose one of the multiple eligibility times as time zero 2. Chose all eligibility times: (each individual at each eligible time as a different individual when emulating target trial)
Individuals do not adhere to sustained treatment strategies	Individuals who initiate study drug at time zero discontinue it after few week ⇒ ITT risk ratio may be null	• Per-protocol effect rather than ITT effect ◦ **Cannot rely on post time zero information** ◦ Censor individuals when their data stop being consistent with the assigned strategy ◦ Need adjustment for Post-time zero censoring (IPW, etc)
Treatment strategies are not uniquely defined at time zero	The outcome is in-hospital mortality, but the intervention is not determined at the time of admission (treatment starts several days later)	1. Randomly assign the individual to one of the strategies 2. Create exact copies of the individual (clone) a. Per protocol estimate b. Need adjustment for Post-time zero censoring (IPW, etc)

Hernán MA, et al. Specifying a target trial prevents immortal time bias and other self-inflicted injuries in observational analyses. J Clin Epidemiol. 2016 Nov;79:70-75.

Target Trial Emulation (Create exact copies of the individual)

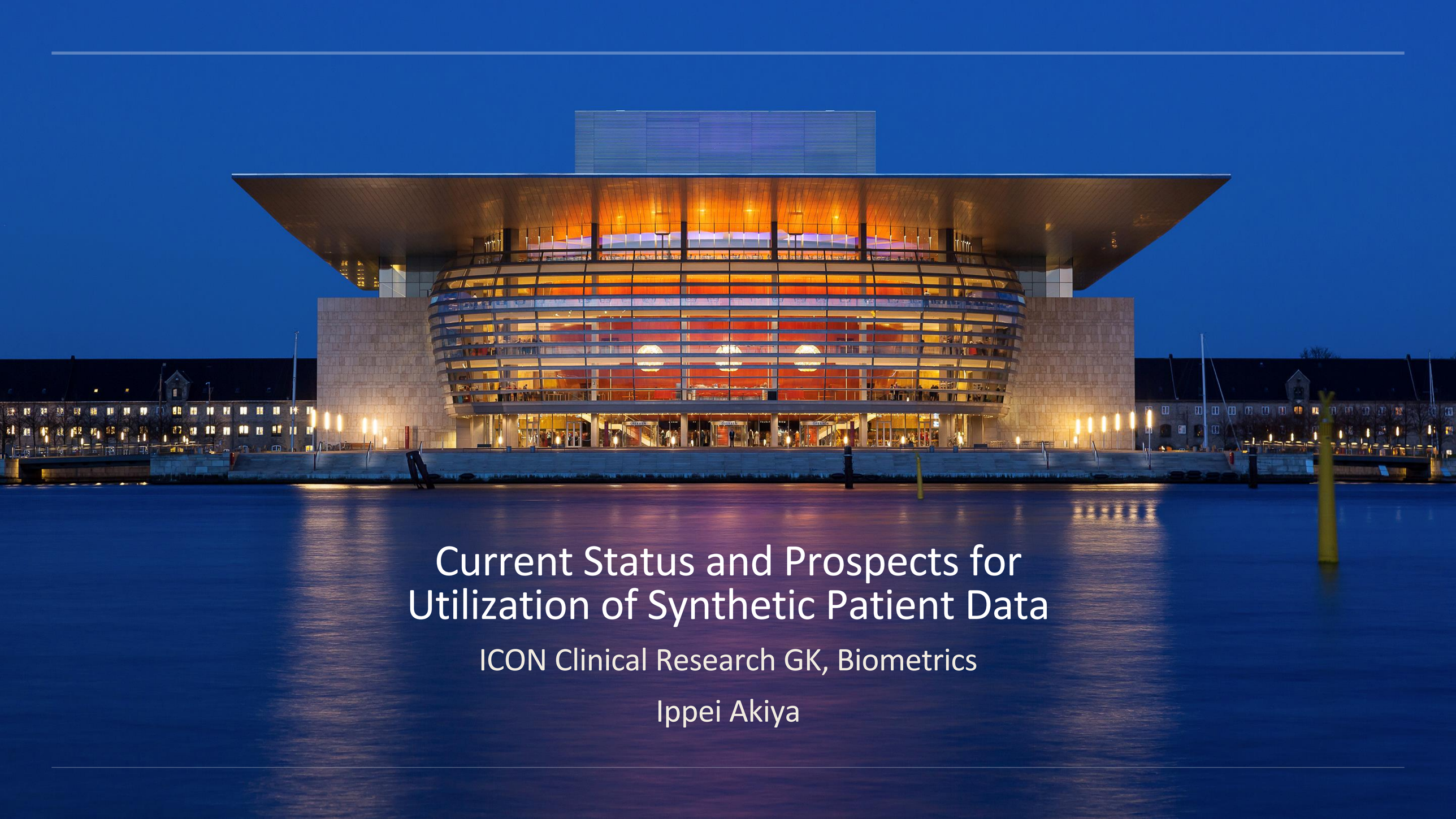


Fu EL, et al. Stopping Renin-Angiotensin System Inhibitors in Patients with Advanced CKD and Risk of Adverse Outcomes: A Nationwide Study. J Am Soc Nephrol. 2021 Feb;32(2):424-435. doi: 10.1681/ASN.2020050682. Epub 2020 Dec 28.

Summary

- RWD studies are expected to utilize drug development process but are prone to various biases
- Target Trial Emulation (TTE) is a framework proposed to minimize these biases by clearly specifying the protocol of an ideal RCT to answer the causal question and emulating it using RWD.
- The target trial emulation framework helps avoid methodological pitfalls, including issues related to time zero setting, and ensures a valid study design
- Appropriately designed RWD studies can yield conclusions comparable to those of emulated RCTs.


DATACK



Current Status and Prospects for Utilization of Synthetic Patient Data

ICON Clinical Research GK, Biometrics

Ippei Akiya

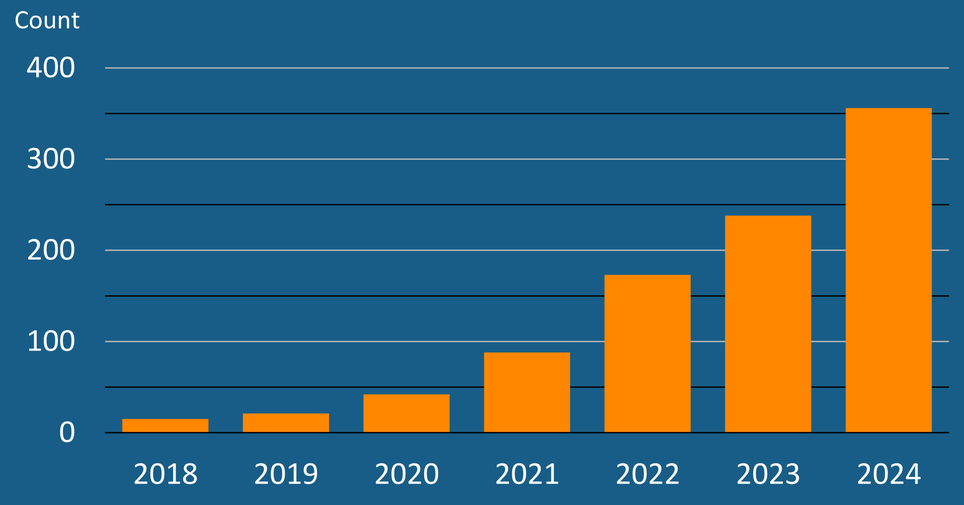
Disclaimer

- The content of this presentation includes my personal research, considerations, and opinions, and is in no way related to the intellectual property rights or business activities of the company I belong to or its clients.
- Furthermore, it does not represent the views of any academic societies or other organizations I am affiliated with.
- The copyrights and intellectual property rights for the software and other products introduced in this presentation belong to their respective owners.
- There are no conflicts of interest (COI) to disclose.

What is synthetic patient data?

- Model-Based Generation Methods
  - These methods do not require individual patient data for model construction.
- Data-Driven Generation Methods << This is the topic of today's discussion.
  - This approach involves generating data using statistical models and machine learning models based on individual patient data.
  - It generates fictitious individual patient data that possesses statistical characteristics similar to real data (a collection of individual patient data).
  - Since the data is synthetic, it does not violate data privacy issues and can be considered a form of anonymization.

- In recent years, there has been a growing interest in synthetic patient data.



Google Scholar 2024/11/3
Search words: "synthetic patient data"

Numerous use cases

- Use cases
  - observational studies
  - test data
  - simulation studies
  - an anonymization method
  - training data in machine learning and AI
  - clinical trials

- The anonymization of patient data has become essential; however, certain anonymization methods have the drawback of not yielding valid analytical results, which has led to increased attention on synthetic patient data.
- Synthetic patient data can generate an unlimited amount of patient data, making it easy to synthesize data according to simulation scenarios, and thus it can be applied in simulation studies.
- Application in clinical trials has not yet begun, but proof of concept (POC) and validation studies are starting to be published in research papers.

Today’s discussion will focus on this topic.

Published on 19.6.2024 in Vol 12 (2024)

Preprints (earlier versions) of this paper are available at https://preprints.jmir.org/preprint/55118, first published December 03, 2023.

Comparison of Synthetic Data Generation Techniques for Control Group Survival Data in Oncology Clinical Trials: Simulation Study

Ippei Akiya[1]; Takuma Ishihara[2]; Keiichi Yamamoto[3]

Abstract

Background:
Synthetic patient data (SPD) generation for survival analysis in oncology trials holds significant potential for accelerating clinical development. Various machine learning methods, including classification and regression trees (CART), random forest (RF), Bayesian network (BN), and conditional tabular generative adversarial network (CTGAN), have been used for this purpose, but their performance in reflecting actual patient survival data remains under investigation.

Objective:
The aim of this study was to determine the most suitable SPD generation method for oncology trials, specifically focusing on both progression-free survival (PFS) and overall survival (OS), which are the primary evaluation end points in oncology trials. To achieve this goal, we conducted a comparative simulation of 4 generation methods, including CART, RF, BN, and the CTGAN, and the performance of each method was evaluated.

Methods:
Using multiple clinical trial data sets, 1000 data sets were generated by using each method for each clinical trial data set and evaluated as follows: (1) median survival time (MST) of PFS and OS; (2) hazard ratio distance (HRD), which indicates the similarity between the actual survival function and a synthetic survival function; and (3) visual analysis of Kaplan-Meier (KM) plots. Each method's ability to mimic the statistical properties of real patient data was evaluated from these multiple angles.

Results:
In most simulation cases, CART demonstrated the high percentages of MSTs for synthetic data falling within the 95% CI range of the MST of the actual data. These percentages ranged from 88.8% to 98.0% for PFS and from 60.8% to 96.1% for OS. In the evaluation of HRD, CART revealed that HRD values were concentrated at approximately 0.9. Conversely, for the other methods, no consistent trend was observed for either PFS or OS. CART demonstrated better similarity than RF, in that CART caused overfitting and RF (a kind of ensemble learning approach) prevented it. In SPD generation, the statistical properties close to the actual data should be the focus, not a well-generalized prediction model. Both the BN and CTGAN methods cannot accurately reflect the statistical properties of the actual data

Citation

Please cite as:

Akiya I, Ishihara T, Yamamoto K
Comparison of Synthetic Data Generation Techniques for Control Group Survival Data in Oncology Clinical Trials: Simulation Study
JMIR Med Inform
2024;12:e55118
doi: 10.2196/55118
PMID: 38722154
PMCID: PMC11196245

Copy Citation to Clipboard

Export Metadata

END for: Endnote
BibTeX for: BibDesk, LaTeX
RIS for: RefMan, Procite, Endnote, RefWorks

Add this article to your Mendeley library

This paper is in the following e-collection/theme issue:

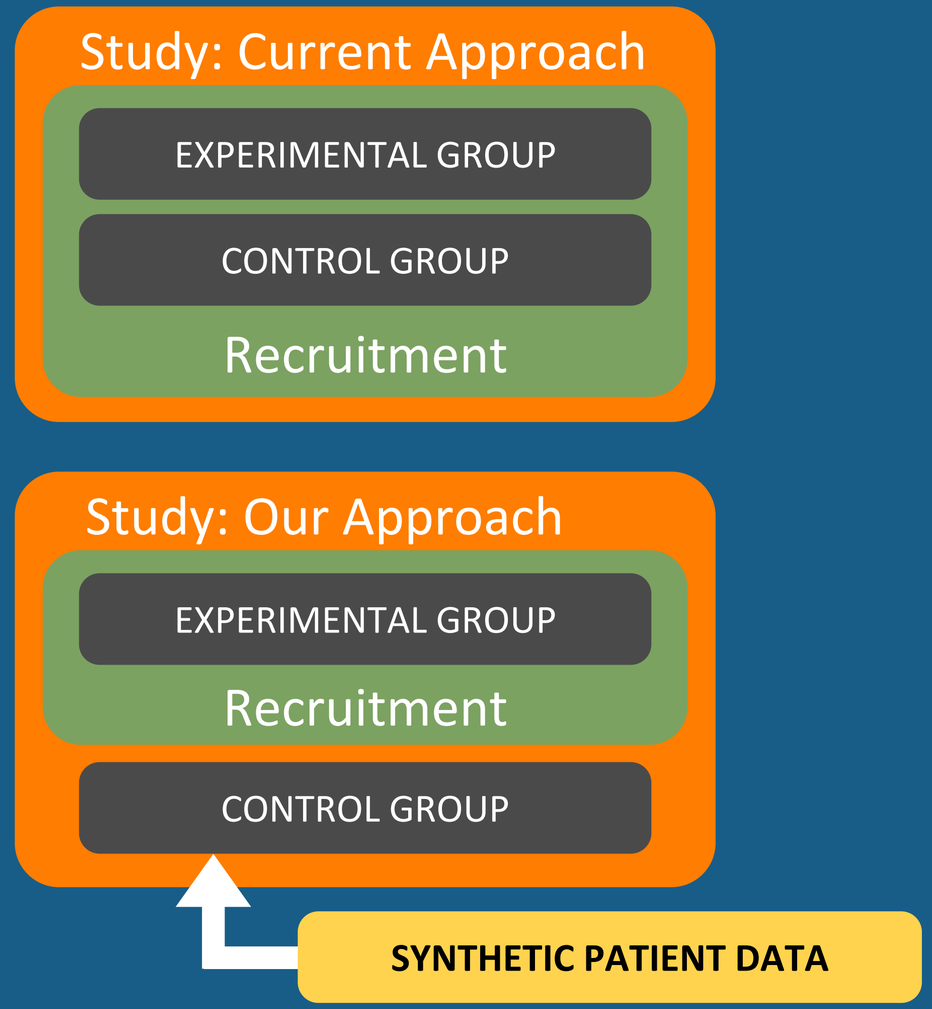
Decision Support for Health Professionals (1137)
Big Data (219)
Methods and Instruments in Medical Informatics (104)
New Technologies (66)
Research Infrastructures and Registries (45)
Tools, Programs and Algorithms (243)
Data Science (226)

Our Paper

- Published in *JMIR Medical Informatics*.
- Objective: Which method is most suitable for synthesizing control group data in oncology clinical trials?
  - CART
  - Random Forest
  - Bayesian Network
  - CTGAN
- Evaluation was performed based on the similarity of Median Survival Time and the survival function.

Motivation

- Many oncology clinical trials struggle to reach target enrollment within the planned recruitment period.
- Trials often face challenges of extending the study period or ending without meeting the target number of participants.
- We believe that it is possible to generate synthetic patient data from control arm data in past clinical trials and use it to establish a control arm for a new clinical trial.



Method (Overview)

- Clinical trial data for solid tumors was obtained from Project Data Sphere, a clinical trial data-sharing site (four trials).
- Synthetic patient data was generated using four methods: CART, Random Forest, Bayesian Network, and CTGAN.
- For each trial, 1,000 synthetic datasets were generated for each method, resulting in a total of 16,000 datasets.
- Evaluation Methods:
  - Median Survival Time
  - Similarity of Kaplan-Meier curves

Evaluation using Median Survival Time

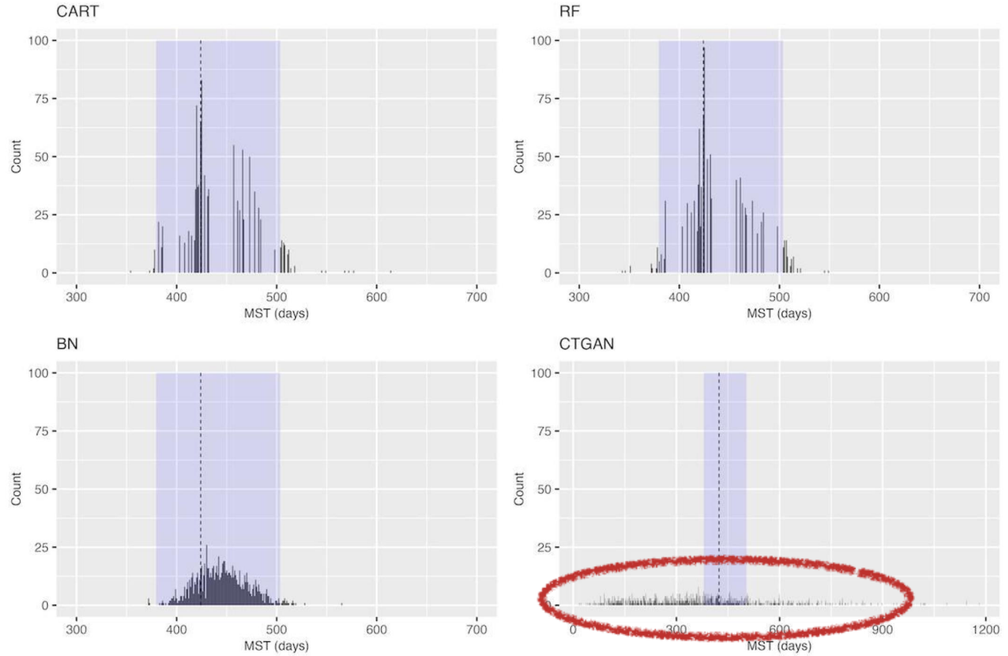
The dashed line indicates the MST of the actual data, and the blue background shows the 95% confidence interval.

In CART and Random Forest, the MST of the synthetic patient data often falls within the 95% confidence interval of the actual data's MST.

In the case of Bayesian Networks, there were instances of significant bias. With CTGAN, the MST resulted in high variance.

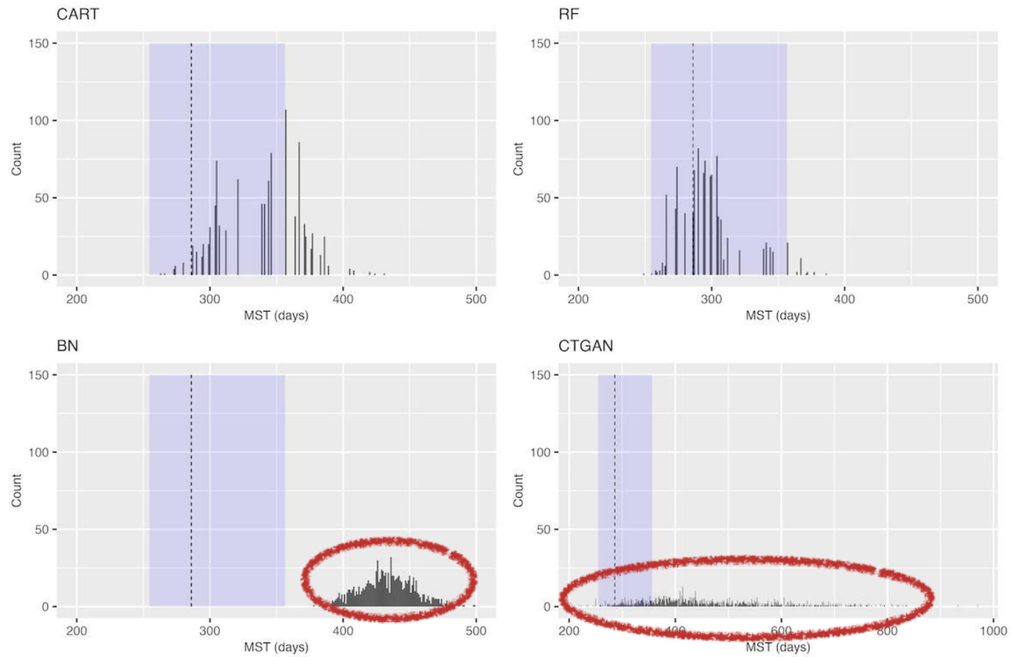
PFS

Figure 1. Histogram of the median survival time of the synthetic data for progression-free survival in the NCT00703326 trial. The dashed vertical line represents the median survival time of the actual data, and the light blue background indicates its 95% CI. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; MST: median survival time; RF: random forest.



OS

Figure 2. Histogram of the median survival time of the synthetic data of overall survival in the NCT00460265 trial. The dashed vertical line represents the median survival time of the actual data, and the light blue background indicates its 95% CI. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; MST: median survival time; RF: random forest.



Evaluation using a Kaplan-Meier (KM) plot

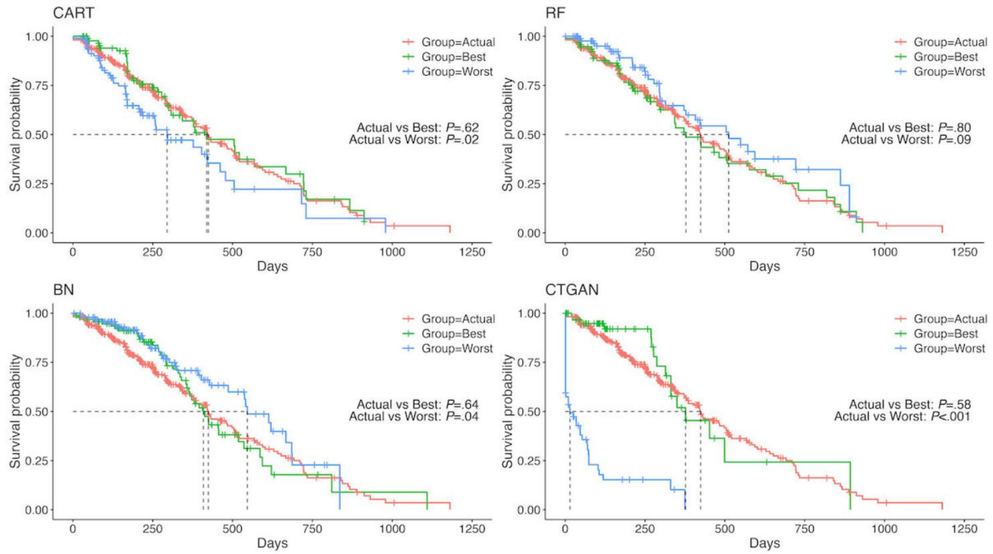
Create KM plots for the following three datasets:

1. Actual data
2. Synthetic patient data with the highest similarity to the survival function (Best)
3. Synthetic patient data with the lowest similarity to the survival function (Worst)

CART and Random Forest generally showed good results.

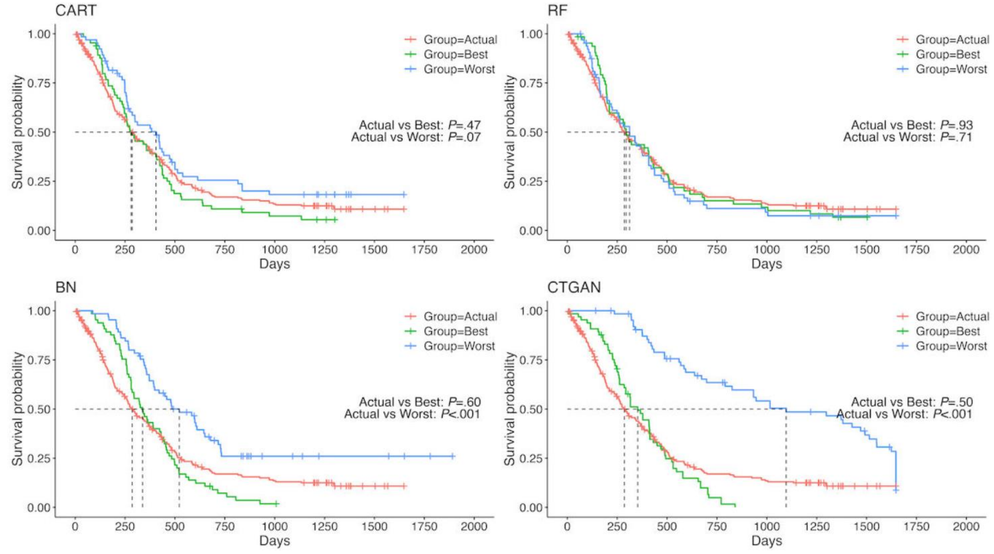
PFS

Figure 3. Kaplan-Meier plots for progression-free survival in the NCT00703326 trial. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; RF: random forest.



OS

Figure 4. Kaplan-Meier plots for overall survival in the NCT00460265 trial. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; RF: random forest.



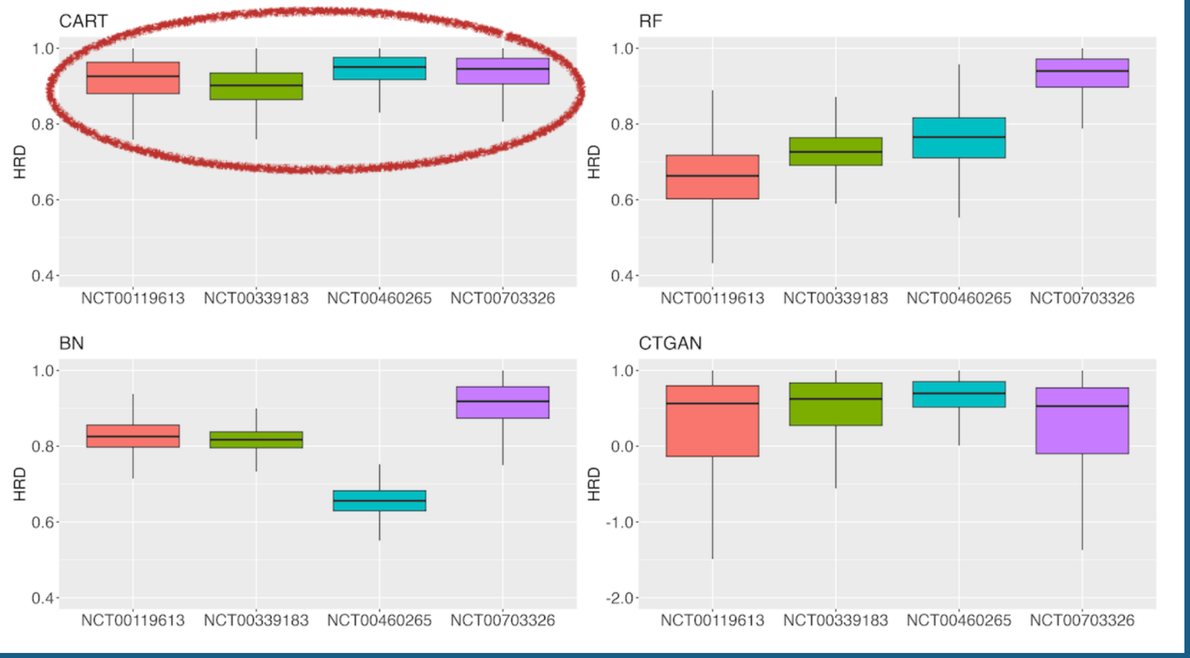
Similarity of survival functions

The closer the hazard ratio is to 1, the more it is determined to be similar to the actual data.

As a result of simulating 1,000 datasets (synthetic patient data), CART tends to be distributed around 0.9 regardless of the test data used.

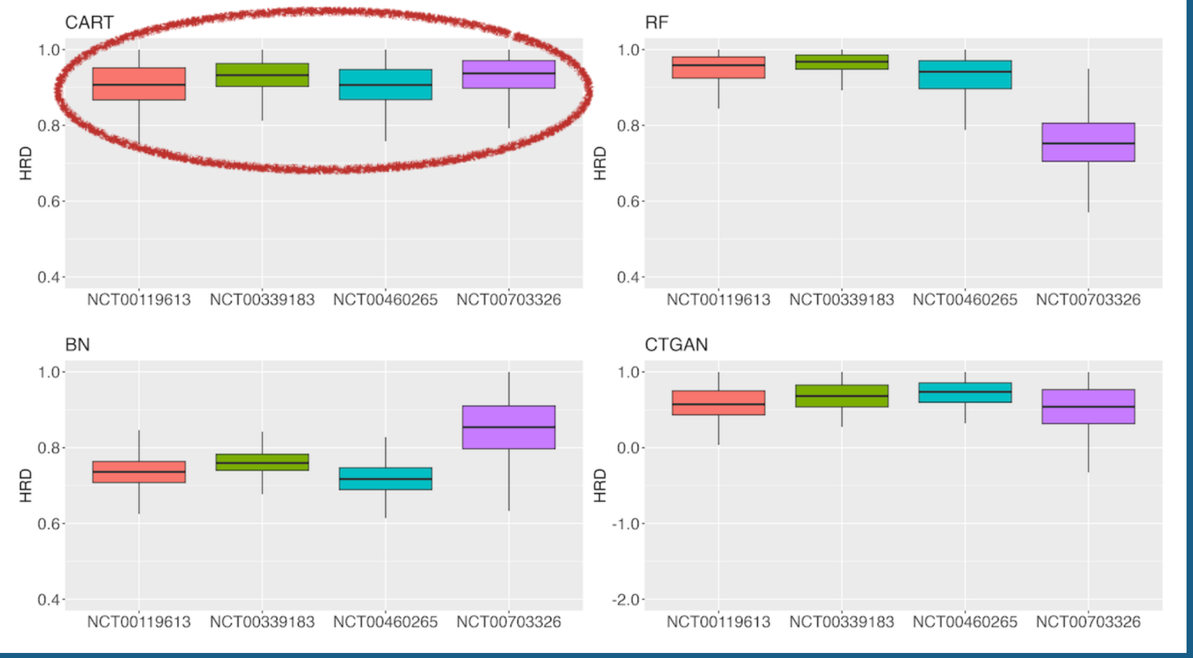
PFS

Figure 5. Box plot of progression-free survival hazard ratio distance (HRD) for each method and clinical trial. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; RF: random forest.



OS

Figure 6. Box plot of overall survival hazard ratio distance (HRD) for each method and clinical trial. BN: Bayesian network; CART: classification and regression tree; CTGAN: conditional tabular generative adversarial network; RF: random forest.



Summary of Results and Discussion

- Regardless of the test data used, CART and Random Forest often showed better results compared to other methods.
  - Bayesian Networks and CTGAN require a relatively large amount of training data for model construction, making them unsuitable for small datasets like clinical trial data.
- Among the methods that showed good results, CART was slightly superior to Random Forest.
  - CART is known to be a learning model that is prone to overfitting, whereas Random Forest employs bootstrap sampling, which has a mechanism to suppress overfitting. This difference is believed to be the reason for the observed outcomes.

- From the perspective of generating synthetic patient data that closely resembles the actual data, methods that tend to overfit are desirable.
- This approach is the exact opposite of the typical machine learning perspective, which aims to build predictive models.
- To achieve overfitting, one approach is to increase the number of predictive factors, but this also requires a large number of patients in the actual data.

Challenges of overfitting

Develop a method to compress data from multiple predictive factors into a single predictive factor.

This approach is expected to enable efficient generation of synthetic patient data that closely resembles actual data, even from small datasets of around a few hundred cases.

Regulatory Trends Regarding Synthetic Patient Data

- It appears that regulatory agencies such as the FDA and PMDA do not mention synthetic patient data in relation to the drug approval review process.
- However, as shown on the right, there are mentions of synthetic patient data in areas outside of drug approval review.

- FDA
  - Addressing the Limitations of Medical Data in AI
  - Leveraging precisionFDA and Synthetic Data to Improve Veteran Healthcare
- EMA
  - New methods for the effective use of real-world data and/or synthetic data in regulatory decision-making and/or in health technology assessment
- PMDA
  - AI を活用したプログラム医療機器に関する報告書

The Paradigm Shift Induced by Synthetic Data

- Examples of Microsoft's Phi-2 (AI)
  - All training data consists of synthetic data.
  - Although it is a relatively small language model (2.7B), it demonstrates performance comparable to that of a large language model with a parameter size 25 times greater.
- Nvidia Nemotron-4 340B
  - A large language model for generating synthetic data for AI training purposes.

- What these examples illustrate is that collecting all training data as real data is not only inefficient but also not necessarily beneficial for enhancing model performance.
- The same can be said for data in healthcare research, right?

It is not just a question of how to collect data, but rather a world that will increasingly focus on how to generate synthetic data.

Thank you for your attention.

DATACK

PHUSE Single Day Events Japan 2024

Supplementary Material

Datack Inc.
Hideki Ninomiya

The burden of RCTs is driving the search for new methodologies

- Challenges of RCTs
  - high cost
  - threatening the completion of a trial
  - increased patient burden
  - ethical concerns

- lin rare, orphan, or very serious drug indications
  - single-arm trials are conducted
    - sometimes employing external controls

Excerpts from PMDA notification in 2024

- ２ 医療用医薬品の一変申請において活用する場合
医療用医薬品の効能・効果又は用法・用量の追加あるいは変更に係る一変申請については、（中略）臨床試験の成績に代えて、**レジストリデータを評価資料として活用できる場合がある。**特に、**臨床試験の実施が容易ではない希少疾病や小児に係る効能・効果、用法・用量の追加等**については、レジストリデータを活用した一変申請の可能性について、検討する意義は大きい。

- ３ 医療用医薬品の電子添文改訂において活用する場合
既存のレジストリのデータから、有効性データが得られていない患者群において臨床試験成績の代わりとなる主要な根拠の結果が得られた場合、又は十分な有効性データが得られていない患者群において「有効性及び安全性に関する試験」を補完する重要な結果が得られた場合等に、（中略）**電子添文の注意喚起の内容等を改訂できる可能性がある。**

医療用医薬品の承認事項一部変更承認申請又は電子化された添付文書改訂においてレジストリデータを活用する際の留意点

１　背景と目的

昨今、本邦において、医療用医薬品の承認申請をはじめとして、Real World Data（以下「RWD」という。）を活用するための検討が更に進められている。RWD の一つであるレジストリデータについては、「「承認申請等におけるレジストリの活用に関する基本的考え方」について」（令和３年３月 23 日付け薬生薬審発 0323 第１号・薬生機審発 0323 第１号厚生労働省医薬・生活衛生局医薬品審査管理課長・医療機器審査管理課長連名通知。以下「基本的考え方通知」という。）により承認申請等への活用に関する考え方を示しているところである。

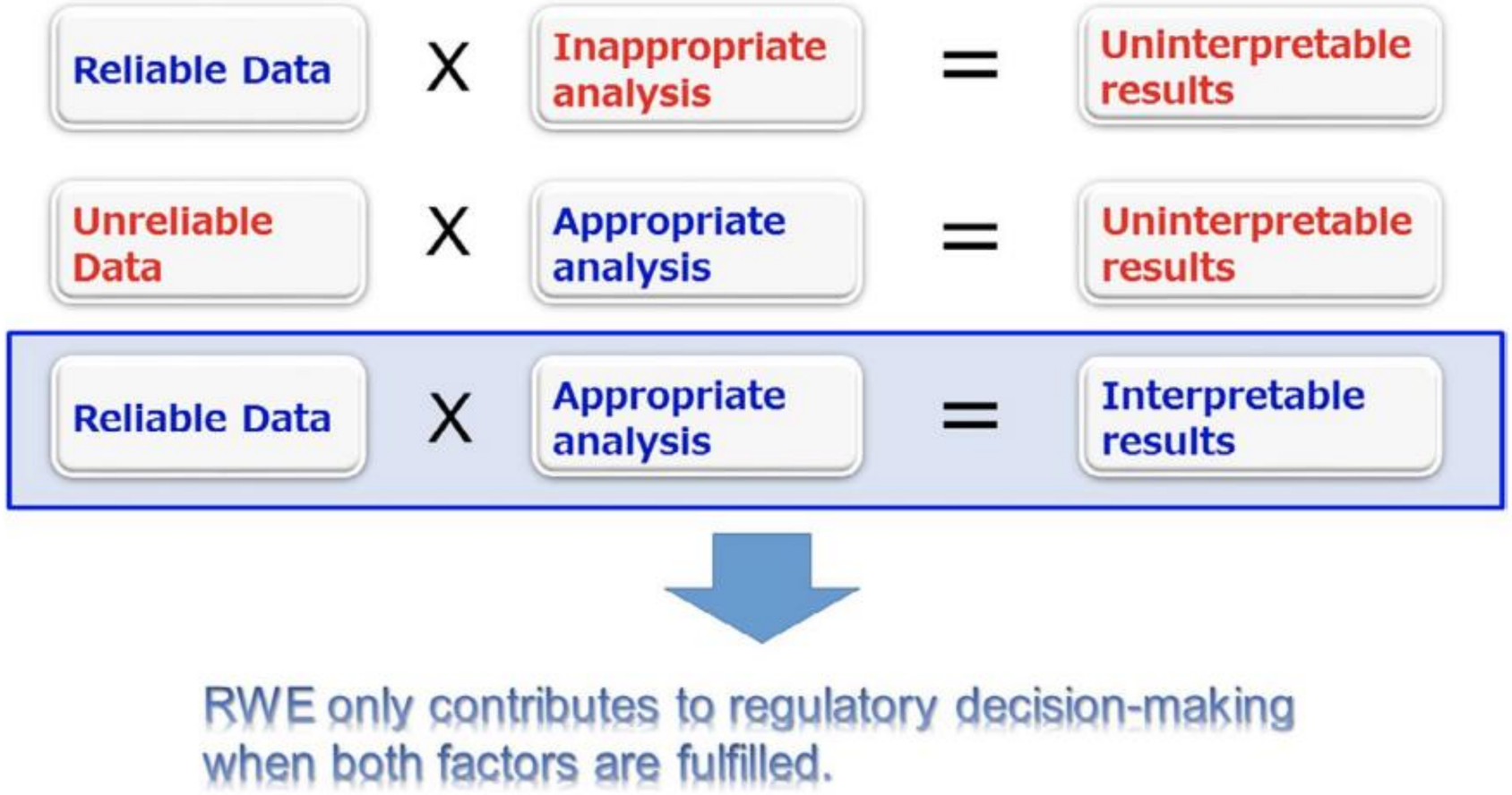
医療用医薬品の製造販売承認を取得するにあたり有効性及び安全性を示すためには、一般的には評価資料となる臨床試験を実施する必要がある。しかしながら、希少疾病のように患者数が極めて少ない等の理由により臨床試験の実施が困難な場合に、レジストリデータを外部対照や臨床試験の代わりとして用いることで医療用医薬品の開発が促進されることが期待される。

基本的考え方通知においても、レジストリデータの活用事例の一つとして、レジストリを臨床試験の補完又は代わりとして承認申請等における有効性及び／又は安全性の評価に用いることが挙げられている。その中で「既承認時に提出された臨床試験の組入れが一部の患者層に限られていたため、用法・用量の記載が臨床試験に基づき設定される又は添付文書において投与対象が制限されたものの、臨床現場では有効性及び安全性に大きな影響がないと考えられ使用されている場合に、適応拡大に係る開発や添付文書の改訂等に際してレジストリデータを用いて有効性及び／又は安全性を評価する場合が考えられる」ことが述べられている。

本通知では、基本的考え方通知に示されたレジストリデータの活用方法のうち、既承認医療用医薬品の効能・効果、用法・用量の追加に係る承認事項一部変更承認申請（以下「一変申請」という。）や電子化された添付文書（以下「電子添文」という。）における効能又は効果に関連する注意、用法及び用量に関連する注意、臨床成績等の改訂（削除、追加、記載内容の変更）において、レジストリデータを活用する際の留意点を示す。

医薬薬審発1004第4号 医療用医薬品の承認事項一部変更承認申請又は電子化された添付文書改訂においてレジストリデータを活用する際の留意点について

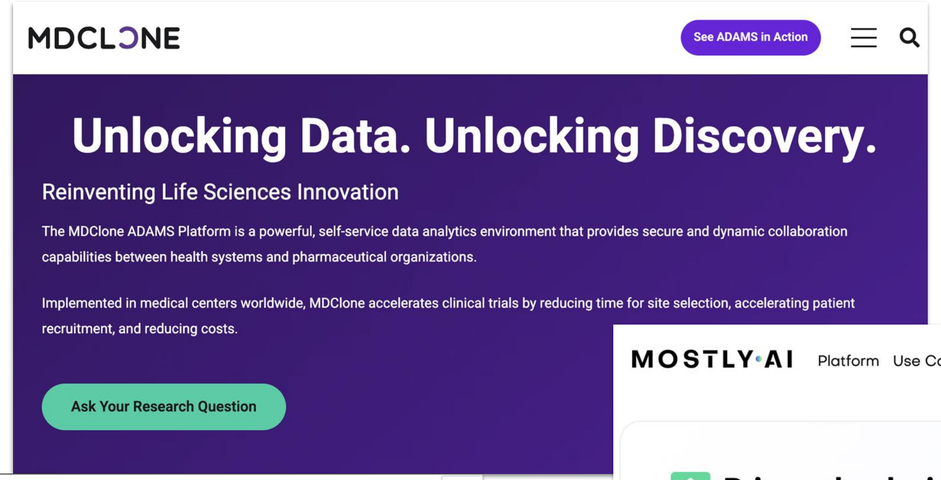
The positioning of today's two lectures

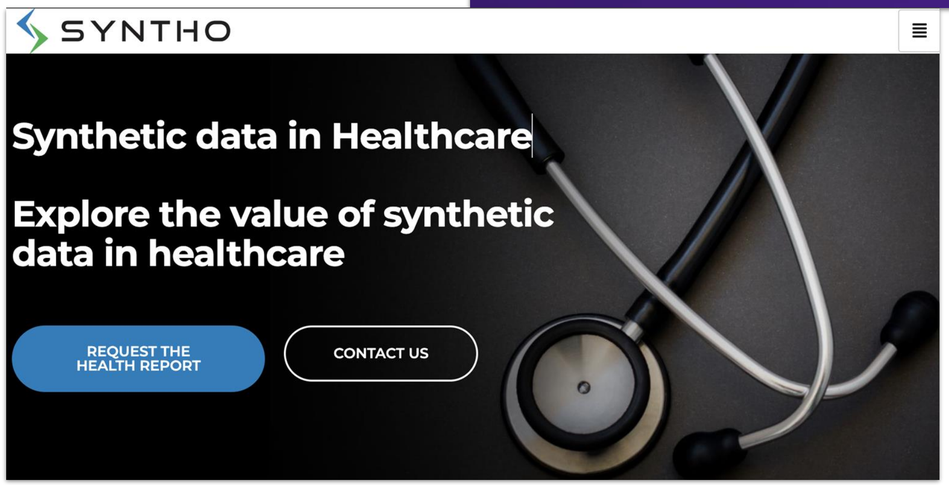


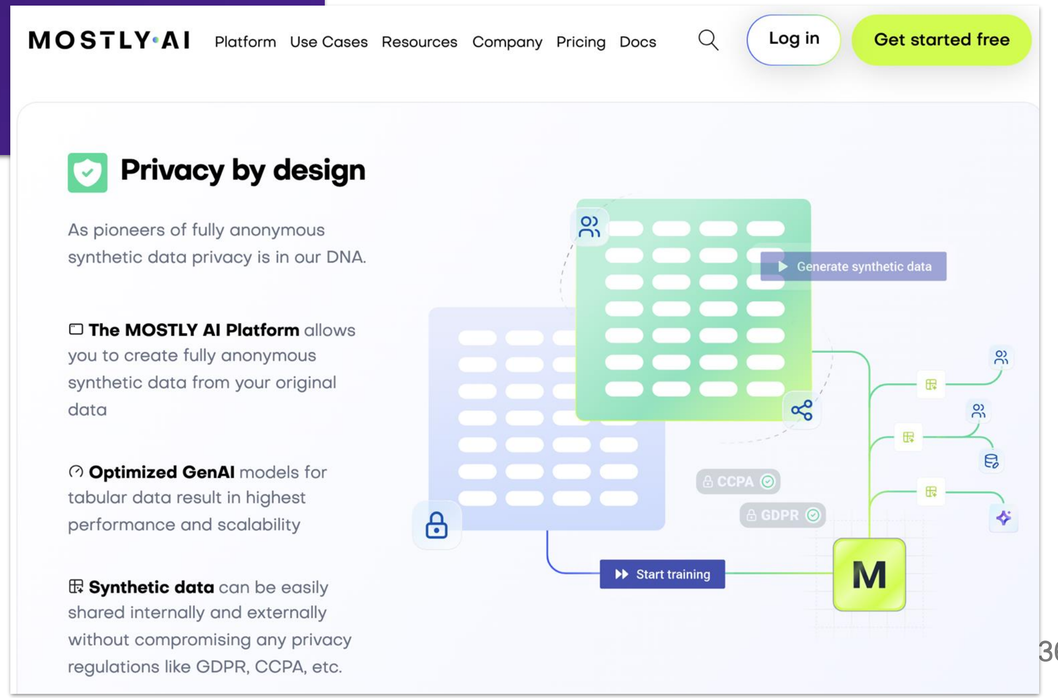
- Today's two topics
  - Target Trial Emulation: minimizing bias as "Appropriate analysis"
  - Synthetic Patient Data: expanding options of "Reliable Data"

Synthetic data are fast gaining traction among healthcare and life science

- Synthetic data is expected to accelerate its use under GDPR in Europe
- There are many growing startups providing synthetic data in the world








DATACK