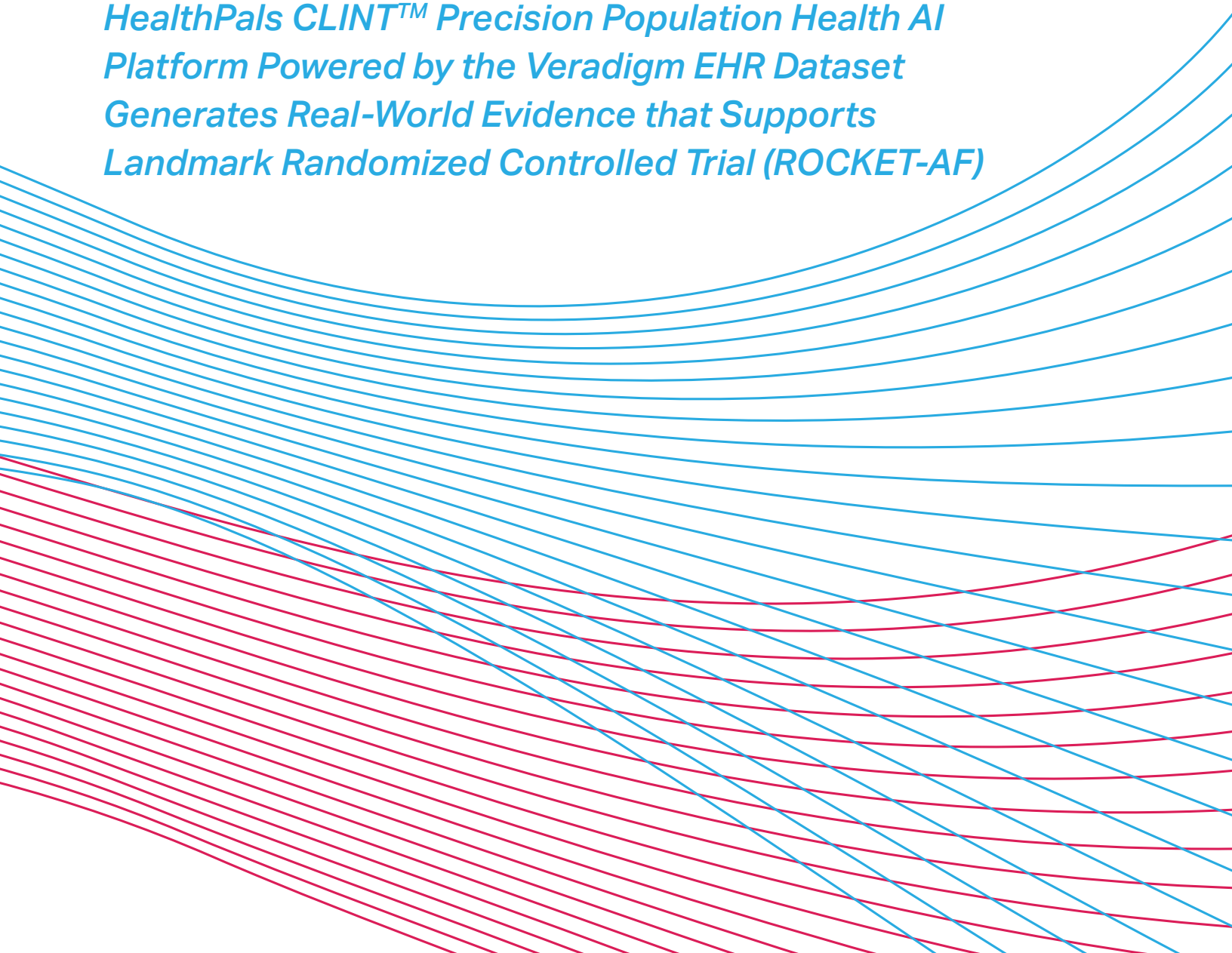
HEALTHPALS | veradigm

WHITEPAPER

First Demonstrated Replication of a Randomized Controlled Trial on Electronic Health Record Data Using the Clinical Intelligence Platform (CLINT™)

HealthPals CLINT™ Precision Population Health AI Platform Powered by the Veradigm EHR Dataset Generates Real-World Evidence that Supports Landmark Randomized Controlled Trial (ROCKET-AF)



EXECUTIVE SUMMARY

We are living in an era of unprecedented public health-related challenges, which have impacted many aspects of our day-to-day lives. These challenges have not spared clinical research, where over 80% of non-COVID related clinical trials have been indefinitely paused or canceled outright [1], and the ability to recruit patients and conduct studies that are still ongoing has been severely hampered. Clinical trials are also increasing in cost, duration, and complexity. Despite the vast resources invested in trial execution, there remains significant risk of not hitting study endpoints. However, many of these challenges can be mitigated through the use of real-world data and modern analytics capabilities.

In an effort to demonstrate the methods by which real-world data may be used to address these challenges, the HealthPals CLINT™ platform was used in conjunction with the Veradigm EHR Health Insights data set to successfully replicate the results of the ROCKET-AF trial, a large, international, randomized controlled trial in the cardiovascular space [2]. The ROCKET-AF study compares the safety and benefit of rivaroxaban versus warfarin in patients with nonvalvular atrial fibrillation. Through this replication effort, we have demonstrated a number of capabilities that provide opportunities to reduce clinical trial cost, duration and regulatory risk.

Key Takeaways

- HealthPals has developed CLINT™, a novel, guideline-encoded population health platform capable of rapidly generating statistical and machine learning-based insights at scale on real-world data (RWD), including EHR and claims data.
- The Veradigm data set is a large, rich, longitudinal data set capable of supporting RWD analyses.
- HealthPals in collaboration with Veradigm is capable of:
  - Quickly confirming the results of a large randomized controlled trial (ROCKET-AF) in RWD
  - Interfacing with clinical trial investigator teams to operationalize key trial objectives, longitudinally tracking outcomes and adjudicating events
  - Generating external control arms (ECA) for clinical trials to support regulatory submission.
- The above capabilities can enable clinical trials to be run with fewer patients, can act to reduce the burden of recruitment, as well as overall clinical trial cost, duration, and risk.

Introduction

RANDOMIZED CONTROL TRIALS: COST, COMPLEXITY, DURATION AND RISK

The Tufts Center for the Study of Drug Development (CSDD) is an academic non-profit think tank at Tufts University in Boston, dedicated to researching drug development. In 2016 the CSDD published an analysis of 106 randomly selected new drugs and found that total capitalized costs of drug development was increasing at a rate of 8.5% above general inflation.

In addition it was found that 57% of all clinical trial protocols in all phases had at least one substantial amendment with the most frequent changes including modifications and revisions to study volunteer demographics and eligibility criteria. The total median direct cost to implement a substantial amendment for Phase II and III protocols is $141,000 and $535,000, respectively [3].

In 2018 the CSDD also reported that rising protocol complexity was hindering study performance, cost and efficiency. They found that Phase III clinical trials have seen the highest increase in complexity during the past 10 years with the total number of endpoints rising 86%. As protocols grew more complex, site initiation and data management cycle times increased [4].

In 2020 the CSDD reported that, despite faster new drug approval phases, clinical trial times are taking longer. Data was analyzed spanning from 2014-2018 and it was found that while the mean approval phase decreased by 1.9 months, the overall trial times increased by 6.7 months [5].

RCTs also carry with them the risk of not hitting the endpoints with regard to safety or efficacy and not receiving regulatory approval. The Tufts 2016 study also showed that of drugs that enter clinical testing, the probability of a drug being approved is 11.83% [3].

IMPACT OF COVID-19 ON CLINICAL TRIALS

In a 2020 article in the Lancet, Aaron van Dorn describes how the COVID-19 pandemic has severely affected the ability to conduct trials in safe and effective ways [1]. Thousands of trials, approximately 80% of non-COVID-19 trials, have been suspended or stopped due to difficulties associated with lockdown restrictions.

Thousands of trials, approximately 80% of non-COVID-19 trials, have been suspended or stopped due to difficulties associated with lockdown restrictions.

Research firm GlobalData tracks the impact of COVID on the number of trials that have been disrupted, delayed, had slow or suspended enrollment. They have found that, while these numbers rose dramatically as a result of COVID-19, adjustments in clinical trial design strategies made by contract service providers and sponsors are enabling some trials to resume.

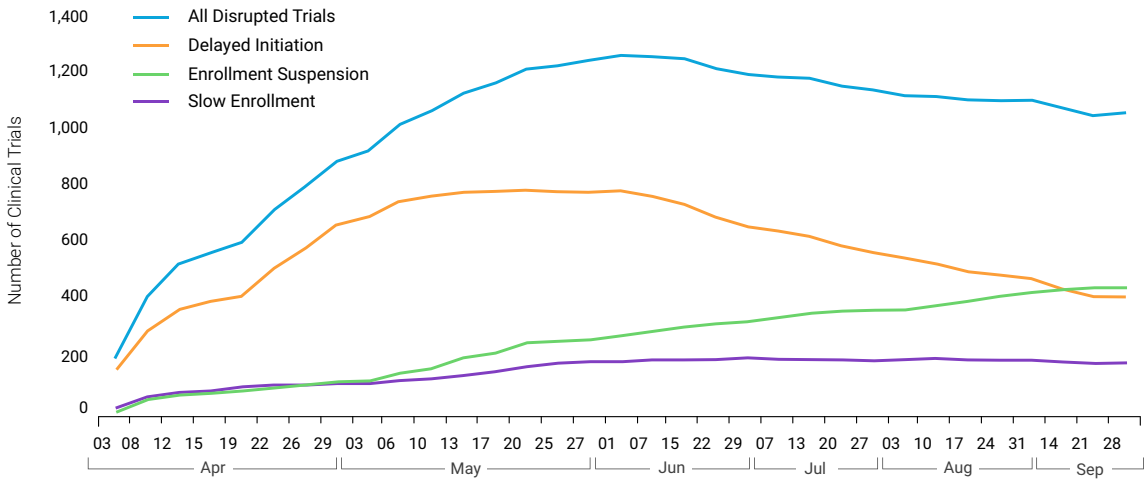


Figure 1. A timeline of clinical trials disrupted due to Covid-19.
Source: GlobalData

INTRODUCTION TO RWD, RWE AND ECA

The FDA defines real-world data (RWD) as:

> Data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources including:
>
> 1. Electronic health records (EHRs)
> 2. Claims and billing activities
> 3. Product and disease registries
> 4. Patient-generated data including in home-use settings
> 5. Data gathered from other sources that can inform on health status, such as mobile devices

Real-world evidence (RWE) is defined as:

> Clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD. RWE can be generated by different study designs or analyses, including but not limited to, randomized trials, including large simple trials, pragmatic trials, and observational studies (prospective and/or retrospective).

External control arms (ECAs) are a specific use case for RWD/RWE in which patient cohorts are derived from external, real-world data to provide a comparison control arm for an experimental arm in a clinical trial. ECAs are matched to experimental arms in such a way as to simulate the effects of randomization by:

- Reducing bias associated with confounding factors by distributing those factors equally across experimental and control groups
- Facilitating causal inference
- Providing the basis for statistical inference

ECAs can be used to reduce the necessary sample size for the study control arm and thereby reduce the duration

and cost of trial associated with patient recruitment. In addition, ECAs can be used to supplement submission to regulatory bodies and help to mitigate the risks associated with regulatory approvals.

The FDA uses RWD and RWE to monitor postmarket safety and adverse events and to make regulatory decisions while medical product developers use RWD and RWE to support clinical trial designs (e.g., large simple trials, pragmatic clinical trials) and generate observational studies to produce innovative, new treatment approaches.

CURRENT EFFORT

CLINT™ (HealthPals, Inc.), a novel, guideline-encoded population health platform capable of generating machine learning-based insights is now producing a diverse array of analyses on RWD in the US. HealthPals is an Innovation Collaborator of the American College of Cardiology (ACC) and is using data from Veradigm, a division of Allscripts and owner of the largest source of de-identified ambulatory patient records available, to produce the most comprehensive and up-to-date trends in the US clinical treatment landscape.

Using CLINT™ and in collaboration with Veradigm, HealthPals successfully duplicated the ROCKET-AF trial using RWD.

Prior efforts have sought to emulate clinical trials using claims data; until now, there have been no successful RCT replication studies based on electronic health record (EHR) data. Using CLINT™ and in collaboration with Veradigm, HealthPals successfully duplicated the ROCKET-AF trial using RWD. The ROCKET-AF inclusion and exclusion criteria were encoded into CLINT™ and control and experimental patients were identified. Control patients were then matched to the experimental patients using four propensity score-based cohort balancing methods: 1:1 patient matching, 2:1 patient matching, inverse probability of treatment weighting (IPTW), and standardized mortality ratio weighting (SMRW). Cohort matching performance was evaluated and longitudinal outcomes were calculated and compared between control and experimental RWD cohorts as well as between RWD and ROCKET-AF RCT cohorts.

Methods

The purpose of the current effort is to demonstrate the RWD research methods used to replicate the ROCKET-AF clinical trial. In addition, this work demonstrates the methods that would be used in the construction of an external control arm (ECA) that could be used to supplement or replace a clinical trial control arm.

DATA SOURCE

The database used for this effort belongs to Veradigm, a division of Allscripts and owner of the largest source of de-identified ambulatory patient records available. Veradigm assets cover over 150M patients and 250,000 clinicians from 35,000 practices. The database is composed of records from primary care (88.5%), cardiology (7.5%), and endocrinology (4%).

COHORT SELECTION

Operationalization

The HealthPals operationalization process leverages EHR field names as well as standardized ICD, LOINC, and NDC codes to capture diagnoses, lab information, and medications respectively. By leveraging standardized codes and names (strings) from the patient data the process allows for a more accurate and broader capture of medical concepts than can be achieved using codes alone. The process requires time, effort and a high level of both clinical and data science expertise.

Figure 2 illustrates an example of how Type 2 Diabetes Mellitus (T2DM) is captured using all available EHR data.

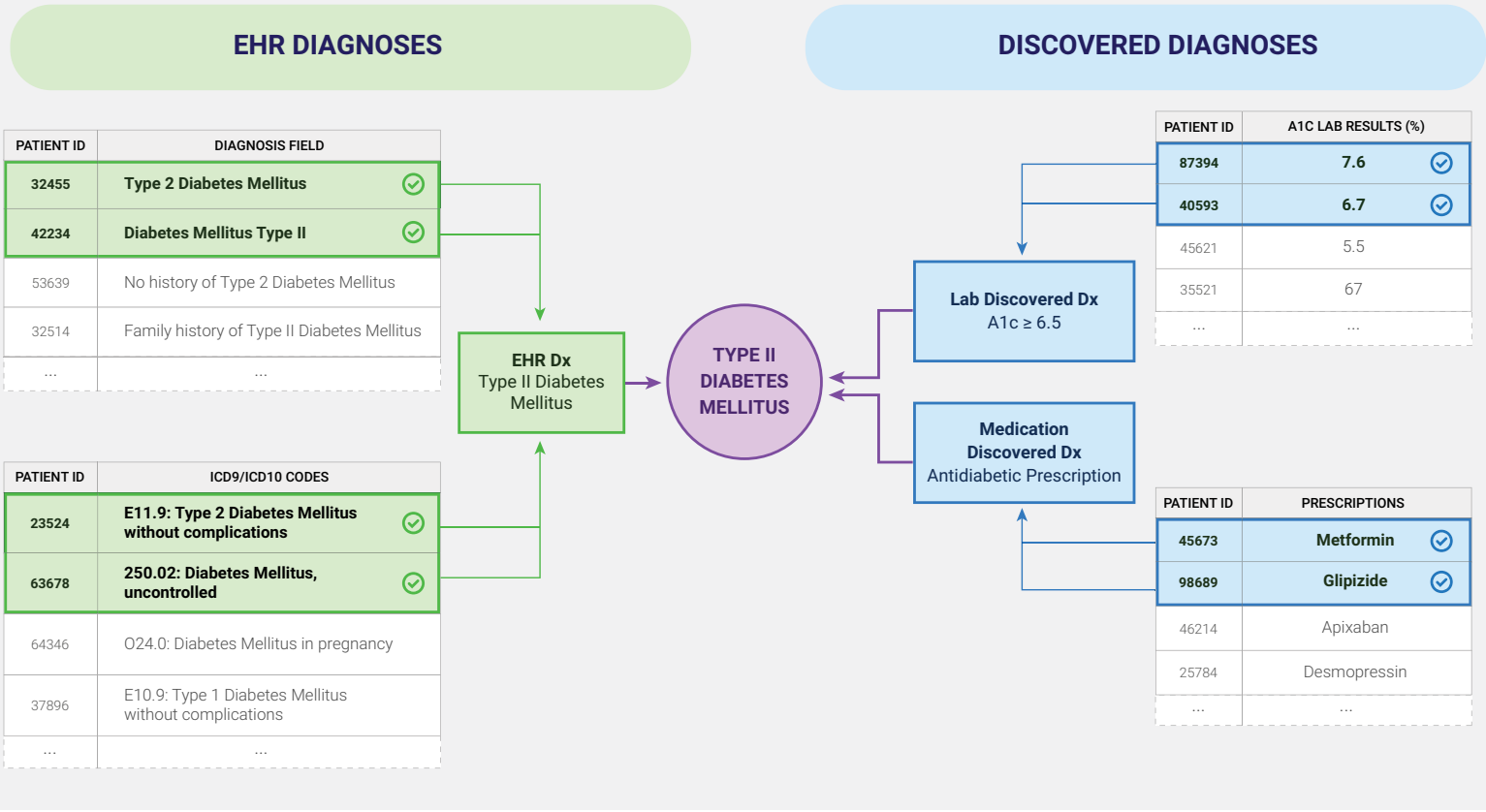


Figure 2. Example of clinical concept operationalization.
Using Type II Diabetes Mellitus as an example, the figure illustrates multiple ways in which EHR data can be used to identify a diagnosis in the patient records.

Application of Inclusion/Exclusion Criteria

ROCKET-AF inclusion and exclusion criteria as well as outcomes were operationalized and mapped to RWD. The inclusion and exclusion criteria were encoded into CLINT™ and applied to the data set. The subset of eligible patients were identified.

Patients who were at any point eligible for the study and were on Rivaroxaban were identified as potential experimental arm patients. Patients who were eligible and on Warfarin were identified as potential control arm patients. To ensure adequate covariate capture, a 365-day wash-in period was enforced: for a given patient, an encounter date was eligible to be an index date for the study if it occurred 365 days later than that patient's first encounter in the dataset. The index date was required to be before a patient's final encounter date in the dataset to allow for follow-up. The index date was defined as the date of a patient's first encounter on which the patient met all ROCKET-AF eligibility criteria and had been prescribed either warfarin or rivaroxaban on or prior

to that encounter date. If the prescription was prior to the encounter date, the warfarin or rivaroxaban prescription was required to be the most recent OAC prescription; i.e., a prescription of apixaban, dabigatran, or edoxaban after the prescription of rivaroxaban/warfarin rendered the patient ineligible.

COHORT BALANCING

Cohort balancing using propensity scores consists of two steps:

1. algorithmically predicting which patients are more or less likely to be assigned to the rivaroxaban arm using labeled rivaroxaban arm and warfarin arm patient data;
2. using the probability associated with this prediction to match and/or weight control patients so that the prognostic variables of the warfarin arm patients are matched to those of the rivaroxaban arm patient population.

Propensity Score Generation

Propensity scores were generated for rivaroxaban and warfarin patients using a logistic regression classifier and prognostic variables (clinical variables believed to be confounders for the outcomes of interest). Missing continuous variables were imputed using the CLINT™ Platform. Imputed values were used only in the propensity score calculation; variable summary statistics and standardized differences were reported only on non-imputed values.

Propensity Score Matching and Weighting

Propensity score matching methods match to each experimental arm patient one or more control arm patients having the nearest propensity scores to that patient within some minimum caliper distance. Matching was done without replacement. Patients were matched at 1:1 and 2:1 ratios using a caliper distance of 0.01.

Propensity score weighting methods assign weights to each patient based on the patient's likelihood of being assigned to the treatment arm in such a way that when the

weighted values of the prognostic variables are compared, these values are balanced between control and experimental arms.

Patients were weighted using inverse probability of treatment weighting (IPTW) with stabilized weights and standardized mortality ratio weighting (SMRW).

IPTW

$$W_{i,T} = \frac{N_T}{N_T + N_C} \cdot \frac{1}{p_i}, \quad W_{i,C} = \frac{N_C}{N_T + N_C} \cdot \frac{1}{1 - p_i}$$

SMRW

$$W_{i,T} = 1, \quad W_{i,C} = \frac{p_i}{1 - p_i}$$

COMPARISON OF METHODS

The efficacy of the four methods' ability to balance cohorts was determined by using the average standardized difference across all variables with standardized difference being defined by the following formula:

$$\text{Standardized Difference} = \frac{\mu_T - \mu_C}{\sqrt{\frac{Var_T + Var_C}{2}}}$$

ANALYSIS

Outcomes

Five event-based outcomes were recorded and analyzed: stroke of any kind, arterial embolism, myocardial infarction, major bleeding, and clinically relevant non-major (CRNM) bleeding. Stroke events included both ischemic and hemorrhagic strokes. Major bleeding was defined as a bleed which occurred in a critical anatomical site: intracranial, subdural, spinal, pericardial, retroperitoneal, articular, ocular, and intramuscular

bleed associated with compartment syndrome. Hemorrhagic strokes were classified both as stroke events and major bleeding events. CRNM bleeding was defined as bleeding that occurred in any of the following sites: esophageal, gastroduodenal, genitourinary, lower gastrointestinal or unspecified gastrointestinal, or other unspecified site. Death information was not available, and was not included as an endpoint.

Similar to ROCKET-AF, the primary endpoint was specified as the composite endpoint of either stroke or arterial systemic embolism. The principal safety endpoint, a composite endpoint of either major or CRNM bleeding, was also analyzed.

To mirror the analyses conducted in the ROCKET-AF trial, analyses were performed in two regimes: the per-protocol population and the intention-to-treat population. In the per-protocol population, patient follow-up was right-censored if the patient switched to an oral anticoagulant other than the one specified by their study arm. For patients in the rivaroxaban arm, a prescription of warfarin, apixaban, dabigatran, or edoxaban constituted a right-censoring event; for patients in the warfarin arm, prescription of rivaroxaban, apixaban, dabigatran, or edoxaban caused right-censoring of future encounters.

In the intention-to-treat population, patients were only right-censored by their last encounter in the dataset. Patients were monitored for outcome events for their entire duration after their index date in the dataset, regardless of whether they remained on the medication specified by their respective arm.

For the best-performing cohort balancing method, we calculated the weighted prevalence of outcomes in both the intention-to-treat and the per-protocol population, and reported outcomes as a rate per 100 patient-years. Within the per-protocol population, we used Cox proportional-hazards models, with rivaroxaban/warfarin arm assignment as the only covariate, to determine hazard ratios, 95% confidence intervals, and P values for all five outcomes of interest. We report cumulative incidence curves for the composite outcome of stroke or systemic embolism, for both the per-protocol and intention-to-treat cohorts.

Results

COHORT SELECTION

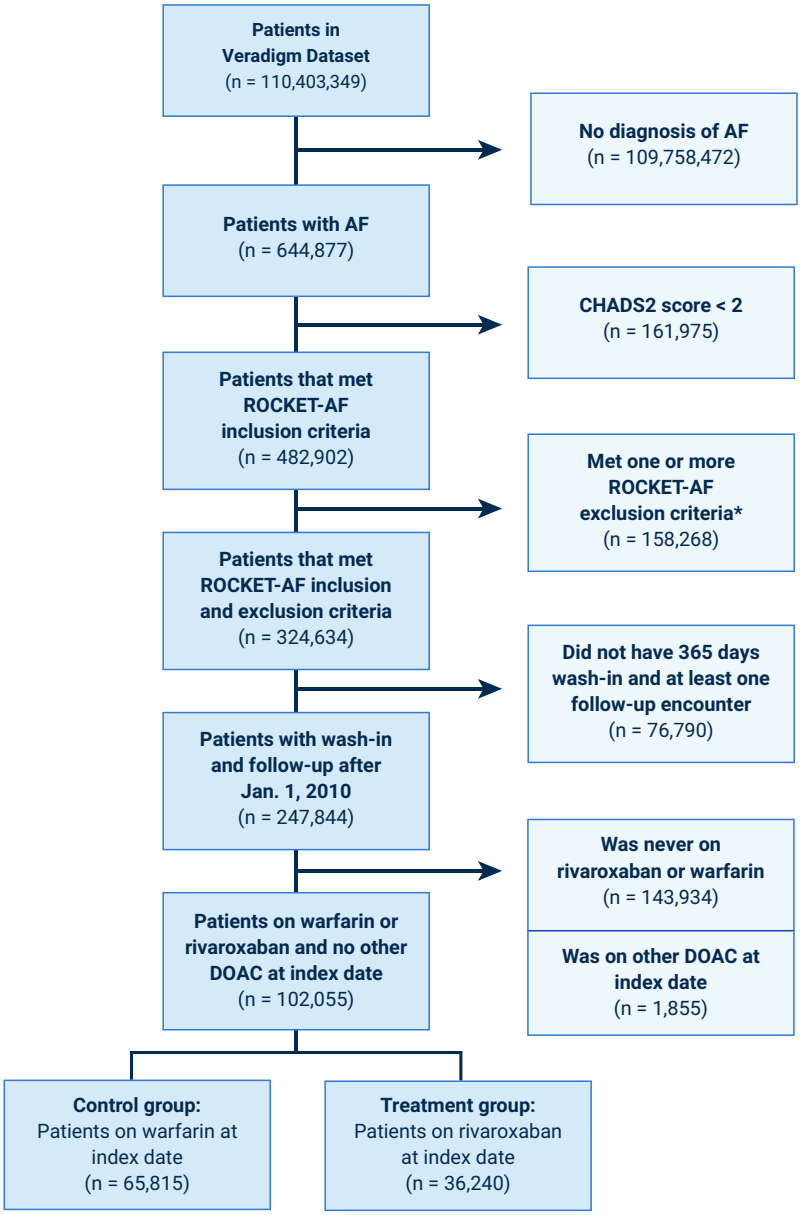


Figure 3: Flowchart cohort diagram detailing the critical filtering steps and the resulting number of patients at each stage of the process.

Abbreviations: AF: Atrial fibrillation; DOAC: direct oral anticoagulant; CHADS2: congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism.

The cohort diagram for the study is shown in Figure 3. From over 110M patients in the Veradigm dataset, 102,055 patients met all the criteria required to be included in the study. Of these, 65,815 (64%) were on warfarin at their index date, and 36,240 were on rivaroxaban.

Characteristic	Rivaroxaban (N= 36,240)	Warfarin (N= 65,815)
Age [years], median (IQR)	76 (69-86)	79 (73-87)
Male sex, N (%)	18,822 (52)	34,629 (53)
BMI [kg/m²], median (IQR)	28.8 (25.5-32.3)	28 (24.9-31.6)
Systolic BP [mmHg], median (IQR)	129 (119-140)	128 (118-140)
Diastolic BP [mmHg], median (IQR)	74 (68-80)	72 (65-80)
Paroxysmal AF, N (%)	25,969 (72)	42,497 (65)
Persistent AF, N (%)	3,639 (10)	4,366 (6.6)
Permanent AF, N (%)	8,107 (22)	20,673 (31)
ACE inhibitors, N (%)	17,689 (49)	33,395 (51)
$CHADS_2$ score, mean ± SD	2.4±0.644	2.48±0.689
Congestive heart failure, N (%)	9,215 (25)	19,178 (29)
Hypertension, N (%)	31,137 (86)	54,769 (83)
Diabetes mellitus, N (%)	19,012 (52)	30,961 (47)
Peripheral vascular disease, N (%)	4,232 (12)	8,511 (13)
Chronic obstructive pulmonary disease, N (%)	5,013 (14)	9,528 (14)
Hemoglobin [g/dL], median (IQR)	21.9 (14.1-23.5)	20.1 (13.5-23.2)
Platelet count [1,000/µL], median (IQR)	187 (133-240)	191 (142-240)
INR, median (IQR)	2.6 (1.1-6.3)	2.5 (1.9-4)

Table 1. Characteristics of the cohorts at index date.

Abbreviations: IQR: interquartile range; BMI: body mass index; BP: blood pressure; AF: atrial fibrillation; ACE: angiotensin-converting enzyme; INR: international normalized ratio; SD: standard deviation; CHADS2: congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism.

Characteristics of the warfarin and rivaroxaban cohorts are described in Table 1. Notable differences include age (rivaroxaban patients had a median of 76, while warfarin patients had a mean age of 79), BMI (28.8 in rivaroxaban patients, 28.0 in warfarin patients), hemoglobin (rivaroxaban patients had a median of 21.9 g/dL compared to warfarin patients' 20.1 g/dL), and breakdown of paroxysmal/persistent/permanent AF (72/10/22 for rivaroxaban patients, 65/6.6/31 for warfarin patients).

MATCHING

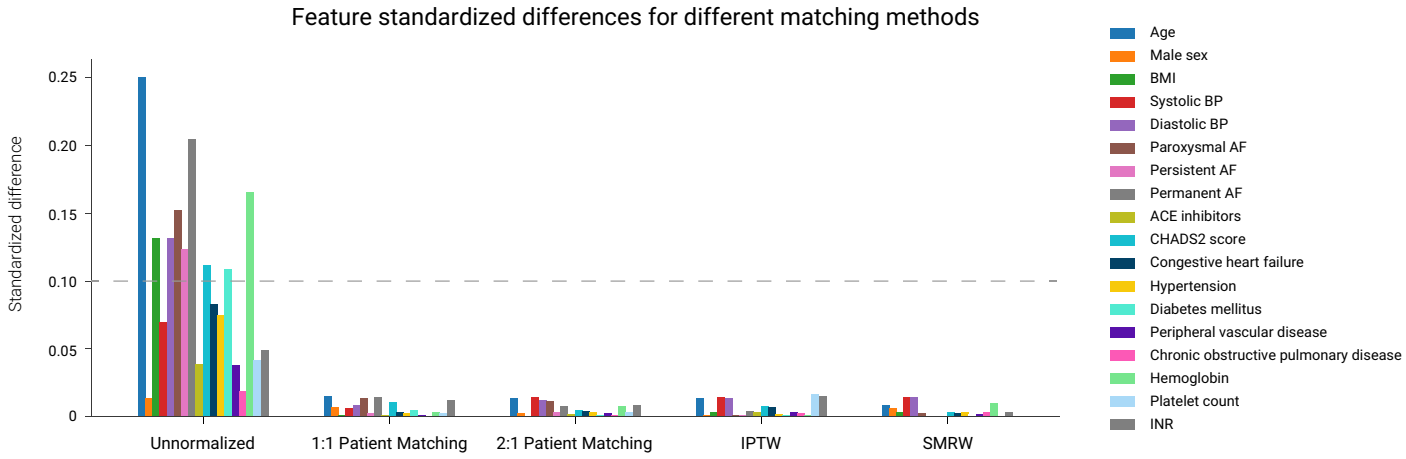


Figure 4. Standardized Differences across prognostic features for different matching methods: The standardized differences were calculated between rivaroxaban and warfarin cohorts prior to balancing (Unnormalized) and for each balancing method. The standardized difference was dramatically reduced using all of the methods explored, with SMRW showing the largest decrease in average standardized difference.

Abbreviations: IPTW: inverse probability of treatment weighting; SMRW: standardized mortality ratio weighting; BMI: body mass index; BP: blood pressure; AF: atrial fibrillation; ACE: angiotensin-converting enzyme; INR: international normalized ratio; CHADS2: congestive heart failure, hypertension, age ≥75 years, diabetes mellitus, prior stroke or transient ischemic attack or thromboembolism.

Each of the four methods demonstrated substantial reduction in standardized differences for all variables. While several variables had a standardized difference of over 0.10 prior to cohort balancing (the average standardized difference across the prognostic features was also above 0.10), no variable had a standardized difference larger than 0.02 after balancing the cohorts with any of the methods.

It was found that SMRW produced the greatest reduction in standardized difference across prognostic variables (average standardized difference of 0.0045, compared to 0.0062 for 1:1 patient matching, 0.0058 for 2:1 patient matching, and 0.0062 for IPTW). For this reason, the subsequent analysis is focused on cohorts which were weighted using SMRW.

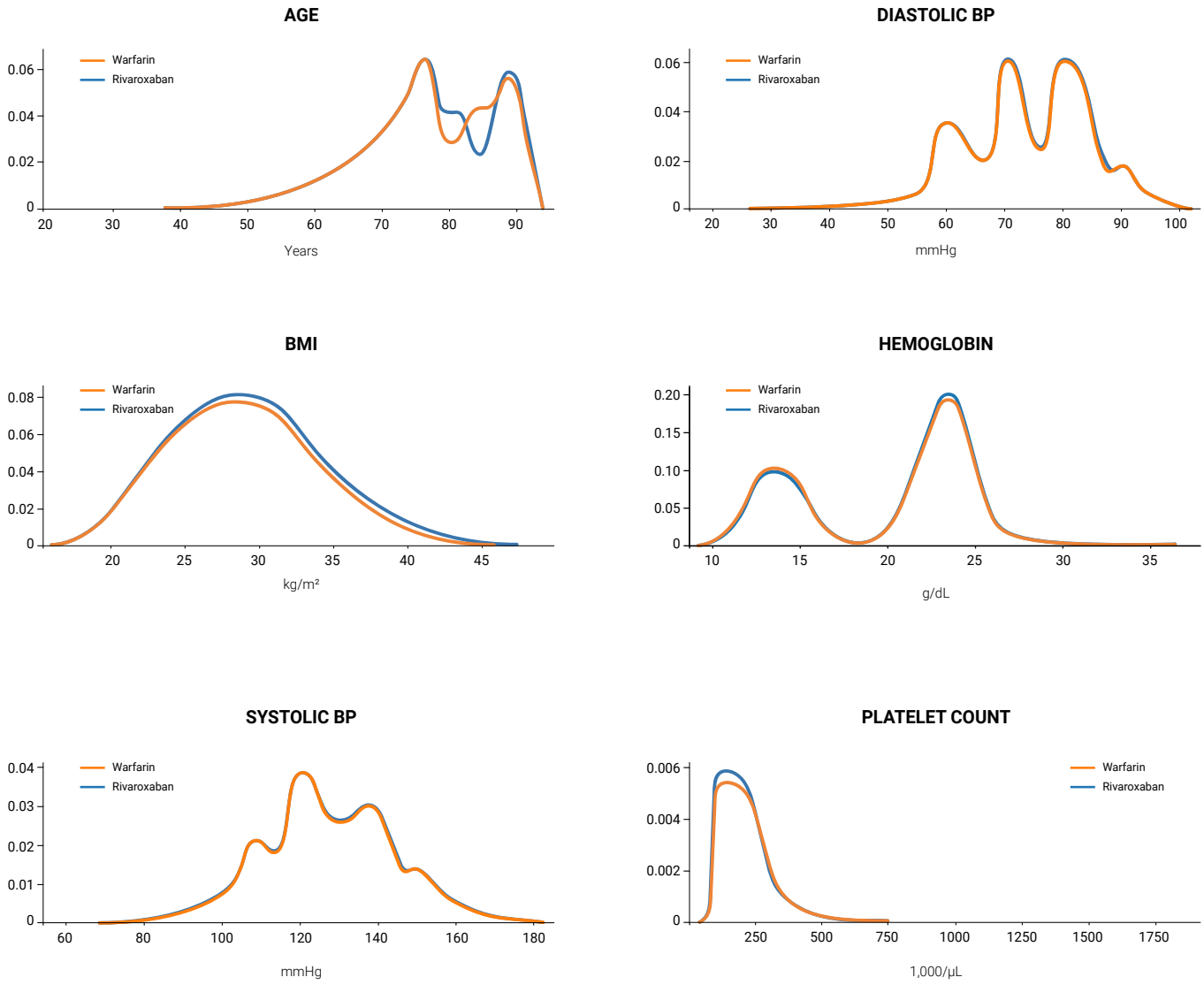


Figure 5. Histograms of continuous variables after weighting using SMRW.

Abbreviations: IPTW: inverse probability of treatment weighting; SMRW: standardized mortality ratio weighting; BMI: body mass index; BP: blood pressure; AF: atrial fibrillation; ACE: angiotensin-converting enzyme; INR: international normalized ratio.

SMRW-based cohort balancing resulted in qualitatively similar distributions for many of the continuous variables, as demonstrated in Fig. 5.

OUTCOMES

	Current Study				ROCKET-AF Study			
	Rivaroxaban Event Rate	Warfarin Event Rate	Hazard Ratio (95% CI)	P Value	Rivaroxaban Event Rate	Warfarin Event Rate	Hazard Ratio (95% CI)	P Value
Primary Endpoint: Stroke/SE	1.50	1.83	0.81 (0.75–0.87)	<.001	1.7	2.2	0.79 (0.65–0.96)	0.002
Stroke	1.19	1.52	0.77 (0.70–0.83)	<.001	1.65	1.96	0.85 (0.70–1.03)	0.092
Systemic embolism	0.33	0.37	0.92 (0.78–1.08)	0.32	0.04	0.19	0.23 (0.09–0.61)	0.003
Myocardial infarction	0.46	0.60	0.74 (0.64–0.84)	<.001	0.91	1.12	0.81 (0.63–1.06)	0.12
Principal Safety Endpoint: Major/ CRNM bleed	4.74	4.18	1.08 (1.03–1.13)	0.001	14.9	14.5	1.03 (0.96–1.11)	0.44
Major bleed	0.48	0.73	0.63 (0.56–0.72)	<.001	3.6	3.4	1.04 (0.90–1.20)	0.58
CRNM bleed	4.34	3.58	1.17 (1.11–1.23)	<.001	11.8	11.4	1.04 (0.96–1.13)	0.35

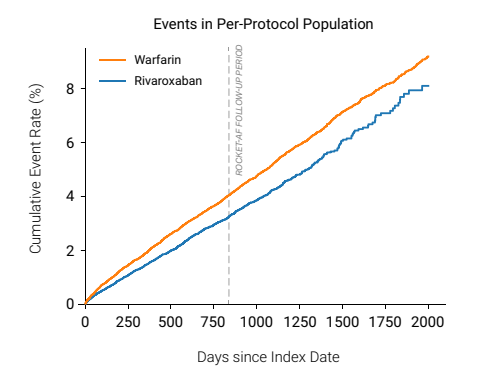
Table 2. Event rates are calculated as the number of events per 100 patient-years. Event rates are calculated in per-protocol population for all outcomes. Hazard ratios are for the rivaroxaban group as compared to the warfarin group.

Abbreviations: CI: confidence interval; SE: systemic embolism; CRNM: clinically relevant nonmajor.

SMRW-weighted rates for each of the five outcomes, calculated in the per-protocol population and normalized to 100 patient-years, are presented in Table 2. For the primary endpoint for stroke or systemic embolism, the hazard ratios for our study were within the 95% confidence intervals of the ROCKET-AF study (ours: HR 0.81, 95% CI 0.75–0.87; ROCKET-AF: HR 0.79, 95% CI 0.65–0.96). The principal safety endpoint was also in concordance (ours: HR 1.08, 95% CI 1.03–1.13; ROCKET-AF: HR 1.03, 95% CI 0.96–1.11), although, due to the increased sample

size in our study, we observed statistically significantly higher rate of bleeding in the rivaroxaban group as compared to the warfarin group. For strokes, systemic embolisms, myocardial infarctions, and CRNM bleeds, our results are directionally consistent with an acceptable range of difference with those of ROCKET-AF. We observed statistically significantly lower rates of strokes, myocardial infarctions, and major bleeds, while the ROCKET-AF study did not establish superiority regarding these outcomes. Additionally, we did not observe a significant difference among systemic embolism rates between the

STROKE / EMBOLISM COMPOSITE ENDPOINT:
CURRENT STUDY



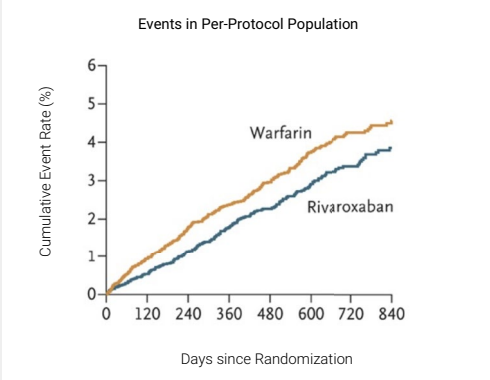
No. at Risk

Rivaroxaban	36240	25925	18436	12760	8107	4790	2241	1136	516
Warfarin	36272	28538	22718	17858	13478	9840	6588	4821	3584

No. at Risk

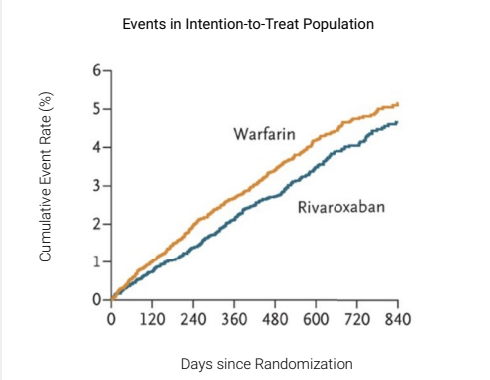
Rivaroxaban	36240	26995	19617	13841	8932	5349	2533	1308	600
Warfarin	36272	29508	23999	19184	14694	10935	7414	5514	4137

STROKE / EMBOLISM COMPOSITE ENDPOINT:
ROCKET-AF STUDY



No. at Risk

Rivaroxaban	6958	6211	5786	5468	4406	3407	2472	1496
Warfarin	7004	6327	5911	5542	4461	3478	2539	1538



No. at Risk

Rivaroxaban	7081	6879	6683	6470	5264	4105	2951	1785
Warfarin	7090	6871	6656	6440	5225	4087	2944	1783

Figure 6. Cumulative stroke and embolism events in RWD (left) compared with cumulative events from the ROCKET-AF study (right) [2].

cohorts of our study, although the ROCKET-AF trial noted significantly lower rates in the rivaroxaban arm. While the ROCKET-AF study reported slightly increased (but not significant) rates of CRNM bleeding in the rivaroxaban population, we observed significantly higher rates of CRNM bleeds in the rivaroxaban population of our study.

Figure 6 shows cumulative stroke and embolism events in our study (left) compared to the cumulative stroke and embolism events from the ROCKET-AF study (right). Notably, the curves in our study extend to 2,000 days after the index date, while the curves in the ROCKET-AF study are censored at 840 days after the randomization date. Both studies demonstrate a consistent divergence in cumulative event rates across the study duration in both the per-protocol populations and the intention-to-treat populations.

Discussion

This is, to our knowledge, the first large-scale replication of a clinical trial using EHR data. Similar studies have replicated trials using insurance claims databases [6,7] using ICD codes alone to map clinical concepts. As insurance claims are centered around billing and reimbursement, claims data has several shortcomings regarding diagnosis capture. For example, there is significant lag between the time at which a diagnosis, prescription or procedure appears in a patient's EHR record and the time at which it appears in claims data. Clinical information that may elucidate a patient's health status but which may not be tied directly to a billed procedure or prescription may not appear in the claims data at all.

Across different disease states, the diagnosis capture rate [using insurance claims data] can fall to as low as 9%. This has noteworthy implications in cohort selection, comorbidity assessment, and outcomes capture.

EHR data is a tremendously valuable raw resource. To extract the maximum value from this resource, the data must be cleaned and mapped to relevant clinical concepts. The effort required in these stages is often greatly underestimated. These processes are complex, time-consuming, and require deeply integrated data science and clinical expertise. Depending on how it is performed, clinical concept mapping can yield varying levels of value. Previous efforts have mapped clinical concepts using ICD codes alone rather than utilizing all

available data. Compared with querying EHR data using all available structured fields, querying claims data using ICD codes alone captures between 33%-87% of patients with a diagnosis of heart failure [8]. Across different disease states, the diagnosis capture rate can fall to as low as 9% [9,10]. This has noteworthy implications in cohort selection, comorbidity assessment, and outcomes capture.

CLINT™'s medical and data-science-based engine can uniquely capture a much larger percentage of otherwise occult diagnoses than what is achievable using ICD codes and claims data alone. This can be accomplished through CLINT™'s enhanced recognition and contextualization of medications, procedures, and laboratory results. The HealthPals medical and technical teams have worked very closely to build into CLINT™ the clinical context that enables identification of diagnoses through either pathognomonic lab results, treatments, or a combination of the two, and to do so for any particular disease state(s)/condition(s). This medical context is also leveraged to identify associated comorbidities to that primary condition (for example: diabetes and chronic kidney disease, or heart failure and end stage liver disease). The result is CLINT™'s ability to identify disease diagnoses and relevant comorbidities with a high degree of fidelity. This diagnostic capability is a differentiating feature of CLINT™.

COHORT SELECTION

The CLINT™ platform maps cardiovascular concepts, diseases, and outcomes at a high level of detail and reliability. This can be leveraged to quickly implement novel and precise inclusion and exclusion criteria to define cohorts of interest. These capabilities also enable calculation of longitudinal outcomes to not only replicate existing clinical trials, but also uncover candidates for indication expansions, strengthen adaptive pathways within trials, monitor post-approval safety, and conduct fully virtual clinical trials.

Our team of clinician experts works directly with study research teams to translate eligibility requirements, outcomes of interest, and relevant clinical concepts to identify RWD patient cohorts (e.g., patients eligible for inclusion in an external control arm). HealthPals supports prospective ECA use cases by providing the study research teams with review/oversight of ECA patient eligibility and access to RWD outcomes via a dynamic dashboard powered by CLINT™.

By leveraging the Veradigm dataset, CLINT™ is capable of running analyses on very large sample sizes. In the current effort, application of ROCKET-AF inclusion and exclusion criteria resulted in the generation of a real-world rivaroxaban experimental arm five times larger than the ROCKET-AF experimental arm (36,240 vs. 7,131) and a real-world warfarin control arm nine times larger than the ROCKET-AF control arm (65,815 vs. 7,133).

COHORT BALANCING

The balancing of cohorts with respect to prognostic variables reduces selection bias, facilitates causal inference, and provides the basis for statistical inference that allows for direct comparison between cohorts derived from RWD; differences observed between cohorts can be attributed to the effect of the treatment being investigated. This same methodology also enables the comparison between clinical trial experimental arms and RWD-derived external control arms.

The methods employed, combined with the detailed capture of clinical concepts, enables balancing of cohorts with regards to prognostic variables. The cohort balancing methodologies reduced the average standardized difference across all prognostic features twenty-fold: from over 0.10 (10%) to below 0.0045 (0.45%).

OUTCOMES

For the survival analysis comparing event rates per 100 patient-years, our results are directionally consistent with the ROCKET-AF study results; patients who received rivaroxaban had fewer strokes or embolisms, and slightly more major/CRNM bleeding events, than patients who received warfarin. Additionally, the results of our study were able to determine statistical significance of the effect of rivaroxaban on the reduction of strokes and myocardial infarction, and on the slight increase in bleeding rates, as compared to warfarin. Due to the smaller sample size of the ROCKET-AF study, these trends were not apparent from the study results. The current study also demonstrated significantly lower rates of major bleeding, and significantly higher rates of CRNM bleeding than the rates reported in the ROCKET-AF study. These differences may be due to major and CRNM bleeding definitions; in our study, major and CRNM bleeding events were divided primarily based on site.

The ROCKET-AF study noted significantly higher rates of major gastrointestinal bleeding among patients on rivaroxaban; in our study, a majority of gastrointestinal bleeding was classified as CRNM bleeding. The learning from these discrepancies is that the source and organization of RWD is a key determinant of the outcomes. One means to improve this analysis would be to add linked claims data to this vast EHR dataset, which would have ensured even more accuracy of captured clinical events; this addition is something HealthPals is actively investigating.

As noted above, the event-rate curves for the primary endpoint of cumulative stroke and embolic events were consistent with the results of the ROCKET-AF study, i.e. the rivaroxaban group displayed lower stroke and stroke/systemic embolism rates than the warfarin group. Importantly, the CLINT™ analysis of RWD also demonstrated a significant reduction in MI in real-world rivaroxaban-treated patients, whereas ROCKET-AF failed to show this statistical difference despite a trend in this direction. Indeed, all primary and secondary endpoints of ROCKET-AF were mirrored in this RWD analysis, trending or showing statistical significance in the same direction as the published RCT, with the exception of major bleeding which was noted above. Clinically, the takeaway from this RWD analysis would lead to the same conclusions as ROCKET-AF did: that rivaroxaban is superior to warfarin in reducing stroke in AF patients.

Since the publishing of ROCKET-AF, rivaroxaban has gone on to obtain nine additional FDA indications for use, including for secondary prevention of cardiovascular events such as MI. In this RWD analysis, not only was the rate of MI found to be significantly lower in the rivaroxaban group, many of these patients (who already had the atrial fibrillation treatment indication) would have qualified for the secondary prevention of MI under this current use label for rivaroxaban. Future HealthPals work will isolate a stable atherosclerotic vascular disease cohort to determine whether they similarly demonstrate a significant reduction in MI and stroke rates–essentially a replication of the COMPASS trial [11]–and whether this type of RWD analysis could have produced sufficient evidence to support a FDA submission for a new treatment indication, without the need for another costly and time-consuming RCT (or at least an RCT of that size and duration). In summary, with regard to identifying additional or repurposing indications for drugs already in the market, this analysis illustrates the power of using a robust and capable analytics engine like CLINT™ to analyze large real-world datasets.

A last point on these results is the length of the observation period. This current RWD effort followed patients for more than 2,000 days after the index date, while the ROCKET-AF study reported only 840 days of outcomes after the randomization date. In the HealthPals results, the protective effect of rivaroxaban was demonstrated to be a full 4 years longer than what was reported in the original RCT. Clinically, this is important information for managing patients confidently over a long period of time, as many drugs exhibit regression to the mean after an initial benefit, whereas this anticoagulant does not. Similarly, this strength of RWD analysis allows for the evaluation of the longevity of the treatment effect of any drug, provided you can track the relevant outcomes for this determination.

Conclusion

This exercise demonstrates that replicating the methodology of a clinical trial using retrospective EHR data in the cardiovascular space yields results similar to those found in the initial study. These results support a broader use of RWD/RWE in regulatory submissions.

The success of RWD analysis depends upon starting with a large, rich, longitudinal data set as the foundation. To extract the most value from that data set, one needs to build on that foundation by mapping all available data elements to clinical concepts. These concepts are required to capture the medical state of a patient, study inclusion and exclusion criteria, and outcomes. This mapping is a process we refer to as operationalization, and requires a great deal of time, effort, and deeply integrated data science and clinical expertise.

To justify comparison of outcomes between cohorts, cohorts must be balanced with respect to prognostic variables. Done correctly, balancing allows for the attribution of any differences between outcomes to the treatment in question. Finally, with cohort-balanced outcomes calculated, clinical expertise is required to interpret the results and properly contextualize them.

Novel insights may be generated from RWD that can prove very valuable in helping an organization better understand their therapy and their target populations. RWD analysis can also be used to guide the design of clinical trials to be more efficient or to explore different indications for a given therapy.

The external control arm (ECA) is a specific use-case for RWD/RWE and can be used to supplement regulatory submissions and reduce regulatory risk. An ECA can also be used to replace some or all of a clinical trial control arm, dramatically reducing the number of patients in the study and thereby significantly lowering the overall study cost.

All of these tools can be brought to bear on the development of new therapies to reduce the duration, sample size, cost, and risk associated with clinical trials.

SUMMARY OF CONCLUSIONS

The HealthPals CLINT™ platform was able to quickly confirm results from a large, 5-year duration RCT.

To our knowledge, this is the first large-scale RCT whose results have been replicated in real-world EHR data.

This was made possible by the deep integration of clinical domain and data expertise of the teams at HealthPals and Veradigm's large, rich, longitudinal dataset.

BIBLIOGRAPHY

1. van Dorn A. COVID-19 and readjusting clinical trials. Lancet. 2020;396: 523–524.
2. Patel MR, Mahaffey KW, Garg J, Pan G, Singer DE, Hacke W, et al. Rivaroxaban versus warfarin in nonvalvular atrial fibrillation. N Engl J Med. 2011;365: 883–891.
3. Amendments reduce number of patients, but at high cost, longer study times. Tufts Center for the Study of Drug Development Impact Report. 1 Jan 2016.
4. Rising Protocol Complexity Is Hindering Performance while Driving Up Cost of Clinical Trials. Tufts Center for the Study of Drug Development Impact Report. 17 Jul 2018.
5. Faster New Drug Approval Phases Are More Than Offset by Longer Clinical Times in U.S. Tufts Center for the Study of Drug Development Impact Report. 16 Jul 2020.
6. Franklin JM, Patorno E, Desai RJ, Glynn RJ, Martin D, Quinto K, et al. Emulating Randomized Clinical Trials with Nonrandomized Real-World Evidence Studies: First Results from the RCT DUPLICATE Initiative. Circulation. 2020. doi:10.1161/CIRCULATIONAHA.120.051718
7. Lip GYH, Keshishian A, Li X, Hamilton M, Masseria C, Gupta K, et al. Effectiveness and safety of oral anticoagulants among nonvalvular atrial fibrillation patients: the ARISTOPHANES study. Stroke. 2018;49: 2933–2944.
8. McCormick N, Lacaille D, Bhole V, Avina-Zubieta JA. Validity of heart failure diagnoses in administrative databases: a systematic review and meta-analysis. PLoS One. 2014;9: e104519.
9. Kern EFO, Maney M, Miller DR, Tseng C-L, Tiwari A, Rajan M, et al. Failure of ICD-9-CM Codes to Identify Patients with Comorbid Chronic Kidney Disease in Diabetes. Health Services Research. 2006. pp. 564–580. doi:10.1111/j.1475-6773.2005.00482.x
10. McCoy TH Jr, Snapper L, Stern TA, Perlis RH. Underreporting of Delirium in Statewide Claims Data: Implications for Clinical Care and Predictive Modeling. Psychosomatics. 2016;57: 480–488.
11. Eikelboom JW, Connolly SJ, Bosch J, Dagenais GR, Hart RG, Shestakovska O, et al. Rivaroxaban with or without Aspirin in Stable Cardiovascular Disease. N Engl J Med. 2017;377: 1319–1330.