Title

Enhancing Evidence Generation by Linking Randomized Clinical Trials to Real-World Data: The INVESTED-Medicare Linkage Study

Journal

Clinical Pharmacology & Therapeutics

ISSN

0009-9236

Authors

Cho, Hanseul
Shin, HoJin
Najafzadeh, Mehdi
et al.

Publication Date

2026-09-06

DOI

10.1002/cpt.70457

Peer reviewed

Enhancing Evidence Generation by Linking Randomized Clinical Trials to Real-World Data: The INVESTED-Medicare Linkage Study

Hanseul Cho[1,2,†], HoJin Shin[1,3,†], Mehdi Najafzadeh[4], Brian Claggett[5], Janinne Ortega-Montiel[1,2], Jun Liu[1], Julie M. Paik[1,2,6], Narlina F. Lalani[7], Marie C. Bradley[8], John Concato[9], Scott D. Solomon[5], Orly Vardeny[10] and Elisabetta Patorno[1,2,*]

Real-world evidence (RWE) derived from real-world data (RWD) can complement randomized controlled trials (RCTs), yet the validity of RWD relative to RCT data remains insufficiently characterized. We obtained *post hoc* consent and linked individual participant data from the US-based INVESTED trial (2016–2019) with Medicare fee-for-service claims (2012–2020) to validate demographic factors, baseline characteristics (using 183-, 365-, and 730-day lookback periods) and outcomes, and to assess post-trial events. Among 5260 trial participants, 126 were enrolled and eligible for linkage. Among 115 participants with demographic information available from Medicare enrollment files, agreement between RCT- and RWD-based demographic factors was high: only one major age discrepancy, 100% agreement for sex, and an overall agreement of 0.89 for race. Participants with Medicare claims data ($n = 65$) were older and more likely to be White compared with the overall RCT population. For the 365-day lookback period, baseline comorbidities showed high sensitivity (median 0.80) and specificity (0.89), as did medication use (sensitivity 1.00, specificity 0.88). Lengthening the lookback period to 730 days increased sensitivity but decreased specificity, whereas shortening to 183 days decreased sensitivity but increased specificity. Clinical outcomes showed high specificity (0.88–0.94) but low sensitivity (0.18–0.50). Among those with Medicare coverage beyond the trial end date ($n = 45$), 22% experienced cardiopulmonary, 18% cardiovascular, and 7% heart failure (HF) hospitalizations, highlighting the value of RWD for extending RCT evidence. Proactive planning of future RCT-RWD linkage initiatives can improve the efficiency of linkage studies, leading to more actionable results.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ Real-world evidence (RWE) derived from real-world data (RWD), such as claims data, can complement randomized controlled trials (RCTs), but the concordance of claims-based patient characteristics and outcomes with trial data remains largely unknown. Linking RCT data with RWD provides an opportunity to evaluate how well RWD captures patient characteristics and outcomes and whether it can extend trial follow-up.

WHAT QUESTION DID THIS STUDY ADDRESS?

☑ How accurately do US Medicare fee-for-service claims capture demographic factors, baseline comorbidities, medication use, and clinical outcomes compared with linked INVESTED trial data? How many participants experienced outcomes after trial follow-up ended? What are the key challenges in RCT-RWD linkage studies?

WHAT DOES THIS STUDY ADD TO OUR KNOWLEDGE?

☑ Claims data showed high agreement with trial data for demographic factors and generally high sensitivity and specificity for baseline comorbidities and medication use. Longer lookback periods increased sensitivity but decreased specificity. Clinical outcomes showed high specificity but low sensitivity. RWD also captured outcomes after trial follow-up. We identified key linkage challenges and propose future directions.

HOW MIGHT THIS CHANGE CLINICAL PHARMACOLOGY OR TRANSLATIONAL SCIENCE?

☑ These findings support claims data for patient characterization and suggest tailoring lookback periods to balance sensitivity and specificity. High outcome specificity suggests relative risk measures may reasonably reflect true effects under nondifferential misclassification, whereas absolute risk differences may be biased toward the null. Capturing post-trial outcomes highlights RWD's value for extending RCT evidence. Proactive linkage planning can improve efficiency, accuracy, and actionability.

Received March 3, 2026; accepted August 12, 2026. doi:10.1002/cpt.70457

[1]Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA; [2]Division of General Internal Medicine & Health Services Research, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [3]Johnson & Johnson Innovative Medicine, Titusville, New Jersey, USA; [4]UCB, Cambridge, Massachusetts, USA; [5]Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA; [6]Division of Nephrology, Department of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, California, USA; [7]VA Health Care System, Minneapolis, Minnesota, USA; [8]Office of Medical Policy, Center for Drug Evaluation and Research, Food and Drug Administration, Silver Spring, Maryland, USA; [9]Department of Medicine, Yale University School of Medicine, New Haven, Connecticut, USA; [10]Department of Medicine, University of Minnesota, Minneapolis, Minnesota, USA. *Correspondence: Elisabetta Patorno (epatorno@mednet.ucla.edu)

[†]These authors contributed equally as co-first authors.

Randomized controlled trials (RCTs) are considered the reference standard for evaluating the efficacy and safety of medical interventions, with randomization balancing confounders and minimizing channeling bias.[1,2] Nonetheless, RCTs cannot address all questions arising from clinical practice due to feasibility and resource constraints, ethical considerations when withholding potentially beneficial treatments, and limitations on timely evidence generation for urgent decisions.[1,3] Moreover, RCTs often have strict eligibility criteria and controlled settings that limit their applicability to broader patient populations, and their follow-up durations may be insufficient for robust assessment of long-term or rare outcomes.[1,2]

Real-world evidence (RWE), which leverages real-world data (RWD) generated by routine healthcare operations, such as administrative claims,[1] can complement RCTs by addressing questions such as comparative effectiveness in diverse populations, head-to-head comparisons between treatments, long-term safety in routine practice, treatment effects in excluded subgroups, and outcomes emerging only with extended follow-up or widespread use.[3,4] RWD can fill these knowledge gaps expeditiously and at relatively lower cost without exposing patients to experimental interventions.[3] The integration of RWD with RCT findings is gaining recognition for enhancing regulatory and clinical decision-making.[4] Despite its potential, the capacity of RWD to accurately capture patient characteristics and outcomes, when compared directly with RCT data, remains largely unexplored.[5] Linking RCT data with RWD presents a unique opportunity to evaluate RWD's potential to complement RCT findings and bridge the gap between controlled trial environments and real-world clinical practice.[5]

In this study, we aimed to link individual participant data from the Influenza Vaccine to Effectively Stop Cardio Thoracic Events and Decompensated Heart Failure (INVESTED) trial with their US Medicare fee-for-service claims data to (1) validate baseline participant characteristics and outcomes measured in claims data against those captured through primary data collection in the trial,[6,7] and (2) assess the capability of RWD to extend the observation period of INVESTED trial participants by capturing RWD-measured outcomes occurring after the trial follow-up end date. Additionally, we aimed to share our experiences in conducting an RCT-RWD linkage, highlighting anticipated challenges and providing insights to enhance the efficiency of the process.

MATERIALS AND METHODS

Data source and study population

RCT. The INVESTED trial was a pragmatic, randomized, double-blind, active comparator study conducted across 157 centers in the United States and Canada over three influenza seasons (2016–2017 to 2018–2019).[6,7] It investigated whether a high-dose trivalent influenza vaccine compared with a standard-dose quadrivalent influenza vaccine could reduce the incidence of all-cause death or cardiopulmonary hospitalization in high-risk participants with cardiovascular disease.[6,7] Eligible participants had a history of hospitalization for myocardial infarction (MI) or heart failure (HF) and at least one of the following additional risk factors: aged ≥ 65 years, current or prior left ventricular ejection fraction < 40%, diabetes, body mass index ≥ 30 kg/m^2, chronic kidney disease, ischemic stroke, peripheral artery disease, current tobacco use, or a hospitalization for MI or HF prior to the index hospitalization.[6,7] Excluded participants had a known allergy, hypersensitivity (anaphylaxis), or Guillain-Barré syndrome related to the influenza vaccine; severe allergy to egg protein; life expectancy < 9 months; prior influenza vaccination during the enrolling season; acute infection requiring antibiotics within 14 days of randomization; fever within 7 days before randomization; pregnancy; or lactation.[6,7]

RWD. We used US fee-for-service Medicare data from 2012 through 2020. US fee-for-service Medicare is a federal health insurance program that covers beneficiaries aged 65 years and older, as well as patients under 65 years with disabilities or end-stage renal disease.[8] This database includes individual-level longitudinal information on demographic factors, diagnoses, procedures, and outpatient prescription dispensings that are recorded during the billing of routine healthcare encounters.[8] We had access to 100% Medicare fee-for-service beneficiaries enrolled in Parts A, B, or D who received a diagnosis of HF from 2012 to 2020 and a diagnosis of MI from 2013 to 2018. Specifically, we had access to information on inpatient hospitalizations, outpatient facility encounters, and physician/professional service claims (with associated ICD-9/10-CM diagnosis and CPT/HCPCS procedure codes) via Parts A and B, and outpatient prescription dispensings via Part D.

Construction of the study cohort

Consenting. Following the completion of the INVESTED trial on July 31, 2019, INVESTED US sites interested in participating in this linkage study obtained approval from their respective Institutional Review Boards for this supplementary investigation. INVESTED trial participants from these US sites were included in this supplementary study if they consented to link their trial data with Medicare claims data. Consent was collected by each site between November 2021 and November 2022. Subsequently, each site collected the Medicare Beneficiary Identifiers (MBIs) of each INVESTED participant who consented to be part of the linkage study.

Linkage. Among the participants who consented to the supplementary study, those who provided a complete MBI became eligible for linkage with Medicare fee-for-service claims data previously acquired by the study investigators. The encrypted list of MBIs was sent to the Research Data Assistance Center (ResDAC), a contractor for the Centers for Medicare & Medicaid Services (CMS) that assists researchers with CMS data.[8,9] ResDAC crosswalked the MBIs with encrypted Beneficiary Identifiers (BENE_IDs), unique identification numbers assigned to each Medicare beneficiary. The resulting crosswalk file was used to identify the trial participants in the Medicare claims data.

Study variables

Baseline participant characteristics. Definitions of baseline characteristics, including demographic factors, such as age, sex, and race, eligibility risk factors, comorbidities, and medications, are detailed in **Table S1**. For validation purposes, we focused on INVESTED trial variables measurable in claims data, excluding the New York Heart Association (NYHA) functional classification, left ventricular ejection fraction, and body mass index.[7] These variables were therefore only reported as trial-derived descriptive characteristics and were not included in the claims-based validation analyses. In the linked Medicare claims, we defined variables analogous to those in the INVESTED trial data using ICD-9/10-CM diagnosis and CPT/HCPCS procedure codes and Part D prescription dispensing records. Each of these sources was assessed during 183, 365 (primary), and 730 days before or on the participant's cohort-entry (trial start) date.[7] These timeframes were selected as they are commonly used to measure baseline information in RWD studies.[5,10,11] For eligibility, the INVESTED trial required a hospitalization for MI within the 12 months before the baseline visit or a hospitalization for HF within the 24 months before baseline. These windows correspond to our RWD lookback windows (365 days for MI and 730 days for HF). Participant-specific trial start and end dates were obtained from the trial data, and all lookback and post-trial follow-up windows were anchored to these dates. Age was calculated from the Medicare date of birth as of the trial start date. In the linked Medicare claims, all demographic data were complete, and the absence of claims for comorbidities and medications was interpreted as the absence of such comorbidities and medications. Therefore, there were no missing values for these baseline characteristics. Among participants included in our study, the INVESTED trial data were complete for all study variables used in the analysis.

Outcomes. All-cause mortality was not considered, as all participants had to be alive to provide consent for this linkage study. Outcomes included cardiopulmonary hospitalization, cardiovascular hospitalization, and HF hospitalization.[6,7] Cardiopulmonary hospitalization was defined as either cardiovascular hospitalization or pulmonary hospitalization.[6,7] Cardiovascular hospitalization was defined as hospitalization with a primary diagnosis of acute MI, unstable angina, cardiac dysrhythmias, HF, stroke, or transient ischemic attack.[6,7] Pulmonary hospitalization was defined as hospitalization with a primary diagnosis of pneumonia, influenza, asthma, chronic obstructive pulmonary disease, acute respiratory failure, or pulmonary embolism.[7] HF hospitalization was defined as hospitalization with a primary diagnosis of HF. Outcomes were defined using validated ICD-9/10-CM diagnosis and procedural code algorithms based on inpatient primary diagnosis codes and were aligned with definitions used in the INVESTED trial (**Table S2**).[11–22] Outcome validation was conducted among linked participants who had Medicare Part A coverage on their outcome event date, or on the trial end date for event-free participants. To confirm that a claims event represented the same hospitalization as the trial-adjudicated event, a claims hospitalization was counted as a match only when its date fell within 10 days of the trial event date.

Statistical analysis

Descriptive statistics were used to compare the distributions of baseline characteristics among trial participants enrolled in the United States, those who consented to participate in this supplementary study, and those who were linked with their Medicare claims data. Binary and categorical variables were summarized using frequencies and percentages; continuous variables were summarized using means and standard deviations (SD).[23,24] Standardized mean differences (SMD) were used to compare the distributions of baseline characteristics.[23,24] Values greater than 0.1 are typically considered indicative of meaningful differences.[23,24]

We used various performance metrics to validate claims-measured variables against trial-reported information. Each performance metric was computed only among participants with information available in the Medicare enrollment file (demographic factors) or continuous enrollment in the relevant Medicare Part(s) (comorbidities, medications and outcomes) over the lookback or follow-up window applicable to that metric, anchored to each participant's randomization date. For binary and categorical variables, we estimated sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and overall agreement (OA) with exact 95% binomial confidence intervals (CIs) using 2×2 tables, with the trial data as the reference standard (**Figure S1**).[25,26] Performance metrics were calculated by restricting comparisons to participants enrolled in relevant parts of the Medicare program: demographic factors (enrollment), comorbidities (Parts A and B), medication dispensing (Part D), and outcomes (Part A).[8,27] Statistical comparisons of sensitivity and specificity between different lookback periods (183 vs. 365 days and 365 vs. 730 days) were performed using the two-sided Wilcoxon signed-rank test for paired data.[28] Finally, among those with Medicare Part A data available after the trial follow-up end date, we assessed the number of participants who experienced RWD-measured outcomes after the conclusion of the trial follow-up period. All analyses were performed using R version 4.0.2.

Ethics approval and data sharing

We used de-identified person-level linked data for this analysis. The study was approved by the Mass General Brigham Institutional Review Board (IRB number 2020P003767) before data analysis, and licensing agreements were in place. Requests to reproduce findings in our data analytics environment will be considered. We used anonymized participant-level INVESTED trial data, governed by the trial's data-use agreements and not available for redistribution. Medicare claims data are governed by CMS Data Use Agreements and may not be made available for sharing. Aggregate study outputs can be made available upon reasonable request and IRB approval.

RESULTS

Cohort construction

Among the 5260 participants in the INVESTED trial, 1676 were excluded as they were from 39 Canadian sites. Additionally, 2526 participants were excluded because 94 US study sites declined participation in this linkage study, and 847 participants from 24 US participating sites were excluded due to participant declinations, death before consent, or inability to reach by phone. A total of 211 participants from these 24 US sites consented to the linkage study. However, 85 participants were further excluded for various reasons, including not providing MBIs, providing incorrect or incomplete MBIs, being enrolled in health insurance other than Medicare fee-for-service, and duplicate entries. Ultimately, among the 211 participants, 126 (59.7%) were eligible for linkage and were included for a crosswalk between MBIs and BENE IDs. Among these, 115 had demographic information, including age, sex, and race, from the Medicare enrollment files, and 65 had Medicare claims. Of those with claims, 45 had Part A coverage for time periods from the trial follow-up end date through 2020 for evaluation of outcomes occurring after the trial follow-up period; 28 had Part D coverage for at least 183 days before trial randomization (26 for 365 days and 25 for 730 days) for validation of claims-measured baseline medication use; and 34 had Parts A and B coverage for at least 183 days before trial randomization (30 for 365 days and 27 for 730 days) for validation of claims-measured baseline comorbidities. **Figure 1** provides a complete description of the cohort flow leading to the final sample sizes.

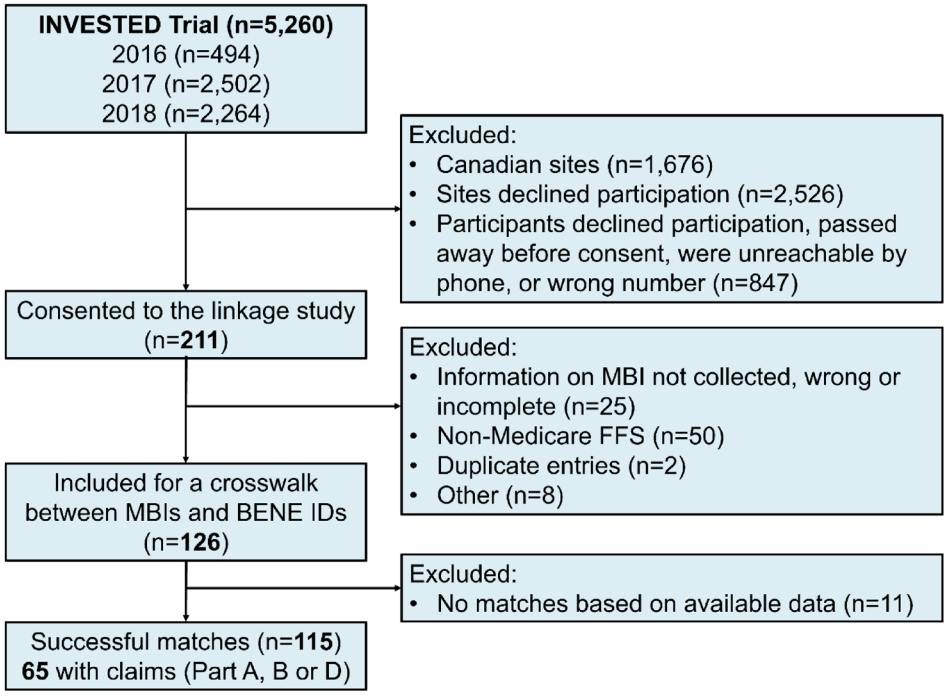

Figure 1 Study cohort flow diagram. Part A, B, and D each refer to Medicare hospital insurance, Medicare medical insurance, and Medicare prescription drug coverage, respectively. BENE ID, Beneficiary Identification Number; FFS, fee-for-service; MBI, Medicare Beneficiary Identifier.

The number of participants contributing to each analysis varied according to the source of data (trial vs. claims data), the Medicare enrollment/coverage, and the health plan continuous enrollment requirements for that specific analysis: **Table 1** describes baseline characteristics of trial participants and is based on trial-collected data, whereas **Tables 2–4** assess agreement between trial and claims data and are based on both trial and claims data. The source of data (trial vs. claims data), required Medicare enrollment/coverage, applicable time window for information measurement, and size of the specific analytic subset for every table/analysis are summarized in **Table S3**.

Demographics and baseline participant characteristics

Compared with the broader US trial participants (Cohort A), those who consented for linkage (Cohort B) were more likely to be recent enrollees, older (SMD = 0.34), and White (SMD = 0.42). They also were more likely to have no or mild symptoms according to the NYHA functional classification (**Table 1**). Compared with the broader US trial participants (Cohort A), those with linked Medicare claims (Cohort C) were less likely to be from Veterans Affairs sites (SMD = 0.24), more likely to be older (SMD = 0.58), and more likely to have no or mild symptoms according to the NYHA functional classification.

Validation of demographic information

Agreement between claims-measured and trial-reported demographic factors was high. Among 115 individuals with demographic data from Medicare enrollment files, we noted 100% match for sex and only one major discrepancy in age. Racial classification (non-White, 6.7%) showed a sensitivity of 1.00 and specificity of 0.88 (**Table 2**).

Validation of baseline comorbidities and medication use

Compared with trial measures, medical history including risk factors and comorbidities measured 365 days before or on cohort entry had a median sensitivity of 0.80 (Q1: 0.63, Q3: 1.00) and a median specificity of 0.89 (0.81, 0.96) (**Tables 3**, **S4**, and **S5**). Baseline medication use measured over the same period had a median sensitivity of 1.00 (0.92, 1.00) and a median specificity of 0.88 (0.73, 0.96) (**Table S6**). Using a two-sided Wilcoxon signed-rank test, we observed that extending the lookback window consistently improved sensitivity but reduced specificity. Sensitivity significantly increased when expanding the window from 183 days to 365 days ($P < 0.001$) and further increased when expanding from 365 days to 730 days ($P = 0.009$). Conversely, specificity was significantly lower at 365 days compared to 183 days ($P < 0.001$) and decreased further at 730 days compared to 365 days ($P < 0.001$).

Validation of outcomes

Outcomes generally showed low sensitivity but high specificity for cardiopulmonary hospitalization (sensitivity, 0.23 (95% CI, 0.05, 0.54); specificity, 0.88 (0.68, 0.97)), cardiovascular hospitalization (0.18 (0.02, 0.52); 0.88 (0.70, 0.98)), and HF hospitalization (0.50 (0.01, 0.99); 0.94 (0.81, 0.99); **Table 4**). Broadening the cardiopulmonary hospitalization definition from the primary discharge diagnosis position to any diagnosis position markedly increased sensitivity (from 0.23 (0.05, 0.54) to 0.62 (0.32, 0.86)), but at the cost of specificity (from 0.88 (0.68, 0.97) to 0.67 (0.45, 0.84); **Table S7**).

Outcomes captured beyond trial follow-up

We identified 45 individuals with Medicare Part A data after the trial follow-up end date. Among these individuals, 10 (22%)

Table 1 Baseline characteristics of US trial participants (Cohort A), eligible for linkage (Cohort B), and with linked claims (Cohort C) based on data from the INVESTED trial

Baseline characteristics[a]	US participants (Cohort A, *N* = 3584)	Participants eligible for linkage (Cohort B, *N* = 126)	Standardized difference (Cohort A vs. B)	Participants with linked Medicare claims (Cohort C, *N* = 65)	Standardized difference (Cohort A vs. C)
Demographic factors					
Age, years	64.84 (12.95)	68.68 (8.95)	0.34	71.11 (8.05)	0.58
Sex					
Male	2586 (72.2)	89 (70.6)	0.04	43 (66.2)	0.13
Female	998 (27.8)	37 (29.4)	0.04	22 (33.8)	0.13
Race					
White	2640 (73.7)	113 (89.7)	0.42	60 (92.3)	0.51
Non-White	944 (26.3)	13 (10.3)	0.42	5 (7.7)	0.51
Ethnicity					
Hispanic/Latino	533 (14.9)	29 (23.0)	0.21	20 (30.8)	0.39
Non-Hispanic/Latino	3051 (85.1)	97 (77.0)	0.21	45 (69.2)	0.39
Enrollment year					
2016	344 (9.6)	6 (4.8)	0.19	2 (3.1)	0.27
2017	1798 (50.2)	55 (43.7)	0.13	26 (40.0)	0.21
2018	1442 (40.2)	65 (51.6)	0.23	37 (56.9)	0.34
Enrollment site[b]					
Veterans Affairs site	1086 (30.3)	46 (36.5)	0.13	13 (20.0)	0.24
New York Heart Association Functional Classification (NYHA)[c]					
NYHA I	382 (14.9)	24 (25.5)	0.27	14 (30.4)	0.38
NYHA II	1284 (50.2)	54 (57.4)	0.14	28 (60.9)	0.22
NYHA III	799 (31.2)	16 (17.0)	0.34	4 (8.7)	0.59
NYHA IV	93 (3.6)	0 (0.0)	NA	0 (0.0)	NA
Medical history related to eligibility criteria					
Hospitalization for HF in past 24 months	2687 (74.9)	96 (76.2)	0.03	47 (72.3)	0.06
Hospitalization for MI in past 12 months	897 (25.0)	30 (23.8)	0.03	18 (27.7)	0.06
Age ≥ 65 years	1928 (53.9)	91 (72.2)	0.39	53 (81.5)	0.62
Current BMI ≥ 30 kg/m^2	1879 (52.6)	70 (55.6)	0.06	36 (55.4)	0.06
Current or past LVEF < 40%	1673 (46.8)	52 (41.3)	0.11	29 (44.6)	0.04
Diagnosis of type 1 or type 2 diabetes	1391 (38.9)	51 (40.5)	0.03	29 (44.6)	0.12
History of chronic kidney disease	1213 (33.9)	40 (31.7)	0.05	19 (29.2)	0.10
Current tobacco smoker	646 (18.1)	17 (13.5)	0.13	6 (9.2)	0.26
History of ischemic stroke	332 (9.3)	14 (11.1)	0.06	6 (9.2)	0
History of peripheral artery disease	166 (4.6)	7 (5.6)	0.05	3 (4.6)	0
Other medical history					
Hypertension	2868 (80.0)	104 (82.5)	0.06	54 (83.1)	0.08
Dyslipidemia	2402 (67.0)	104 (82.5)	0.36	54 (83.1)	0.38
Percutaneous coronary intervention	1244 (34.7)	42 (33.3)	0.03	28 (43.1)	0.17
Atrial fibrillation	1320 (36.8)	48 (38.1)	0.03	24 (36.9)	0.00
Coronary artery bypass grafting	704 (19.6)	35 (27.8)	0.19	21 (32.3)	0.29
Chronic obstructive pulmonary disease	775 (21.6)	28 (22.2)	0.01	13 (20.0)	0.04
Implantable cardioverter-defibrillator	834 (23.3)	26 (20.6)	0.06	14 (21.5)	0.04
Asthma	435 (12.1)	14 (11.1)	0.03	4 (6.2)	0.21

(Continued)

Table 1 (Continued)

Baseline characteristics[a]	US participants (Cohort A, *N* = 3584)	Participants eligible for linkage (Cohort B, *N* = 126)	Standardized difference (Cohort A vs. B)	Participants with linked Medicare claims (Cohort C, *N* = 65)	Standardized difference (Cohort A vs. C)
Maintenance medications for those with MI as index event					
Statins	817 (91.1)	28 (93.3)	0.08	17 (94.4)	0.13
Aspirin	808 (90.1)	24 (80.0)	0.29	15 (83.3)	0.20
Beta-blockers	793 (88.4)	24 (80.0)	0.23	14 (77.8)	0.29
Maintenance medications for those with HF as index event					
Beta-blockers	2241 (83.7)	89 (92.7)	0.28	44 (93.6)	0.32
Diuretics	2136 (79.8)	80 (83.3)	0.09	35 (74.5)	0.13
ACEi or ARB or ARNI	1756 (65.6)	68 (70.8)	0.11	32 (68.1)	0.05
Mineralocorticoid receptor antagonist	852 (31.8)	26 (27.1)	0.10	11 (23.4)	0.19
Digoxin	266 (9.9)	12 (12.5)	0.08	7 (14.9)	0.15

ACEi, angiotensin-converting enzyme inhibitors; ARB, angiotensin II receptor blockers; ARNI, angiotensin receptor–neprilysin inhibitor; BMI, body mass index; HF, heart failure; LVEF, left ventricular ejection fraction; MI, myocardial infarction; NYHA, New York Heart Association.
[a]Mean (standard deviation) for continuous variables, number of participants (percentage among whole cohort) for categorical or binary variables. All characteristics are drawn from the INVESTED trial data. Cohort A, full trial population; Cohort B, participants eligible for linkage (consented, with a complete beneficiary identifier; *N* = 126); Cohort C, linked participants with at least one Medicare claim during 2012–2020 (*N* = 65). [b]INVESTED enrolled participants at 157 sites across the United States and Canada, organized into Veterans Affairs, Canadian, other US non–Veterans Affairs, and PCORnet performance networks; enrolling sites were generally cardiology specialty clinics. [c]Proportion among individuals with available data.

Table 2 Validation of RWD-measured demographic information against RCT-reported information

	Prevalence	Overall agreement	Sensitivity	Specificity	PPV	NPV
Age	112 out of 115 are consistent; 1 person with a > 10 year discrepancy, 2 people with 1year difference in age					
Sex (male)	73.0% male	1.00	1.00	1.00	1.00	1.00
		(0.97, 1.00)	(0.96, 1.00)	(0.89, 1.00)	(0.96, 1.00)	(0.89, 1.00)
Race (non-white)	6.7% non-white	0.89	1.00	0.88	0.41	1.00
		(0.81, 0.94)	(0.66, 1.00)	(0.80, 0.93)	(0.21, 0.64)	(0.96, 1.00)

Analysis restricted to the 115 linked participants with demographic information available in the Medicare enrollment files. NPV, negative predictive value; PPV, positive predictive value.

experienced RWD-measured cardiopulmonary hospitalizations, eight individuals (18%) experienced cardiovascular hospitalizations, and three individuals (7%) experienced HF hospitalizations after the trial follow-up end date. Median post-trial follow-up duration was 519 days (IQR 366–534; range 31–875).

DISCUSSION

Our study provides important insights into the integration of RWD with RCT data, highlighting both the potential and limitations of claims data as a complementary resource to enhance regulatory decision-making and clinical research. We observed a high level of agreement between claims-measured and trial-reported demographic variables, baseline comorbidities, and medication use, with sensitivity and specificity varying according to the look-back periods used for covariate measurement. We also observed high specificity of claims-based measurements to capture clinical outcomes, together with low sensitivity. RWD captured a substantial number of outcomes occurring after the end of the trial follow-up period. These findings underscore the need for careful interpretation of RWD-based studies, in light of the opportunities and challenges it provides.

Validation results

The high agreement between claims-measured and trial-reported demographic variables (sex, age, and race) confirms the reliability of claims data for capturing these characteristics. Although race and ethnicity in the trial data were self-reported, race and ethnicity in Medicare enrollment files were derived from original Social Security Administration self-reported variables with historically limited categories and have been further modified using the RTI race code, which applies an imputation algorithm.[29–32] Future linkage studies should further characterize how self-reported race and ethnicity collected in trials diverge from race and ethnicity measures recorded in claims data and other RWD, and use these findings to inform strategies for improving the accuracy of race and ethnicity classification in RWD-based research.

Table 3 Validation of RWD-measured baseline medical history and medication use against RCT-reported information

Characteristic	RCT prevalence	RWD prevalence[a]	Overall agreement	Sensitivity	Specificity	PPV	NPV
Risk factors (eligibility-related medical history)							
Hospitalization for HF in past 24 months	0.78 (0.58, 0.91)	0.37 (0.19, 0.58)	0.37 (0.19, 0.58)	0.33 (0.15, 0.57)	0.50 (0.12, 0.88)	0.70 (0.35, 0.93)	0.18 (0.04, 0.43)
Hospitalization for MI in past 12 months	0.43 (0.25, 0.63)	0.23 (0.10, 0.42)	0.80 (0.61, 0.92)	0.54 (0.25, 0.81)	1.00 (0.80, 1.00)	1.00 (0.59, 1.00)	0.74 (0.52, 0.90)
Age ≥65 years	0.93 (0.78, 0.99)	0.93 (0.78, 0.99)	1.00 (0.88, 1.00)	1.00 (0.88, 1.00)	1.00 (0.16, 1.00)	1.00 (0.88, 1.00)	1.00 (0.16, 1.00)
Current BMI ≥30 kg/m^2	0.57 (0.37, 0.75)	0.47 (0.28, 0.66)	0.77 (0.58, 0.90)	0.71 (0.44, 0.90)	0.85 (0.55, 0.98)	0.86 (0.57, 0.98)	0.69 (0.41, 0.89)
Diagnosis of type 1 or type 2 diabetes	0.47 (0.28, 0.66)	0.57 (0.37, 0.75)	0.90 (0.73, 0.98)	1.00 (0.77, 1.00)	0.81 (0.54, 0.96)	0.82 (0.57, 0.96)	1.00 (0.75, 1.00)
History of ischemic stroke	0.13 (0.04, 0.31)	0.17 (0.06, 0.35)	0.90 (0.73, 0.98)	0.75 (0.19, 0.99)	0.92 (0.75, 0.99)	0.60 (0.15, 0.95)	0.96 (0.80, 1.00)
History of peripheral artery disease	0.07 (0.01, 0.22)	0.30 (0.15, 0.49)	0.77 (0.58, 0.90)	1.00 (0.16, 1.00)	0.75 (0.55, 0.89)	0.22 (0.03, 0.60)	1.00 (0.84, 1.00)
History of chronic kidney disease	0.33 (0.17, 0.53)	0.33 (0.17, 0.53)	0.87 (0.69, 0.96)	0.80 (0.44, 0.97)	0.90 (0.68, 0.99)	0.80 (0.44, 0.97)	0.90 (0.68, 0.99)
Current or past LVEF <40%	0.37 (0.20, 0.56)	0.17 (0.06, 0.35)	0.67 (0.47, 0.83)	0.27 (0.06, 0.61)	0.89 (0.67, 0.99)	0.60 (0.15, 0.95)	0.68 (0.46, 0.85)
Current tobacco smoker	0.07 (0.01, 0.22)	0.33 (0.17, 0.53)	0.73 (0.54, 0.88)	1.00 (0.16, 1.00)	0.71 (0.51, 0.87)	0.20 (0.03, 0.56)	1.00 (0.83, 1.00)
Comorbidities (other medical history)							
Atrial fibrillation	0.37 (0.20, 0.56)	0.37 (0.20, 0.56)	1.00 (0.88, 1.00)	1.00 (0.72, 1.00)	1.00 (0.82, 1.00)	1.00 (0.72, 1.00)	1.00 (0.82, 1.00)
Asthma	0.07 (0.01, 0.22)	0.07 (0.01, 0.22)	0.93 (0.78, 0.99)	0.50 (0.01, 0.99)	0.96 (0.82, 1.00)	0.50 (0.01, 0.99)	0.96 (0.82, 1.00)
Coronary artery bypass grafting	0.27 (0.12, 0.46)	0.23 (0.10, 0.42)	0.83 (0.65, 0.94)	0.63 (0.24, 0.91)	0.91 (0.71, 0.99)	0.71 (0.29, 0.96)	0.87 (0.66, 0.97)
Chronic obstructive pulmonary disease	0.20 (0.08, 0.39)	0.30 (0.15, 0.49)	0.90 (0.73, 0.98)	1.00 (0.54, 1.00)	0.88 (0.68, 0.97)	0.67 (0.30, 0.93)	1.00 (0.84, 1.00)
Dialysis	0.03 (0.00, 0.17)	0.03 (0.00, 0.17)	1.00 (0.88, 1.00)	1.00 (0.03, 1.00)	1.00 (0.88, 1.00)	1.00 (0.03, 1.00)	1.00 (0.88, 1.00)
Deep vein thrombosis	0.07 (0.01, 0.22)	0.03 (0.00, 0.17)	0.90 (0.73, 0.98)	0.00 (0.00, 0.84)	0.96 (0.82, 1.00)	0.00 (0.00, 0.98)	0.93 (0.77, 0.99)
Dyslipidemia	0.80 (0.61, 0.92)	0.90 (0.73, 0.98)	0.77 (0.58, 0.90)	0.92 (0.73, 0.99)	0.17 (0.00, 0.64)	0.81 (0.62, 0.94)	0.33 (0.01, 0.91)
Prior influenza vaccine	0.67 (0.47, 0.83)	0.27 (0.12, 0.46)	0.47 (0.28, 0.66)	0.30 (0.12, 0.54)	0.80 (0.44, 0.97)	0.75 (0.35, 0.97)	0.36 (0.17, 0.59)
Heart failure	0.23 (0.10, 0.42)	0.70 (0.51, 0.85)	0.47 (0.28, 0.66)	0.86 (0.42, 1.00)	0.35 (0.16, 0.57)	0.29 (0.11, 0.52)	0.89 (0.52, 1.00)
Hypertension	0.80 (0.61, 0.92)	0.53 (0.34, 0.72)	0.67 (0.47, 0.83)	0.63 (0.41, 0.81)	0.83 (0.36, 1.00)	0.94 (0.70, 1.00)	0.36 (0.13, 0.65)
Implantable cardioverter-defibrillator	0.33 (0.17, 0.53)	0.33 (0.17, 0.53)	1.00 (0.88, 1.00)	1.00 (0.69, 1.00)	1.00 (0.83, 1.00)	1.00 (0.69, 1.00)	1.00 (0.83, 1.00)
Malignancy	0.10 (0.02, 0.27)	0.10 (0.02, 0.27)	0.93 (0.78, 0.99)	0.67 (0.09, 0.99)	0.96 (0.81, 1.00)	0.67 (0.09, 0.99)	0.96 (0.81, 1.00)
Old myocardial infarction	0.20 (0.08, 0.39)	0.20 (0.08, 0.39)	0.87 (0.69, 0.96)	0.67 (0.22, 0.96)	0.92 (0.73, 0.99)	0.67 (0.22, 0.96)	0.92 (0.73, 0.99)
Pacemaker implanted	0.20 (0.08, 0.39)	0.43 (0.25, 0.63)	0.77 (0.58, 0.90)	1.00 (0.54, 1.00)	0.71 (0.49, 0.87)	0.46 (0.19, 0.75)	1.00 (0.80, 1.00)
Percutaneous coronary intervention	0.57 (0.37, 0.75)	0.50 (0.31, 0.69)	0.67 (0.47, 0.83)	0.65 (0.38, 0.86)	0.69 (0.39, 0.91)	0.73 (0.45, 0.92)	0.60 (0.32, 0.84)
Pneumococcal vaccine	0.63 (0.44, 0.80)	0.17 (0.06, 0.35)	0.40 (0.23, 0.59)	0.16 (0.03, 0.40)	0.82 (0.48, 0.98)	0.60 (0.15, 0.95)	0.36 (0.18, 0.57)
Pulmonary embolism	0.03 (0.00, 0.17)	0.07 (0.01, 0.22)	0.90 (0.73, 0.98)	0.00 (0.00, 0.98)	0.93 (0.77, 0.99)	0.00 (0.00, 0.84)	0.96 (0.82, 1.00)

(Continued)

Table 3 (Continued)

Characteristic	RCT prevalence	RWD prevalence[a]	Overall agreement	Sensitivity	Specificity	PPV	NPV
Sleep apnea	0.47 (0.28, 0.66)	0.47 (0.28, 0.66)	0.87 (0.69, 0.96)	0.86 (0.57, 0.98)	0.88 (0.62, 0.98)	0.86 (0.57, 0.98)	0.88 (0.62, 0.98)
Stroke	0.17 (0.06, 0.35)	0.17 (0.06, 0.35)	0.93 (0.78, 0.99)	0.80 (0.28, 0.99)	0.96 (0.80, 1.00)	0.80 (0.28, 0.99)	0.96 (0.80, 1.00)
Medication use							
ACE inhibitors	0.46 (0.27, 0.67)	0.50 (0.30, 0.70)	0.88 (0.70, 0.98)	0.92 (0.62, 1.00)	0.86 (0.57, 0.98)	0.85 (0.55, 0.98)	0.92 (0.64, 1.00)
Beta-blockers	0.92 (0.75, 0.99)	1.00 (0.87, 1.00)	0.92 (0.75, 0.99)	1.00 (0.86, 1.00)	0.00 (0.00, 0.84)	0.92 (0.75, 0.99)	N/A
Antiarrhythmics	0.15 (0.04, 0.35)	0.15 (0.04, 0.35)	1.00 (0.87, 1.00)	1.00 (0.40, 1.00)	1.00 (0.85, 1.00)	1.00 (0.40, 1.00)	1.00 (0.85, 1.00)
Anticoagulants	0.15 (0.04, 0.35)	0.15 (0.04, 0.35)	1.00 (0.87, 1.00)	1.00 (0.40, 1.00)	1.00 (0.85, 1.00)	1.00 (0.40, 1.00)	1.00 (0.85, 1.00)
Antiplatelets	0.58 (0.37, 0.77)	0.69 (0.48, 0.86)	0.88 (0.70, 0.98)	1.00 (0.78, 1.00)	0.73 (0.39, 0.94)	0.83 (0.59, 0.96)	1.00 (0.63, 1.00)
COPD and asthma medications	0.23 (0.09, 0.44)	0.23 (0.09, 0.44)	0.92 (0.75, 0.99)	0.83 (0.36, 1.00)	0.95 (0.75, 1.00)	0.83 (0.36, 1.00)	0.95 (0.75, 1.00)
Statins	0.92 (0.75, 0.99)	0.92 (0.75, 0.99)	0.85 (0.65, 0.96)	0.92 (0.73, 0.99)	0.00 (0.00, 0.84)	0.92 (0.73, 0.99)	0.00 (0.00, 0.84)
ARBs	0.35 (0.17, 0.56)	0.46 (0.27, 0.67)	0.88 (0.70, 0.98)	1.00 (0.66, 1.00)	0.82 (0.57, 0.96)	0.75 (0.43, 0.95)	1.00 (0.77, 1.00)
Aspirin	0.77 (0.56, 0.91)	0.04 (0.00, 0.20)	0.27 (0.12, 0.48)	0.05 (0.00, 0.25)	1.00 (0.54, 1.00)	1.00 (0.03, 1.00)	0.24 (0.09, 0.45)
Calcium channel blockers	0.23 (0.09, 0.44)	0.38 (0.20, 0.59)	0.85 (0.65, 0.96)	1.00 (0.54, 1.00)	0.80 (0.56, 0.94)	0.60 (0.26, 0.88)	1.00 (0.79, 1.00)
Digoxin	0.19 (0.07, 0.39)	0.19 (0.07, 0.39)	1.00 (0.87, 1.00)	1.00 (0.48, 1.00)	1.00 (0.84, 1.00)	1.00 (0.48, 1.00)	1.00 (0.84, 1.00)
Diuretic (other than MRA)	0.73 (0.52, 0.88)	0.81 (0.61, 0.93)	0.92 (0.75, 0.99)	1.00 (0.82, 1.00)	0.71 (0.29, 0.96)	0.90 (0.70, 0.99)	1.00 (0.48, 1.00)
Hydralazine	0.04 (0.00, 0.20)	0.08 (0.01, 0.25)	0.96 (0.80, 1.00)	1.00 (0.03, 1.00)	0.96 (0.80, 1.00)	0.50 (0.01, 0.99)	1.00 (0.86, 1.00)
Insulin	0.35 (0.17, 0.56)	0.35 (0.17, 0.56)	0.85 (0.65, 0.96)	0.78 (0.40, 0.97)	0.88 (0.64, 0.99)	0.78 (0.40, 0.97)	0.88 (0.64, 0.99)
Long-acting nitrates	0.12 (0.02, 0.30)	0.46 (0.27, 0.67)	0.65 (0.44, 0.83)	1.00 (0.29, 1.00)	0.61 (0.39, 0.80)	0.25 (0.05, 0.57)	1.00 (0.77, 1.00)
Mineralocorticoid receptor antagonists	0.27 (0.12, 0.48)	0.35 (0.17, 0.56)	0.92 (0.75, 0.99)	1.00 (0.59, 1.00)	0.89 (0.67, 0.99)	0.78 (0.40, 0.97)	1.00 (0.80, 1.00)
Warfarin	0.15 (0.04, 0.35)	0.23 (0.09, 0.44)	0.92 (0.75, 0.99)	1.00 (0.40, 1.00)	0.91 (0.71, 0.99)	0.67 (0.22, 0.96)	1.00 (0.83, 1.00)

Analyses on medical history were restricted to participants with continuous Medicare Part A and Part B enrollment over the lookback window (n=30 for the 365-day window), anchored to each participant's trial start date. Analyses on medication use were restricted to participants with continuous Medicare Part D enrollment over the lookback window (n=26 for the 365-day window), anchored to each participant's trial start date. ACE, angiotensin-converting enzyme; ARBs, angiotensin II receptor blockers; COPD, chronic obstructive pulmonary disease; FN, false negative; FP, false positive; HF, heart failure; LVEF, left ventricular ejection fraction; MI, myocardial infarction; MRA, mineralocorticoid receptor antagonists; NPV, negative predictive value; NYHA, New York Heart Association; PPV, positive predictive value; RCT, randomized controlled trial; RWD, real-world data; TN, true negative; TP, true positive.

[a]Prevalence of claims-measured medical history and medication use was assessed during a 365-day baseline assessment period.

Table 4 Validation of RWD-measured outcomes against RCT-reported outcomes

Outcome	Overall agreement	Sensitivity	Specificity	PPV	NPV
Cardiopulmonary hospitalizations	0.65 (0.47, 0.80)	0.23 (0.05, 0.54)	0.88 (0.68, 0.97)	0.50 (0.12, 0.88)	0.68 (0.49, 0.83)
Cardiovascular hospitalizations	0.68 (0.50, 0.82)	0.18 (0.02, 0.52)	0.88 (0.70, 0.98)	0.40 (0.05, 0.85)	0.72 (0.53, 0.86)
Heart failure hospitalizations	0.92 (0.78, 0.98)	0.50 (0.01, 0.99)	0.94 (0.81, 0.99)	0.33 (0.01, 0.91)	0.97 (0.85, 1.00)

Analysis restricted to linked participants with Medicare Part A who had Medicare Part A coverage on their outcome event date, or the trial end date for event-free participants. FN, false negative; FP, false positive; NPV, negative predictive value; PPV, positive predictive value; TN, true negative; TP, true positive; RCT, randomized controlled trial; RWD, real-world data.

Baseline medication use showed high sensitivity and specificity, consistent with the expected accuracy of claims data in capturing medication dispensings for billing purposes. However, baseline comorbidities measured 365 days before or on cohort entry showed moderate to high sensitivity, suggesting that claims-measured comorbidities may not always correctly estimate disease burden.[18] Varying measurement periods indicated that a shorter period of 183 days yielded lower sensitivity but higher specificity, whereas a longer period of 730 days showed higher sensitivity but lower specificity.

The findings regarding outcome measures are particularly noteworthy. Although claims data showed high specificity for cardiopulmonary, cardiovascular, and HF hospitalizations, the low sensitivity indicates that claims data could often underreport certain health outcomes, which is consistent with findings from previous research.[18] The high specificity of claims-based outcome ascertainment and the correspondingly low false positive rate support the validity of relative-effect measures, conditional on nondifferential outcome misclassification.[33,34] In particular, when specificity is near perfect and sensitivity is nondifferential across treatment groups, relative risk measures remain approximately unbiased, although absolute risk differences are biased toward the null.[33,34] Accordingly, findings from claims-based studies should be interpreted with the understanding that relative measures may reasonably reflect true associations, whereas absolute risk differences are likely to underestimate the true magnitude of risk.[33] Several established approaches can mitigate outcome under-ascertainment: composite outcome definitions that increase sensitivity; supplementation of claims with electronic health record, registry, or laboratory data; quantitative bias-analysis methods that correct estimates under an explicit misclassification model[35–37]; and refinement and re-validation of claims-based algorithms in the target population. Future linkage studies could evaluate how these approaches improve outcome ascertainment.

The PPV for outcomes was modest, indicating that a meaningful proportion of events recorded in claims were not confirmed in the trial data. One potential reason behind this discordance is that claims may capture clinically genuine events arising in routine care that did not meet the trial's adjudicated endpoint definitions or were not ascertained at the trial's scheduled study contacts. This discordance warrants further investigation in future linkage studies, for example with corroboration through linkage of administrative claims with electronic health records, and may also indicate that claims could complement trial data by capturing events occurring outside trial ascertainment.

The validation metrics reported here should be interpreted in light of the trade-offs imposed by the claims-based definitions we applied. Multiple algorithmic definitions can be used to define comorbidities and outcomes—with variations in the diagnostic, procedural, and medication codes included, the diagnosis position considered (e.g., primary vs. any-listed), the care settings included (e.g., inpatient vs. outpatient), and the number of qualifying encounters required (e.g., any vs. at least two encounters)—and each choice trades off sensitivity, specificity, and PPV. For example, expanding the cardiopulmonary hospitalization definition from the primary discharge-diagnosis position to any diagnosis position identified many events that had previously been classified as false negatives, thereby substantially improving sensitivity, while reducing specificity.

Considering these trade-offs, the choice of lookback window and algorithmic definition for variables should be tailored to the analytic objective and the prevalence and chronicity of the condition of interest. Higher-specificity definitions may be preferable when false positives are most consequential, such as for outcomes used to estimate causal effects or criteria used to define cohort eligibility, whereas higher-sensitivity definitions may be preferable when missed events are most consequential, such as for safety-signal screening.[38–41] Longer measurement periods or algorithms with greater sensitivity may also be preferable for confounding adjustment in causal-inference analyses, because residual confounding from incompletely captured covariates may pose greater concern than the modest misclassification introduced by a small number of false positive comorbidity flags.[38–41]

Long-term follow-up via RCT-RWD linkage

The observation that 22% of 45 linked participants with post-trial Medicare Part A coverage experienced cardiopulmonary hospitalizations after the conclusion of the active trial highlights the critical value of RWD for extending follow-up periods of trial participants. This finding aligns with a recent scoping review, which identified post-trial long-term follow-up as the most frequent use case for RCT-RWD linkage, accounting for 31% of identified studies.[5] Linking clinical trials to routinely collected data sources enables the capture of patient trajectories and outcomes beyond the temporal and resource constraints of traditional trial designs.[5] Post-trial outcomes captured in claims should be interpreted with

caution, given the high specificity but modest PPV of claims-based outcome ascertainment.

Challenges and future directions

Operational challenges. This study faced several operational challenges that highlighted the complexities of conducting a supplementary investigation beyond the scope of the main trial. Because linkage activities were initiated after completion of the main trial, re-engaging study sites and re-contacting participants were challenging. Substantial attrition occurred through multiple stages, including site non-participation, participant declinations, and unsuccessful linkage attempts. Sites were often less responsive or available once the primary research activities of the main trial concluded, making it difficult to coordinate additional study-related tasks. The shift in focus away from the main trial objectives led to delays and complications in executing supplementary study activities. Obtaining re-consent from participants after the conclusion of the main trial was particularly challenging. Participants were often unreachable at the provided phone numbers, a situation exacerbated by the concomitant COVID-19 pandemic. The INVESTED trial enrolled high-risk participants with cardiovascular disease; consequently, many participants may have passed away by the time investigators attempted to recontact them for re-consent. The requirement for participants to survive until the time of re-consent precluded our ability to validate all-cause mortality and may have affected the observed performance of claims-based measures for other outcomes, as participants who survived could have had different health profiles and healthcare utilization patterns compared with those who were deceased.[42]

Scope and completeness of claims data. The study's linkage was limited to trial participants enrolled in fee-for-service Medicare insurance. Those who had Medicare Advantage or other health insurance coverage were excluded, potentially limiting the generalizability of the findings. Even when trial participants were successfully linked to their fee-for-service Medicare claims, insurance coverage was often incomplete. Periods with limited Medicare coverage—in which participants were enrolled for only part of the relevant lookback or follow-up window or for only a subset of the Medicare Parts (A, B, or D)—impact the completeness of the claims data, and thus the size of the analytic subsets available for analysis. Additionally, post-trial outcome capture is bounded by the December 2020 administrative end of the available claims data and by each participant's Medicare coverage span; events after this date are unobservable, and participants whose coverage began later in the post-trial period contribute less observation time.

Selection and generalizability of the linked cohort. These post-trial linkage processes introduced selection into the linked cohort. The linked cohort differed systematically from the broader trial population, being older, more likely to be White, and less likely to be from Veterans Affairs sites. Generalizability of our validation findings to the full INVESTED population and other populations warrants caution, as differences in insurance continuity, enrollment churn, healthcare utilization, and coding patterns may affect validation metrics.[43,44] Both RCT and linked RWD populations in this study substantially under-represent historically minoritized racial and ethnic groups. This under-representation restricts the applicability of our validation metrics to more diverse populations. Larger linked cohorts are needed to evaluate validation metrics within sociodemographic subgroups defined by race and ethnicity, age, and sex, so that the accuracy of claims-based ascertainment can be characterized across diverse populations.

Proactive planning for linkage. These challenges underscore the need for proactive planning in RCT-RWD linkage studies.[5] Re-consenting participants years after trial completion poses ethical and operational challenges and is best avoided by planning linkage from trial inception. We recommend that future clinical trials considering the use of linked RCT-RWD infrastructure: (1) plan for linkage at the trial design stage and incorporate consent for secondary use of trial data for linkage to RWD into the initial informed-consent process, or at least initiate linkage before study close, to avoid the challenges of re-consenting participants or re-engaging trial study sites; (2) collect and verify, at initial consent or as early as possible, the identifiers needed for deterministic linkage or linkage based on tokenization[45]—such as participant names, dates of birth, insurance information, or national identifiers—so that participants can be linked from the analytic index date forward and identifier errors can be corrected while the trial is active; and (3) consider a diverse range of RWD sources—including Medicare fee-for-service and Medicare Advantage, Medicaid, commercial insurance claims, electronic health records, disease registries, laboratory data, and the National Death Index—to increase coverage, granularity, and the likelihood of successful linkage, thereby increasing the linked sample size and enhancing the robustness of study findings. By addressing these issues, researchers can improve the efficiency and accuracy of RCT-RWD linkage studies, leading to more actionable results.

Limitations

Several limitations should be considered when interpreting these findings. The small validation cohort produced wide confidence intervals and limited statistical precision for several validation metrics, especially those for low-prevalence variables or sparse endpoints; therefore, these estimates should be interpreted with caution and confirmed in substantially larger linked cohorts. In addition, although we used trial-reported outcomes as the reference standard, some clinical events identified in claims data might have actually occurred but were not captured in the trial process.[46,47] Therefore, some claims-based outcomes classified as false positives may have represented clinically genuine events, highlighting the potential value of RCT-RWD linkage for improving outcome ascertainment in traditional clinical trials. In the INVESTED trial, outcomes were adjudicated by a blinded Clinical Endpoints Committee, whereas claims-measured outcomes were identified using diagnostic and procedural codes. These inherent differences in outcome ascertainment may have introduced outcome misclassification. Although we used previously validated claims-based outcome definitions, careful evaluation of the alignment

between claims-based coding systems (e.g., ICD codes) and the classification systems used in the trial is important when assessing the validity of claims-derived outcomes. Despite these limitations, our study provides valuable empirical evidence on the validity of claims-based measures and practical insights for conducting RCT-RWD linkage studies.

CONCLUSION

Linking RCT data with RWD, such as administrative claims, can enhance evidence generation by extending follow-up for long-term safety and effectiveness outcomes among trial participants and by augmenting trial data with RWD, offering a promising avenue to inform medication approval and label updates, including secondary indications. Our study provides a validation framework—quantifying sensitivity, specificity, and predictive values of claims-based covariates and outcomes against an adjudicated trial reference—that can guide future RCT-RWD linkage studies to investigate potential measurement challenges, identify calibration needs, and interpret findings accordingly. We found that claims data accurately captured demographic factors and baseline medication use, while claims-measured clinical outcomes showed high specificity but overall low sensitivity compared with trial-reported information. RWD also captured a substantial number of outcomes occurring after the trial follow-up end date, highlighting the value of RWD for extending the observation period of trial participants. These results need to be interpreted in light of the limited and selected study population and the operational challenges faced in conducting the RCT-RWD linkage. Proactive planning of future RCT-RWD linkage studies with a more refined and efficient approach, as outlined in the current investigation, can improve linkage efficiency, measurement validity, and interpretability, ultimately leading to more actionable evidence.

SUPPORTING INFORMATION

Supplementary information accompanies this paper on the *Clinical Pharmacology & Therapeutics* website (www.cpt-journal.com).

ACKNOWLEDGMENTS

None.

FUNDING

This work was supported by a research grant from the Food and Drug Administration (5U01FD007213). Dr Patorno was additionally supported by research grants from the National Institute of Diabetes and Digestive and Kidney Diseases (R01DK138036) and the Patient-Centered Outcomes Research Institute (DB-2020C2-20326).

CONFLICT OF INTEREST

Dr Shin completed this work while employed at Brigham and Women's Hospital. Dr Shin is currently employed by Johnson & Johnson Innovative Medicine and has no conflicts of interest to disclose. The current employer had no role in the study design; data collection, analysis, or interpretation; manuscript preparation; or the decision to submit for publication. Dr Najafzadeh works for UCB and declares no conflict of interest related to this study. Dr Claggett has consulted for statistics for Alnylam, Cardurion, Corvia, CVRX, Cytokinetics, Intellia, Novartis, and Rocket. Dr Solomon has received research grants from Actelion, Alnylam, Amgen, AstraZeneca, Bellerophon, Bayer, BMS, Celladon, Cytokinetics, Eidos, Gilead, GSK, Ionis, Lilly, Mesoblast, MyoKardia, NIH/NHLBI, Neurotronik, Novartis, NovoNordisk, Respicardia, SanofiPasteur, Theracos, Us2.AI, and has consulted for Abbott, Action, Akros, Alnylam, American Regent, Amgen, Anacardio, Arena, AstraZeneca, Bayer, Boehringer-Ingelheim, BMS, Cardiac Dimensions, Cardior, Cardurion, CellProThera, Corvia, Cytokinetics, Daiichi-Sankyo, Dinaqor, GSK, Janssen, Lexicon, Lilly, Merck, Moderna, MyoKardia, Novartis, Puretech Health, Quantum Genomics, Roche, Sanofi Pasteur, Sarepta, Tenaya, Theracos, and Tremeau. Dr Vardeny receives research support from AstraZeneca, Bayer, and Cardurion and has consulted for Moderna and Sanofi Pasteur. Dr Patorno was principal investigator of research grants to Brigham and Women's Hospital from AstraZeneca, Bayer, and Boehringer-Ingelheim, not related to the topic of the submitted work. She has consulted for IQVIA and receives royalties from UpToDate. All other authors declared no competing interests for this work.

AUTHOR CONTRIBUTIONS

H.C. and H.S. wrote the manuscript; H.C., H.S., M.N., S.D.S., M.C.B., J.C., O.V., and E.P. designed the research; H.C., H.S., O.V., S.D.S., B.C., N.F.L., J.O.-M., M.C.B., J.C., J.M.P., and E.P. performed the research; H.C., H.S., and J.L. analyzed the data.

DISCLAIMER

The views expressed are the authors' and not necessarily those of the United States Food and Drug Administration or the Department of Health and Human Services.

1. Concato, J. & Corrigan-Curay, J. Real-world evidence – where are we now? *N. Engl. J. Med.* **386**, 1680–1682 (2022).
2. Fernainy, P., Cohen, A.A., Murray, E., Losina, E., Lamontagne, F. & Sourial, N. Rethinking the pros and cons of randomized controlled trials and observational studies in the era of big data and advanced methods: a panel discussion. *BMC Proc.* **18**(Suppl 2), 1 (2024).
3. Patorno, E. Well-conducted real-world evidence studies can complement essential evidence from clinical trials. *Circulation* **150**, 1412–1415 (2024).
4. Jansen, M.S. *et al*. Multiple perspectives on the need for real-world evidence to inform regulatory and health technology assessment decision-making: scoping review and stakeholder interviews. *Pharmacoepidemiol. Drug Saf.* **34**, e70074 (2025).
5. NajafZadeh, M. *et al*. Linkage of clinical trial data to routinely collected data sources: a scoping review. *JAMA Netw. Open* **8**, e257797 (2025).
6. Vardeny, O. *et al*. Effect of high-dose trivalent vs standard-dose quadrivalent influenza vaccine on mortality or cardiopulmonary hospitalization in patients with high-risk cardiovascular disease: a randomized clinical trial. *JAMA* **325**, 39–49 (2021).
7. Vardeny, O. *et al*. High-dose influenza vaccine to reduce clinical outcomes in high-risk cardiovascular patients: rationale and design of the INVESTED trial. *Am. Heart J.* **202**, 97–103 (2018).
8. Mues, K.E. *et al*. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clin. Epidemiol.* **9**, 267–277 (2017).
9. Leonard, C.E. *et al*. The quality of Medicaid and Medicare data obtained from CMS and its contractors: implications for pharmacoepidemiology. *BMC Health Serv. Res.* **17**, 304 (2017).
10. Fang, Y.E., Paik, J.M., Ortega-Montiel, J., Tesfaye, H., Wexler, D.J. & Patorno, E. Risk of acute pancreatitis and biliary events after initiation of Incretin-based medications in patients with type 2 diabetes. *Diabetes Care* **48**, 2127–2137 (2025).

11. Bea, S. *et al*. Oral anticoagulation and risk of adverse clinical outcomes in venous thromboembolism. *JAMA Intern. Med.* **185**, 837–846 (2025).
12. Andrade, S.E. *et al*. A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol. Drug Saf.* **21 Suppl 1**(Suppl 1), 100–128 (2012).
13. Aronsky, D., Haug, P.J., Lagor, C. & Dean, N.C. Accuracy of administrative data for identifying patients with pneumonia. *Am. J. Med. Qual.* **20**, 319–328 (2005).
14. Birman-Deych, E., Waterman, A.D., Yan, Y., Nilasena, D.S., Radford, M.J. & Gage, B.F. Accuracy of ICD-9-CM codes for identifying cardiovascular and stroke risk factors. *Med. Care* **43**, 480–485 (2005).
15. Jensen, P.N., Johnson, K., Floyd, J., Heckbert, S.R., Carnahan, R. & Dublin, S. A systematic review of validated methods for identifying atrial fibrillation using administrative data. *Pharmacoepidemiol. Drug Saf.* **21 Suppl 1**, 141–147 (2012).
16. Jones, N., Schneider, G., Kachroo, S., Rotella, P., Avetisyan, R. & Reynolds, M.W. A systematic review of validated methods for identifying acute respiratory failure using administrative and claims data. *Pharmacoepidemiol. Drug Saf.* **21**(Suppl 1), 261–264 (2012).
17. Kiyota, Y., Schneeweiss, S., Glynn, R.J., Cannuscio, C.C., Avorn, J. & Solomon, D.H. Accuracy of Medicare claims-based diagnosis of acute myocardial infarction: estimating positive predictive value on the basis of review of hospital records. *Am. Heart J.* **148**, 99–104 (2004).
18. Kumamaru, H. *et al*. Validity of claims-based stroke algorithms in contemporary Medicare data: reasons for geographic and racial differences in stroke (REGARDS) study linked with medicare claims. *Circ. Cardiovasc. Qual. Outcomes* **7**, 611–619 (2014).
19. Mapel, D.W. *et al*. Development and validation of a healthcare utilization-based algorithm to identify acute exacerbations of chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.* **16**, 1687–1698 (2021).
20. Snyder, B.M. *et al*. Validation of international classification of diseases criteria to identify severe influenza hospitalizations. *Influenza Other Respi. Viruses* **16**, 371–375 (2022).
21. Ye, Y., Larrat, E.P. & Caffrey, A.R. Algorithms used to identify ventricular arrhythmias and sudden cardiac death in retrospective studies: a systematic literature review. *Ther. Adv. Cardiovasc. Dis.* **12**, 39–51 (2018).
22. Andersson, T., Isaksson, A., Khalil, H., Lapidus, L., Carlberg, B. & Söderberg, S. Validation of the Swedish National Inpatient Register for the diagnosis of pulmonary embolism in 2005. *Pulm. Circ.* **12**, e12037 (2022).
23. Austin, P.C. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat. Med.* **28**, 3083–3107 (2009).
24. Austin, P.C. Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Commun. Stat. Simul. Comput.* **38**, 1228–1234 (2009).
25. Hammill, B.G. *et al*. Fitness of real-world data for clinical trial data collection: results and lessons from a HARMONY outcomes ancillary study. *Clin. Trials* **19**, 655–664 (2022).
26. Cohen, J.F. *et al*. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open* **6**, e012799 (2016).
27. Mallett, S., Halligan, S., Thompson, M., Collins, G.S. & Altman, D.G. Interpreting diagnostic accuracy studies for patient care. *BMJ* **345**, e3999 (2012).
28. Hollander, M., Wolfe, D.A. & Chicken, E. The one-sample location problem. In *Nonparametric Statistical Methods* 3rd edn. 39–114 (John Wiley & Sons, Inc., Hoboken, NJ, 2015).
29. Jarrín, O.F., Nyandege, A.N., Grafova, I.B., Dong, X. & Lin, H. Validity of race and ethnicity codes in Medicare administrative data compared with gold-standard self-reported race collected during routine home health care visits. *Med. Care* **58**, e1–e8 (2020).
30. Imai, K., Olivella, S. & Rosenman, E.T.R. Addressing census data problems in race imputation via fully Bayesian improved surname geocoding and name supplements. *Sci. Adv.* **8**, eadc9824 (2022).
31. Eicheldinger, C. & Bonito, A. More accurate racial and ethnic codes for Medicare administrative data. *Health Care Financ. Rev.* **29**, 27–42 (2008).
32. Elliott, M.N., Fremont, A., Morrison, P.A., Pantoja, P. & Lurie, N. A new method for estimating race/ethnicity and associated disparities where administrative records lack self-reported race/ethnicity. *Health Serv. Res.* **43**(5 Pt 1), 1722–1736 (2008).
33. Copeland, K.T., Checkoway, H., McMichael, A.J. & Holbrook, R.H. Bias due to misclassification in the estimation of relative risk. *Am. J. Epidemiol.* **105**, 488–495 (1977).
34. Jurek, A.M., Greenland, S., Maldonado, G. & Church, T.R. Proper interpretation of non-differential misclassification effects: expectations vs observations. *Int. J. Epidemiol.* **34**, 680–687 (2005).
35. Lash, T.L., Fox, M.P., MacLehose, R.F., Maldonado, G., McCandless, L.C. & Greenland, S. Good practices for quantitative bias analysis. *Int. J. Epidemiol.* **43**, 1969–1985 (2014).
36. Greenland, S. & Kleinbaum, D.G. Correcting for misclassification in two-way tables and matched-pair studies. *Int. J. Epidemiol.* **12**, 93–97 (1983).
37. Brenner, H. & Gefeller, O. Use of the positive predictive value to correct for disease misclassification in epidemiologic studies. *Am. J. Epidemiol.* **138**, 1007–1015 (1993).
38. Weinstein, E.J., Ritchey, M.E. & Lo Re, V. 3rd. Core concepts in pharmacoepidemiology: validation of health outcomes of interest within real-world healthcare databases. *Pharmacoepidemiol. Drug Saf.* **32**, 1–8 (2023).
39. Nakasian, S.S., Rassen, J.A. & Franklin, J.M. Effects of expanding the look-back period to all available data in the assessment of covariates. *Pharmacoepidemiol. Drug Saf.* **26**, 890–899 (2017).
40. Connolly, J.G., Schneeweiss, S., Glynn, R.J. & Gagne, J.J. Quantifying bias reduction with fixed-duration versus all-available covariate assessment periods. *Pharmacoepidemiol. Drug Saf.* **28**, 665–670 (2019).
41. Brunelli, S.M. *et al*. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiol. Drug Saf.* **22**, 542–550 (2013).
42. Al-Shahi, R., Vousden, C. & Warlow, C. Bias from requiring explicit consent from all participants in observational research: prospective, population based study. *BMJ* **331**, 942 (2005).
43. Lin, K.J., Singer, D.E., Glynn, R.J., Murphy, S.N., Lii, J. & Schneeweiss, S. Identifying patients with high data completeness to improve validity of comparative effectiveness research in electronic health records data. *Clin. Pharmacol. Ther.* **103**, 899–905 (2018).
44. Goldstein, B.A., Bhavsar, N.A., Phelan, M. & Pencina, M.J. Controlling for informed presence bias due to the number of health encounters in an electronic health record. *Am. J. Epidemiol.* **184**, 847–855 (2016).
45. Bernstam, E.V. *et al*. Real-world matching performance of deidentified record-linking tokens. *Appl. Clin. Inform.* **13**, 865–873 (2022).
46. St Peter, W.L., Liu, J., Weinhandl, E.D. & Fan, Q. Linking centers for Medicare & Medicaid Services data with prospective DCOR trial data: methods and data comparison results. *Hemodial. Int.* **12**, 480–491 (2008).
47. Hlatky, M.A. *et al*. Use of Medicare data to identify coronary heart disease outcomes in the Women's Health Initiative. *Circ. Cardiovasc. Qual. Outcomes* **7**, 157–162 (2014).