


# Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research

Elisabetta Patorno <sup>1</sup>, Sebastian Schneeweiss,<sup>1</sup> Mary G George,<sup>2</sup> Xin Tong,<sup>2</sup> Jessica M Franklin,<sup>1</sup> Ajinkya Pawar,<sup>1</sup> Helen Mogun,<sup>1</sup> Lidia M V R Moura,<sup>3</sup> Lee H Schwamm<sup>4</sup>

**To cite:** Patorno E, Schneeweiss S, George MG, *et al.* Linking the Paul Coverdell National Acute Stroke Program to commercial claims to establish a framework for real-world longitudinal stroke research. *Stroke & Vascular Neurology* 2022;**7**: e001134. doi:10.1136/svn-2021-001134

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/svn-2021-001134>).

Received 27 May 2021  
Accepted 12 October 2021  
Published Online First  
8 November 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Elisabetta Patorno;  
epatorno@bwh.harvard.edu

## ABSTRACT

**Background** Non-interventional large-scale research on real-world patients who had a stroke requires the use of multiple data sources ensuring access to longitudinal data from large populations with clinically-detailed information. We sought to establish a framework for longitudinal research on patients hospitalised with stroke by linking information-rich, deidentified inpatient data from the Paul Coverdell National Acute Stroke Program (PCNASP) to commercial and Medicare Advantage longitudinal claims data.

**Methods** All stroke admissions in PCNASP between 2008 and 2015 were evaluated for linkage to longitudinal claims from a commercial insurer using an algorithm based on six available common data fields (patient age, gender, admission date, discharge date, discharge diagnosis and state) and a hospital match. We evaluated the linkage quality (via the percentage of unique records in the linked dataset) and the representativeness of the linked population. We also described medical history, stroke severity and patterns of medication use among the PCNASP-claims linked cohort.

**Results** The linkage produced uniqueness equal to 99.1%. We identified 5644 linked and 98 896 unlinked patients who had an ischaemic stroke hospitalisation in claims data. Linked patients were younger than unlinked (69.7 vs 72.5 years), but otherwise similar by medical history, prestroke medication use or lab values. Stroke severity was mild and most patients were discharged home. Prestroke and discharge use of antihypertensive and statins in the PCNASP were greater than their use as measured by filled prescriptions in claims.

**Conclusions** High-quality linkage between the PCNASP and commercial claims data is feasible. This linkage identified differences between reported or recommended versus actual out-of-hospital medication utilisation, highlighting the importance of longitudinal data availability for research aimed to improve the care of patients who had a stroke.

## INTRODUCTION

Several classes of medications have shown to be effective in managing stroke risk factors and secondary stroke prevention,<sup>1 2</sup> up to approximately 80% as compared with no treatment.<sup>3</sup> Healthcare utilisation databases can be a useful tool to study the use and the comparative safety

and effectiveness of therapeutics in routine care of patients who had ischaemic stroke, and thus complement information from randomised controlled trials (RCTs), which although generally considered the gold standard for establishing the causal relationship between interventions and patient outcomes, are often costly, take a long time to complete, and are often applicable to only a narrow patient population.<sup>4-6</sup> These databases allow for the creation of a continuous record of hospitalisations, outpatient care, and medication use, but they often lack sufficient detailed information on critical clinical characteristics such as disease severity.<sup>7</sup> Furthermore, most administrative datasets do not contain in-hospital drug use information. By contrast, clinical stroke registries are generally rich in clinical detail but lack longitudinal data and are deidentified abstractions of the medical record without explicit patient consent. The Paul Coverdell National Acute Stroke (PCNASP) was established by the Centers for Disease Control and Prevention in 2001 to collect data on the quality of care provided to patients who had a stroke from the initial emergency response through hospital discharge with the goal of improving the quality of hospital-based acute stroke care.<sup>8-10</sup>

In such a context, the linkage with alternative data sources such as clinical or quality improvement registries, to confirm clinical diagnoses that may not be accurately recorded in claims and to collect information on disease severity and inpatient medications and procedures, provides an opportunity for clinical research on the use and the effects of medications in large real-world patient populations.

While a body of literature exists on Medicare fee for service claims linked to clinical stroke registries,<sup>11-17</sup> little is known about commercially insured patients under age 65 and those in Medicare Advantage plans. In

a cohort study of patients who had ischaemic stroke, we sought to determine: (1) the feasibility of linking patients from an inpatient deidentified stroke registry (ie, the PCNASP) to a commercial claims dataset with longitudinal information on inpatient and outpatient care and out-of-hospital filled prescriptions, (2) the representatives of the linked versus unlinked claims patients and (3) the reliability of registry-derived clinical inpatient and outpatient information such as prestroke and postdischarge medication use by comparing it to administrative claims data on filled prescriptions in the linked population.

## METHODS

### Data sources

#### Inpatient stroke data from the PCNASP

Information on hospitalised patients who had an ischaemic stroke was available via PCNASP between January 2008 and September 2015. The PCNASP was established in 2001 and from 2007 to 2011 collected data from acute care hospitals in six states (Georgia, Massachusetts, Michigan, Minnesota, Ohio, and North Carolina), increasing to 11 states in 2012–2015 (Arkansas, California, Georgia, Iowa, Massachusetts, Michigan, Minnesota, New York, North Carolina, Ohio and Wisconsin).<sup>18</sup> PCNASP includes patients aged  $\geq 18$  years with a clinical diagnosis of acute ischaemic stroke, intracerebral haemorrhage, subarachnoid haemorrhage or transient ischaemic attack (TIA) or an International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code indicative of a stroke or TIA.<sup>19</sup> PCNASP collects several in-hospital data elements including stroke subtype, National Institutes of Health (NIH) Stroke Scale Score, and prescribed preventative pharmacological therapy at discharge, for example, antihypertensive, statin and antithrombotic treatments.<sup>18</sup>

#### Claims data

Longitudinal information on commercially insured patients was collected from the Clinformatics Data Mart (OptumInsight, Eden, Prairie, Minnesota, USA) between 2008 and 2015, a US-based healthcare insurance dataset including deidentified administrative claims for over 14 million persons annually (hereinafter referred to as Optum). The patients in this dataset are commercial health plan members and Medicare Advantage members (approximately 25%–30%) from all 50 states. For each enrollee, the dataset includes demographic information, health plan enrollment status, inpatient and outpatient medical encounters coded via ICD-9CM and Current Procedural Terminology (CPT)-4 classifications, and filled prescriptions, including National Drug Code numbers, quantity dispensed, and days' supply. Claims data are deterministically linked to laboratory test results provided by two national laboratory providers, with results for outpatient laboratory tests available for approximately a third of beneficiaries.

### Data linkage strategy

In the absence of direct patient identifiers, we used the following six available common data fields to link the PCNASP and the Optum dataset: patient age, gender, admission date, discharge date, discharge diagnosis and state. In order to improve linkage validity, we required a hospital match to occur first.<sup>20</sup> We established a hospital 'crosswalk' by matching hospitalisations from the inpatient PCNASP data with hospitalisations from the claims, on the basis of exact values for the six common data fields. We restricted the crosswalk to hospitals with at least five exact patient level matches. The hospital identified in claims data that contained the majority of exactly matched hospitalisations for any given PCNASP hospital was assumed to be the correct link for that PCNASP hospital.

Within each hospital, we calculated the percentage of records that were unique after implementing the linkage, as previously described.<sup>20</sup> This was defined as follows:

$$\text{Uniqueness} = 1 - \frac{(\text{N multiple records in claims} + \text{N multiple records in PCNASP})}{\text{N linked records}} * 100$$

where,

N multiple records in claims=number of multiple records in claims that linked to the same record in PCNASP,

N multiple records in PCNASP=number of multiple records in PCNASP that linked to the same record in claims,

N linked records=total number of records in claims and PCNASP for which linkage was possible.

We linked PCNASP and Optum dataset using increasingly stringent criteria with respect to matching by age and dates of admission or discharge, and used the strictest linkage rule (uniqueness equal to 99.1%), to identify a study population of linked patients who had ischaemic stroke for which we had high confidence that the linkage accurately identified the same patient in claims and PCNASP.

### Study population

Within the linked study population, we restricted to patients who had a first index ischaemic stroke hospitalisation and six or more months of continuous health plan enrollment in claims prior to that hospital admission. Within the same time period, we also identified patients who had a first index ischaemic stroke hospitalisation in claims and no linkage with the PCNASP, and applied the same inclusion criteria as for linked patients. To identify patients who had ischaemic stroke in Optum linked and unlinked patients, we used a definition previously validated against medical records, based on a primary discharge diagnosis of ischaemic stroke (ICD-9 codes 433.x1, 434.xx (excluding 434.x0), or 436.xx),<sup>21 22</sup> during the index hospitalisation.

### Characteristics of linked and unlinked claims-based populations with ischaemic stroke

To assess the representativeness of the claims-based population with linkage to the PCNASP, with respect to the

claims-based population without linkage to the PCNASP, we compared baseline patient characteristics between Optum enrollees with ischaemic stroke who were linked and not linked to the PCNASP data. Baseline patient characteristics of interest were measured in claims during the 6 months preceding the index stroke hospitalisation and included demographic information, comorbid conditions, use of medications, measures of healthcare utilisation, and, for a subset of the study cohort, selected baseline laboratory test results. We also described characteristics of the index hospitalisation including length of stay, discharge status, and in-hospital death, when available.

### Characterisation of cohorts of patients who had an ischaemic stroke through PCNASP and claims-based longitudinal information

Within the linked study population, we assessed common fields pertinent to medical history and use of medications prior to the index stroke hospitalisation as measured in both PCNASP and claims. History of comorbid conditions at baseline was assessed through ICD-9 diagnostic and procedural codes in claims, and through recorded information on past medical history in PCNASP. Active use of antihypertensive and lipid-lowering medication treatment was measured via filled prescriptions during the 90 days prior to the stroke hospitalisation in claims, and patient's reported medication use on admission in PCNASP. Baseline stroke severity was captured by NIH Stroke Scale (NIHSS) scores and by degree of independence in ambulation at discharge, both recorded in the PCNASP.

### Statistical analysis

To evaluate the representativeness of the linked population, we examined descriptive statistics for a range of claims-based characteristics of patients linked to the PCNASP versus unlinked patients. We quantified the differences between linked and unlinked patients via standardised differences, using the conventional definition of meaningful differences as values greater than 0.1.<sup>23</sup> Standardised differences compare the difference in means in units of the pooled SD and, unlike p values, are not influenced by sample size. Thus, in the context of a large sample, such as the current study, standardised differences are the preferred tool to identify meaningful differences in covariates.

To describe characteristics of linked patients who had an ischaemic stroke through PCNASP and claims-based longitudinal information, we evaluated frequencies and percentages for binary variables; and means (SD) and medians (IQR) for continuous variables. For the information captured by both PCNASP and claims within the linked population, we evaluated the presence of any meaningful discordance via McNemar test for paired nominal data and also provided absolute per cent differences. Percentages, means and medians for PCNASP variables were calculated using only the data of stroke hospitalisations that occurred in the time periods for which the

specific information was collected (online supplemental material table 1).

Within the linked study population, we also identified a subset of patients who were directly discharged home after the index stroke hospitalisation, in order to reliably measure dispensing of filled prescription medications after the stroke hospitalisation in claims data, since patients do not have individual pharmacy claims while in a postacute care facility. In this population, we described the frequency of antihypertensive and statin treatment prescribed at discharge based on information from the PCNASP, with the frequency of filled prescriptions for antihypertensive and statin medications in the 90 days after discharge from the same hospitalisation.

In analyses stratified by age (<65 and ≥65 years) and by coverage (commercial and Medicare Advantage), we evaluated the representativeness of the linked population with stroke, described patients' characteristics as captured by PCNASP and by longitudinal claims, and compared the frequency of antihypertensive and statin treatment prescribed at discharge based on information from the PCNASP, with the frequency of filled prescriptions for antihypertensive and statin medications in the 90 days after discharge from the same hospitalisation.

### RESULTS

There were 32 991 571 patients contained in the claims dataset and 574 586 hospitalisations in PCNASP from 2008 to 2015. After applying a strict linkage rule based on exact matching of linking fields (table 1, linkage step 5), we successfully linked 10 079 hospitalisations among 9548 unique patients in Optum to hospitalisations in the PCNASP.

When we further restricted Optum linked and unlinked patients to those with a primary discharge diagnosis of ischaemic stroke during the index hospitalisation and with six or more months of continuous enrollment prior to the hospital admission, there were 5644 linked and 98 896 unlinked patients in claims data, that is, information from the PCNASP was available for 5.4% of Optum patients who had an ischaemic stroke (figure 1). After the linkage, most of the data derived from acute care hospitalisations in Georgia, North Carolina, Ohio and Minnesota.

Claims-based patient characteristics between enrollees for whom PCNASP information was available (linked enrollees) versus not available (unlinked enrollees) were balanced with most standardised differences <0.1 (table 2).

In both cohorts within claims data, PCNASP-linked and unlinked patients had similar gender distribution (approximately 50% women), burden of comorbidities as measured by the combined comorbidity index, a claims-based score with lower values associated with lower mortality risk and higher values associated with higher mortality risk,<sup>24</sup> and medication use prior to the hospitalisation. Minor imbalances were noted for a few characteristics. Compared with patients without PCNASP linkage, PCNASP-linked patients in Optum were slightly younger (69.2 vs 72.5 years), had a

**Table 1** Linkage performance in claims dataset of 32 991 571 claims encounters

Linkage step	Linkage rule*†	Number of flexible fields for matching	Total linked Hospitalisations	Uniqueness‡
1	Age at admission ( $\pm 1$ year), date of admission ( $\pm 1$ day), date of discharge ( $\pm 1$ day), sex, state, and primary diagnosis, by hospital ID	3	17 850	89.6
2	Age at admission, date of admission ( $\pm 1$ day), date of discharge, sex, state, and primary diagnosis, by hospital ID	1	10 917	98.0
3	Age at admission, date of admission, date of discharge ( $\pm 1$ day), sex, state, and primary diagnosis, by hospital ID	1	10 445	98.0
4	Age at admission ( $\pm 1$ year), date of admission, date of discharge, sex, state, and primary diagnosis, by hospital ID	1	14 004	98.0
5	Age at admission, date of admission, date of discharge, sex, state, and primary diagnosis, by hospital ID	0	10 079	99.1

\*Unless otherwise specified, linkage rules employed exact matching of linking fields.

†Optum Clinformatics only includes year of birth for commercial beneficiaries, thus the linking field age at admission was always allowed to differ by 1 year.

‡Uniqueness =  $(1 - (N \text{ multiple hospitalisations in claims} + N \text{ multiple hospitalisations in registry}) / (N \text{ linked hospitalisations})) * 100$

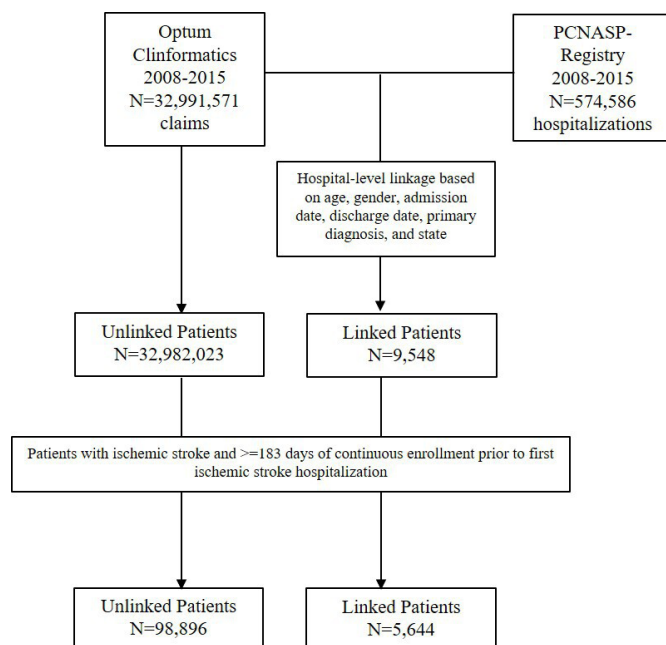
higher number of physician visits at baseline and a higher baseline total cholesterol and haemoglobin A1c level, and were more frequently discharged to home healthcare. During the 6 months prior to the stroke hospitalisation, prescriptions were filled for an antihypertensive in over 60% of patients, a statin in ~35% and an anticoagulant or antiplatelet in ~10%. In addition, measures of healthcare utilisation, baseline laboratory test results and characteristics of the index stroke hospitalisation were similar, with ~20%

of patients experiencing a hospitalisation in the prior 6 months, of ~6 days length. Claims-based patient characteristics between PCNASP-linked and unlinked patients were also balanced with most standardised differences  $< 0.1$ , when we stratified by age and coverage type (online supplemental material tables 2 and 3).

When we explored the concordance of baseline medical history in the PCNASP-linked patients between the claims and PCNASP sources, we found discordance was overall low and identified only a few variables for which discordance was higher as measured by McNemar test and absolute per cent differences (table 3).

In particular, the prevalence of history of diabetes mellitus, carotid stenosis, peripheral vascular disease, congestive heart failure and depression was higher in claims. Conversely, the prevalence of history of atrial fibrillation, obesity and cigarette smoking was higher in the PCNASP compared with claims, consistent with expected under-recording practice for these variables in claims data. Stroke severity was often mild or moderate (ie, mean initial NIHSS was 6.2 with a median of 3), and 50.4% patients were able to ambulate independently with or without a device at discharge. When we stratified by age and coverage type (online supplemental material tables 4 and 5), findings remained consistent, except for a higher prevalence of hypertension and chronic kidney disease in the PCNASP compared with claims among patients younger than 65 and among commercially insured patients.

Patients filled prescriptions for antihypertensive and lipid-lowering medications in the 90-day period prior to the stroke hospitalisation less frequently than they reported these medications as being currently taken at admission (figure 2). Among PCNASP-linked patients discharged home, antihypertensive treatment was prescribed at discharge in 81.6% of



**Figure 1** Identification of linked and unlinked study cohorts resulting from data linkage between ischaemic stroke admissions in Paul Coverdell National Acute Stroke Program (PCNASP)—registry and ischaemic stroke admissions in claims dataset.

**Table 2** Comparison of claims variables between beneficiaries in Optum with versus without successful Paul Coverdell National Acute Stroke Program-linkage

Patient characteristics	Linked (n=5644) N (%)	Unlinked (n=98896) N (%)	Standardised difference
<b>Demographics</b>			
Age, mean (SD)	69.7 (11.9)	72.5 (12.3)	-0.23
Female	2797 (49.6)	51 821 (52.4)	-0.06
<b>Comorbidities*</b>			
Combined comorbidity index, mean (SD)	1.4 (2.5)	1.5 (2.4)	-0.03
Hypertension	3866 (68.5)	64 675 (65.4)	0.07
Diabetes	1974 (35.0)	31 964 (32.3)	0.06
Dyslipidaemia	2710 (48.0)	43 584 (44.1)	0.08
Prior ischaemic stroke	1192 (21.1)	17 852 (18.1)	0.08
Prior transient ischaemic attack	526 (9.3)	8802 (8.9)	0.01
History of atrial fibrillation	820 (14.5)	17 384 (17.6)	-0.08
Carotid stenosis	411 (7.3)	6342 (6.4)	0.03
Prior haemorrhagic stroke	83 (1.5)	1067 (1.1)	0.03
Ischaemic heart disease or procedure	1341 (23.8)	22 974 (23.2)	0.01
Peripheral vascular disease (PVD) or PVD surgery	497 (8.8)	10 024 (10.1)	-0.05
Congestive heart failure	786 (13.9)	14 487 (14.7)	-0.02
Chronic kidney disease	699 (12.4)	13 338 (13.5)	-0.03
Chronic obstructive pulmonary disease	694 (12.3)	12 772 (12.9)	-0.02
Pneumonia	346 (6.1)	5890 (6.0)	0.01
Dementia	575 (10.2)	12 087 (12.2)	-0.06
Cancer or history of malignant neoplasm	661 (11.7)	11 328 (11.5)	0.01
Obesity	811 (14.4)	13 357 (13.5)	0.02
Smoking	455 (8.1)	6411 (6.5)	0.06
Alcohol abuse or dependence	76 (1.4)	1221 (1.2)	0.01
Drug abuse or dependence	61 (1.1)	991 (1.0)	0.01
Depression	556 (9.9)	9194 (9.3)	0.02
<b>Medication use*</b>			
Any antihypertensives†	3437 (60.9)	62 519 (63.2)	-0.05
Angiotensin-converting enzyme inhibitors	1556 (27.6)	28 447 (28.8)	-0.03
Angiotensin II receptor blockers	738 (13.1)	12 944 (13.1)	0.00
Beta-blockers	1980 (35.1)	35 563 (36.0)	-0.02
Calcium channel blockers	1361 (24.1)	23 941 (24.2)	0.00
Thiazide diuretics	1147 (20.3)	18 151 (18.4)	0.05
Loop diuretics	837 (14.8)	15 521 (15.7)	-0.02
Other antihypertensives	458 (8.1)	8045 (8.1)	0.00
Nitrates and other antianginal therapies	389 (6.9)	6088 (6.2)	0.03
Antiarrhythmics	105 (1.9)	2044 (2.1)	-0.01
Digoxin	176 (3.1)	4329 (4.4)	-0.07
Any lipid-lowering agents‡	2144 (38.0)	36 294 (36.7)	0.03
Statins	2016 (35.7)	33 977 (34.4)	0.03
Other lipid-lowering agents	340 (6.0)	5231 (5.3)	0.03
Antiplatelets§	694 (12.3)	10 480 (10.6)	0.05
Anticoagulants¶	491 (8.7)	9163 (9.3)	-0.02
Warfarin	383 (6.8)	7778 (7.9)	-0.04
Direct oral anticoagulants	94 (1.7)	1198 (1.2)	0.04
Heparin, LMWH or fondaparinux	66 (1.2)	935 (1.0)	0.02

Continued

Table 2 Continued

Patient characteristics	Linked (n=5644) N (%)	Unlinked (n=98896) N (%)	Standardised difference
Insulin	622 (11.0)	8937 (9.0)	0.07
Non-insulin glucose-lowering medications	1034 (18.3)	17 429 (17.6)	0.02
Antidepressants**	1145 (20.3)	19 146 (19.4)	0.02
<b>Measures of healthcare utilisation*</b>			
Any hospitalisation, %	1215 (21.5)	17 462 (17.7)	0.10
Number of any hospitalisation, mean (SD)	0.3 (0.7)	0.3 (0.7)	0.08
Any hospitalisation within prior 30 days, %	467 (8.3)	6426 (6.5)	0.07
N hospital days, mean (SD)	2.0 (6.1)	1.6 (5.6)	0.07
Number of emergency department visits, mean (SD)	0.8 (2.0)	0.7 (1.9)	0.09
Number of any physician visit, mean (SD)	12.8 (17.8)	10.5 (15.7)	0.14
Total N distinct pharmacological agents prescribed, mean (SD)	6.5 (6.0)	6.3 (5.6)	0.04
<b>Laboratory values*</b>			
LDL, mg/dL, mean (SD)	105.3 (56.9)	106.9 (44.7)	-0.03
LDL, N (%)	123 (2.2)	2471 (2.5)	-0.02
Total cholesterol, mg/dL, mean (SD)	199.6 (61.9)	193.3 (48.8)	0.11
Total cholesterol, N (%)	120 (2.1)	2514 (2.5)	-0.03
INR, mean (SD)	1.6 (0.9)	1.7 (0.9)	-0.02
INR, N (%)	19 (0.3)	384 (0.4)	-0.01
Creatinine, mg/dL, mean (SD)	1.1 (0.5)	1.1 (2.1)	-0.05
Creatinine, N (%)	171 (3.0)	3393 (3.4)	-0.02
HbA1c, %, mean (SD)	8.2 (2.5)	7.8 (2.2)	0.15
HbA1c, N (%)	90 (1.6)	1530 (1.6)	0.00
<b>Characteristics of stroke hospitalisation</b>			
Length of stay of index hospitalisation, mean (SD)	5.7 (4.7)	5.9 (6.8)	-0.05
<b>Discharge status</b>			
Home	2691 (47.7)	41 191 (41.7)	0.12
Home healthcare	680 (12.1)	10 376 (10.5)	0.05
Rehabilitation facility	480 (8.5)	8596 (8.7)	-0.01
Skilled nursing facility	935 (16.6)	19 766 (20.0)	-0.09
Other acute inpatient facility	331 (5.9)	6701 (6.8)	-0.04
Hospice	190 (3.4)	4098 (4.1)	-0.04
In-hospital mortality††	33 (0.6)	516 (0.5)	0.01

Values are N (%) unless otherwise specified.

\*Unless otherwise specified, measured during the 6 months preceding the index stroke hospitalisation.

†Includes ACE-inhibitors, ARBs, beta-blockers, calcium channel blockers, thiazide diuretics, loop diuretics, and other antihypertensives.

‡Includes statins or other lipid-lowering medications.

§Includes aspirin, clopidogrel, prasugrel, ticagrelor, dipyridamole, aspirin-dipyridamole, ticlopidine.

¶Includes warfarin, dabigatran, rivaroxaban, apixaban, edoxaban, heparin, dalteparin, enoxaparin, tinzaparin, fondaparinux, argatroban, desirudin, lipirudin.

\*\*Includes selective serotonin reuptake inhibitors (SSRIs) and non-SSRI antidepressants.

††Information on mortality is available in Optum through linkage with the Social Security Administration Death Master File. This capture is limited by a policy change in 2011 concerning the extent of the Social Security Administration disclosure of death records received from states (important notice: change in public death master file records. 2011; <https://classic.ntis.gov/assets/pdf/import-change-dmf.pdf>).

ARBs, angiotensin II receptor blockers; HbA1c, haemoglobin A1c; INR, international normalised ratio; LDL, low-density lipoprotein; LMWH, low-molecular-weight heparin.

cases; however, claims for prescriptions filled post discharge were found in only 61.9%. Similar patterns were observed for statin treatment (prescribed at discharge in 84.7% of cases and filled in the first 90 days post discharge in 61.1% of cases). Findings were similar when we stratified by age and coverage type (online supplemental material figures 1 and 2).

## DISCUSSION

In a large cohort of patients hospitalised with ischaemic stroke, we found that a reliable linkage between the PCNASP, a rich inpatient stroke registry, and commercial claims data using indirect identifiers was feasible, and permitted combining detailed disease markers of the acute stroke care episode with longitudinal postdischarge care

**Table 3** Assessment of discordance among selected variables from the Paul Coverdell National Acute Stroke Program (PCNASP) and claims data among linked patients

	PCNASP-based variables (n=5644)	Claims-based variables (n=5644)	McNemar's test p value	Absolute difference
<b>Comorbidities*, N (%)</b>				
Hypertension	4021 (71.5)	3866 (68.5)	0.00	3.0
Diabetes mellitus	1780 (31.8)	1974 (35.0)	0.00	3.2
Dyslipidaemia	2590 (46.1)	2710 (48.0)	0.01	1.9
Prior ischaemic stroke	1075 (21.9)	1192 (21.1)	0.12	0.8
Prior transient ischaemic attack	406 (8.1)	526 (9.3)	0.08	1.2
History of atrial fibrillation	1155 (20.5)	820 (14.5)	0.00	6.0
Carotid stenosis	222 (4.0)	411 (7.3)	0.00	3.3
Ischaemic heart disease	1323 (23.6)	1341 (23.8)	0.70	0.2
Peripheral vascular disease	315 (5.6)	497 (8.8)	0.00	3.2
Congestive heart failure	449 (8.0)	786 (13.9)	0.00	5.9
Chronic kidney disease	138 (11.4)	699 (12.4)	0.00	1.0
Obesity	624 (39.6)	811 (14.4)	0.00	25.2
Smoking	1024 (18.3)	455 (8.1)	0.00	10.2
Drug or alcohol abuse	32 (2.8)	61 (1.1)	1.00	1.7
Depression	174 (5.6)	556 (9.9)	0.00	4.3
<b>Stroke severity and functional information at discharge†</b>				
NIH Stroke Scale				
Mean (SD)	6.2 (7.2)	N/A	N/A	N/A
Median (IQR)	3 (1–8)	N/A	N/A	N/A
Ambulatory status at discharge				
Able to ambulate independently with or without device, N (%)	2698 (50.4)	N/A	N/A	N/A

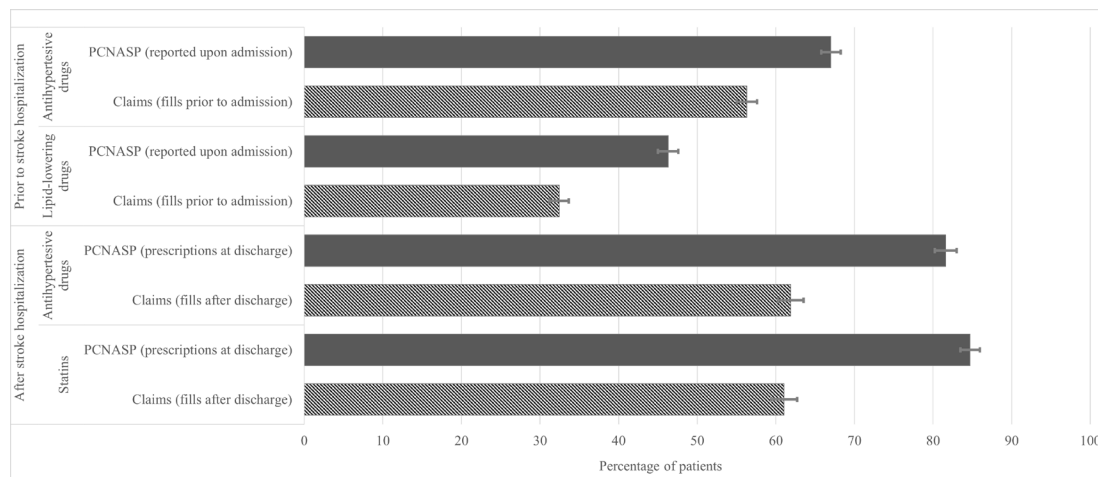
\*Comorbidities in the PCNASP are based on recorded medical history during stroke hospitalisation; comorbidities for in claims are based on International Classification of Diseases, Ninth Revision diagnoses recorded during the 183-day period prior to the stroke hospitalisation. Comorbidities in the PCNASP were characterised by varying level of missingness: information on prior ischaemic stroke, transient ischaemic attack, obesity, chronic kidney disease, drug or alcohol abuse, and depression was missing for 13.0%, 11.6%, 72.1%, 78.6%, 79.6% and 44.7%, respectively; information on the other comorbidities was missing for <1%.

†Information on National Institutes of Health (NIH) Stroke Scale and Ambulatory status at discharge was missing for 28.9% and 5.2% patients, respectively.  
N/A, not applicable.

in a vulnerable population of patients at high risk of recurrent stroke. Despite small differences in a few variables, the PCNASP-claims linked subset appeared to be overall representative of the general claims-based population with stroke. Detailed data on stroke severity (NIHSS), and ambulatory status at discharge were available for most linked patients in the PCNASP, providing crucial information for predictors and risk adjustment in clinical and pharmacoepidemiological analyses addressing the postdischarge period. Medication information on antihypertensive and lipid-lowering treatments from the PCNASP was discordant with claims-based drug utilisation patterns outside of the hospital, with reported use of medications on admission and prescriptions at discharge largely overestimating the real-world use of medications as measured by filled prescriptions, which highlights the role of a longitudinal framework based on dispensing information from claims to accurately assess the

use of medications outside of the hospital. Use of medications as measured by drug dispensing in claims is known to strongly correlate with medication use by patients, in contrast to prescribing information, typically captured in medical records, which may overestimate medication use by patients who fail to fill their prescriptions, resulting in substantial bias.<sup>25</sup>

Real World Evidence (RWE), the understanding of causal treatment effects from electronic data generated by the routine provision of care, has gained much attention from regulators, payers and physician groups.<sup>26–30</sup> RWE is thought to complement essential evidence on the efficacy of medications that we gain from RCTs, by providing information on their safety and effectiveness in clinical practice.<sup>5 6</sup> Generated evidence needs to be internally valid and generalisable to an identifiable target population in order to be actionable.<sup>31–33</sup> A valid evaluation



**Figure 2** Antihypertensive and lipid-lowering medication use prior and subsequent to index stroke hospitalisation as measured in Paul Coverdell National Acute Stroke Program (PCNASP) and in claims at 90 days prestroke and 90 days post discharge. PCNASP (reported on admission): based on medications reported on admission in PCNASP registry. Claims (fills prior to admission): based on prescription medications filled in the 90 days prior to the stroke hospitalisation in claims. Antihypertensive and lipid-lowering medication use prior to admission in the PCNASP was missing for <1%. PCNASP (prescription at discharge): based on medications prescribed at discharge in PCNASP among patients discharged home. Antihypertensive and statin prescription at discharge in the PCNASP was missing for 11.3% and 1.7%, respectively. Claims (fills after discharge): based on prescription medications filled in the 90 days following discharge as recorded in claims among patients discharged home. To evaluate prescription medications filled following discharge in claims, analyses were limited to patients with continuous enrollment for the 90 days after the stroke hospitalisation. Antihypertensive drugs: includes ACE-inhibitors, angiotensin II receptor blockers (ARBs), beta-blockers, calcium channel blockers, thiazide diuretics, loop diuretics and other antihypertensives. Lipid-lowering drugs: includes statins or other lipid-lowering medications. PCNASP included information on the use of lipid-lowering drugs on admission and on the prescription of statins at discharge.

of patients who had a stroke in routine setting of care requires the joint contribution of data sources that can ensure the access to large populations with complete healthcare longitudinal information and rich clinical descriptors. Our study provides evidence that the creation of such a research framework is feasible and can provide a valid platform in which to address a range of clinical and pharmacoepidemiological research questions. These data resources can help understand why in some individuals evidence-based treatments fail to prevent recurrent events and can help identify during the hospitalisation those at greatest risk for non-adherence in the time period after discharge. The observation of important differences between reported or prescribing information and out-of-hospital medication dispensing for two mainstay treatments for stroke prevention highlights the challenges of assuming medication adherence subsequent to acute stroke hospitalisation based on discharge prescriptions. These findings are in line with results from previous studies assessing primary medication non-adherence in the USA, with estimated pooled primary non-adherence rates of 16% for antihypertensive medications and 25% for lipid-lowering medications,<sup>34</sup> and with rates as high as 28% for both classes in the primary care setting.<sup>35</sup> Our findings are also consistent with prior studies showing that more than half of patients stop taking their prescribed secondary prevention medications, including antihypertensive and lipid-lowering medication, 1–2 years after an incident stroke.<sup>36–39</sup> In a study evaluating 1-year

self-report of persistence and adherence to medications prescribed to patients after stroke discharge, up to one-third of patients who had a stroke discontinued one or more secondary prevention medications within 1 year of hospital discharge.<sup>37</sup> Of note, self-discontinuation of medications was uncommon, and several potentially modifiable patient, provider, and system-level factors associated with persistence and adherence may be targets for future interventions. The proposed research framework can hopefully help identify modifiable elements that can be the basis for taking quality improvement interventions to the next level, such as intensive education, remediation of social determinants of health, and better coordination with primary care after discharge. This framework will also provide the opportunity to study stroke care during readmissions as well as the use of other major medication classes, for example, antiplatelets, anticoagulants and glucose-lowering medications.

Another relevant example of a successful large-scale linkage between administrative data and alternative data sources with additional clinical insight on patients who had a stroke is the linkage between Medicare fee-for-service part A and B claims and the in-hospital Get With The Guidelines (GWTG)—Stroke Registry.<sup>20 40</sup> This enriched data source allowed investigators to follow Medicare fee-for-service beneficiaries who were linked to the GWTG-Stroke Registry for a range of claims-based clinical outcomes after acute ischaemic stroke hospitalisations.<sup>11–18</sup> However, even in this important example,



out-of-hospital medication information was limited to the documentation of treatment reported at hospital admission and drug prescriptions at discharge as recorded in the GWTC-Stroke Registry, without longitudinal data on prescriptions actively filled by the patient prior or subsequent to the stroke hospitalisation.<sup>11 12</sup> With the exception of over-the-counter medications and self-pay, our study could rely on complete information on filled prescriptions outside of the hospital. Another strength of this study is that we were able to provide information on a population enrolled in commercial or Medicare Advantage plans, which is a crucial complementary data source to those traditionally available from Medicare fee-for-service, Medicaid and Veterans Affairs.

Our study has limitations. First, as the data from claims and the PCNASP were deidentified in accordance with the data use agreement of these data sources, we could not confirm our linkage through the use of personal identifiers. However, our linkage was built on an established strategy successfully implemented in a similar setting, which demonstrated the feasibility of a reliable linkage between claims and an inpatient registry.<sup>20</sup> Second, information from the PCNASP was only available for 5.4% of Optum patients who had an ischaemic stroke. This is not evidence of poor performance of the linkage strategy, but it is likely driven by (1) the complete absence of information from US states represented in the Optum database but not participating to the PCNASP, (2) the limited presence of Optum enrollees in some of the US states that mostly contributed to the PCNASP registry during the study period and (3) the limited participation of hospitals to the PCNASP registry in some of the US states mostly represented in the Optum database. Nevertheless, we found that the Optum population with ischaemic stroke that was treated at hospitals participating to the PCNASP registry had similar characteristics to the remainder of the Optum population with ischaemic stroke that was treated at hospitals that did not participate to the PCNASP registry, suggesting that our findings likely apply to the broader Optum population with stroke. Similarly, information from the Optum database was only available for 1.7% of PCNASP patients, which could limit the generalisability of our findings to the remainder of the PCNASP population. This was not assessable as we did not have access to the unlinked PCNASP population. Third, PCNASP-based information on several variables was characterised by a considerable amount of missingness (online supplemental material table 6), in particular for the variables collected only later in the study period (eg, obesity, chronic kidney disease, drug or alcohol abuse, depression). Finally, claims data capture actual medication dispensing patterns and therefore provide a more complete and reliable approximation of medication use than electronic health records, which only capture prescriptions.<sup>25</sup> However, claims data do not capture actual medication use, and still have limitations in measuring drug use in certain settings, for example, over-the-counter medication use. Lastly, claims

do not provide information on the reasons patients are non-adherent or may discontinue medications, which could be due to financial hardship, forgetfulness or countermanding physician orders.

In conclusion, in a large cohort of patients who had a stroke, we found that a reliable linkage between the PCNASP and commercial claims data using indirect identifiers was feasible, representative and permitted combining detailed disease markers of the acute stroke care episode with reliable postdischarge longitudinal information in a vulnerable population of patients at high risk of recurrent stroke. This enriched data source will provide important insights into postdischarge evaluation of medication use and outcomes, ultimately improving the care of patients who had a stroke.

#### Author affiliations

<sup>1</sup>Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

<sup>2</sup>Division for Heart Disease and Stroke Prevention, NCCDPHP, Centers for Disease Control and Prevention, Atlanta, Georgia, USA

<sup>3</sup>Epilepsy and Neurophysiology Division, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

<sup>4</sup>Fireman Vascular Center and Stroke Division, Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA

**Acknowledgements** We thank Moa Lee and Julie A. Barberio for their contribution to early stages of this research.

**Contributors** EP, SS and LHS were involved in all parts of the study. MGG, XT, JMF and LMVRM were involved in designing the study and revising the manuscript. PA and HM were involved in data analysis and revising the manuscript. EP, SS and LHS are the guarantors.

**Funding** This study was funded by the Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA. EP was supported by a career development grant K08AG055670 from the National Institute on Aging. LM was supported by a career development grant 5K08AG05338002 from the National Institute on Aging.

**Disclaimer** The findings and conclusions in this report are those of the authors and do not necessarily represent the opinion of the Centers for Disease Control and Prevention.

**Competing interests** EP is co-investigator of an investigator-initiated grant to the Brigham and Women's Hospital from Boehringer-Ingelheim, not related to the topic of the submitted work. SS is the principal investigator of investigator-initiated grants to the Brigham and Women's Hospital from Bayer, Vertex, and Boehringer Ingelheim unrelated to the topic of this study. He is a consultant to WHISCON and to Aetion, a software manufacturer of which he owns equity. His interests were declared, reviewed, and approved by the Brigham and Women's Hospital and Partners HealthCare System in accordance with their institutional compliance policies. LHS is principal investigator of NIH-NINDS grants (U24NS107243), co-investigator of other grants (PCORI R-1609-35995, NIH-NINDS R01NS111952, NIH R01AG062770), and reports the following relationships relevant to research grants or companies that manufacture products for thrombolysis or thrombectomy even if the interaction involves non-thrombolysis products: scientific consultant regarding trial design and conduct to Genentech (late window thrombolysis) and Member of steering committee (TIMELESS NCT03785678); consultant on user interface design and usability to Lifelime; stroke systems of care consultant to the Massachusetts Dept of Public Health; member of a Data Safety Monitoring Boards (DSMB) for Penumbra (MIND NCT03342664); Serving as National PI for Medtronic (Stroke AF NCT02700945); National Co-PI, late window thrombolysis trial, NINDS (P50NS051343, MR WITNESS NCT01282242. LMVRM reports is principal investigator of a NIH-NIA grant (5K08AG053380-02), principal investigator of an investigator-initiated grant sponsored by the Epilepsy Foundation of America (60300-EFA-PCO-000-19-01), co-principal investigator of a CDC grant (SIP 20-007), co-investigator of other grants (NIH-NIA 5R01AG062282-02, NIH-NIA

2P01AG032952-11, NIH- NIA 3R01AG062282-03S1), and reports no conflict of interest.

**Patient consent for publication** Not applicable.

**Ethics approval** The Institutional Review Boards of the Mass General Brigham and the Centers for Disease Control and Prevention approved the study (#2014P002184). A signed data use agreement was in place. Informed consent was waived for study participants.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** No data are available. Data are not available as subjected to specific data use agreement that does allow data sharing.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Elisabetta Paterno <http://orcid.org/0000-0002-8809-9898>

## REFERENCES

- Sacco RL, Adams R, Albers G, *et al.* Guidelines for prevention of stroke in patients with ischemic stroke or transient ischemic attack: a statement for healthcare professionals from the American heart Association/American stroke association Council on stroke: co-sponsored by the Council on cardiovascular radiology and intervention: the American Academy of Neurology affirms the value of this guideline. *Stroke* 2006;37:577–617.
- Rothwell PM, Algra A, Amarenco P. Medical treatment in acute and long-term secondary prevention after transient ischaemic attack and ischaemic stroke. *Lancet* 2011;377:1681–92.
- Hackam DG, Spence JD. Combining multiple approaches for the secondary prevention of vascular events after stroke: a quantitative modeling study. *Stroke* 2007;38:1881–5.
- Schneeweiss S. A basic study design for expedited safety signal evaluation based on electronic healthcare data. *Pharmacoepidemiol Drug Saf* 2010;19:858–68.
- Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol* 2016;183:758–64.
- Franklin JM, Glynn RJ, Martin D, *et al.* Evaluating the use of nonrandomized real-world data analyses for regulatory decision making. *Clin Pharmacol Ther* 2019;105:867–77.
- Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. *J Clin Epidemiol* 2005;58:323–37.
- Reeves MJ, Arora S, Broderick JP, *et al.* Acute stroke care in the US: results from 4 pilot prototypes of the Paul Coverdell National acute stroke Registry. *Stroke* 2005;36:1232–40.
- Tong X, Wiltz JL, George MG, *et al.* A decade of improvement in Door-to-Needle time among acute ischemic stroke patients, 2008 to 2017. *Circ Cardiovasc Qual Outcomes* 2018;11:e004981.
- Tong X, George MG, Yang Q, *et al.* Predictors of in-hospital death and symptomatic intracranial hemorrhage in patients with acute ischemic stroke treated with thrombolytic therapy: Paul Coverdell acute stroke Registry 2008–2012. *Int J Stroke* 2014;9:728–34.
- Xian Y, Wu J, O'Brien EC, *et al.* Real world effectiveness of warfarin among ischemic stroke patients with atrial fibrillation: observational analysis from patient-centered research into outcomes stroke patients prefer and effectiveness research (prosper) study. *BMJ* 2015;351:h3786.
- O'Brien EC, Greiner MA, Xian Y, *et al.* Clinical effectiveness of statin therapy after ischemic stroke: primary results from the statin therapeutic area of the patient-centered research into outcomes stroke patients prefer and effectiveness research (prosper) study. *Circulation* 2015;132:1404–13.
- Fonarow GC, Liang L, Thomas L, *et al.* Assessment of Home-Time after acute ischemic stroke in Medicare beneficiaries. *Stroke* 2016;47:836–42.
- Song S, Fonarow GC, Olson DM, *et al.* Association of get with the Guidelines-Stroke program participation and clinical outcomes for Medicare beneficiaries with ischemic stroke. *Stroke* 2016;47:1294–302.
- Reeves MJ, Fonarow GC, Xu H, *et al.* Is Risk-Standardized in-hospital stroke mortality an adequate proxy for Risk-Standardized 30-day stroke mortality data? findings from get with the Guidelines-Stroke. *Circ Cardiovasc Qual Outcomes* 2017;10:e003748.
- Kaufman BG, O'Brien EC, Stearns SC, *et al.* The Medicare shared savings program and outcomes for ischemic stroke patients: a retrospective cohort study. *J Gen Intern Med* 2019;34:2740–8.
- Kaufman BG, Shah S, Hellkamp AS, *et al.* Disease burden following non-cardioembolic minor ischemic stroke or high-risk TIA: a GWTC-Stroke study. *J Stroke Cerebrovasc Dis* 2020;29:105399.
- CDC State Heart Disease and Stroke Prevention Programs. Paul Coverdell National acute stroke Registry. Available: [https://www.cdc.gov/dhdsp/programs/stroke\\_registry.htm](https://www.cdc.gov/dhdsp/programs/stroke_registry.htm) [Accessed 3 Dec 2017].
- George MG, Tong X, McGruder H, *et al.* Paul Coverdell National Acute Stroke Registry Surveillance - four states, 2005-2007. *MMWR Surveill Summ* 2009;58:1–23.
- Hammill BG, Hernandez AF, Peterson ED, *et al.* Linking inpatient clinical registry data to Medicare claims data using indirect identifiers. *Am Heart J* 2009;157:995–1000.
- Andrade SE, Harrold LR, Tjia J, *et al.* A systematic review of validated methods for identifying cerebrovascular accident or transient ischemic attack using administrative data. *Pharmacoepidemiol Drug Saf* 2012;21 Suppl 1:100–28.
- McCormick N, Bhole V, Lacaile D, *et al.* Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One* 2015;10:e0135834.
- Austin PC. Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Stat Med* 2009;28:3083–107.
- Gagne JJ, Glynn RJ, Avorn J, *et al.* A combined comorbidity score predicted mortality in elderly patients better than existing scores. *J Clin Epidemiol* 2011;64:749–59.
- Lin KJ, Glynn RJ, Singer DE, *et al.* Out-of-system care and recording of patient characteristics critical for comparative effectiveness research. *Epidemiology* 2018;29:356–63.
- Bonamici S. H.R.34 - 114th Congress (2015–2016): 21st Century Cures Act, 2016. Available: <https://www.congress.gov/bill/114th-congress/house-bill/34> [Accessed 9 Aug 2018].
- Published US Food and Drug Administration. Prescription drug user fee act (PDUFA) – PDUFA VI: fiscal years 2018–2022. Available: <https://www.fda.gov/forindustry/userfees/prescriptiondruguserfee/ucm446608.htm> [Accessed 9 Aug 2021].
- Sherman RE, Anderson SA, Dal Pan GJ, *et al.* Real-World Evidence - What Is It and What Can It Tell Us? *N Engl J Med* 2016;375:2293–7.
- Jarow JP, LaVange L, Woodcock J. Multidimensional Evidence Generation and FDA Regulatory Decision Making: Defining and Using "Real-World" Data. *JAMA* 2017;318:703–4.
- US Food and Drug Administration. *Framework for FDA's Real-World Evidence Program*. Washington, DC: US FDA, 2018. <https://www.fda.gov/media/120060/download>
- Camm AJ, Coleman CI, Larsen TB, *et al.* Understanding the value of real-world evidence: focus on stroke prevention in atrial fibrillation with rivaroxaban. *Thromb Haemost* 2018;118:S45–60.
- Franklin JM, Schneeweiss S. When and how can real world data analyses substitute for randomized controlled trials? *Clin Pharmacol Ther* 2017;102:924–33.
- Schneeweiss S, Paterno E. Conducting real-world evidence studies on the clinical outcomes of diabetes treatments. *Endocr Rev* 2021;12:bnab007.
- Cheen MHH, Tan YZ, Oh LF, *et al.* Prevalence of and factors associated with primary medication non-adherence in chronic disease: a systematic review and meta-analysis. *Int J Clin Pract* 2019;73:e13350.
- Fischer MA, Stedman MR, Lii J, *et al.* Primary medication non-adherence: analysis of 195,930 electronic prescriptions. *J Gen Intern Med* 2010;25:284–90.
- Kiran A, Viscoli CM, Furie KL, *et al.* Adherence to study drug in a stroke prevention trial? *J Stroke Cerebrovasc Dis* 2020;29:105048.
- Bushnell CD, Olson DM, Zhao X, *et al.* Secondary preventive medication persistence and adherence 1 year after stroke. *Neurology* 2011;77:1182–90.
- Glader E-L, Sjölander M, Eriksson M, *et al.* Persistent use of secondary preventive drugs declines rapidly during the first 2 years after stroke. *Stroke* 2010;41:397–401.
- Mouradian MS, Majumdar SR, Senthilselvan A, *et al.* How well are hypertension, hyperlipidemia, diabetes, and smoking managed after a stroke or transient ischemic attack? *Stroke* 2002;33:1656–9.
- Reeves MJ, Fonarow GC, Smith EE, *et al.* Representativeness of the get with the Guidelines-Stroke registry: comparison of patient and hospital characteristics among Medicare beneficiaries hospitalized with ischemic stroke. *Stroke* 2012;43:44–9.