



Development and validation of an algorithm to identify aspirin use from electronic health records and claims data

Valerie Haley PhD¹, Michael A Head MS¹, Sravanthy Myneni MS¹, Shiva K Vojjala MS¹, Emily Durden PhD², Steven Caproni PharmD², Lynda Lisabeth PhD³, Seemant Chaturvedi MD FAHA⁴, Kaitlyn Hopkins BS¹, Vincent Willey PharmD BCACP¹

¹Carelon Research, Wilmington, DE; ²Bayer HealthCare Pharmaceuticals, Whippany, NJ; ³Department of Epidemiology, Univ. of Michigan School of Public Health and Stroke Program, Univ. of Michigan Medical School, Ann Arbor, MI; ⁴University of Maryland School of Medicine, Baltimore, MD

Objectives

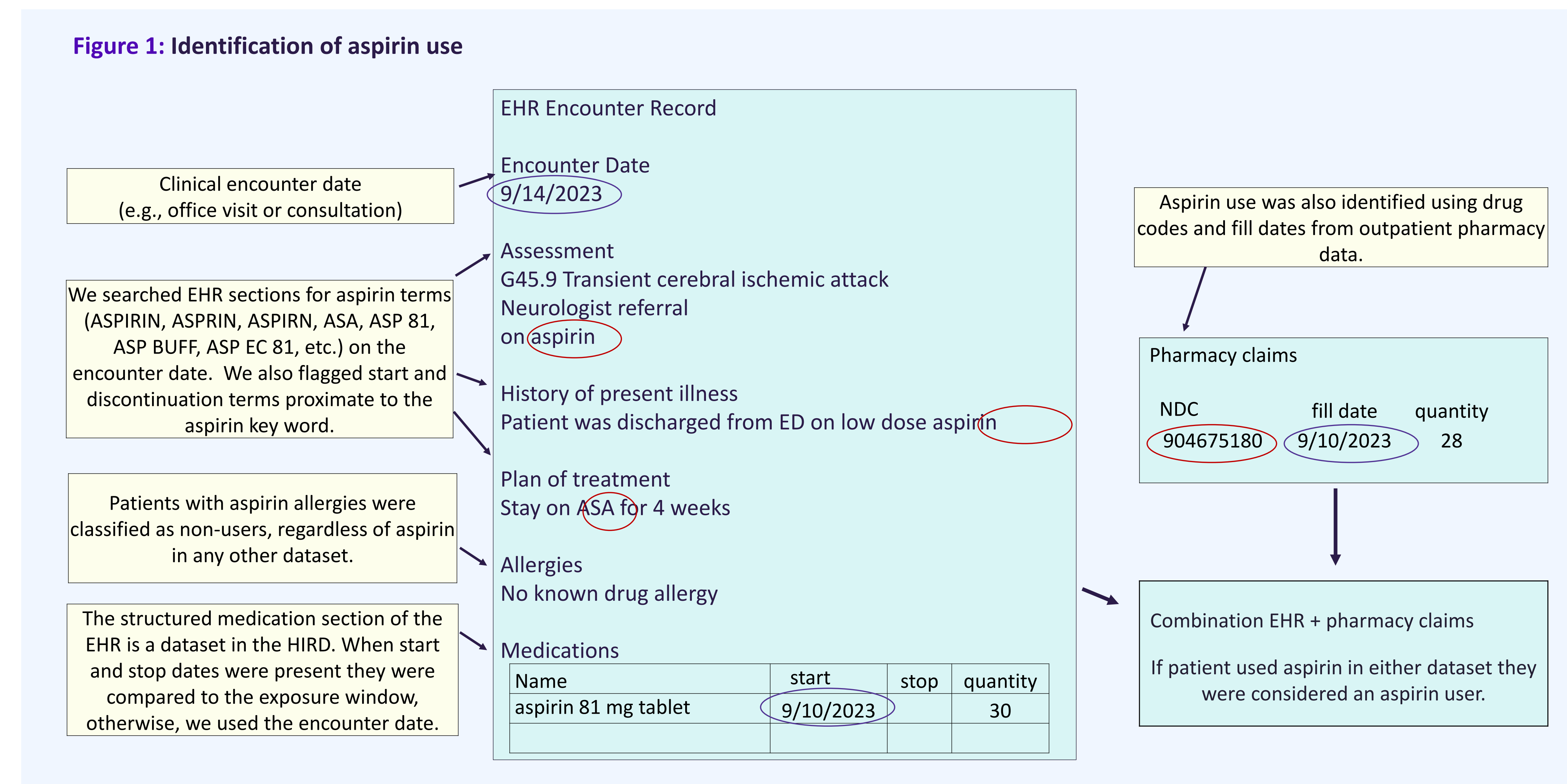
- Because millions of adults take low-dose aspirin for primary and secondary prevention of cardiovascular disease (CVD), including myocardial infarction and ischemic stroke, accurately measuring aspirin exposure is important in CVD research.
- We developed and validated an algorithm to identify aspirin use from electronic health records (EHR) and pharmacy claims data.

Methods

We studied 9,480 U.S. adults in the Healthcare Integrated Research Database (HIRD®) who:

- had a first hospitalization or emergency department visit for transient ischemic attack (TIA) or non-cardioembolic ischemic stroke between 2016 and 2024;
- were continuously enrolled in a Commercial or Medicare Advantage health plan for at least one year prior to TIA/stroke and 90 days post-discharge; and
- had EHR clinical encounter data generated during the aspirin exposure window (TIA/stroke onset to 90 days post-discharge).

We searched for aspirin use within the exposure window using EHR and claims data as described in Figure 1.



Abbreviations: EHR, electronic health record; HIRD, Healthcare Integrated Research Database; ED, emergency department; mg, milligram.

We selected a stratified random sample of 100 patients; data were stratified by EHR algorithm results to increase the number of records with no aspirin identified. We compared the EHR result to manual EHR review of the records to validate that we correctly captured the information in the EHR.

We compared the prevalence of aspirin use based on presence in EHR, claims, or the combination of both. In addition, we described the type of antiplatelet therapy used by the cohort: single antiplatelet therapy (SAPT: aspirin, cilostazol, clopidogrel, prasugrel, or ticagrelor alone, or switch between two P2Y12 inhibitors); dual antiplatelet therapy (DAPT: aspirin + one P2Y12 inhibitor), multiple/other, or none.

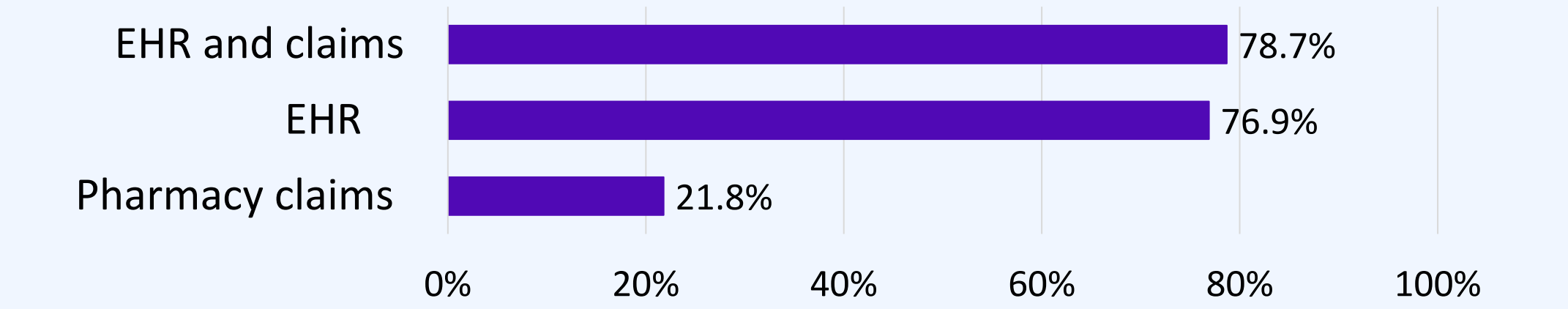
Results

Figure 2: Comparison of EHR algorithm with manual review

Carelon Algorithm		Manual review of EHR	
		Aspirin	No aspirin
Aspirin		64	1
No aspirin		2	33

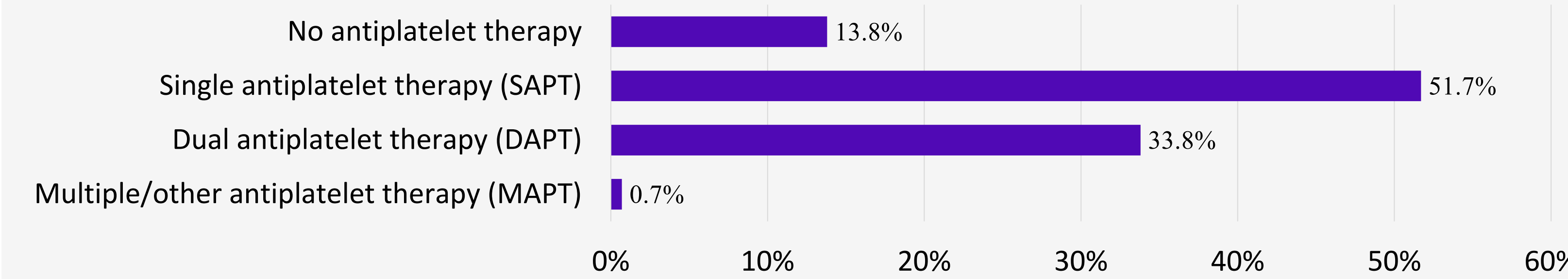
The algorithm had a positive predictive value of 98.5% (95% confidence interval 91.7 to 100%) and negative predictive value of 94.3% (80.8 to 99.3%).

Figure 3: Prevalence of aspirin use by data source



The prevalence of aspirin treatment among TIA/stroke patients within 90 days post-discharge was 78.7% based on EHR and pharmacy claims data combined, compared to 21.8% based on claims alone.

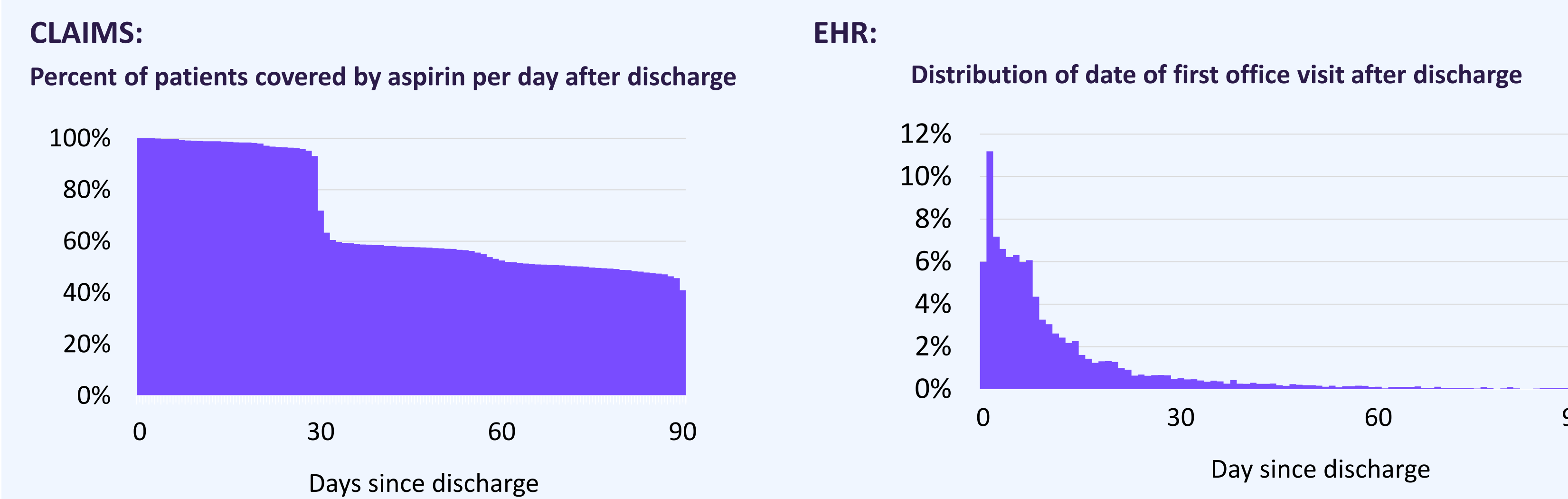
Figure 4: Antiplatelet therapy type



No antiplatelet therapy was found for 13.8% of patients. SAPT was used by 51.7%. DAPT was used by 33.8%.

Type of SAPT	Type of DAPT
Aspirin (85.6%)	Aspirin + Clopidogrel (96.9%)
Clopidogrel (13.8%)	Aspirin + Ticagrelor (2.3%)
Other (0.7%)	Other (0.8%)

Figure 5: Comparison of start/stop date availability in EHR and Claims



In claims data, start (fill date) and stop (from days supply) are always available.

In EHR data, aspirin use is known on encounter dates.

- 75% of individuals met with a physician within 14 days after discharge; 93% had more than one visit.
- In unstructured EHR, 2% had aspirin discontinuation phrases (e.g., discontinue, pause, stop, not using) during the exposure window.
- In structured EHR, 4% had aspirin stop date during the exposure window.

Overall, aspirin start and stop dates were incomplete in the combined data, so we classified individuals as exposed or not exposed.

Conclusions

- Integration of EHR data and administrative claims data is an effective method to identify exposure to aspirin and other over-the-counter medications commonly used for treatment and provides a more complete treatment history, facilitating clinically relevant analyses where treatment is assessed.
- The method can be used to reduce misclassification of SAPT vs. DAPT treatment, compared to studies based only on claims.
- The 57-percentage point increase in the prevalence of aspirin users suggests that use of EHR data is vital to perform clinically relevant analyses in certain therapeutic areas.

Limitations

- The presence of a claim for a filled prescription does not indicate that the medication was consumed or that it was taken as prescribed.
- While aspirin use was accurately classified using documentation found in the EHRs, underestimation of exposure is possible if clinicians associated with the available EHRs were unaware or didn't document aspirin use.
- Specific start and end dates for aspirin use could not be identified.

Disclosures

This work was supported by Elevance Health/Carelon Research through a collaborative partnership with Bayer and was conducted as an adjunct to a related project funded by Bayer. V.H., M.H., S.M., S.V., K.H., and V.W. are employees of Elevance Health and may own shares in the company. E.D. and S.C. are employees of Bayer and may own shares or share options in the company. L.L. receives research funding from the National Institutes of Health and receives consulting fees from Bayer. S.C. is a consultant to Bayer and Novartis.

Poster presented at ISPOR May 2026, Philadelphia PA