International Journal of Population Data Science

Data resource profile: COVid VAXines effects on the aged (COVVAXAGE)

Kaleen N. Hayes[1,*], Daniel A. Harris[1], Andrew R. Zullo[1,2,3], Djeneba Audrey Djibo[4], Renae L. Smith-Ray[5], Michael S. Taitel[5], Tanya G. Singh[5], Cheryl McMahill-Walraven[4], Preeti Chachlani[1], Katherine J. Wen[1,6], Ellen P. McCarthy[7,8], Stefan Gravenstein[1,3,9], Sean McCurdy[10], Kristina E. Baird[10], Daniel Moran[10], Derek Fenson[10], Yalin Deng[1], and Vincent Mor[1,3]

[1]Department of Health Services, Policy & Practice, Brown University School of Public Health, Providence, RI, USA, 02903

[2]Department of Epidemiology, Brown University School of Public Health, Providence, RI, USA, 02903

[3]Center of Innovation in Long-Term Services and Supports, Providence Veterans Affairs Medical Center, Providence, RI, USA, 02903

[4]CVS Health Clinical Trial Services, Safety, Surveillance & Collaboration, Blue Bell, PA, USA, 19422

[5]Walgreens Center for Health & Wellbeing Research, Walgreen Company, Deerfield, IL, USA, 60015

[6]Department of Medicine, Health, and Society, Vanderbilt University, Nashville, TN, USA, 37240

[7]Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA, 02131

[8]Division of Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA, 02215

[9]Division of Geriatrics, Warren Alpert Medical School of Brown University, Providence, Rhode Island, USA, 02903

[10]Acumen, LLC, Burlingame, CA, USA, 94010

Abstract

Background
To improve the assessment of COVID-19 vaccine use, safety, and effectiveness in older adults and persons with complex multimorbidity, the COVid VAXines Effects on the Aged (COVVAXAGE) database was established by linking CVS Health and Walgreens pharmacy customers to Medicare claims.

Methods
We deterministically linked CVS Health and Walgreens customers who had a pharmacy dispensation/encounter paid for by Medicare to Medicare enrollment and claims records. Linked data include U.S. Medicare claims, Medicare enrollment files, and community pharmacy records. The data currently span 01/01/2016 to 08/31/2022. "Research-ready" files were created, with weekly indicators for vaccinations, censoring, death, enrollment, demographics, and comorbidities. Data are updated quarterly.

Results
As of November 2022, records for 27,086,723 CVS Health and 23,510,025 Walgreens unique customer IDs were identified for potential linkage. Approximately 91% of customers were matched to a Medicare beneficiary ID (95% for those aged 65 years or older). In the final linked cohort, there were 38,250,873 unique beneficiaries representing ~60% of the Medicare population. Among those alive and enrolled in Medicare as of January 1, 2020 (n = 33,721,568; average age = 73 years, 74% White, 51% Medicare Fee-for-Service, and 11% dual-eligible for Medicaid), the average follow-up time was 130 weeks. The cohort contains 16,021,055 beneficiaries with evidence a first COVID-19 vaccine dose. Data are stored on the secure Medicare & Medicaid Resource Information Center Health & Aging Data Enclave.

Data access
Investigators with funded or in-progress funding applications to the National Institute on Aging who are interested in learning more about the database should contact Dr Vincent Mor [Vincent_mor@brown.edu] and Dr Kaleen Hayes [kaley_hayes@brown.edu]. A data dictionary can be provided under reasonable request.

Conclusions
The COVVAXAGE cohort is a large and diverse cohort that can be used for the ongoing evaluation of COVID-19 vaccine use and other research questions relevant to the Medicare population.

Keywords
mRNA vaccines; Aged; COVID-19*/prevention & control; 2019-nCoV Vaccine; mRNA-1273; BNT162 Vaccine

*Corresponding Author:
Email Address: kaley_hayes@brown.edu (Kaleen N. Hayes)

Key features

- Vaccines against COVID-19 are effective and safe; however, their uptake, effectiveness, and safety are understudied in older adults and those with multimorbidity and dementia. Rigorous observational research using large and representative cohorts is needed to inform policy, clinical guidelines, and future vaccine development for these populations.

- The COVid VAXines Effects on the Aged (COVVAXAGE) database was created to characterize the use, safety, and effectiveness of COVID-19 vaccines, including booster doses, for U.S. older adults and individuals with complex multimorbidity.

- CVS Health and Walgreens customer records were deterministically linked to Medicare data using encrypted identifiers. Research-ready files with indicators for vaccinations, sociodemographics, and chronic conditions were prepared that can be linked to other Medicare claims files (e.g., Medicare Part D for prescription claims).

- COVVAXAGE contains longitudinal data on vaccinations, diagnoses, and healthcare use for 38,250,873 Medicare beneficiaries. These data are unique in that they are more representative of Medicare beneficiaries in the U.S. and larger than existing health administrative data sources for observational research.

- The cohort is large and diverse in ages (e.g., >2.6 million individuals aged ≥85 years), sex (>56% female), and races and ethnicities (e.g., >25% non-White), with an average follow-up time of 130 weeks for each beneficiary.

- Investigators funded by the National Institute on Aging (NIA) who establish relevant data use agreements can access these data. Investigators with funded or who have in-progress funding applications to the NIA who are interested in learning more about the database should reach out to Drs Vincent Mor [Vincent_mor@brown.edu] and Kaleen Hayes [kaley_hayes@brown.edu].

Introduction

Since COVID-19 vaccines came to market in December 2020, over 58 million older adults have received at least one dose of a COVID-19 vaccine [1]. Evidence from clinical trials demonstrated that both mRNA vaccines (BNT162b2 or mRNA-1273) are highly effective and associated with few serious safety concerns; however, older adults and those with complex multimorbidity were generally excluded from these studies. As physiological changes associated with aging and frailty [2] can alter the response to vaccines [3, 4], rigorous observational research is needed to inform policy, clinical guidelines, and future vaccine development for these populations. Although some observational research has explored vaccine safety and effectiveness in real-world contexts [5, 6], many have a lack of clinical detail; limitations in generalizability, particularly to multimorbid older adults;

and limited long-term follow-up. Moreover, despite the large sample sizes of some prior studies, they may not have been sufficiently powered to detect extremely rare, yet important, outcomes and potential adverse effects (e.g., Guillain-Barre syndrome) [5]. Older adults who were understudied in clinical trials are among those with the highest risk of adverse events from COVID-19 [7–9].

The COVid VAXines Effects on the Aged (COVVAXAGE) database was created to characterize the use, safety, and effectiveness of COVID-19 vaccines for older adults and to meet evidence needs for public health decision making. This project was funded as two administrative supplements to a large cooperative agreement (U54AG063546) to examine how COVID-19 has affected the older adult population, particularly those living with Alzheimer's disease and related dementias (ADRD). We established a public-private partnership between CVS Health, Walgreens, and academic institutions funded by the National Institute on Aging (NIA) to link customer data from CVS Health and Walgreens to Medicare enrollment and claims information. This process created a national, longitudinal, population-based cohort and near-real-time database to study COVID-19 vaccine related questions in older adults, especially those with frailty, multimorbidity, and ADRD. The aims of the COVVAXAGE database were to enable studies that:

1) Determine which Medicare beneficiaries, including those with dementia and/or are racial/ethnic minorities, are less likely to be completely vaccinated for COVID-19.

2) Examine and compare rates of suspected adverse events possibly attributed to COVID-19 vaccines, overall and among subgroups of Medicare beneficiaries.

3) Analyze rates of breakthrough COVID-19 infections and related diagnoses among Medicare beneficiaries who received COVID-19 vaccinations.

For example, our research team has used the COVVAXAGE database to directly compare the safety and effectiveness of mRNA vaccines across frailty levels [10] and to understand racial and ethnic disparities in the receipt of booster COVID-19 vaccines [11]. The specific aim of this paper is to describe the data linkage process and characteristics of the final linked cohort for future investigators to use.

Methods

Population and data sources

The population of interest for the cohort was CVS Health or Walgreens pharmacy customers who were Medicare beneficiaries. Medicare is the U.S. federal health insurance program that provides coverage to people aged 65 years or older, people younger than age 65 who have certain disabilities, and people with end-stage renal disease [12]. Medicare is the primary source of healthcare insurance coverage for older adults in the U.S., and administrative data from the program are commonly used for observational research [13]. Additional details on the Medicare program that are relevant for epidemiologic and health services research have been published elsewhere [13]. This population was targeted

for large, population-based COVID-19 vaccine monitoring because in 2021 CVS Health and Walgreens administered approximately 59 million and 55 million COVID vaccines, respectively [14, 15]. CVS Health and Walgreens pharmacies (over 18,000 stores total) are located in every US state and are accessed by older adult populations that are representative of the US Medicare population in demographics (e.g., age, sex, rurality, socioeconomic status) and clinical characteristics [12, 16, 17]. Therefore, data from customers from these pharmacies can be considered population-based.

For the initial linkage, CVS Health and Walgreens provided data capturing customer demographic information (e.g., age, sex, store location), pharmacy records (including payor), and vaccination records (including reports submitted to the Centers for Disease Control and Prevention [CDC] regarding vaccination information and customer demographics) for customers filling a prescription or receiving a vaccine between January 1, 2019 to August 31, 2022 (CVS Health) and January 1, 2018 to June 30, 2021 (Walgreens). Both CVS Health and Walgreens sent data pertaining to any customer who filled a prescription or received an immunization that was paid for by Medicare. In addition, CVS sent records for any customer aged 65 years or older at the time of the prescription or immunization, regardless of whether they had a prescription paid for by Medicare, to capture persons who potentially enrolled later in follow-up. For a description of all Medicare claims files used once CVS Health and Walgreens customers were linked, see the Supplement.

CVS Health and Walgreens each send weekly vaccination record updates to Acumen LLC (henceforth "Acumen"). Acumen acquires updated Medicare files (e.g., new Part A & B claims) as they become available through the Medicare & Medicaid Resource Information Center (MedRIC), usually monthly depending on the file. Acumen re-performs the match approximately every two months to account for these data updates from CMS and pharmacy data sources to investigate whether they improve the match rate; however, match rates have been nearly identical between iterations (proportion of matched CVS Health/Walgreens customer IDs ranging from 94 to 96% for customers aged 65 years or older).

Acumen is a contractor funded by NIA to provide linked CMS data of NIA-supported surveys and studies to researchers (GS10F0133S/NIA NIH HHS/United States). Acumen received data files from each entity and performed the data linkage, beneficiary de-identification, analysis file creation, and variable coding. All Medicare files were provided directly to Acumen by MedRIC, which is a subsidiary of Acumen.

Data linkage methods and privacy protection

Acumen deterministically linked CVS Health and Walgreens customer IDs to Medicare enrollment data to identify preliminary matches. All persons with a pharmacy customer ID (i.e., had a record of a drug dispensing or vaccine receipt through CVS Health or Walgreens that was billed to Medicare) were eligible for matching. All demographic variables had a >99% completeness (populated) rate in the vaccine and drug history files in the CVS Health and Walgreens customer data, except ZIP code in the Walgreens drug history data (98% completeness rate). Prior to the matching process, we applied the following steps to clean data related to first and last names in the CVS Health and Walgreens customer data: 1) remove all non-alpha characters (e.g., hyphens, spaces); 2) remove suffixes such as "JR.," "SR," "I-IV"; 3) removed variations of "deceased" (e.g., "dead," "decease"); 4) convert text to uppercase.

A match was considered preliminary if a customer ID and Medicare beneficiary ID had a matching last name, date of birth, and sex. Medicare demographic information across all years was used for the matching process. Each preliminary match was then classified as *strong* if at least one of the following also matched: first name, state, or zip code. Otherwise, the match was classified as weak and then excluded from the datasets. If a pharmacy customer ID had a strong match as defined above with more than one Medicare beneficiary ID, the Medicare ID with the most matched variables (first name, state, zip code) was selected as the matching ID. If multiple beneficiary IDs were an equally strong match, the pharmacy customer ID was dropped. If multiple pharmacy customer IDs (either from the same pharmacy data source or separately from CVS Health and Walgreens) matched to the same Medicare ID, then the pharmacy IDs were collapsed into one person. Therefore, the final study population only comprised customer IDs with one unambiguous, strong match to a Medicare beneficiary ID. For each match, Acumen created a unique encrypted beneficiary ID that could be used to link across all Medicare and pharmacy datasets. The data in the present study are for a match performed on November 18, 2022, with data from all sources up to and including August 2022. The date of an individual's entry to the cohort was the first date of Medicare enrollment within the study period (January 1, 2016 to most recent data available), and individuals were permitted to "enter" the open cohort at any time. Follow-up for each beneficiary continues until the individual dies, disenrolls from Medicare (the individual can re-enter the cohort upon re-enrollment), has a missing or invalid address, or the end of the study data (presently August 31, 2022). A diagram of the data linkage process is presented in Figure 1.

CMS, CVS Health, and Walgreens data are disclosed to the research team under the strict terms of respective Data Use Agreements, which require stringent privacy protections. Several efforts to minimize the risk of breach of patient confidentiality were implemented. Access to names is strictly limited to pre-approved members of the Acumen team who are directly involved in the initial data cleaning and linkage activities. No names or social security numbers are included in the analytic data. While a synthetic beneficiary ID is used, exact dates of service and provider identification are available in the files allowing for detailed temporal and geographic analysis; thus, only the beneficiary ID is truly de-identified. All analysts are required to undergo privacy training and sign a Data User and Analyst Access Policy that includes acknowledgement of detailed privacy considerations, including the CMS and NIA cell suppression policies. Acumen created a secure data management and analysis platform called the Health and Aging Data (HaAD) Enclave through which researchers access and analyze data. The HaAD Enclave is a secure computing desktop (a digital interface) for studies on aging, healthcare, and health outcomes. The desktop provides access to sensitive CMS datasets linked to NIA-sponsored

Figure 1: COVid VAXines effects on the aged (COVVAXAGE) data linkage illustration
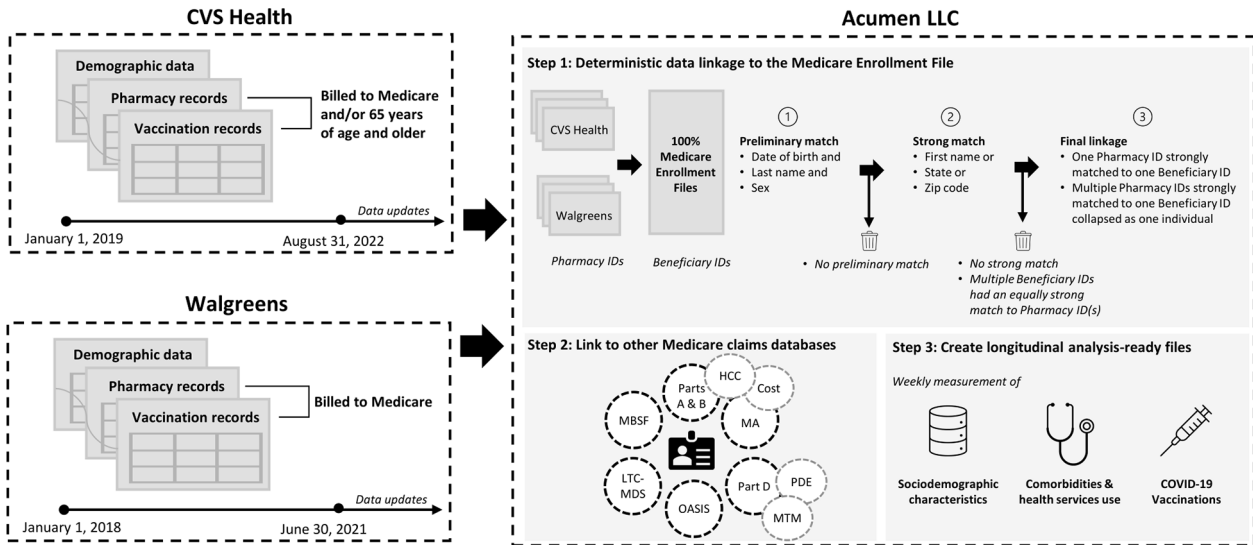
HCC – Hierarchical Condition Category; MA – Medicare Advantage (Encounter Claims); LTC-MDS – Long-term Care Minimum Dataset; MBSF – Medicare Beneficiary Summary File; MTM – Medication Therapy Management; OASIS - Outcome and Assessment Information Set; PDE – Part D Event.

study or survey datasets. The HaAD Enclave complies with over 450 Federal Information Security Modernization Act of 2014 (FISMA) safeguards and has a CMS authority to operate (ATO) in place. Prior to each instance of accessing the analytic data through the HaAD Enclave, an analyst must acknowledge and agree to federal privacy and security requirements. All output, including analytic results, is reviewed to ensure compliance with minimum cell size according to confidentiality requirements set forth by the NIA ($n \leq 25$).

Variables and final dataset creation

Using data available from CMS and pharmacy files, Acumen created variables representing demographic and clinical information for each person in the cohort. COVID-19 vaccination events were captured using CPT billing codes for vaccine administration that are specific to each vaccine product and dose number (Supplementary Table 1). We defined the date of vaccination as the date of billing for the vaccine administration. We supplemented these vaccine administration data with CVS Health and Walgreens pharmacy records that captured vaccinations that were not billed to Medicare. These data therefore can measure precise products, doses, and timing of COVID-19 vaccines for the cohort throughout follow-up.

Other variables include indicators for censoring, death, enrollment status, demographics (e.g., 5-digit zip code for home address), comorbidities defined using hierarchical condition category (HCC) coding [18] and validated algorithms (e.g., the Claims-Based deficit-accumulation Frailty Index [2], the Bynum-Standard ADRD algorithm [19], comorbidity indices [20–22]) that leverage International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM), Common Procedural Terminology (CPT), and Healthcare Common Procedure Coding System (HCPCS) codes. The final datasets are provided in "person-week" format, with one row describing time-varying events (vaccinations, comorbidities)

for each member of the cohort in every calendar week of the study. The original CMS and pharmacy files (e.g., Medicare Part B claims, CVS Health prescription drug fill records) can be accessed and linked through the encrypted beneficiary ID created by Acumen. These files can be accessed to create additional variables, like hospital discharge diagnoses or location of pharmacies visited.

In general, using these data, researchers can identify when and where a vaccination event occurred, characterize beneficiaries' demographic and clinical information, and measure clinical events as they are captured through Medicare claims and associated diagnoses in the days and months before and after vaccination. The large cohort thereby facilitates the study of important study important questions regarding health events and outcomes possibly associated with COVID-19 vaccination over long periods, including the study of rare events (e.g., Guillain Barre syndrome), in a way that minimizes false detection rate due to the routinely collected nature of Medicare claims.

Results

Data linkage performance and final cohort size

Records for 27,086,723 CVS Health and 23,510,025 Walgreens unique customer IDs were sent to Acumen for linkage. Overall, 88.7% (n = 24,016,176) and 95.8% (n = 22,519,900) of respective CVS Health and Walgreens customers were matched to a Medicare beneficiary ID and included in the final linked cohort. When restricting to customers aged 65 years or older, 94.3% (n = 20,906,184) of CVS Health and 96.0% (n = 18,311,833) of Walgreens customer IDs were matched, respectively. In total, there were 38,250,873 unique beneficiaries in the final linked cohort (around 60% of the Medicare population [12]; and 28,251,114 of these beneficiaries were aged 65 years or older, representing around

Table 1: Sociodemographic and clinical characteristics of CVS Health and Walgreens customers linked to Medicare claims in the COVid VAXines Effects on the Aged (COVVAXAGE) database ($N = 38,250,873$)

Characteristics	Overall, at first week of Medicare enrollment (cohort entry) N(%)[a]	Medicare-enrolled (FFS, MA or Part A only) on Jan 1, 2020, N(%)[b]	Continuously enrolled in FFS for 12 months prior to Jan 1, 2020, N(%)[c]
Unique Beneficiaries alive and with a valid address at first observation, N (%)	38,116,281 (100.0%)[a]	33,721,568 (100.0%)	15,796,838 (100.0%)
Customer ID Source			
CVS Health	16,472,713 (43.22%)	14,169,265 (42.02%)	6,960,205 (44.06%)
Walgreens	15,398,313 (40.40%)	13,743,620 (40.76%)	6,077,249 (38.47%)
Both CVS Health and Walgreens	6,245,255 (16.38%)	5,808,683 (17.23%)	2,759,384 (17.47%)
Age			
Age in years, mean (SD)	70.47 (9.53)	72.89 (9.70)	74.00 (10.02)
<65 years	9,865,167 (25.88%)	3,093,191 (9.17%)	1,254,970 (7.94%)
65 to <85 years	25,563,987 (67.07%)	27,170,424 (80.57%)	12,567,075 (79.55%)
85 years or older	2,687,127 (7.05%)	3,457,953 (10.25%)	1,974,793 (12.50%)
Sex			
Female	21,620,577 (56.72%)	19,233,556 (57.04%)	9,124,115 (57.76%)
Race and Ethnicity			
White	28,266,584 (74.16%)	25,026,691 (74.22%)	12,756,507 (80.75%)
Black	4,060,716 (10.65%)	3,571,648 (10.59%)	1,264,157 (8.00%)
Asian	986,871 (2.59%)	813,744 (2.41%)	314,106 (1.99%)
Hispanic	1,037,050 (2.72%)	890,492 (2.64%)	250,922 (1.59%)
Native American	81,946 (0.21%)	74,987 (0.22%)	42,105 (0.27%)
Other	386,614 (1.01%)	286,621 (0.85%)	132,499 (0.84%)
Missing	946,988 (2.48%)	772,749 (2.29%)	358,831 (2.27%)
US Geographic Region			
Northeast	7,568,995 (19.86%)	6,600,384 (19.57%)	3,206,020 (20.30%)
Midwest	8,151,583 (21.39%)	7,193,431 (21.33%)	3,393,492 (21.48%)
South	14,850,441 (38.96%)	13,316,708 (39.49%)	6,319,981 (40.01%)
West	7,182,637 (18.84%)	6,280,357 (18.62%)	2,857,433 (18.09%)
Other	362,625 (0.95%)	330,688 (0.98%)	19,912 (0.13%)
Medicare Enrollment			
Fee-for-Service	20,748,530 (54.43%)	17,310,134 (51.33%)	15,796,838 (100.0%)
Medicare Advantage	12,612,717 (33.09%)	14,295,599 (42.39%)	0 (0.00%)
Full Dual Medicaid enrollment eligible[d] [23]			
Full Dual Medicaid Eligible	4,200,266 (11.02%)	3,676,886 (10.90%)	1,773,067 (11.22%)
History of Comorbidities[e]			
Alzheimer's Disease and Related Dementias[f] [19]			1,148,252 (7.27%)
Acute Myocardial Infarction			312,807 (1.98%)
Cancer			2,254,141 (14.27%)
Chronic Obstructive Pulmonary Disease			1,978,697 (12.53%)
Diabetes			4,137,977 (26.19%)
Heart Failure			2,019,764 (12.79%)
Mental Health Condition[g]			2,120,501 (13.42%)
Chronic Renal Disease			966,920 (6.12%)
Ischemic Stroke			502,591 (3.18%)
Claims-Based Frailty Index (CFI)[e,h] [2]			
Non-frail (CFI < 0.15)			8,090,013 (51.21%)

Continued

Table 1: Continued

Characteristics	Overall, at first week of Medicare enrollment (cohort entry) N(%)[a]	Medicare-enrolled (FFS, MA or Part A only) on Jan 1, 2020, N(%)[b]	Continuously enrolled in FFS for 12 months prior to Jan 1, 2020, N(%)[c]
Pre-Frail (CFI $\geq$ 0.15 and $<$ 0.25)			6,466,047 (40.93%)
Frail (CFI $\geq$ 0.25)			1,240,778 (7.85%)
Combined Comorbidity Index[e,i] [20, 22]			
Combined Comorbidity Index, mean (SD)			1.77 (2.70)

BID: beneficiary ID; FFS – fee-for-service; MA – Medicare Advantage; SD = standard deviation.

[a]Column 1 represents those alive and enrolled in Medicare at the first week of observation within our data (Jan 01, 2016 to Aug 31, 2022). All characteristics were measured as of beneficiary's first week of Medicare enrollment. Individuals were permitted to enter the cohort at any date. N = 134,592 (0.3%) beneficiaries died or had an invalid address before their first observable week of Medicare enrollment.

[b]Column 2 represents those with any Medicare enrollment on January 1, 2020 (individuals most likely to be included in studies of COVID-19 vaccination and infection).

[c]Column 3 represents those with 12 months or more of Medicare FFS enrollment as of January 1, 2020 (individuals most likely to be included in studies that require measurement of a person's clinical information through Medicare claims, which are not currently available for those in the Medicare Advantage program).

[d]Eligible for full state Medicaid benefits in addition to Medicare coverage, versus partial dual Medicaid/Medicare eligibility (qualifies for a Medicare Savings Program) or no dual eligibility [23].

[e]Comorbidities and frailty scores are only reported for beneficiaries with 12-months of continuous Medicare FFS enrollment due to incomplete claims for MA beneficiaries during the study period.

[f]Measured using the 1-year Bynum Standard Algorithm [19].

[g]Includes psychosis, schizophrenia, bipolar disorder, major depressive disorder, substance use disorders.

[h]Ranges from 0.0 to 1.0, with higher scores indicating greater frailty [2].

[i]Ranges from −2 to 26, with higher scores indicating a greater comorbidity burden [20, 22].

70% of the Medicare population in this age group [12]); of note, some CVS Health and Walgreens customer IDs were matched to the same Medicare beneficiary ID. Detailed information on results from the matching process is available in the Supplementary Tables 2.1–2.3. Characteristics of all matched individuals and subsets restricted by Medicare enrollment status as of January 1, 2020 (to illustrate the characteristics of beneficiaries most likely to be included in studies related to COVID-19 and vaccination) are presented in Table 1.

Among all persons with an unmatched CVS customer ID, 89.9% were unmatched because no potential Medicare beneficiary candidates were identified, and 10.1% were unmatched due to multiple unresolved strong Medicare ID matches. Among persons with an unmatched Walgreens ID, 82.7% were unmatched due to a lack of potential Medicare beneficiary matches, and 17.3% were unmatched due to unresolved strong matches. Comparison of demographic characteristics of individuals with matched versus unmatched CVS Health and Walgreens customer IDs are presented in Supplementary Table 3. Characteristics between matched and unmatched IDs were compared using absolute standardized mean differences (SMDs), with an SMD greater than 0.10 representing a potentially meaningful difference between groups [24]. Overall, persons with matched IDs were slightly older and more likely to be female compared to those who were unmatched. We attempted to also compare the distributions of race and ethnicity between individuals with matched and unmatched customer IDs but were unable due to a high proportion of missing race/ethnicity data in

pharmacy data files for customers without at least one vaccine record.

Cohort characteristics and follow-up

Table 1 presents demographic and clinical characteristics of the cohort overall and by type of Medicare enrolment as of January 1, 2020 (just prior to the onset of the COVID-19 pandemic). In the overall cohort, the average age was 70.5 (SD 9.5) years, with $n = 2,687,127$ being aged 85 years or older. Fifty-seven percent of the cohort was female, 74% were White, and 11% were Black. Among those alive, enrolled in Medicare, and with a valid address as of January 1, 2020 ($n = 33,721,568$), the average follow-up time was 130 weeks (SD 23, median 137 weeks), with data available maximally until August 2022; of these, 1,114,394 (3.30%) died within 1 year (to December 31, 2020).

We identified 16,021,055 beneficiaries with a first dose of a COVID-19 vaccine, as evidenced through a CVS Health or Walgreens record for a vaccination or through a Medicare claim containing a CPT code for a COVID-19 vaccine (Supplementary Table 1). Of these, overall, 65% had a CMS claim only, 7% had only a pharmacy record, and 28% had the vaccine recorded in both data sources. The proportion of vaccinated beneficiaries who were identified using Medicare claims vs. pharmacy claims changed over time, from a low of 4.8% with a pharmacy record only in February 2021 to a high of 17.6% in August 2021, Figure 2. Excluding those with 2 recorded first doses of the COVID-19 vaccine (n=6,241),

Figure 2: Percentages of Medicare beneficiaries in the COVid VAXines Effects on the Aged (COVVAXAGE) database with a first dose of a COVID-19 vaccination* identified via a Medicare claim only, a pharmacy record only, or both, over time
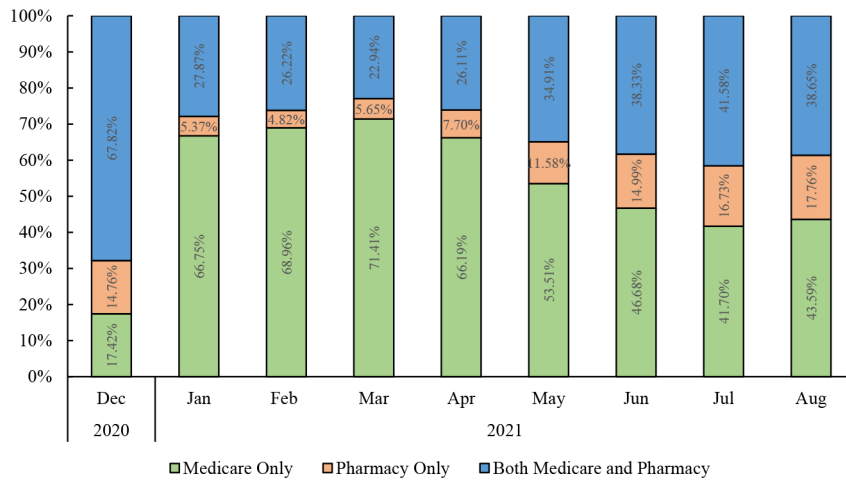
*Exact counts of beneficiaries and proportions over time are available in Supplementary Table 4.

52.9% received BNT162b2, 43.1% received mRNA-1273, and 4.1% received JNJ-78436735.

Discussion

We linked customers from CVS Health and Walgreens to Medicare data to create a near-real-time database of clinical data for over 38 million Medicare beneficiaries in the U.S. (approximately 60% of the U.S. Medicare population, whereas many samples for research purposes are restricted to 20% random samples of beneficiaries) [12]. The COVVAXAGE database is being used to examine the uptake, safety, and effectiveness of COVID-19 vaccinations among a large and generalizable population of older adults in the U.S. The deterministic linkage rate was high, with around 95% of customer IDs from each pharmacy data source being matched to a Medicare beneficiary for those aged 65 years or older. Among these customers, we were able to identify 16,021,055 beneficiaries with a recorded first dose of a COVID-19 vaccination.

The clinical and vaccine-related information contained within this database is important for the ongoing evaluation of COVID-19 vaccine use and effects among older adults in the U.S. Similar population-based cohorts with aligned aims have been created for other countries, including Canada [25] and the U.K. [26]. First, the large sample size and longitudinal capture of health services use increases the statistical precision needed to characterize understudied subgroups (e.g., persons 90+ years, women, persons with multimorbidity, and racial and ethnic subgroups) and investigate the incidence of rare clinical events possibly associated with COVID-19 vaccines (e.g., Guillain-Barre and myocarditis) [6]. The clinical trial for BNT162b2 included just under 16,000 adults 55 years and older (approximately 8,000 per experimental group) [27]. Although recent observational studies have used large real-world data sources to increase the representation of older age groups, other challenges to their generalizability remain, such as the use of a predominantly male sample [5, 28]. The current linked, near-real-time cohort of older adults helps to overcome many of these challenges and meet the demands

of public health and clinical decision making by including populations typically excluded from clinical trials, particularly persons with dementia. Additionally, as this cohort contains time-varying clinical and sociodemographic information, target trial emulation and advanced analytic techniques to mitigate bias can be readily implemented. Finally, the use of pharmacy records in conjunction with Medicare billing data (the traditional source of vaccine information in many large healthcare administrative databases) allowed us to identify over one million additional beneficiaries with a first dose of a COVID-19 vaccine.

Linking data from CVS Health and Walgreens, the two largest private pharmacy providers in the United States, to Medicare claims also permits the implementation of novel research investigations and knowledge translation beyond the use and effects of COVID-19 vaccines. For example, the database can facilitate investigations into the role and impact of community pharmacies on COVID-19 vaccine access and uptake. Additionally, the formal sharing of resources across public and private institutions directly facilitates knowledge translation and more rapid insights for knowledge-users. Indeed, as community pharmacies are a unique and rapidly expanding access point to health care, including the administration of COVID-19 vaccines and testing, connections to organizational leadership within large pharmacy companies has the potential to improve knowledge dissemination. Although public-private partnerships have notable considerations and limitations [29, 30], the CDC supports their formation, and prior public-private data linkages during the 2009 H1N1 pandemic held notable value [31].

Limitations

Although the current database addresses several existing gaps in the infrastructure to evaluate COVID-19 vaccine safety and effectiveness, there are limitations. Most importantly, we are not able to capture vaccinations received through mass vaccinations sites or other locations where vaccine receipt was not captured by CMS health administrative data sources (e.g., a billing to Medicare). Nevertheless, we identified over

16 million beneficiaries with at least one dose of a COVID-19 vaccine, representing a large population that is well suited to examine vaccine safety and effectiveness, particularly for booster vaccinations in the era of emerging new variants. Secondly, though our overall match rate was high, some CVS Health and Walgreens customers we were not matched to a Medicare beneficiary ID. We assume these non-linkages occur at random given that the characteristics of the final, matched cohort are similar to that of the general Medicare population (e.g., 74% vs. 76% White, 11% vs. 11% Black, 56% vs. 55% female, 33% vs 36% Medicare Advantage beneficiaries) [12]. However, if certain groups were more likely to be excluded due to not being matched (e.g., those more likely to have SARS-CoV-2 infection), selection bias and reduced generalizability are potential limitations of the database [32]. Further, the database is unable to capture SARS-CoV-2 infections that are not measured as COVID-19 diagnoses in administrative data. Therefore, the database is likely to underdiagnose COVID-19 disease, and measured cases are likely more serious than the average SARS-CoV-2 infection. Finally, the current database does not include up-to-date clinical encounter information for Medicare Advantage beneficiaries due to a lag in availability of Medicare Advantage Encounter data. Despite the lag, the nature of COVID-19 vaccination billing and the Medicare enrolment files afford access to demographic, enrolment, and vaccination data that remain useful for many studies of these beneficiaries.

Investigators funded by the NIA who establish data use agreements with each of the data partners who contribute to COVVAXAGE can access these data. Investigators with funded or who have in-progress funding applications to the NIA who are interested in learning more about the database should reach out to Vincent Mor [Vincent_mor@brown.edu] or Kaleen Hayes [kaley_hayes@brown.edu].

Acknowledgements

Funding

Ethics approval

Brown University's Institutional Review Board approved the study and waived the requirement for informed consent (Protocol # 2103002950). We established DUAs with CVS Health, Walgreens, and MedRIC for CMS data files). When updated Medicare files became available (e.g., 2022 Part B Carrier files), DUA amendments were needed to add these files.

Author contributions

KNH: Conceptualization, formal analysis, methodology, project administration, visualization, investigation, supervision, writing - original draft, writing - review & editing.

DAH: Conceptualization, methodology, project administration, writing - review & editing.

ARZ: Conceptualization, methodology, funding acquisition, supervision, writing - review & editing.

DAD: Methodology, data curation, writing - review & editing.

RLS: Methodology, data curation, writing - review & editing, supervision.

MST: Data curation, writing - review & editing, supervision.

TGS: Data curation, writing - review & editing, supervision.

CM: Data curation, writing - review & editing, supervision.

PC: Conceptualization, methodology, formal analysis, writing – reviewing & editing.

KJW: Conceptualization, methodology, writing - review & editing.

EPM: Conceptualization, methodology, funding acquisition, writing - review & editing.

SG: Conceptualization, methodology, funding acquisition, writing - review & editing.

SM: Methodology, data curation, writing - review & editing.

KEB: Methodology, data curation, writing - review & editing.

DM: Methodology, data curation, writing - review & editing.

DF: Methodology, data curation, writing - review & editing.

VM: Conceptualization, methodology, funding acquisition, supervision, writing - review & editing.

Conflict of interest

Kaleen Hayes has received grant funding paid directly to Brown University for collaborative research from Insight Therapeutics and Sanofi Pasteur for research on complex insulin regimens. Kaleen Hayes also serves as a consultant for the Canadian Agency for Drugs and Technologies in Health. Andrew Zullo has received grant funding paid directly to Brown University by Sanofi for collaborative research on the epidemiology of infections and vaccinations among nursing home residents and infants. Stefan Gravenstein is a recipient of support from the U.S. Department of Veterans Affairs and investigator-initiated grants to Brown University and Lifespan from the National Institute of Allergy and Infectious Diseases (NIAID) to study influenza vaccine and COVID-19 in the nursing home, Pfizer to study pneumococcal vaccines, and from Sanofi Pasteur and Seqirus to study influenza vaccines. Stefan Gravenstein also performs consulting work for Icosavax, Janssen, Merck, Moderna, Novavax, Pfizer, Sanofi, Seqirus, and Vaxart; has served on the speaker's bureaus for Seqirus, Janssen and Sanofi; and was paid to chair data safety monitoring boards from Longeveron and SciClone. Acumen, LLC (Sean McCurdy, Kristina Baird, Daniel Moran, Derek Fenson) has received federal funding from the National Institutes of Health, the Centers for Medicare & Medicaid Services, and the U.S.

Food and Drug Administration to study vaccine safety and to provide vaccination surveillance support. Renae Smith-Ray, Michael Taitel, and Tanya Singh are employees of Walgreens and have received funding from Moderna and Pfizer to study vaccine uptake and effectiveness. Djeneba Audrey Djibo and Cheryl McMahill-Walraven are full-time employees of CVS Health and conduct work for government, public, and private organizations, including pharmaceutical companies, as part of their employment. All other authors have no COI to report.

Publication consent

We have approval to publish this data resource article and share the steps of obtain data access for external investigators.

References

1. COVID-19 Vaccinations in the United States [Internet]. [cited 2023 Feb 8]; Available from: https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-people-onedose-pop-total.

2. Kim DH, Schneeweiss S, Glynn RJ, Lipsitz LA, Rockwood K, Avorn J. Measuring Frailty in Medicare Data: Development and Validation of a Claims-Based Frailty Index. J Gerontol A Biol Sci Med Sci. 2018 Jun 14;73(7):980–7. https://doi.org/10.1093/gerona/glx229

3. Semelka CT, DeWitt ME, Blevins MW, Holbrook BC, Sanders JW, Alexander-Miller MA. Frailty impacts immune responses to Moderna COVID-19 mRNA vaccine in older adults. Immun Ageing. 2023 Jan 17;20(1):4. https://doi.org/10.1186/s12979-023-00327-x

4. Shapiro JR, Sitaras I, Park HS, Aytenfisu TY, Caputo C, Li M, et al. Association of Frailty, Age, and Biological Sex With Severe Acute Respiratory Syndrome Coronavirus 2 Messenger RNA Vaccine-Induced Immunity in Older Adults. Clin Infect Dis. 2022 Aug 15;75(Suppl 1):S61–71. https://doi.org/10.1093/cid/ciac397

5. Dickerman BA, Madenci AL, Gerlovin H, Kurgansky KE, Wise JK, Figueroa Muñiz MJ, et al. Comparative Safety of BNT162b2 and mRNA-1273 Vaccines in a Nationwide Cohort of US Veterans. JAMA Intern Med. 2022 Jul 1;182(7):739. https://doi.org/10.1001/jamainternmed.2022.2109

6. Moll K, Lufkin B, Fingar KR, Ke Zhou C, Tworkoski E, Shi C, et al. Background rates of adverse events of special interest for COVID-19 vaccine safety monitoring in the United States, 2019–2020. Vaccine. 2023 Jan;41(2):333–53. https://doi.org/10.1016/j.vaccine.2022.11.003

7. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Underlying Medical Conditions Associated with Higher Risk for Severe COVID-19: Information for Healthcare Professionals [Internet]. 2023 Feb. Available from: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/underlyingconditions.html#:~:text=Age%20remains%20the%20strongest%20risk,increasing%20markedly%20with%20increasing%20age.

8. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature. 2020 Aug;584(7821):430–6. https://doi.org/10.1038/s41586-020-2521-4

9. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? Aging (Albany NY). 2020 May 29;12(10):9959–81. https://doi.org/10.18632/aging.103344

10. Harris DA, Hayes KN, Zullo AR, Mor V, Chachlani P, Deng Y et al. Comparative risks of potential adverse events following COVID-19 mRNA vaccination among older US adults. JAMA Netw Open. 2023. https://doi.org/10.1001/jamanetworkopen.2023.26852

11. Hayes K, Harris DA, Zullo A, Chachlani P, Wen K, Smith-Ray R, et al. Racial and ethnic disparities in COVID-19 booster vaccination among U.S. older adults differ by geographic region and Medicare enrollment. Frontiers in Public Health. 2023 Aug 10;11:1243958. https://doi.org/10.3389/fpubh.2023

12. Medicare Beneficiary Enrollment Trends and Demographic Characteristics [Internet]. 2022 Mar 2 [cited 2023 Mar 5]; Available from: https://aspe.hhs.gov/sites/default/files/documents/f81aafbba0b331c71c6e8bc66512e25d/medicare-beneficiary-enrollment-ib.pdf.

13. Mues KE, Liede A, Liu J, Wetmore JB, Zaha R, Bradbury BD, 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. 2017;9:267–77. https://doi.org/10.2147/CLEP.S105613

14. 2021 CVS Health COVID-19 Response [Internet]. [cited 2023 Mar 5]; Available from: https://www.cvshealth.com/content/dam/enterprise/cvs-enterprise/pdfs/2021/cvs-health-covid-19-response-report-2021.pdf

15. COVID-19 FAQs [Internet]. 2022 Aug 10 [cited 2023 Mar 5]; Available from: https://news.walgreens.com/our-stories/covid-19-stories/covid-19-faq.htm#:~:text=We%20remain%20committed%20to%20providing,oral%20antivirals%20to%20eligible%20individuals.

16. CVS 2022 Annual Report. Making healthier happen [Internet]. Available from: https://s2.q4cdn.com/447711729/files/doc_financials/2022/ar/CVS-Health-2022-Annual-Report.pdf.

17. Walgreens Boots Alliance, Inc. 2022 Annual Report [Internet]. Available from: https://www.annualreports.com/HostedData/AnnualReports/PDF/NASDAQ_WBA_2022.pdf.

18. Risk Adjustment |CMS [Internet]. [cited 2022 Dec 8]; Available from: https://www.cms.gov/Medicare/Health-Plans/MedicareAdvtgSpecRateStats/Risk-Adjustors.

19. McCarthy EP, Chang CH, Tilton N, Kabeto MU, Langa KM, Bynum JPW. Validation of Claims Algorithms to Identify Alzheimer's Disease and Related Dementias. J Gerontol A Biol Sci Med Sci. 2022 Jun 1;77(6):1261–71. https://doi.org/10.1093/gerona/glab373

20. Gagne JJ, Glynn RJ, Avorn J, Levin R, Schneeweiss S. A combined comorbidity score predicted mortality in elderly patients better than existing scores. J Clin Epidemiol. 2011 Jul;64(7):749–59. https://doi.org/10.1016/j.jclinepi.2010.10.004

21. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. Journal of Chronic Diseases. 1987 Jan;40(5):373–83. https://doi.org/10.1016/0021-9681(87)90171-8

22. Sun JW, Rogers JR, Her Q, Welch EC, Panozzo CA, Toh S, et al. Adaptation and Validation of the Combined Comorbidity Score for ICD-10-CM. Med Care. 2017 Dec;55(12):1046–51. https://doi.org/10.1097/MLR.0000000000000824

23. Dually Eligible Individuals - Categories [Internet]. [cited 2023 Mar 5]; Available from: https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-and-Medicaid-Coordination/Medicare-Medicaid-Coordination-Office/Downloads/MedicareMedicaidEnrolleeCategories.pdf.

24. Austin PC. Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research. Communications in Statistics - Simulation and Computation. 2009 May 14;38(6):1228–34. https://doi.org/10.1080/03610910902859574

25. Nasreen S, Calzavara A, Buchan SA, Thampi N, Johnson C, Wilson SE, et al. Background incidence rates of adverse events of special interest related to COVID-19 vaccines in Ontario, Canada, 2015 to 2020, to inform COVID-19 vaccine safety surveillance. Vaccine. 2022 May 26;40(24):3305–12. https://doi.org/10.1016/j.vaccine.2022.04.065

26. Vasileiou E, Shi T, Kerr S, Robertson C, Joy M, Tsang R, et al. Investigating the uptake, effectiveness and safety of COVID-19 vaccines: protocol for an observational study using linked UK national data. BMJ Open. 2022 Feb 14;12(2):e050062. https://doi.org/10.1136/bmjopen-2021-050062

27. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med. 2020 Dec 31;383(27):2603–15. https://doi.org/10.1056/NEJMoa2034577

28. Dickerman BA, Gerlovin H, Madenci AL, Kurgansky KE, Ferolito BR, Figueroa Muñiz MJ, et al. Comparative Effectiveness of BNT162b2 and mRNA-1273 Vaccines in U.S. Veterans. N Engl J Med. 2022 Jan 13;386(2):105–15. https://doi.org/10.1056/NEJMoa2115463

29. Parker LA, Zaragoza GA, Hernández-Aguado I. Promoting population health with public-private partnerships: Where's the evidence? BMC Public Health. 2019 Dec;19(1):1438. https://doi.org/10.1186/s12889-019-7765-2

30. Hernandez-Aguado I, Zaragoza GA. Support of public-private partnerships in health promotion and conflicts of interest. BMJ Open. 2016 Apr 18;6(4):e009342. https://doi.org/10.1136/bmjopen-2015-009342

31. Salmon D, Yih WK, Lee G, Rosofsky R, Brown J, Vannice K, et al. Success of program linking data sources to monitor H1N1 vaccine safety points to potential for even broader safety surveillance. Health Aff (Millwood). 2012 Nov;31(11):2518–27. https://doi.org/10.1377/hlthaff.2012.0104

32. Griffith GJ, Morris TT, Tudball MJ, Herbert A, Mancano G, Pike L, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. Nat Commun. 2020 Nov 12;11(1):5749. https://doi.org/10.1038/s41467-020-19478-2

33. National Center for Health Statistics. International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) [Internet]. 2023 Jun. Available from: https://www.cdc.gov/nchs/icd/icd-10-cm.htm.

Abbreviations

ADRD:	Alzheimer's disease and related dementias
ATO:	Authority to operate
CDC:	Centers for Disease Control and Prevention
CMS:	Centers for Medicare and Medicaid Services
CPT:	Common Procedural Terminology
COVVAXAGE:	COVid VAXines Effects on the Aged
DUA:	Data use agreement
HaAD:	Health and Aging Data
HCPCS:	Healthcare Common Procedure Coding System
HCC:	Hierarchical Condition Category
ICD-10-CM:	International Classification of Diseases, Tenth Revision, Clinical Modification
ID:	identification
IRF-PAI:	Inpatient Rehab Facility-Patient Assessment Instrument
LTC-MDS:	Long-term Care Minimum Dataset
MA:	Medicare Advantage (Encounter Claims)
MBSF:	Common Medicare Environment/Medicare Beneficiary Summary File
MedRIC:	Medicare & Medicaid Resource Information Center
MTM:	Medication Therapy Management
NIA:	National Institute on Aging
OASIS:	Outcome and Assessment Information Set
PDE:	Medicare Part D Event

Supplementary Appendix

Methods Supplementary

Data sources and Medicare claims files

The following files from the Center for Medicare and Medicaid Services (CMS) were the primary data used after the CVS and Walgreens customer data were matched to a Medicare beneficiary (January 2016 to August 2022): Common Medicare Environment/Medicare Beneficiary Summary File (MBSF; demographics, enrollment, cost/utilization, history of chronic conditions); Medicare Parts A and B claims (e.g., inpatient, outpatient, carrier, long-term care claims, which include vaccination billings that would occur outside of the pharmacy setting), Medicare Advantage Encounter Data (outpatient/carrier encounters for Medicare Advantage enrollees; available through December 2020), and Part D event data (prescription drug claims). Medicare claims data are administrative billing data generated through healthcare encounters (e.g., physician visits, hospitalizations) that can be used to derive longitudinal information about vaccine receipt, diagnoses, medication use, facilities use, healthcare utilization, and other clinical variables. Diagnoses are provided via International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) codes [33]. Procedures (including vaccination events, diagnostic testing, etc.) are captured via Common Procedural Terminology (CPT) codes. Medical devices and supply use are captured using Healthcare Common Procedure Coding System (HCPCS) codes. Other CMS files available for linkage include the Outcome and Assessment Information Set (OASIS; 2016-2020), the Inpatient Rehab Facility-Patient Assessment Instrument (IRF-PAI; 2016-2020), and the Long-term Care Minimum Data Set (MDS (2016–2022).

Supplementary Table 1: Common Procedural Terminology (CPT) Codes used to Identify COVID-19 vaccinations in Medicare claims data

Vaccine type, Dose	Manufacturer	Dose	CPT Code
BNT162b2	Pfizer, BioNTech	Dose 1	0001A
mRNA-1273	Moderna	Dose 1	0011A
JNJ-78436735	Johnson & Johnson	Dose 1	0034A
BNT162b2	Pfizer, BioNTech	Dose 2	0002A
mRNA-1273	Moderna	Dose 2	0012A
BNT162b2	Pfizer, BioNTech	Dose 3	0003A
BNT162b2	Pfizer, BioNTech	Booster Dose	0004A
BNT162b2	Pfizer, BioNTech	Booster Dose	0124A
JNJ-78436735	Johnson & Johnson	Booster Dose	0031A
mRNA-1273	Moderna	Booster Dose	0134A
mRNA-1273	Moderna	Dose 3	0013A
mRNA-1273	Moderna	Booster Dose	0064A

Supplementary Table 2.1: Number of CVS Health IDs by match/no-match across all records (vaccine and drug history)

Type of record	Match type	All CVS Health IDs		CVS Health IDs 65+	
		#	%	#	%
Total		27,086,723	100.0%	22,165,224	100.0%
All Records	Match	24,016,176	88.7%	20,906,184	94.3%
	One CVS Health ID to One BID	20,138,112	74.3%	17,574,851	79.3%
	Many CVS Health IDs to One BID	1,541,987	5.7%	1,340,133	6.0%
	Many BIDs to One CVS Health ID, Resolved	2,336,077	8.6%	1,991,200	9.0%
	No Match	3,070,547	11.3%	1,259,040	5.7%
	Potential Candidates not found	2,761,956	10.2%	1,105,319	5.0%
	Many BIDs to One CVS Health ID, Unresolved	308,591	1.1%	153,721	0.7%

Supplementary Table 2.2: Number of Walgreens IDs by match/no-match across all records (vaccine and drug history)

Type of record	Match type	All Walgreens IDs #	All Walgreens IDs %	Walgreens IDs 65+ #	Walgreens IDs 65+ %
Total		*23,510,025*	*100.0%*	*19,069,353*	*100.0%*
All Records	Match	22,519,900	95.8%	18,311,833	96.0%
	One WAG ID to One BID	20,216,716	86.0%	16,412,633	86.1%
	Many WAG IDs to One BID	1,424,560	6.1%	1,164,851	6.1%
	Many BIDs to One WAG ID, Resolved	878,624	3.7%	734,349	3.9%
	No Match	990,125	4.2%	757,520	4.0%
	Potential Candidates not found	818,512	3.5%	618,630	3.2%
	Many BIDs to One WAG ID, Unresolved	171,613	0.7%	138,890	0.7%

BID – Medicare beneficiary identifier. WAG – Walgreens.

Supplementary Table 2.3: Summary of matched Medicare beneficiary IDs (BIDs) in the COVid VAXines Effects on the Aged (COVVAXAGE) database alive and with a valid address at first observable week of Medicare enrollment*

Data source	BIDs Total n	BIDs Total %	BIDs Age 65+ n	BIDs Age 65+ %
Total	38,116,281	100.00%	28,251,114	100.00%
CVS Health customer only	16,472,713	43.22%	12,573,226	44.51%
Walgreens customer only	15,398,313	40.40%	10,600,585	37.52%
Both CVS Health and Walgreens	6,245,255	16.38%	5,077,303	17.97%

BID – Medicare beneficiary identifier.
*$N = 134,592$ beneficiaries died or had an invalid address before their first observable week of Medicare enrollment.

Supplementary Table 3: Characteristics of Matched versus Unmatched IDs (age 65+ only) from CVS Health and Walgreens Databases

	Matched to medicare Beneficiary	Not matched	aSMD
CVS Health			
Average Age in years, mean (SD)	73.8 (7.2)	73.0 (8.1)	0.09
% Female	57%	39%	0.37
Walgreens			
Average Age in years, mean (SD)	74.0 (7.1)	74.0 (7.6)	0.12
% Female	59%	39%	0.41

aSMD-absolute standardized mean difference. An aSMD > 0.10 (10%) is considered to be a large enough difference to be potentially important [24].

Supplementary Table 4: Counts and percentages of beneficiaries with a first dose of a COVID-19 vaccination identified via a medicare claim only, a pharmacy record only, or both, over time in the COVid VAXines effects on the aged (COVVAXAGE) database

year	Month	Medicare only, n	Pharmacy only, n	Both medicare & pharmacy, n	Total N	% Medicare*	% Pharmacy*	% Both*
2020	Dec	47,026	39,832	183,081	269,939	17.42%	14.76%	67.82%
2021	Jan	3,427,518	275,920	1,431,237	5,134,675	66.75%	5.37%	27.87%
	Feb	3,349,877	234,168	1,273,435	4,857,480	68.96%	4.82%	26.22%
	Mar	2,006,305	158,765	644,657	2,809,727	71.41%	5.65%	22.94%
	Apr	768,656	89,367	303,228	1,161,251	66.19%	7.70%	26.11%
	May	363,536	78,693	237,167	679,396	53.51%	11.58%	34.91%
	Jun	127,969	41,096	105,090	274,155	46.68%	14.99%	38.33%
	Jul	139,006	55,760	138,600	333,366	41.70%	16.73%	41.58%
	Aug	218,393	89,011	193,662	501,066	43.59%	17.76%	38.65%

*percentage of total vaccines administered in that month.