



Published in final edited form as:

Expert Rev Vaccines. 2025 December ; 24(1): 230–241. doi:10.1080/14760584.2025.2483719.

Strengths and limitations of non-survey-based data sources for assessing adult vaccination coverage in the United States

Matthew F. Daley^{a,b}, Kamonthip J. Homdayjanakul^c, Laura P. Hurley^{d,e}, Peng-Jun Lu^f, Yuping Tsai^f, Carla L. Black^f, Suchita Patel^f, James A. Singleton^f, Lori A. Crane^{c,d}

^aInstitute for Health Research, Kaiser Permanente Colorado, Aurora, CO, USA

^bDepartment of Pediatrics, University of Colorado School of Medicine, Aurora, CO, USA

^cDepartment of Community and Behavioral Health, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

^dAdult and Child Consortium for Health Outcomes Research and Delivery Science, University of Colorado Anschutz Medical Campus and Children's Hospital Colorado, Aurora, CO, USA

^eDivision of General Internal Medicine, Denver Health, Denver, CO, USA

^fImmunization Services Division, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract

Introduction: Non-survey-based data sources (e.g. electronic health records, administrative claims) have been used to estimate vaccination coverage among US adults. However, these data sources were not collected for research or surveillance purposes and may have substantial limitations. The objectives of this narrative review were to: 1) identify published studies that used non-survey-based data sources to estimate adult vaccination coverage for one or more routinely recommended vaccines; and 2) summarize the strengths and limitations of these data sources for coverage assessments.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

CONTACT Matthew F. Daley, matthew.f.daley@kp.org, Institute for Health Research, Kaiser Permanente Colorado, 16601 E. Centretch Pkwy, Aurora, CO 80011, USA.

Author contributions

Conception and study design: MF Daley, KJ Homdayjanakul, LP Hurley, P-J Lu, Y Tsai, CL Black, S Patel, JA Singleton, and LA Crane; Analysis: MF Daley, KJ Homdayjanakul, LP Hurley, and LA Crane; Interpretation of data: MF Daley, KJ Homdayjanakul, LP Hurley, P-J Lu, Y Tsai, CL Black, S Patel, JA Singleton, and LA Crane; Drafting of the manuscript: MF Daley; Reviewing and revising manuscript for intellectual content: KJ Homdayjanakul, LP Hurley, P-J Lu, Y Tsai, CL Black, S Patel, JA Singleton, and LA Crane; Final approval of the version to be published: MF Daley, KJ Homdayjanakul, LP Hurley, P-J Lu, Y Tsai, CL Black, S Patel, JA Singleton, and LA Crane. All authors agree to be accountable for this work.

Declaration of interest

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

Reviewer disclosures

Peer reviewers on this manuscript have no relevant financial or other relationships to disclose.

Areas covered: Thirty-four publications derived from 9 data sources were reviewed: 16 publications were in a general population (i.e. defined by age), 12 were among pregnant women, and 6 were among individuals with chronic health conditions. While several data sources used continuous health insurance enrollment to define the study population, doing so limited generalizability to stably insured populations. Methods for obtaining race and ethnicity data were complex and potentially subject to bias. None of the reviewed studies presented any formal assessment of vaccine data validity.

Expert opinion: While multiple non-survey-based data sources have been used to assess adult vaccination coverage in the United States, important limitations exist, including related to generalizability, data validity, and risk of bias.

Keywords

Adult vaccination; immunization; vaccine; surveillance; coverage

1. Introduction

The adult immunization schedule in the United States has become increasingly complex as new vaccines are added for the general population or specific risk groups [1-3]. While new vaccines represent new opportunities for disease prevention, adult vaccination programs have historically received less attention than childhood programs, public health resources are more limited to promote vaccination in adults, and vaccination coverage for recommended vaccines in adults remains low [4-7]. Vaccination coverage estimates can help guide public health resources to areas of greatest need [7]. Historically, coverage in US adults has been assessed through national surveys, such as the National Health Interview Survey (NHIS) [7], the Behavioral Risk Factor Surveillance Survey (BRFSS) [8], the National Immunization Survey Adult COVID Module [9,10], and Internet panel surveys [11]. Although survey-based assessments have many strengths and remain the foundation for adult coverage surveillance at the Centers for Disease Control and Prevention (CDC), important limitations include: low survey response rates; nonresponse bias; reliance on self-reported vaccination status; small sample size for subgroup analyses; survey costs; and timeliness of data for many survey platforms [7,12].

Over the past decade, non-survey-based data sources have increasingly been used to assess adult vaccination coverage [13-16]. Many of these data sources rely on electronic data created through documentation of care provided (i.e. electronic health records [EHR]) or payment for care received (i.e. administrative claims). Using EHR and claims for this purpose is appealing as these data sources rely on already-existing data, are potentially more timely and less expensive than surveys, and usually cover large populations across multiple states [13-16]. Jurisdiction-specific immunization information systems (IIS) can also be used to assess population-level coverage: IIS are confidential, population-based, computerized records of vaccine doses administered in specific geographic regions in the United States [17]. IIS were initially focused on pediatric rather than adult populations, and not all vaccination providers report to a jurisdictional IIS [17]. Additionally, laws and policies regarding the submission of vaccination data to an IIS vary by jurisdiction, which may impact the completeness of available vaccine data. While non-survey-based data

sources have the potential to make important contributions to coverage assessments, the strengths and limitations of these data sources have not been fully described [18].

Additionally, the national landscape is rapidly changing, with continued adoption of EHR in diverse settings [19], expansion of insurer-linked consolidated electronic databases utilized for research purposes [20], the pandemic-related requirement that COVID-19 vaccination sites electronically report vaccine administration data to the federal government, ideally through jurisdictional IIS [18], and new vaccine recommendations for adults [21,22]. In this context it is important to evaluate the utility of non-survey-based methods for estimating adult vaccination coverage. No other studies have comprehensively evaluated this question. The objectives of this narrative review were to 1) identify published studies that used non-survey-based data sources to estimate adult vaccination coverage for one or more vaccines; and 2) assess the strengths and limitations of these data sources for coverage estimation.

2. Methods

2.1. Overview

In January through May 2022, we searched the PubMed and Google Scholar electronic databases to identify non-survey-based data sources for estimating adult vaccination coverage in a US-based population. A first search was performed to identify data sources, and a second search was performed to collect all published articles that used the identified data sources. Using a standardized data collection instrument, three investigators independently reviewed all articles, rating each data source according to predefined core attributes (e.g. representativeness, data quality). In addition, the strengths and limitations of each data source were summarized across several use cases.

2.2. Study eligibility criteria

We required that eligible studies be published in English and provide an estimate of vaccination coverage in an adult (i.e. aged 18 years or older) population located in the United States. We limited our review to studies reporting coverage for vaccines licensed in the United States and listed on the current US adult vaccination schedule [1-3]; this approach was intentionally broad and included vaccines recommended for a general population (i.e. defined only by age) or recommended for individuals with additional risk factors or other indications. While we did not restrict the publication years in our search strategy, the earliest study that met study eligibility criteria was published in 2007, and we did not include studies published after the search process was completed in May 2022. We included studies that generated coverage estimates for children and adolescents if the data also included adults aged 18 years or older. We did not include studies that exclusively used survey-based data, because the intent of this review was to evaluate non-survey sources. We did not include studies presented in abstract form only, such as for a scientific conference. Finally, we did not include studies with numerator data only, meaning studies that only presented vaccine doses administered without denominator data.

2.3. Search strategy

The search of published medical literature was accomplished in two phases. The purpose of the first phase was to develop a comprehensive list of data source types for any article that included an adult vaccination coverage estimate. The search was conducted along two axes: for axis one, we used the keyword terms ‘vaccine,’ ‘vaccination,’ and ‘immunization.’ For axis two, we used the keyword terms ‘coverage,’ ‘estimate,’ ‘estimation,’ and ‘uptake.’ For each axis and each database, the keywords were combined with the ‘OR’ operator as well as the ‘AND’ operator. For retrieved articles, the title, abstract, and if necessary full text were reviewed for relevance to our research question. To keep the search strategy as broad and comprehensive as possible, the keyword term ‘adult’ was not included in the search; while many child-only publications were returned in the initial search, these were readily excluded upon review of article title, abstract, and text.

Once saturation was reached for the data sources identified from the first phase, a second phase was conducted; this phase focused on identifying any additional articles that used these data sources to estimate vaccination coverage. This search was also conducted along two axes: for axis one, we used the keyword terms ‘vaccine,’ ‘vaccination,’ and ‘immunization.’ For axis two, we used the keyword terms ‘registry,’ ‘immunization information system,’ ‘IIS,’ ‘Vaccine Safety Datalink,’ ‘VSD,’ ‘EHR,’ ‘IQVIA,’ ‘PRISM,’ ‘Optum,’ ‘MarketScan,’ ‘claims database,’ ‘multi-state Medicaid,’ ‘Medicaid,’ and ‘Medicare.’ Axis keywords were combined using the ‘OR’ and ‘AND’ operators.

2.4. Study appraisal and synthesis methods

Prior to reviewing the retrieved articles, we developed a list of key attributes and use cases against which each article was rated. The key attributes, designed to assess the fitness of the data source for generating adult vaccination coverage estimates, were identified based upon published literature and guidelines, including the IIS Data Quality Blueprint produced by the CDC [23,24]. Five main attributes were assessed for each data source: representativeness, timeliness, completeness, data accessibility for research and public health purposes, and data validity. We developed a 5-point Likert-type scale for each attribute: a rating of 5 indicated the data source reached the highest standard for this attribute with few limitations; 4 indicated the data source had many strengths but some limitations for broad use or representativeness; 3 indicated the data source was adequate for the attribute but had important limitations; 2 indicated the data source had many limitations for this attribute and limited circumstances in which use could be justified; and 1 indicated that the data source should not be used based on limitations with respect to this attribute. To further guide our rating process, we developed a detailed definition for each score when applied to each specific attribute; these are provided in Table 1.

As an additional process separate from the ratings described above, we developed four use cases and then assessed the utility of each data source for each use case. The use cases included: 1) ability to generate vaccination coverage estimates for adult populations at increased risk for severe disease and disease complications (e.g. pregnant women, adults with chronic medical conditions); 2) ability to generate coverage estimates for more than

one vaccine; 3) ability to generate coverage estimates for multi-dose vaccine series; and 4) ability to generate vaccination coverage estimates for sociodemographic groups of public health importance (e.g. by race and ethnicity).

The study appraisal process was conducted by three investigators (LAC, LPH, MFD; these investigators have training in internal medicine, pediatrics, and public health, and all have expertise in immunization services delivery research). Each article was independently reviewed by two investigators, using a standardized data collection instrument. Each data source type was judged based on the information provided in the published manuscript(s) relevant to each data source; we did not directly contact the study authors or the organizations that maintain the data sources for any unpublished information regarding the data sources. Ratings were compared between the independent raters; when ratings were discordant, all three investigators met, re-reviewed the articles, and reached a consensus regarding the rating.

2.5. Research compliance

This research involved a review of previously published articles. This does not meet the definition of human subject research, and review by an institutional review board does not apply.

3. Results

3.1. Data sources identified

We identified 9 data source types for published non-survey-based studies of adult vaccination coverage in the United States: all-payer claims databases [25]; IIS [26,27]; IQVIA [14]; MarketScan [13,28-36]; single-state Medicaid [37]; fee-for-service Medicare [38-44]; Optum [15,45-47]; the Post-licensure Rapid Immunization Safety Monitoring (PRISM) program [48]; and the Vaccine Safety Datalink (VSD) [16,49-54]. Thirty-four publications [13-16,25-54] were included in our review, with a range of one to 10 publications per data source type; the characteristics of these data sources are described in Table 2. Some of these data sources integrated data from multiple underlying systems. For example, studies using MarketScan may have included health care claims from large employers and health plans [13,28,29,31,34-36], as well as multi-state Medicaid claims data [30,32]. One study using Optum included claims and EHR data [47] whereas others relied primarily on claims [15,45,46]. Although PRISM vaccine data was primarily claims-based, jurisdictional IIS data were also included in the single study we reviewed [48]. Finally, VSD data were primarily EHR-derived, but claims and IIS data supplemented EHR data in published studies [16,49-54].

Regarding population size and geographical distribution (also shown in Table 2), all 9 data sources were large and covered multiple states. However, currently there is no national all-payer claims database, and no national IIS. For these two data sources, each state or jurisdiction operated independently, and the reviewed studies used data from single US states [25-27]. Other data sources (IQVIA, MarketScan, Optum) reported the inclusion of

data from all 50 US states, although it was difficult to determine from published studies the population size contributed from each state.

3.2. Populations and vaccines assessed

As shown in Table 3, the identified studies of adult vaccination coverage from non-survey sources included a wide array of age groups, populations, and vaccines assessed. We identified the greatest number of published studies for MarketScan, Medicare, VSD, and Optum [13,15,16,28-36,38-47,49-54]; while fewer published studies used all-payer claims databases, IIS, IQVIA, or PRISM [14,25-27,48]. We identified 14 publications among a general population [13-15,26,28,29,35,39-41,43,48,49] (the term ‘general population’ refers to a population not selected based on high-risk criteria other than age), 12 among pregnant women [25,27,30-32,36,37,50-54], and 6 among individuals with chronic health conditions [33,34,38,45-47]. Additionally, two studies from the Medicare data source included general (i.e. ≥ 65 years) and higher risk (i.e. those < 65 years with permanent disabilities, end-stage renal disease) populations [42,44]. Pregnant women were identified using birth certificate records, claims-based diagnosis and procedural codes (i.e. *International Classification of Diseases, 9th and 10th Revision* and *Current Procedural Terminology* codes) and EHR-based diagnosis and procedural codes [25,27,30-32,36,37,50-54]. The studies among adults with chronic health conditions also relied upon claims-based diagnosis codes to identify the populations of interest [33,34,45-47]. Most studies among pregnant women and adults with chronic health conditions did not report on the accuracy of the methods used to identify the populations of focus. The one exception was VSD studies among pregnant women, in which the algorithm used to identify pregnant women was validated against manual medical record review [55,56].

In terms of the different vaccines assessed, we identified coverage studies of most vaccines routinely recommended for adults (see Table 3), including vaccines recommended based on age and those recommended based on additional risk factors or other indications [2]. Most published studies acknowledged the difficulty in estimating the completeness of the vaccine data available. Strategies used to mitigate the problem of missing vaccine data when using administrative claims data included requiring a certain duration of health insurance enrollment (e.g. requiring one year’s enrollment), although this approach could limit generalizability because individuals with unstable insurance coverage would be excluded. Additionally, vaccination that did not generate a health insurance claim, such as vaccination in a workplace or vaccination that occurred prior to health insurance enrollment, may have been missing. Vaccination coverage estimates from the reviewed studies were often lower than national estimates based on surveys. For example, in a study using Optum data, Li and colleagues reported influenza vaccination coverage among adults of 42.3% during August to December 2019, and 45.9% during the same period of 2020 [15]; this compares to national influenza coverage estimates from BRFSS of 48.4% in 2019–2020 and 50.2% in 2020–2021 [57,58]. In a study using MarketScan data to assess maternal Tdap and influenza coverage, Ghaswalla and colleagues [30] reported lower coverage for both vaccines than national survey-based estimates, particularly for pregnant women enrolled in Medicaid rather than private health insurance [59,60].

3.3. Qualitative rating of data sources

A qualitative rating of the attributes of non-survey-based data sources for assessing adult vaccination coverage is presented in Table 4. The table highlights our observation that these data sources, when used for adult vaccination coverage estimates, have varying strengths and limitations across the attributes assessed. Data sources which rely heavily upon claims data such as IQVIA, MarketScan, and Optum [13-15,28-36,45-47] were rated more highly for data accessibility and timeliness and were able to estimate coverage by presence of a chronic health condition or by health status. However, the completeness and validity of vaccine data from these sources were rated less highly. The timeliness and validity of vaccine data from IIS were rated more highly, but only two published studies were exclusively IIS-based [26,27] (as opposed to using IIS data to supplement other data sources). The VSD was the only identified data source that relied primarily on EHR-derived data [16,49-54]; the timeliness and validity of vaccine data was rated highly, and the VSD was able to identify pregnant cohorts [50-54] using validated EHR-based algorithms [55,56]. Only one study was identified from PRISM that reported coverage estimates [48], and the study was conducted to inform future vaccine safety studies rather than adult vaccination coverage surveillance. As such, it was difficult to rate the attributes of PRISM for use in coverage surveillance.

3.4. Data source utility for different use cases

The examination of data sources with respect to four use cases is presented in Table 5. While most of the nine data sources could be used to estimate coverage for multiple vaccines given simultaneously (such as Tdap and influenza), these data sources were found to have important limitations for assessing coverage by health status (e.g. pregnancy), for multi-dose series (e.g. HPV vaccine), and by sociodemographic characteristics such as race and ethnicity. The Medicare fee-for-service and Medicaid data sources appeared to have fewer limitations when estimating coverage by sociodemographic characteristics, whereas the VSD and Optum databases had fewer limitations when estimating coverage by presence of a chronic medical condition. Each of the nine data sources had substantial limitations depending on the use case.

4. Discussion

In this narrative review of non-survey data sources for assessing adult vaccination coverage for routinely recommended vaccines in the United States, we assessed 9 data source types contributing more than 30 peer-reviewed publications from 2007 through 2022. Studies reported coverage for a wide range of the vaccines currently recommended for adults, with multiple studies examining coverage in populations defined by their health status (e.g. pregnancy or presence of a chronic health condition). The number of data sources and publications, and the diversity of vaccines and populations examined, suggest that non-survey data sources are viable resources for estimating adult vaccination coverage at the national or sub-national level. However, it is important to highlight that these data sources have distinct strengths and limitations. Additionally, when reviewing manuscripts for this study, detailed information on the limitations of various data sources was often missing. Finally, the merits of one data source compared to others varied depending upon the research questions being posed within individual studies.

While some studies made inferences about data quality, none of the reviewed studies presented any formal assessment of vaccination data validity. In the context of claims- and EHR-based vaccine administration data, false negatives (i.e. an individual who appeared unvaccinated but had received the vaccine of interest) may occur for vaccines administered outside the health system, and are arguably more likely to occur than false positives (i.e. an individual who appeared vaccinated who truly was not), and false negatives would underestimate vaccination coverage. The likelihood of missing vaccination information probably differs by vaccine and data source, however. For example, influenza vaccine, relatively inexpensive and often available in workplaces, may be more likely misclassified than other vaccines, because vaccination may not be billed to insurance [61,62]. In some settings with EHRs, vaccination data in EHRs are supplemented by vaccination data from jurisdictional IIS [63], making these systems less prone to missing vaccination information than claims-based systems. In the reviewed studies, the accuracy of vaccination coverage estimates is often inferred by comparing study estimates to those based on national surveys; while this approach is understandable, survey-derived estimates are also susceptible to error, such as error created by biases inherent in sampling, response, and recall of vaccination. Validating the vaccination data in non-survey databases would entail triangulating vaccination data across data sources (such as comparisons between claims, EHR, IIS, and self-report), manually reviewing individual EHR records to look for coding errors, or surveying a sample of individuals from these data sources. While these validation studies would be challenging to conduct, they would greatly enhance understanding of the validity of non-survey-based data sources.

In the reviewed studies, it was also difficult to assess the validity of the denominator used to calculate coverage, particularly for studies in general populations (i.e. those not identified according to health status). For example, when using the IQVIA database to estimate pneumococcal conjugate vaccine uptake among adults aged ≥ 65 years, a ‘proprietary algorithm’ was used to identify a ‘stable panel,’ from which coverage estimates were extrapolated to the US population [14]. Li and colleagues used continuous insurance enrollment to define their population denominator in an Optum-based study of influenza vaccination coverage; a table was provided comparing the study population demographics to the US population, further aiding study interpretability [15]. When estimating live attenuated zoster vaccination coverage for adults aged ≥ 50 years using MarketScan, Zhang and colleagues also used continuous health insurance enrollment to define their population denominator; while this may represent an accurate and interpretable denominator, few other details were provided [13]. While limiting vaccination coverage assessments to adults with continuous health insurance may reduce data missingness (e.g. for vaccines received or for health status), doing so limits generalizability to groups with stable insurance over multiple years; the estimates generated would not reflect vaccination coverage among uninsured or incarcerated adults, for example.

Persistent racial and ethnic disparities in adult vaccination coverage have been documented [7,64], and reducing these disparities is an important national public health goal. Surveys such as the NHIS are critical to understanding these disparities, because self-reported vaccination status is collected alongside individual-level data on race, ethnicity, and socioeconomic factors such as household income and educational attainment [7,64]. For

several of the data sources reviewed here, the methods for obtaining race and ethnicity data were complex and opaque, and potentially subject to bias. In a study assessing racial and ethnic disparities in influenza vaccination coverage, Li and colleagues used Optum claims data linked to ‘additional sources’ in ‘proprietary analytic and demographic models’ to obtain race and ethnicity data; individuals excluded from analyses due to missing demographic information were more likely from racial and ethnic minority groups [15]. In a study examining racial and ethnic disparities in pneumococcal conjugate vaccination coverage among adults aged ≥ 65 years, McLaughlin and colleagues used IQVIA claims data linked to the Experian database to obtain individual-level sociodemographic information; Experian data ‘come primarily from consumer participation in the credit market’ [14]. The authors acknowledged that older adults living in poverty or with less formal education may be underrepresented in the credit market, leading to differential exclusion of these populations in the IQVIA-Experian linked dataset used for their analyses [14]. These examples illustrate the challenges and potential direction of bias when using non-survey-based data sources to monitor disparities: if racial and ethnic minority groups are underrepresented in analytic datasets, the extent of disparities in vaccine coverage may be underestimated. Finally, while Medicare and EHR-based systems collect data on race and ethnicity, the completeness and accuracy of the race and ethnicity data may create challenges [65], and these data sources have limitations to generalizability (Medicare beneficiaries are predominantly ≥ 65 years of age, VSD members are predominantly privately insured and reside in a limited number of US states) [16,40,66].

Because adults with chronic health conditions are particularly vulnerable to certain vaccine-preventable diseases, monitoring coverage in these populations is important to help guide interventions and other public health activities. The reviewed studies demonstrated substantial capacity for assessing vaccination coverage in these special populations, including among adults with asplenia [45], complement deficiencies [46], immunocompromise [47], sickle cell disease [33], and end-stage renal disease [38]. The conditions of interest were most often identified using International Classification of Diseases, Ninth or Tenth Revision (ICD-9, ICD-10) diagnosis codes or procedure codes; however, the accuracy of using these codes to identify the study cohort was not validated. Additionally, the nature of these studies increases the likelihood of missing historical vaccination: the conditions of focus may have been present for many years and prompted vaccination years earlier. While these studies often required one year or more of continuous health insurance enrollment to be included in the study cohort, vaccination prior to insurance enrollment may have been missing. Non-survey data sources have also been used extensively to examine vaccination coverage among pregnant women [25,27,30-32,36,37,50-54]. While issues of cohort identification and missing historical vaccination may be less problematic in studies of pregnant women, issues of generalizability arise. For example, while the VSD has developed and validated an algorithm to identify pregnant women [55,56], and supplements EHR-based data with vaccine data from jurisdictional IIS [63], vaccination coverage in the VSD population likely does not reflect coverage in other populations (such as pregnant women with public health insurance or intermittent private insurance) [66].

The findings in this narrative review should be interpreted in the context of several limitations. First, we limited our review to manuscripts published in peer-reviewed journals;

we did not directly interrogate the data sources themselves. While the CDC has developed publicly facing ‘dashboards’ for preliminary influenza vaccination coverage estimates [67], and state immunization programs have used jurisdictional IIS for adult vaccination coverage reports, these types of assessments were not included in our review. Second, we did not review coverage studies based on place of employment (e.g. coverage among health care personnel) or residence (e.g. coverage among residents of nursing homes). Third, although our search strategy was broad and iterative, it is possible that we did not identify all published manuscripts and all relevant data sources for inclusion in our review. Related, we have not included any manuscripts published after our literature search was completed in May 2022. Fourth, because we were not able to identify an already-existing schema for comparing and rating the reviewed data sources, we developed and defined a set of key study attributes relevant to adult vaccination coverage. While the schema that we developed was useful in guiding our review, the process and the resulting ratings were subjective and qualitative in nature. It is certainly possible that other researchers would have prioritized other key attributes or rated the reviewed data sources differently. However, we used a process with multiple reviewers and adjudication of discrepancies in judgment between reviewers. Fifth, when evaluating different data sources, we were reliant upon the study methods described in the manuscripts, and at times relevant details were not included. In such circumstances, our assessment of key attributes was inferred from knowledge about the limitations of certain types of data; for example, vaccination data from claims-based sources was generally viewed as less complete than the vaccination data from EHR-based sources such as the VSD, recognizing that exceptions may exist. Sixth, only a single publication was identified for each of several data source types (e.g. all-payer claims databases [25], IQVIA [14], single-state Medicaid [37], and PRISM [48]), which limited the assessment of the attributes of these data sources. Finally, this review focused on published studies regarding adult vaccination coverage, and the findings and limitations do not apply to using these data sources (e.g. MarketScan or Optum) for other purposes.

5. Conclusions

In conclusion, while multiple non-survey data sources have been used to assess adult vaccination coverage in the United States, across a range of populations and vaccines, important limitations exist related to generalizability, data quality, and the resulting risk of bias. In the context of four use cases of critical public health importance, no single data source had high utility with minimal limitations. Therefore, it is important to carefully consider the strengths and limitations of these data sources, which were not developed for research or surveillance purposes, when answering a particular assessment or research question. To improve the validity of coverage estimates, it may be necessary to synthesize findings across more than one data source.

6. Expert opinion

National surveys remain the cornerstone for adult vaccination coverage surveillance in the United States, and have many advantages compared to non-survey-based data sources, which are not collected for research or surveillance purposes. Critically, well-resourced national surveys can employ complex sampling and weighting strategies to strengthen the

representativeness and generalizability of survey findings. Additionally, surveys can collect respondent-reported data on socioeconomic factors such as race, ethnicity, education, and income, which are important for identifying disparities in coverage. Data on current health status and respondent attitudes and knowledge about vaccines can also be collected through surveys; these data are important to design interventions to increase vaccine acceptance and coverage, including among individuals with risk-based indications for vaccination.

As described in this narrative review, more than 30 peer-reviewed publications have used non-survey-based data sources to estimate adult coverage in a US population. The sheer quantity of publications does not necessarily provide insight into the quality or utility of these data sources, however. While these data sources may mitigate some of the limitations of surveys, such as nonresponse bias, recall bias, and challenges with timeliness, non-survey-based data sources introduce other significant limitations. To illustrate this point, we provide three examples related to the numerator (i.e. the vaccines received), the denominator, and the representativeness of the study population. In studies reliant on health insurance claims to estimate vaccination coverage, not all vaccines administered will generate a claim, resulting in an underestimation of the true coverage. In studies using statelevel IIS to estimate coverage, defining an adult population denominator is challenging due to people moving out of the jurisdiction without being removed from the denominator; this presumably leads to an inflated denominator which would bias results toward an underestimation of true coverage. With respect to representativeness, studies relying on EHR data are typically restricted to a stably insured population; as such, coverage estimates may be higher than the true population-wide coverage. The purpose of this review has been to not only detail the extent of coverage-related publications using non-survey-based data sources but interrogate the limitations of these data sources.

How could the contribution of non-survey-based data sources to estimating adult vaccination coverage be strengthened? First, there is a need to fully understand how these data sources are constructed, how coverage estimates are calculated, and to what extent these estimates reflect true coverage in a population. Our narrative review found a lack of clarity in describing the construction and representativeness of the populations included. Methods sections were frequently brief and opaque regarding how the cohort was defined, the denominator calculated, and what geographic areas contributed to the data sources. Regarding the completeness and accuracy of the underlying data, systematic data quality assessments are needed. For example, a manual review of records could be conducted to verify vaccination status for a sample of records, and a sample of individuals could be contacted to verify their vaccination status. Researchers using claims or EHR-based systems could also link records with their jurisdictional IIS (while maintaining privacy and confidentiality), again with the goal of conducting a more formal analysis of the accuracy of vaccine administration data.

Reinforcement of these expectations could be addressed through the review process at scientific journals. At the time manuscripts are submitted to journals, peer reviewers and editors could request a detailed description of how the cohort was developed and the denominator determined, and a quantitative assessment of the representativeness of the study population. Regarding data quality issues, reviewers and editors could hold studies

using non-survey-based sources to similar standards for data validity that they apply to studies using surveys. At a minimum, studies reporting coverage estimates based on non-survey-based data sources should transparently describe the potential biases inherent in the data sources and analyses in relation to the research question being addressed. While non-survey-based data sources have the potential to provide rapid coverage assessments in large populations, the promise of these data sources may not be fully realized until more foundational work is completed around data quality and population representativeness.

Acknowledgments

The authors MF Daley, LP Hurley, and LA Crane, had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding

The authors were supported by the Centers for Disease Control and Prevention, through the Rocky Mountain Prevention Research Center, Cooperative Agreement [U48DP006399], Special Interest Project 20–008. Co-authors from the Centers for Disease Control and Prevention were involved in design and conduct of the study, interpretation of data, and review and approval of the manuscript. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abbreviations

ACIP	Advisory Committee on Immunization Practices
BRFSS	Behavioral Risk Factor Surveillance Survey
CDC	Centers for Disease Control and Prevention
HER	electronic health records
IIS	immunization information systems
NHIS	National Health Interview Survey
PRISM	Post-licensure Rapid Immunization Safety Monitoring
VSD	Vaccine Safety Datalink

References

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Murthy N, Wodi AP, Bernstein H, et al. Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older — United States, 2022. *MMWR Morb Mortal Wkly Rep.* 2022;71(7):229–233. doi: 10.15585/mmwr.mm7107a1 [PubMed: 35176010]
2. Murthy N, Wodi AP, McNally V, et al. Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older — United States, 2023. *MMWR Morb Mortal Wkly Rep.* 2023;72(6):141–144. doi: 10.15585/mmwr.mm7206a2 [PubMed: 36757861]

3. Murthy N, Wodi AP, McNally VV, et al. Advisory committee on immunization practices recommended immunization schedule for adults aged 19 years or older - United States, 2024. *MMWR Morb Mortal Wkly Rep.* 2024;73(1):11–15. [PubMed: 38206880]
4. Gellin BG, Shen AK, Fish R, et al. The national adult immunization plan: strengthening adult immunization through coordinated action. *Am J Prev Med.* 2016;51(6):1079–1083. [PubMed: 27236479]
5. Roper L, Hall MAK, Cohn A. Overview of the United States' immunization program. *J Infect Dis.* 2021;224(12 Suppl 2):S443–s451. doi: 10.1093/infdis/jiab310 [PubMed: 34590134]
6. Kolobova I, Nyaku MK, Karakusevic A, et al. Vaccine uptake and barriers to vaccination among at-risk adult populations in the US. *Hum Vaccin Immunother.* 2022;18(5):2055422. doi: 10.1080/21645515.2022.2055422 [PubMed: 35536017]
7. Lu PJ, Hung MC, Srivastav A, et al. Surveillance of vaccination coverage among adult populations -United States, 2018. *MMWR Surveill Summ.* 2021;70(3):1–26.
8. Centers for Disease Control and Prevention. Vaccination coverage among adults. [cited 2023 Mar 23]. Available from: <https://www.cdc.gov/vaccines/imz-managers/coverage/adultvaxview/data-reports/general-population/index.html>
9. Lu PJ, Zhou T, Santibanez TA, et al. COVID-19 bivalent booster vaccination coverage and intent to receive booster vaccination among adolescents and adults — United States, November–December 2022. *MMWR Morb Mortal Wkly Rep.* 2023;72(7):190–198. doi: 10.15585/mmwr.mm7207a5 [PubMed: 36795677]
10. Kriss JL, Hung MC, Srivastav A, et al. COVID-19 vaccination coverage, by race and ethnicity — National immunization survey adult COVID module, United States, December 2020–November 2021. *MMWR Morb Mortal Wkly Rep.* 2022;71(23):757–763. doi: 10.15585/mmwr.mm7123a2 [PubMed: 35679179]
11. Kahn KE, Razzaghi H, Jatlaoui TC, et al. Flu, Tdap, and COVID-19 vaccination coverage among pregnant women-United States. 2022 Apr [cited 2022 Nov 5]. Available from: <https://www.cdc.gov/flu/fluview/pregnant-women-apr2022.htm>
12. Meghani M, Razzaghi H, Kahn KE, et al. Surveillance systems for monitoring vaccination coverage with vaccines recommended for pregnant women, United States. *J Womens Health (Larchmt).* 2023;32(3):260–270. [PubMed: 36884385]
13. Zhang D, Johnson K, Newransky C, et al. Herpes zoster vaccine coverage in older adults in the U.S. 2007–2013. *Am J Prev Med.* 2017;52(1):e17–e23. [PubMed: 28340974]
14. McLaughlin JM, Swerdlow DL, Khan F, et al. Disparities in uptake of 13-valent pneumococcal conjugate vaccine among older adults in the United States. *Hum Vaccin Immunother.* 2019;15(4):841–849. doi: 10.1080/21645515.2018.1564434 [PubMed: 30676236]
15. Li K, Yu T, Seabury SA, et al. Trends and disparities in the utilization of influenza vaccines among commercially insured US adults during the COVID-19 pandemic. *Vaccine.* 2022;40(19):2696–2704. doi: 10.1016/j.vaccine.2022.03.058 [PubMed: 35370018]
16. Pingali C, Meghani M, Razzaghi H, et al. COVID-19 vaccination coverage among insured persons aged ≥16 years, by race/ethnicity and other selected characteristics — eight integrated health care organizations, United States, December 14, 2020–May 15, 2021. *MMWR Morb Mortal Wkly Rep.* 2021;70(28):985–990. doi: 10.15585/mmwr.mm7028a1 [PubMed: 34264911]
17. Groom H, Hopkins DP, Pabst LJ, et al. Immunization information systems to increase vaccination rates: a community guide systematic review. *J Public Health Manag Pract.* 2015;21(3):227–248. doi: 10.1097/PHH.000000000000069 [PubMed: 24912082]
18. Shen AK, Sobczyk EA, Coyle R, et al. How ready was the US vaccination infrastructure and network of immunization information systems for COVID-19 vaccination campaigns: recommendations to strengthen the routine vaccination program and prepare for the next pandemic. *Hum Vaccin Immunother.* 2022;18(5):2088010. doi: 10.1080/21645515.2022.2088010 [PubMed: 35796624]
19. Centers for Disease Control and Prevention. Electronic medical records/electronic health records (EMRs/EHRs). [cited 2023 Jun 2]. Available from: <https://www.cdc.gov/nchs/fastats/electronic-medical-records.htm>

20. Moll K, Lufkin B, Fingar KR, et al. Background rates of adverse events of special interest for COVID-19 vaccine safety monitoring in the United States, 2019–2020. *Vaccine*. 2023;41(2):333–353. [PubMed: 36404170]
21. Melgar M, Britton A, Roper LE, et al. Use of respiratory syncytial virus vaccines in older adults: recommendations of the advisory committee on immunization practices — United States, 2023. *MMWR Morb Mortal Wkly Rep*. 2023;72(29):793–801. doi: 10.15585/mmwr.mm7229a4 [PubMed: 37471262]
22. Fleming-Dutra KE, Jones JM, Roper LE, et al. Use of the Pfizer respiratory syncytial virus vaccine during pregnancy for the prevention of respiratory syncytial virus-associated lower respiratory tract disease in infants: recommendations of the advisory committee on immunization practices - United States, 2023. *MMWR Morb Mortal Wkly Rep*. 2023;72(41):1115–1122. doi: 10.15585/mmwr.mm7241e1 [PubMed: 37824423]
23. Lu PJ, Santibanez TA, Williams WW, et al. Surveillance of influenza vaccination coverage—United States, 2007–08 through 2011–12 influenza seasons. *MMWR Surveill Summ*. 2013;62(4):1–28.
24. Centers for Disease Control and Prevention. Immunization information systems (IIS) data quality blueprint. [cited 2023 Jul 20]. Available from: <https://www.cdc.gov/vaccines/programs/iis/downloads/Data-Quality-Blueprint-508.pdf>
25. Koepke R, Kahn D, Petit AB, et al. Pertussis and influenza vaccination among insured pregnant women - Wisconsin, 2013–2014. *MMWR Morb Mortal Wkly Rep*. 2015;64(27):746–750. [PubMed: 26182193]
26. Enger KS, Stokley S. Meningococcal conjugate vaccine uptake, measured by Michigan's immunization registry. *J Adolesc Health*. 2007;40(5):398–404. doi: 10.1016/j.jadohealth.2006.11.141 [PubMed: 17448396]
27. Koepke R, Schauer SL, Davis JP. Measuring maternal Tdap and influenza vaccination rates: comparison of two population-based methods. *Vaccine*. 2017;35(18):2298–2302. doi: 10.1016/j.vaccine.2017.03.024 [PubMed: 28341114]
28. Dunne EF, Stokley S, Chen W, et al. Human papillomavirus vaccination of females in a large health claims database in the United States, 2006–2012. *J Adolesc Health*. 2015;56(4):408–413. [PubMed: 25797632]
29. Gargano JW, Zhou F, Stokley S, et al. Human papillomavirus vaccination in commercially-insured vaccine-eligible males and females, United States, 2007–2014. *Vaccine*. 2018;36(23):3381–3386. doi: 10.1016/j.vaccine.2018.03.045 [PubMed: 29735321]
30. Ghaswalla P, Poirrier JE, Packnett ER, et al. Maternal immunization in the U.S.: a nationwide retrospective cohort study. *Am J Prev Med*. 2019;57(3):e87–e93. [PubMed: 31427034]
31. Hong K, Lindley MC, Zhou F. Coverage and timing of influenza vaccination among privately insured pregnant women in the United States, 2010–2018. *Public Health Rep*. 2022;137(4):739–748. doi: 10.1177/00333549211026779 [PubMed: 34161183]
32. Moll K, Wong HL, Fingar K, et al. Vaccine exposure during pregnancy among privately and publicly insured women in the United States, 2016–2018. *Vaccine*. 2021;39(41):6095–6103. doi: 10.1016/j.vaccine.2021.08.091 [PubMed: 34507857]
33. Payne AB, Adamkiewicz TV, Grosse SD, et al. Influenza vaccination rates and hospitalizations among Medicaid enrollees with and without sickle cell disease, 2009–2015. *Pediatr Blood Cancer*. 2021;68(12):e29351. [PubMed: 34542932] • A multi-state Medicaid claims database was used to assess influenza vaccination coverage among persons with and without sickle cell disease. Three or more claims for a diagnosis of sickle cell disease were required to meet the study definition of sickle cell disease. Claims were also used to identify influenza vaccination.
34. Petigara T, Zhang D. Pneumococcal vaccine coverage in adults aged 19–64 years, newly diagnosed with chronic conditions in the U.S. *Am J Prev Med*. 2018;54(5):630–636. doi: 10.1016/j.amepre.2018.01.033 [PubMed: 29551328]
35. Prabhu VS, Bansal N, Liu Z, et al. HPV vaccination uptake and administration from 2006 to 2016 in a commercially insured population of the United States. *BMC Public Health*. 2021;21(1):1629. doi: 10.1186/s12889-021-11664-1 [PubMed: 34488705]

36. Zhou F, Xu J, Black CL, et al. Trends in Tdap vaccination among privately insured pregnant women in the United States, 2009–2016. *Vaccine*. 2019;37(14):1972–1977. doi: 10.1016/j.vaccine.2019.02.042 [PubMed: 30826146]
37. Housey M, Zhang F, Miller C, et al. Vaccination with tetanus, diphtheria, and acellular pertussis vaccine of pregnant women enrolled in Medicaid–Michigan, 2011–2013. *MMWR Morb Mortal Wkly Rep*. 2014;63(38):839–842. [PubMed: 25254561]
38. Shen AK, Kelman JA, Warnock R, et al. Beneficiary characteristics and vaccinations in the end-stage renal disease medicare beneficiary population, an analysis of claims data 2006–2015. *Vaccine*. 2017;35(52):7302–7308. doi: 10.1016/j.vaccine.2017.10.105 [PubMed: 29132990]
39. Shen AK, Warnock R, Kelman JA. Driving immunization through the medicare annual wellness visit: a growing opportunity. *Vaccine*. 2017;35(50):6938–6940. [PubMed: 29089193]
40. Black CL, Williams WW, Warnock R, et al. Pneumococcal vaccination among medicare beneficiaries occurring after the advisory committee on immunization practices recommendation for routine use of 13-valent pneumococcal conjugate vaccine and 23-valent pneumococcal polysaccharide vaccine for adults aged ≥65 years. *MMWR Morb Mortal Wkly Rep*. 2017;66(27):728–733. [PubMed: 28704347]
41. Shen AK, Warnock R, Chu S, et al. Receipt of other routinely recommended vaccines relative to receipt of seasonal influenza vaccines: trends from medicare administrative data, 2013–2015. *Vaccine*. 2018;36(30):4399–4403. doi: 10.1016/j.vaccine.2018.06.020 [PubMed: 29909132]
42. Shen AK, Warnock R, Brereton S, et al. Influenza vaccination coverage estimates in the fee-for service medicare beneficiary population 2006–2016: using population-based administrative data to support a geographic based near real-time tool. *Hum Vaccin Immunother*. 2018;14(8):1848–1852. doi: 10.1080/21645515.2018.1462067 [PubMed: 29641277]
43. Shen AK, Warnock R, Selna W, et al. Vaccination among medicare-fee-for service beneficiaries: characteristics and predictors of vaccine receipt, 2014–2017. *Vaccine*. 2019;37(9):1194–1201. [PubMed: 30683507] • Data for Medicare fee-for-service beneficiaries were used to assess coverage for multiple vaccines among Medicare recipients. The authors identified disparities in coverage by race and ethnicity and by urban versus rural location.
44. Cho BH, Weinbaum C, Tsai Y, et al. Influenza vaccine uptake and missed opportunities among the medicare-covered population with high-risk conditions during the 2018 to 2019 influenza season: a retrospective cohort study. *Ann Intern Med*. 2022;175(1):1–10. doi: 10.7326/M21-1550 [PubMed: 34781717]
45. Ghaswalla PK, Bengtson LGS, Marshall GS, et al. Meningococcal vaccination in patients with newly diagnosed asplenia in the United States. *Vaccine*. 2021;39(2):272–281. [PubMed: 33309081]
46. Marshall GS, Ghaswalla PK, Bengtson LGS, et al. Low meningococcal vaccination rates among patients with newly diagnosed complement component deficiencies in the United States. *Clin Infect Dis*. 2022;75(1):155–158. [PubMed: 34718466] • A national health care claims database was used to assess meningococcal vaccination rates among patients with complement component deficiencies. The process used to identify complement component deficiencies did not appear to be validated, and it was challenging to assess the representativeness of the study population.
47. Vietri J, Harnett J, Emir B, et al. Uptake of 13-valent pneumococcal conjugate vaccine among US adults aged 19 to 64 years with immunocompromising conditions. *Hum Vaccin Immunother*. 2020;16(1):161–168. doi: 10.1080/21645515.2019.1632683 [PubMed: 31343949]
48. Baker MA, Nguyen M, Cole DV, et al. Post-licensure rapid immunization safety monitoring program (PRISM) data characterization. *Vaccine*. 2013;31(Suppl 10):K98–112. [PubMed: 24331080]
49. Schmidt MA, Gold R, Kurosky SK, et al. Uptake, coverage, and completion of quadrivalent human papillomavirus vaccine in the vaccine safety datalink, July 2006–June 2011. *J Adolesc Health*. 2013;53(5):637–641. doi: 10.1016/j.jadohealth.2013.08.002 [PubMed: 24138765]
50. Groom HC, Henninger ML, Smith N, et al. Influenza vaccination during pregnancy: influenza seasons 2002–2012, vaccine safety datalink. *Am J Prev Med*. 2016;50(4):480–488. doi: 10.1016/j.amepre.2015.08.017 [PubMed: 26526159]

51. Groom HC, Irving SA, Koppolu P, et al. Uptake and safety of hepatitis B vaccination during pregnancy: a vaccine safety datalink study. *Vaccine*. 2018;36(41):6111–6116. doi: 10.1016/j.vaccine.2018.08.074 [PubMed: 30194002]
52. Groom HC, Smith N, Irving SA, et al. Uptake and safety of hepatitis a vaccination during pregnancy: a vaccine safety datalink study. *Vaccine*. 2019;37(44):6648–6655. [PubMed: 31548013]
53. Kharbanda EO, Vazquez-Benitez G, Lipkind H, et al. Receipt of pertussis vaccine during pregnancy across 7 vaccine safety datalink sites. *Prev Med*. 2014;67:316–319. doi: 10.1016/j.ypmed.2014.05.025 [PubMed: 24952094]
54. Razzaghi H, Meghani M, Pingali C, et al. COVID-19 vaccination coverage among pregnant women during pregnancy — eight integrated health care organizations, United States, December 14, 2020–May 8, 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(24):895–899. doi: 10.15585/mmwr.mm7024e2 [PubMed: 34138834] •• Based upon a validated algorithm, electronic health record data were used to identify pregnant women. Data from electronic health records, health insurance claims, and jurisdictional immunization information systems were used to assess coverage with COVID-19 vaccines during pregnancy.
55. Naleway AL, Gold R, Kurosky S, et al. Identifying pregnancy episodes, outcomes, and mother-infant pairs in the vaccine safety datalink. *Vaccine*. 2013;31(27):2898–2903. [PubMed: 23639917]
56. Naleway AL, Crane B, Irving SA, et al. Vaccine safety datalink infrastructure enhancements for evaluating the safety of maternal vaccination. *Ther Adv Drug Saf*. 2021;12:20420986211021233. [PubMed: 34178302]
57. Centers for Disease Control and Prevention. Flu vaccination coverage, United States, 2019–20 influenza season. [cited 2023 Nov 5]. Available from: <https://www.cdc.gov/flu/fluview/coverage-1920estimates.htm>
58. Centers for Disease Control and Prevention. Flu vaccination coverage, United States, 2020–21 influenza season. [cited 2023 Nov 5]. Available from: <https://www.cdc.gov/flu/fluview/coverage-2021estimates.htm>
59. Kahn KE, Black CL, Ding H, et al. Influenza and Tdap vaccination coverage among pregnant women - United States, April 2018. *MMWR Morb Mortal Wkly Rep*. 2018;67(38):1055–1059. [PubMed: 30260946]
60. Ding H, Black CL, Ball S, et al. Influenza vaccination coverage among pregnant women — United States, 2016–17 influenza season. *MMWR Morb Mortal Wkly Rep*. 2017;66(38):1016–1022. doi: 10.15585/mmwr.mm6638a2 [PubMed: 28957044]
61. Daley MF, Reifler LM, Shoup JA, et al. Influenza vaccination among pregnant women: self-report compared with vaccination data from electronic health records, 2018–2020 influenza seasons. *Public Health Rep*. 2023;138(3):456–466. doi: 10.1177/00333549221099932 [PubMed: 35674233]
62. Sy LS, Liu IL, Solano Z, et al. Accuracy of influenza vaccination status in a computer-based immunization tracking system of a managed care organization. *Vaccine*. 2010;28(32):5254–5259. doi: 10.1016/j.vaccine.2010.05.061 [PubMed: 20554065]
63. Groom HC, Crane B, Naleway AL, et al. Monitoring vaccine safety using the vaccine safety datalink: assessing capacity to integrate data from immunization information systems. *Vaccine*. 2022;40(5):752–756. doi: 10.1016/j.vaccine.2021.12.048 [PubMed: 34980508]
64. Lu PJ, O'Halloran A, Williams WW, et al. Racial and ethnic disparities in vaccination coverage among adult populations in the U.S. *Am J Prev Med*. 2015;49(6 Suppl 4):S412–425. doi: 10.1016/j.amepre.2015.03.005 [PubMed: 26297451]
65. Eicheldinger C, Bonito A. More accurate racial and ethnic codes for medicare administrative data. *Health Care Financ Rev*. 2008;29(3):27–42. [PubMed: 18567241]
66. Sukumaran L, McCarthy NL, Li R, et al. Demographic characteristics of members of the vaccine safety datalink (VSD): a comparison with the United States population. *Vaccine*. 2015;33(36):4446–4450. doi: 10.1016/j.vaccine.2015.07.037 [PubMed: 26209836]
67. Centers for Disease Control and Prevention. Influenza vaccinations administered in pharmacies and physician medical offices, adults, United States. [cited 2023 Nov 26]. Available from: <https://www.cdc.gov/flu/fluview/dashboard/vaccination-administered.html>

Article highlights

- This narrative review examined the strengths and limitations of using non-survey-based data sources to assess US adult vaccination coverage.
- Multiple data sources were identified across a range of populations and vaccines.
- Important limitations exist related to generalizability, missing data, and risk for bias.

Table 1. Key attributes assessed during the study appraisal and synthesis, with a description for each score when applied to specific attributes.

Attribute	Rating	Description
Representativeness	5	Approximates population-based sampling; missing not at random is minimized
	4	Most populations represented, but missing some populations of public health importance
	3	While some populations represented, missing substantial populations of public health importance
	2	Limited populations are included
Timeliness	1	Convenience sample, not representative of any populations
	5	Within a month of vaccine administration for vaccines given annually or during an outbreak setting; other vaccines within a year
	4	Within 1–11 months of vaccine administration for vaccines given annually or during an outbreak setting; other vaccines within a year
	3	Within 12–36 months of vaccine administration
Completeness	2	Within 3–5 years of vaccine administration
	1	Not available until >5 years after vaccine administration
	5	All routinely recommended vaccines and all vaccine doses for multi-dose series, regardless of when given
	4	All routinely recommended vaccines, with some capacity for historical vaccines (before enrollment in system) through an IIS or documented based on available vaccine records
	3	For routinely recommended vaccines given annually, high likelihood of completeness for those enrolled in system but limitations beyond this
	2	Claims only; limitations for capturing vaccines administered before enrollment in system; subject to missing doses given without claims generated or due to data entry errors in claims
Accessibility for use	1	Major limitations to completeness
	5	Publicly available for free or for a small fee to cover costs of data transfer
	4	Publicly available but payment required
	3	Access requires approval of research question or data analysis plans, but process can be completed in a timely manner
Data validity	2	Access requires approval and process is long, resulting in long delays in publishing data
	1	Access only available for study investigators with no broader access to data
	5	Published evidence of vaccine record validity, demonstrating high data quality by formal assessment
	4	Published evidence of vaccine record validity, demonstrating reasonable data quality by formal assessment
	3	No formal evidence of vaccine record validity; some inferences about data validity can be made based on data source (for example, from EHR or healthcare data system)
	2	No formal evidence of vaccine record validity; misclassification is likely (for example through a claims database) and data entry errors are expected
	1	No formal evidence of vaccine record validity; based on data source, no confidence in validity, and expect data quality to be poor

Abbreviations: EHR, electronic health records; IIS, immunization information system.

Table 2. Characteristics of non-survey-based data sources used for published studies assessing adult vaccination coverage in the United States.

Data source ^{a,b}	Reference number(s)	Description	Overall size ^c	Geography ^{d,e}
All-payer claims databases	[25]	State-based; intended to capture health care claims from most payers within a state, including from Medicaid and Medicare	Dependent upon state population and proportion of payers contributing	State-based; unclear how many states have data available for research
IIS	[26,27]	Confidential computerized databases that record all vaccine doses administered by participating providers to persons living within a given geopolitical area; separate IIS by state and other jurisdictions	Dependent upon state population and proportion of providers contributing	All 50 states plus additional jurisdictions
IQVIA	[14]	Health care claims from office-based encounters, pharmacies, and hospital-based encounters; includes claims from commercial insurance plans, Medicaid, Medicare Part D, and cash-only transactions	>1 billion electronic medical claims annually	All 50 states
MarketScan	[13,28-36]	Health care claims from large employers and health plans; may include multi-state Medicaid	~180 million unique individuals	All 50 states (commercial); up to 13 states (Medicaid)
Medicaid	[37]	State-based; based on health care claims submitted to Medicaid; each state administers a separate Medicaid program with state-specific eligibility criteria	Dependent on state population and Medicaid eligibility criteria	Individual states
Medicare fee-for-service	[38-44]	National; based on health care claims submitted to Medicare; eligibility includes adults ≥65 years of age, individuals <65 years of age with permanent disabilities, and individuals < 65 with end-stage renal disease	~33 million unique individuals	All 50 states
Optum	[15,45-47]	Health care claims from a large managed care company; includes Medicare Advantage; may include EHR data	~9 million unique individuals	All 50 states and District of Columbia
PRISM	[48]	Primary function is to study vaccine safety; health care claims from three national health insurers and vaccine data from eight IIS	~107 million unique individuals	Not specified; claims data presumed to be national
VSD	[16,49-54]	Primary function is to study vaccine safety; 8–12 participating sites; EHR represent primary data source; IIS and claims data supplement EHR data	~10–12 million unique individuals	Wisconsin, Minnesota, Washington, Oregon, California, Colorado

Abbreviations: EHR, electronic health records; IIS, immunization information systems; PRISM, Post-licensure Rapid Immunization Safety Monitoring; VSD, Vaccine Safety DataLink.

^aPresented alphabetically by data source.

^bExpanded names for the data sources include: MarketScan commercial claims and encounters database; MarketScan multi-state Medicaid databases; Medicare fee-for-service administrative claims; IQVIA administrative medical and prescription claims data.

^cOverall size derived from published articles describing adult vaccination coverage; reported size varied in different publications from same data source; largest reported size noted here.

^dSome data sources may also include vaccination data from the District of Columbia, Guam, Puerto Rico, and the US Virgin Islands.

^ePublications from several data sources (all payers claims database [25], IIS [26,27], and state Medicaid [37]) did not include data from more than one state.

Table 3.
Published studies assessing adult vaccination coverage in the United States, using non-survey-based data sources.

Data source ^{a,b}	Study ^c	Reference number	Age group	Population studied ^d	Vaccines assessed
All-payer claims database	Koepke et al. (2015)	[25]	11–44 yrs	Pregnant women	Influenza, Tdap
IIS	Enger, Stokley (2007)	[26]	<20 yrs	General population	Meningococcal conjugate
	Koepke, Schauer, Davis (2017)	[27]	Not specified	Pregnant women	Influenza, Tdap
IQVIA	McLaughlin et al. (2019)	[14]	≥65 yrs	General population	Pneumococcal conjugate
MarketScan	Dunne et al. (2015)	[28]	11–26 yrs	General population	HPV
	Zhang et al. (2017)	[13]	≥50 yrs	General population	Herpes zoster
	Petigara, Zhang (2018)	[34]	19–64 yrs	Chronic health conditions	Pneumococcal polysaccharide
	Gargano et al. (2018)	[29]	11–26 yrs	General population	HPV
	Zhou et al. (2019)	[36]	15–49 yrs	Pregnant women	Tdap
	Ghaswalla et al. (2019)	[30]	15–44 yrs	Pregnant women	Tdap, influenza
	Hong et al. (2021)	[31]	15–49 yrs	Pregnant women	Influenza
	Prabhu et al. (2021)	[35]	9–26 yrs	General population	HPV
	Moll et al. (2021)	[32]	12–55 yrs	Pregnant women	Tdap, influenza, HPV, MMR
Medicaid (individual state; also linked to state IIS)	Payne et al. (2021)	[33]	0–64 yrs	Sickle cell disease	Influenza
	Housey et al. (2014)	[37]	≥18 yrs	Pregnant women	Tdap
Medicare fee-for-service	Shen et al. (2017)	[38]	All ages	End-stage renal disease	Influenza, pneumococcal polysaccharide, pneumococcal conjugate, hepatitis B
	Shen et al. (2017)	[39]	Not specified	Medicare beneficiaries; otherwise not specified	Influenza, pneumococcal conjugate
	Black et al. (2017)	[40]	≥65 yrs	General population	Pneumococcal polysaccharide, pneumococcal conjugate
	Shen et al. (2018)	[41]	≥65 yrs	General population	Influenza, pneumococcal polysaccharide, pneumococcal conjugate, herpes zoster, tetanus, hepatitis B
	Shen et al. (2018)	[42]	All ages	General population ≥65 yrs; those <65 yrs with permanent disabilities, end-stage renal disease	Influenza
	Shen et al. (2019)	[43]	≥65 yrs	General population	Influenza, pneumococcal polysaccharide, pneumococcal conjugate, herpes zoster

Data source ^{a,b}	Study ^c	Reference number	Age group	Population studied ^d	Vaccines assessed
Optum	Cho et al. (2022)	[44]	≥19 yrs	General population ≥65 yrs with permanent disabilities, end-stage renal disease	Influenza
	Vietri et al. (2020)	[47]	19–64 yrs	Immuno-compromised	Pneumococcal conjugate
	Marshall et al. (2021)	[46]	≥2 yrs	Complement component deficiencies	MCV4, MenB
	Ghaswalla et al. (2021)	[45]	≥2 yrs	Asplenia	Meningococcal (ACWY, MenB)
	Li et al. (2022)	[15]	≥18 yrs	General population	Influenza
PRISM	Baker et al. (2013)	[48]	All ages	General population	DTaP, hepatitis A, hepatitis B, Hib, HPV, influenza, IPV, meningococcal conjugate (ACWY), MMR, pneumococcal conjugate, pneumococcal polysaccharide, rotavirus, Td, Tdap, varicella, zoster
VSD	Schmidt et al. (2013)	[49]	9–26 yrs	General population	HPV
	Kharbanda et al. (2014)	[53]	14–49 yrs	Pregnant women	Tdap
	Groom et al. (2016)	[50]	Mean age 29.9 yrs (range not specified)	Pregnant women	Influenza
	Groom et al. (2018)	[51]	12–55 yrs	Pregnant women	Hepatitis B
	Groom et al. (2019)	[52]	12–55 yrs	Pregnant women	Hepatitis A
	Razzaghi et al. (2021)	[54]	18–49 yrs	Pregnant women	COVID-19
	Pingali et al. (2021)	[16]	≥16 yrs	General population	COVID-19

Abbreviations: IIS, immunization information systems; PRISM, Post-licensure Rapid Immunization Safety Monitoring; VSD, Vaccine Safety Datalink; DTaP, diphtheria and tetanus toxoids and acellular pertussis; Hib, *Haemophilus influenzae* type b; HPV, human papillomavirus; IPV, inactivated poliovirus; MCV4, quadrivalent meningococcal conjugate; MenB, serogroup B meningococcal conjugate; MMR, measles, mumps, rubella; Tdap, tetanus and reduced diphtheria and acellular pertussis; Td, tetanus and diphtheria toxoid.

^aPresented alphabetically by data source.

^bPublications from several data sources (all payers claims database [25], IIS [26,27], and state Medicaid [37]) did not include data from more than one state.

^cWithin data source, presented chronologically.

^dThe term general population refers to a population not selected based on high-risk criteria other than age.

Table 4. Rating of attributes of non-survey-based data sources for assessing adult vaccination coverage, on Likert-type scale of 1 (lowest) to 5 (highest) rating.

Data source ^{a,b}	Representa-tiveness	Timeliness	Complete-ness	Data access-ibility	Data validity
All-payer claims databases	3	3	2	5	2
IIS	3	5	3	Unclear	3
IQVIA	4	3	3	4	2
MarketScan	3	3	2	4	2
Medicaid (individual state)	2	3	3	5	2
Medicare fee-for-service	4	3	2	5	2
Optum	3	3	3	3	3
PRISM	3	4	4	Unclear	3
VSD	3	5	4	2	3

Abbreviations: IIS, immunization information systems; PRISM, Post-licensure Rapid Immunization Safety Monitoring; VSD, Vaccine Safety Datalink.

^aPresented alphabetically by data source.

^bPublications from several data sources (all payers claims database [25], IIS [26,27], and state Medicaid [37]) did not include data from more than one state.

Table 5. Utility of non-survey-based data sources for assessing adult vaccination coverage in four types of use.

Data source ^a	Use cases				Estimate coverage by relevant sociodemographic characteristics
	Estimate coverage by presence of chronic medical conditions or by health status (e.g. pregnancy)	Assess coverage for multiple single-dose vaccines simultaneously (e.g. Tdap and influenza)	Assess coverage for a multi-dose vaccine series (e.g. HPV)		
All-payer claims databases	Yes, with important limitations	Yes	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations
IIS	No	Yes	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations
IQVIA	Yes, with important limitations	Yes	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations
MarketScan	Yes, with important limitations	Yes	Yes, with important limitations	No	No
Medicaid (individual state)	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations	Yes	Yes
Medicare fee-for-service	Yes, with important limitations	Yes	Yes, with important limitations	Yes	Yes
Optum	Yes	Yes	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations
PRISM	Yes, with important limitations	Yes	Yes, with important limitations	No	No
VSD	Yes	Yes	Yes, with important limitations	Yes, with important limitations	Yes, with important limitations

Abbreviations: IIS, immunization information systems; PRISM, Post-licensure Rapid Immunization Safety Monitoring; VSD, Vaccine Safety Datalink; HPV, human papillomavirus; Tdap, tetanus and diphtheria toxoid and acellular pertussis.

^aPresented alphabetically by data source.