







Article

Self-Reported SARS-CoV-2 Vaccination Is Consistent with Electronic Health Record Data among the COVID-19 Community Research Partnership

Ashley H. Tjaden ^{1,*}, Lida M. Fette ¹, Sharon L. Edelstein ¹, Michael Gibbs ², Amy N. Hinkelman ³, Michael Runyon ², Roberto P. Santos ⁴, William S. Weintraub ⁵, Joshua Yukich ⁶, Diane Uschner ¹ and the COVID-19 Community Research Partnership Study Group [†]

- ¹ The Biostatistics Center, George Washington University, Rockville, MD 20852, USA; lbecerra@bsc.gwu.edu (L.M.F.); sharone@bsc.gwu.edu (S.L.E.); duschner@bsc.gwu.edu (D.U.)
- ² Department of Emergency Medicine, Atrium Health, Charlotte, NC 28204, USA; michael.gibbs@atriumhealth.org (M.G.); michael.runyon@atriumhealth.org (M.R.)
- ³ School of Osteopathic Medicine, Campbell University, Lillington, NC 27546, USA; hinkelman@campbell.edu
- ⁴ Pediatric Infectious Diseases, University of Mississippi Medical Center, Jackson, MS 39216, USA; rsantos@umc.edu
- ⁵ Population Health Research, MedStar Health Research Institute, Georgetown University, Washington, DC 20057, USA; william.s.weintraub@medstar.net
- ⁶ School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70118, USA; jyukich@tulane.edu
- * Correspondence: ahogan@bsc.gwu.edu; Tel.: +1-301-881-9260 (ext. 8031)
- † Membership of the COVID-19 Community Research Partnership Study Group is provided in Appendix A.



Citation: Tjaden, A.H.; Fette, L.M.; Edelstein, S.L.; Gibbs, M.; Hinkelman, A.N.; Runyon, M.; Santos, R.P.; Weintraub, W.S.; Yukich, J.; Uschner, D.; et al. Self-Reported SARS-CoV-2 Vaccination Is Consistent with Electronic Health Record Data among the COVID-19 Community Research Partnership. *Vaccines* **2022**, *10*, 1016. <https://doi.org/10.3390/vaccines10071016>

Academic Editor: Vincenzo Baldo

Received: 24 May 2022

Accepted: 21 June 2022

Published: 24 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Introduction: Observational studies of SARS-CoV-2 vaccine effectiveness depend on accurate ascertainment of vaccination receipt, date, and product type. Self-reported vaccine data may be more readily available to and less expensive for researchers than assessing medical records. **Methods:** We surveyed adult participants in the COVID-19 Community Research Partnership who had an authenticated Electronic Health Record (EHR) (N = 41,484) concerning receipt of SARS-CoV-2 vaccination using a daily survey beginning in December 2020 and a supplemental survey in September–October 2021. We compared self-reported information to that available in the EHR for the following data points: vaccine brand, date of first dose, and number of doses using rates of agreement and Bland–Altman plots for visual assessment. Self-reported data was available immediately following vaccination (in the daily survey) and at a delayed interval (in a secondary supplemental survey). **Results:** For the date of first vaccine dose, self-reported “immediate” recall was within ± 7 days of the date reported in the “delayed” survey for 87.9% of participants. Among the 19.6% of participants with evidence of vaccination in their EHR, 95% self-reported vaccination in one of the two surveys. Self-reported dates were within ± 7 days of documented EHR vaccination for 97.6% of the “immediate” surveys and 92.0% of the “delayed” surveys. Self-reported vaccine product details matched those in the EHR for over 98% of participants for both “immediate” and “delayed” surveys. **Conclusions:** Self-reported dates and product details for COVID-19 vaccination can be a good surrogate when medical records are unavailable in large observational studies. A secondary confirmation of dates for a subset of participants with EHR data will provide internal validity.

Keywords: SARS-CoV-2; vaccination; epidemiology

1. Introduction

Observational studies of SARS-CoV-2 vaccine effectiveness (VE) and time to seroconversion or seroreversion depend on accurate ascertainment of vaccination receipt, date, and product type. While medical records of vaccination are typically considered the gold

standard, self-reported vaccine data may be more readily available and less expensive within the context of an observational study.

In their guidance on how to assess COVID-19 VE using observational studies, the WHO advised against the use of self-reported COVID-19 vaccination data as the sole source indicating whether a person is vaccinated for primary analyses, due to recall bias and lack of product details [1]. Self-reported dates are subject to recall bias, but Electronic Health Record (EHR) data may contain vaccination status only on a subset of participants and may also be erroneous [2,3]. In the United States, COVID-19 vaccines have been made available in both medical and non-medical settings, including hospitals, pharmacies, public health departments, and mass vaccination sites, and the information may not be reliably recorded in a participant's EHR; thus, self-reported data may be the only source for obtaining real-time vaccination information in a disease surveillance study.

Researchers have assessed the validity of self-reported influenza vaccine status in various populations and have found self-reported influenza vaccine status to be a very sensitive and moderately specific indicator of actual vaccine status [4,5]. Other studies have found slightly lower, yet still high, sensitivity and specificity for pneumococcal vaccine status in adults [5–8]. Influenza vaccination agreement is higher when self-reporting on the current influenza season compared to the past season suggesting that there may be a difference in immediate and delayed self-reporting of vaccination [9]. Accurate self-report of other vaccinations, including hepatitis A, may be lower and the accuracy of self-report varies by race/ethnicity and age [4,10].

Previous research has focused on agreement of vaccine receipt but not on date of vaccine receipt or product [4–10]. To our knowledge, studies on the accuracy of self-reported COVID-19 vaccination receipt, date, number of doses, or product type are not yet publicly available.

Study Aims

The goal of this analysis was to assess the level of agreement of self-reported vaccination receipt, number of doses, date(s) received, and vaccine product from two separate surveys (immediate and delayed recall) with vaccination information from participant EHRs. We hypothesized strong agreement between the two based on previous research of self-reported vaccination information for other vaccines.

2. Methods

2.1. Study Sample

The COVID-19 Community Research Partnership (CCRP) is a prospective, multi-site cohort syndromic COVID-19 surveillance study of a convenience sample of adults (18+ years) enrolled from April 2020 through June 2021 primarily through direct email outreach at 10 healthcare systems from the mid-Atlantic and southeastern United States (<http://www.covid19communitystudy.org/>, 14 March 2022) [11]. Data were collected via a secure, Health Insurance Portability and Accountability Act of 1996 (HIPAA)-compliant, online platform through March 2022. All participants provided informed consent, and Institutional Review Board (IRB) approval was provided by the Wake Forest School of Medicine IRB. Study sites included: Atrium Health (Charlotte, NC, USA), Campbell University School of Osteopathic Medicine (Lillington, NC, USA), Medstar Health (Columbia, MD, USA), New Hanover Regional Medical Center (Wilmington, NC, USA), Tulane University (New Orleans, LA, USA), University of Maryland, Baltimore (Baltimore, MD, USA), University of Mississippi Medical Center (Jackson, MS, USA), Vidant Health (Greenville, NC, USA), Wake Forest Baptist Health (Winston-Salem, NC, USA), and WakeMed Health and Hospitals (Raleigh, NC, USA). For a subset of participants, EHR data were available.

2.2. Ascertainment of Vaccination Date

At baseline, participants self-reported demographic characteristics including age, sex, race, ethnicity, educational level, healthcare worker status, and county of residence.

We classified counties of residence as urban, rural, or suburban based on population density estimates. COVID-19 symptoms and personal behaviors were self-reported in daily electronic surveillance surveys. Beginning in December 2020, a question was added to the daily electronic surveillance survey asking participants “Have you received a vaccine for COVID-19, since the last survey?” Those responding in the affirmative were asked to enter the vaccination date, the dose number (dose 1 or dose 2), and the vaccine brand (Box A1). For the first month of the daily vaccine survey question, the date of receipt was not included; thus, we used the date the survey was completed as a surrogate for date of receipt. Additionally, we developed an algorithm to identify data entry errors in the daily survey (e.g., year < 2020) and again used the date the survey was entered as the vaccine date in these cases. A secondary supplemental survey (Box A2) was sent to participants in September–October 2021 to ask whether they had received a COVID-19 vaccine, the date of each dose, the dose number (allowing more than two doses), and the vaccine brand. The daily survey was used to evaluate self-reported, immediate recall while the secondary supplemental survey was used for self-reported, delayed recall. In addition, a subset of participants had a record of COVID-19 vaccination in their EHR.

2.3. Participants

Of the 66,403 adult participants enrolled in the CCRP, 46,228 (69.6%) had a data entry on or after 14 December 2020 (first date of COVID-19 vaccine availability in the U.S.). 41,484 adult participants (18+ years) enrolled in the CCRP with COVID-19 vaccination data (receipt, date, and brand) provided on one or more vaccine doses from at least one source (i.e., authenticated EHR, immediate recall survey response, or delayed recall survey response) were included in the current analysis (Figure 1). For comparisons of sources, participants with at least two sources of vaccination data (immediate recall self-report, delayed recall self-report, or EHR data) were included in the comparison of vaccination information. Participants missing the date of dose 1 (i.e., only have an entry labeled as dose 2) on the daily survey or with implausible dates/dates with obvious typos (e.g., dates prior to 1 December 2020 for participants not in a clinical trial and dates after the date of entry) on the supplemental survey were excluded.

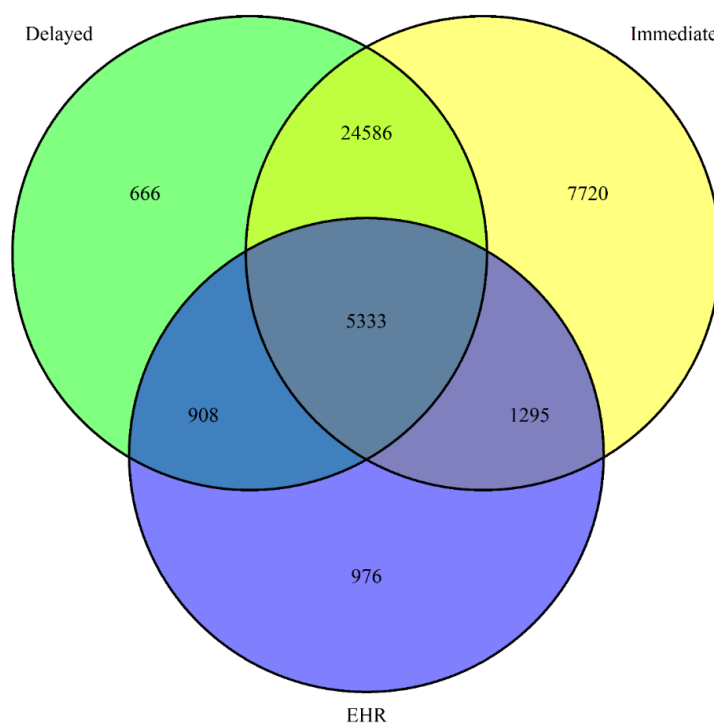


Figure 1. Venn diagram depicting number of participants with COVID-19 vaccine information available from each source.

Included in the primary analysis are 32,122 participants (Table 1) with at least two sources of vaccine information for comparison. In total, 29,919 participants are included in the comparison of the two self-reported sources (immediate and delayed recall); 6628 participants are included in the comparison of immediate recall self-report and EHR vaccine information; and 6241 participants are included in the comparison of delayed recall self-report and EHR vaccine information (Table 2).

Table 1. Characteristics of participants.

| Characteristic | N (%) |
|-----------------------------------|----------------|
| Number of participants | 32,122 |
| Age (years) | 54.2 ± 14.5 |
| Sex | |
| Female | 22,300 (69.4%) |
| Male | 9819 (30.6%) |
| Unknown | 3 (0.0%) |
| Healthcare Worker Occupation | |
| No | 24,493 (76.2%) |
| Yes | 7629 (23.8%) |
| Race/Ethnicity | |
| American Indian or Alaskan Native | 54 (0.2%) |
| Asian or Pacific Islander | 634 (2.0%) |
| Black or African American | 1888 (5.9%) |
| Hispanic or Latino | 758 (2.4%) |
| Mixed Ethnicity | 426 (1.3%) |
| Not Specified | 238 (0.7%) |
| White (not Hispanic or Latino) | 28,124 (87.6%) |
| Region | |
| Deep South | 365 (1.1%) |
| Mid-Atlantic | 11,541 (36.0%) |
| South East | 20,216 (62.9%) |
| County Classification | |
| Rural | 6323 (19.7%) |
| Suburban | 7436 (23.1%) |
| Urban | 18,362 (57.2%) |

2.4. Statistical Analysis

We compared vaccine receipt, number of doses, date of the first dose, and vaccine brand received from self-report (both immediate and delayed recall) to each other and to those available in the EHR. We calculated agreement of receipt, number of doses and vaccine brand (proportion of participants with matching report). We visually assessed Bland–Altman plots comparing the date of dose 1 (as number of days from 14 December 2020) from the various sources of vaccination dates, representing the mean difference and the 90% agreement interval around the mean difference. Lastly, we compared the presence of vaccine receipt in EHR compared with self-report among all participants with information about receipt and date of at least one dose of a COVID-19 vaccine from at least one source (N = 41,488). R version 4.0.13 was used for all statistical analyses.

Table 2. Comparisons of vaccine information from multiple data sources (N = 32,122).

| | Self-Report | Self-Report vs. EHR | |
|---------------------------------|------------------------------|---------------------|----------------|
| | Immediate vs. Delayed Recall | Immediate Recall | Delayed Recall |
| N | 29,919 | 6628 | 6241 |
| N Excluded | 1653 | 567 | 1072 |
| Reason for Exclusion | | | |
| Possible Booster | 295 * | 567 ± | 729 ± |
| Same Date as Entry | 1358 † | - | 343 † |
| N for Comparison | 28,266 | 6061 | 5169 |
| Number of Doses | | | |
| Match | 26,684 (94.4%) | 5742 (94.7%) | 5050 (97.7%) |
| Mismatch | 1582 (5.6%) | 319 (5.3%) | 119 (2.3%) |
| Dose 1 Date Absolute Difference | | | |
| Median [Q1, Q3] | 0.0 [0.0, 1.0] | 0.0 [0.0, 1.0] | 0.0 [0.0, 1.0] |
| Mean ± SD | 5.7 ± 28.0 | 1.5 ± 10.6 | 2.5 ± 14.8 |
| Dose 1 Date Difference | | | |
| Exact Match | 16,850 (59.6%) | 3198 (52.8%) | 3510 (67.9%) |
| ±3 Days | 5812 (20.6%) | 2679 (44.2%) | 990 (19.2%) |
| ±4–7 Days | 2165 (7.7%) | 39 (0.6%) | 251 (4.9%) |
| ±8–14 Days | 1383 (4.9%) | 31 (0.5%) | 190 (3.7%) |
| ±15–30 Days | 1261 (4.5%) | 87 (1.4%) | 174 (3.4%) |
| ±>30 Days | 795 (2.8%) | 27 (0.4%) | 54 (1.0%) |
| Product | | | |
| Match | 27,751 (98.2%) | 5967 (98.4%) | 5125 (99.1%) |
| Mismatch | 515 (1.8%) | 94 (1.6%) | 44 (0.9%) |

* Participants were excluded because their only dose in the daily survey (“immediate” recall) appeared to be their booster dose (e.g., it is after their delayed recall doses 1 and 2, or there is only one dose recorded in the daily/“immediate” recall survey, and the dose is reported to be received after 1 October 2021). † Delayed recall survey indicated dose 1 as the same date they took the survey likely due to an error with the date picker selecting “today’s” date by default. In most cases for these participants, the dates of doses 1, 2, and 3 were all entered as “today’s” date. ± Participants were excluded because their only vaccine date in EHR appeared to be their booster dose (e.g., does not match their self-reported dose 1 or 2, only one dose in EHR, and only EHR dose is after 1 October 2021).

3. Results

3.1. Comparison of Two Sources of Self-Reported Vaccine Information: Immediate vs. Delayed Recall

Prior to comparing the vaccine information, using Bland–Altman plot analysis, we identified 1653 participants to exclude from further analysis (Table 2 and Figure 2A). For most participants, the difference between delayed and immediate report of vaccine information is low, indicated by the fact that most points are included in the agreement interval around zero (green lines). Two point clouds outside the agreement interval at ~100 and ~200 days mean difference in days after 14 December 2020 reflect disagreement, potentially due to participants erroneously entering their booster date as their first dose, or due to entering the day of taking the survey as their date of vaccination. After excluding the flagged participants, 28,266 were available for comparison (Table 2). The mean number of days between the initial survey (immediate recall) and supplemental survey (delayed recall) was 205.0 ± 42.6 days (mean ± SD). The number of doses self-reported immediately and self-reported delayed matched for 94.4% (N = 26,684). Self-reported immediate recall date of first dose of vaccine was within ±7 days of the delayed recall date for 87.9% of

participants (Table 2 and Figure 2B). As depicted in Figure 1B, a majority of participants were within the 90% agreement interval around the mean difference (green lines), but there are groupings of participants with large differences where one dose may be mislabeled as a second or third dose on a subsequent survey. Self-reported vaccine product (for the first dose recorded) from the delayed recall survey matches the immediate recall survey for 98.2% of participants.

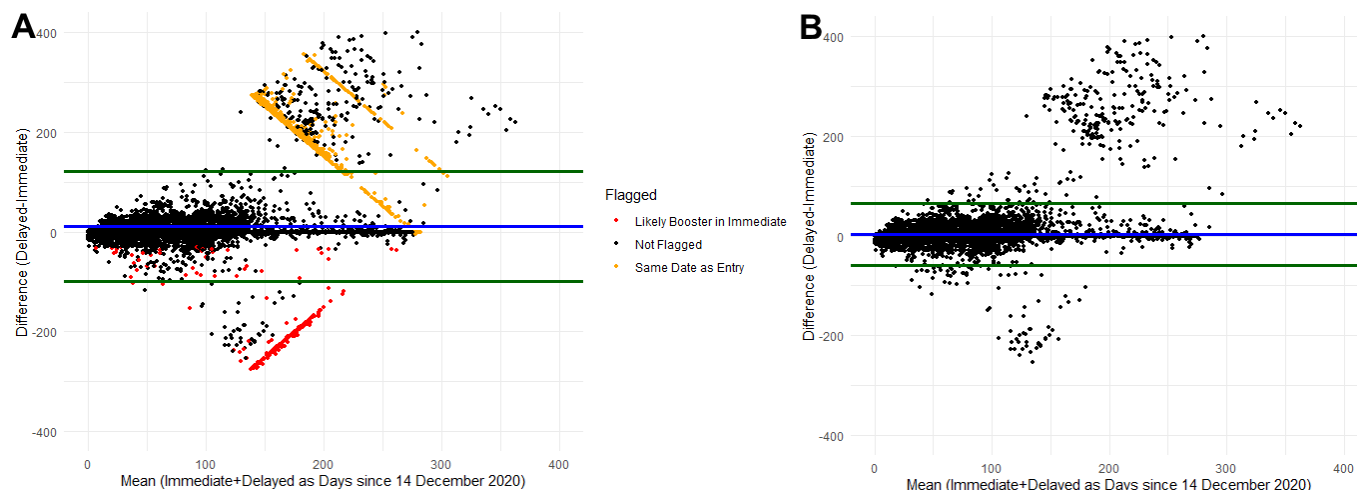


Figure 2. Bland–Altman plot comparing immediate self-report and delayed self-report of COVID-19 vaccination. Blue line represents the mean difference and green lines represent the 90% agreement interval around the mean difference. (A) represents data prior to cleaning identifying likely entry errors. (B) represents the data after removing participants with identified errors.

3.2. Comparison of Self-Reported Immediate Recall to EHR Vaccine Information

Prior to comparing the vaccine information, we flagged 567 participants (Table 2 and Figure 3A). After excluding the flagged participants, 6061 were available for comparison (Table 2). The number of doses self-reported and recorded in EHR matched for 94.7%, and among participants who had two doses recorded in their EHR ($N = 5826$), 96.5% ($N = 5621$) participants also self-reported two doses. Self-reported date of first dose of vaccine was within ± 7 days of the EHR date for 97.6% of participants (Table 2 and Figure 3B). As depicted in Figure 3B, after excluding the flagged participants, nearly all participants are within the 90% agreement interval around the mean difference (green lines) suggesting that the difference between self-reported date of the first dose and the first date listed in EHR is low. Self-reported vaccine product (for the first dose recorded) from the delayed recall survey matched the EHR product for 98.4% of participants. A majority of participants self-reported receiving the Pfizer (BNT16262) vaccine product (87%), 10% reported receiving Moderna (mRNA-1273), and then the remainder reported another brand.

3.3. Comparison of Self-Reported Delayed Recall to EHR Vaccine Information

Prior to comparing the vaccine information, we flagged 1072 participants (Table 2 and Figure 4A). After excluding the flagged participants, 5169 were available for comparison (Table 2). The number of doses self-reported and recorded in EHR matched for 97.7% of participants and among participants who had two doses recorded in their EHR ($N = 5018$), 99.8% ($N = 5006$) of participants also self-reported two doses. Self-reported date of first dose of vaccine was within ± 7 days of the EHR date for 92% of participants (Table 2 and Figure 4B). As illustrated in Figure 4B, after excluding the flagged participants, the majority of participants are within the 90% agreement interval around the mean difference (green lines) with a subset of participants in the top-center that may have mislabeled doses and participants in the bottom-center that have dates similar to the date the supplemental survey was released. Self-reported vaccine product (for the first dose recorded) from the delayed recall survey matched the EHR product for 99.1% of participants. Similar to

the immediate recall, a majority of participants self-reported receiving the Pfizer vaccine product (87%), 11% reported receiving Moderna, and the remainder reported receiving another brand.

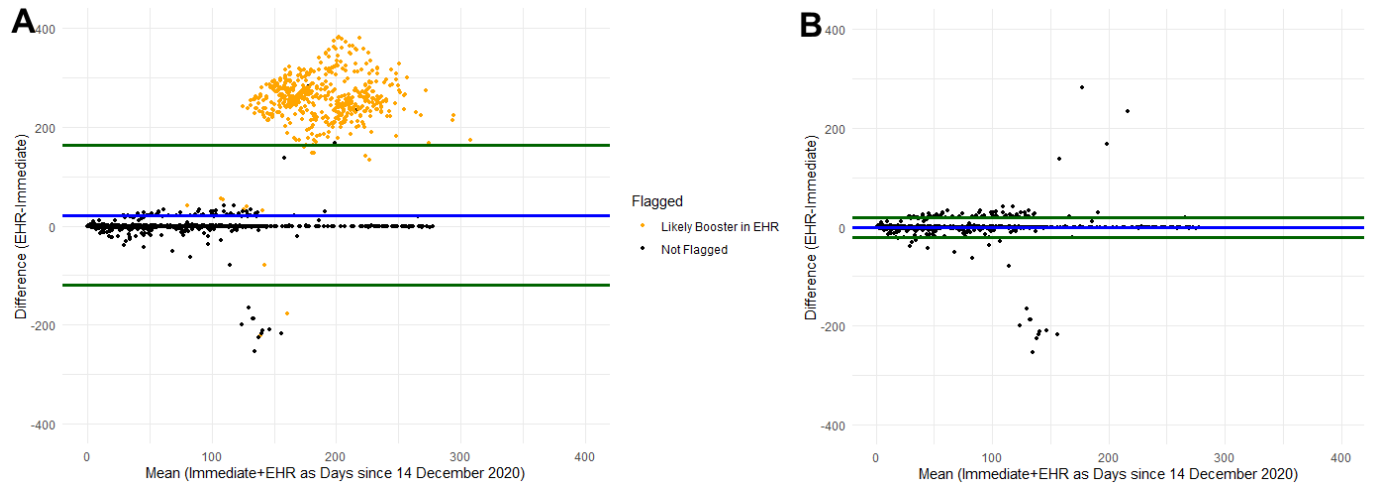


Figure 3. Bland–Altman plot comparing immediate self-report and EHR date of dose 1 for COVID-19 vaccination. Blue line represents the mean difference and green lines represent the 90% agreement interval around the mean difference. (A) represents data prior to cleaning identified likely errors. (B) represents the data after removing participants with identified errors.

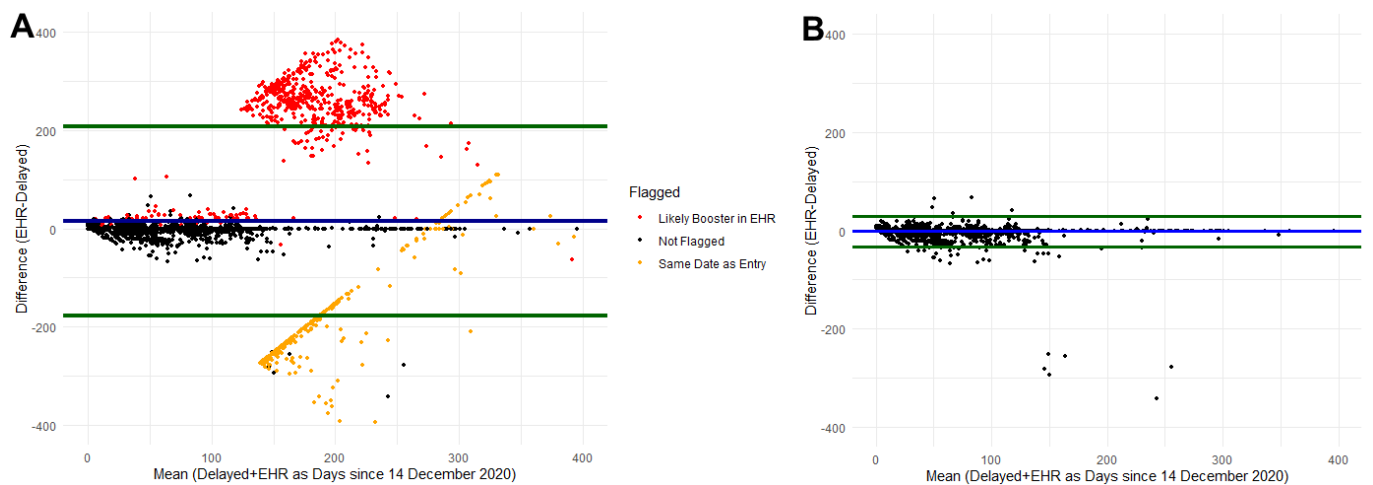


Figure 4. Bland–Altman plot comparing delayed self-report and EHR date of dose 1 for COVID-19 vaccination. Blue line represents the mean difference and green lines represent the 90% agreement interval around the mean difference. (A) represents data prior to cleaning identified likely errors. (B) represents the data after removing participants with identified errors.

3.4. Comparison of Vaccine Receipt

As a secondary analysis, we compared the presence of vaccine receipt in the EHR with self-report among all participants with information about receipt and date of at least one dose of a COVID-19 vaccine from at least one source ($N = 41,484$). Among participants with at least one survey entry after 1 March 2021 when vaccines were widely available ($N = 39,735$), 19.6% ($N = 7786$) of participants had evidence of vaccination in their EHR, and 94.9% of them self-reported vaccination in one of the two surveys. Among participants with at least one survey entry after the supplemental survey was sent on 16 September ($N = 32,103$), 19.6% of participants ($N = 6303$) had evidence of vaccination in their EHR and 99% of them self-reported vaccination in the supplemental survey (Appendix A Table A1).

4. Discussion

In the CCRP cohort, we found strong agreement between self-reported and EHR information pertaining to receipt of the SARS-CoV-2 vaccinations including number of doses, date of receipt of the first dose, and product details. In our cohort, there was high agreement between immediate and delayed recall self-reported vaccination information in terms of number of doses reported (94% agreement), vaccine product (98% agreement), and date of first vaccine dose (88% within ± 7 days). While number of doses and vaccine product had similar agreement with EHR vaccination information for the immediate and delayed recall self-report, the immediate recall date of dose 1 was more closely aligned with the date found in EHR, with 98% of participants within ± 7 days for immediate recall compared to 92% for delayed recall. This difference is likely explained by the timing between vaccination and report. On average, 205 days passed between the daily survey (immediate recall) and supplemental survey (delayed recall). While the coverage of SARS-CoV-2 vaccination data in participant electronic medical records was only available for about a fifth of participants with linked EHR, nearly all of them (95%) also self-reported vaccination. Among participants with data in two sources, comparison of the details (number of doses, dates, and product) revealed the strength of self-reported COVID-19 vaccination data.

To our knowledge, this is the first study to evaluate the accuracy of self-reported COVID-19 vaccination information. Researchers have shown that in test-negative studies of VE, self-reported influenza vaccination results in biased estimates, suggesting the importance of validating self-reported dates [12,13]. However, they also found that the greatest impact for influenza VE estimation occurred when vaccination coverage was low. About half of Americans typically receive an influenza vaccine annually [14] compared with over 80% of Americans having received at least one dose of a SARS-CoV-2 vaccine [15]. This difference in vaccine coverage (50% for influenza vs. 80% for COVID-19) taken together with our findings of high agreement of self-report and EHR vaccine information suggests that the use of self-reported SARS-CoV-2 vaccination may have less of an impact on biasing VE estimates than observed with influenza.

While EHR vaccination information may be considered the gold standard, we note that the lack of documentation of vaccination in EHR likely does not indicate that no vaccination had been obtained. In the CCRP, only a subset of our participants had documentation of SARS-CoV-2 vaccination in their medical records. Moreover, vaccination dates in EHR do not guarantee delivery of the vaccine by the healthcare system and may be data entry from a patient's vaccination card (including the possibility of forged vaccination cards) or patient recall during subsequent routine medical visits. Data entry errors are also possible in medical records.

The COVID-19 vaccines are unique from other vaccinations (e.g., influenza) in their distribution at largely mass vaccination sites and pharmacies (e.g., CVS, Walgreens) rather than from one's primary care provider or employer [16,17]. The rollout of the vaccines in December 2020 and early 2021 was a memorable milestone in the pandemic which may have made the date of receipt of first dose more unforgettable than an annual influenza vaccine. Additionally, the vaccine brands were widely discussed by the media, making them part of popular culture and easily recognizable [18].

The CCRP comparison of self-reported and EHR vaccination information had several strengths including the large number of diverse participants with self-reported and EHR data available for comparison. Many participants self-reported their vaccine data twice: once in real time as vaccines were being administered and once several months later for confirmation. Our study has a number of important limitations. The CCRP is a U.S. healthcare-system-based convenience sample of participants who are predominantly non-Hispanic White and about a quarter are healthcare workers. These results may not be consistent in populations of varying race/ethnicity, occupation, education level, health literacy, or region. Importantly, the results are relevant only to the U.S. population where the details of COVID-19 vaccines have been discussed extensively in the media and may

not be generalizable to other regions with different media landscapes, and vaccine attitudes and availability. Additionally, there were significant differences between participants with and without documentation of vaccination in EHR: female participants, healthcare workers, participants in the South East, and participants residing in urban areas were more likely to have vaccine information in their EHR (Appendix A Table A2). There were also differences by enrollment site. This difference in availability of EHR data highlights the importance of accepting self-reported vaccine dates particularly where EHR vaccination data may be sparse. Thus, we did not compute sensitivity or specificity of self-reported receipt of vaccination. Further, while EHR may be convenient and inexpensive in some cases, the observed limited coverage of COVID-19 vaccination in the EHR makes its utility limited. Additionally, due to the structure of the vaccination data we received from EHR (e.g., unnumbered doses), it was difficult to identify the dose number especially when one or more doses were likely missing. The self-reported dates for CCRP were also flawed. Initially, we did not ask the date of receipt for vaccination, and later the data entry system allowed the date field to be left blank. Insufficient data checks were in place to prevent the entry of nonsensical dates (e.g., typos, multiple doses labeled dose 1, dose 2 occurring before dose 1, etc.). For the immediate recall survey, we used the date the daily survey was entered as a surrogate for the date of receipt. For the supplemental survey (delayed recall), the date picker that allowed users to enter a date either through text input, or by choosing a date from the calendar defaulted to “today’s” date which led to issues with data quality. For the CCRP analyses, we took advantage of having three sources of vaccination data and compared the three to identify the dates most likely to be correct for each participant, taking into account agreement between more than one source, timing between doses, and timing of entry of immediate recall survey. Lastly, while vaccine product agreement was high, the majority of participants received the Pfizer vaccine increasing the chance for random agreement.

5. Conclusions

Our findings suggest that self-reported vaccination information regarding dates of administration and product details can be an effective surrogate when medical records are unavailable in large observational studies. There are limitations to using medical records for COVID-19 vaccination as they do not always note if vaccination actually occurred; thus, self-reported dates may provide more data and be the simplest option for obtaining vaccination status and details. It is important to have robust data entry rules in place to avoid errors and when possible, to collect data in real time about vaccine receipt as the self-reported data may be of higher quality than delayed recall. A secondary confirmation of vaccine data on a subset of participants using EHR will provide evidence of validity.

Author Contributions: Conceptualization and methodology A.H.T., L.M.F., S.L.E., D.U.; methodology, A.H.T., L.M.F., S.L.E., D.U.; formal analysis, A.H.T.; data curation, A.H.T., L.M.F., S.L.E., M.G., A.N.H., M.R., R.P.S., W.S.W., J.Y., D.U.; writing—original draft preparation, A.H.T., L.M.F., S.L.E., D.U.; writing—review and editing, All authors. All authors have read and agreed to the published version of the manuscript.

Funding: This publication was supported by the Centers for Disease Control and Prevention (CDC) [contract #75D30120C08405] and the CARES Act of the U.S. Department of Health and Human Services (HHS) [Contract # NC DHHS GTS #49927]. Fifty percent of the current project was funded by the CDC/HHS award and 50% by the CARES Act/HHS award. The Partnership is listed in [clinicaltrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT04342884) (NCT04342884).

Financial Disclosure: No financial disclosures were reported by the authors of this paper.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and the study was reviewed and approved by the Wake Forest Institutional Review Board (IRB), which served as the central IRB for this study (See 45 C.F.R. part 46; 21 C.F.R. part 56). The study is registered with [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT04342884), NCT04342884.

Informed Consent Statement: All participants in the COVID-19 Community Research Partnership provided consent for participation.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Data Sharing: Results of the COVID-19 CRP are being disseminated on the study website (<https://www.covid19communitystudy.org/>, 14 March 2022) as well as in publications and presentations in medical journals and at scientific meetings. At end of the study, the databases will be made publicly available in a de-identified manner according to CDC and applicable U.S. Federal policies.

Acknowledgments: The COVID-19 Community Research Partnership gratefully acknowledges the commitment and dedication of the study participants. Programmatic, laboratory, and technical support was provided by Vysnova Partners, Inc., Javara, Inc., Oracle Corporation, LabCorp, Scanwell Health, and Neoteryx. This publication was supported by the Centers for Disease Control and Prevention (CDC) [contract #75D30120C08405] and the CARES Act of the U.S. Department of Health and Human Services (HHS) [Contract # NC DHHS GTS #49927]. Fifty percent of the current project was funded by the CDC/HHS award and 50% by the CARES Act/HHS award. The Partnership is listed in clinicaltrials.gov (NCT04342884). 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, HHS, or the U.S. Government. A complete list of Study Sites, investigators, and staff can be found in Appendix A.

Conflicts of Interest: No conflict of interest were reported by the authors of this paper.

Appendix A

Table A1. Comparing vaccine self-reported vaccine receipt by EHR vaccine receipt (among participants with information about receipt and date of at least one dose of a COVID-19 vaccine from at least one source, N = 41,488).

| <i>Among Participants with at Least One Survey Entry after 1 March 2021 (N = 39,735)</i> | | |
|---|--------------|-----------------|
| | EHR + | EHR – |
| N | 7786 | 31,949 |
| Immediate Recall | | |
| Self-Report + | 6479 (83.2%) | 31,302 (98.0%) |
| Self-Report – | 1307 (16.8%) | 647 (2.0%) |
| Delayed Recall | | |
| Self-Report + | 6236 (80.1%) | 25,171 (78.8%) |
| Self-Report – | 1550 (19.9%) | 6778 (21.2%) |
| Either Self-Report | | |
| Self-Report + | 7386 (94.9%) | 31,949 (100.0%) |
| Self-Report – | 400 (5.1%) | 0 (0.0%) |
| <i>Among participants with at least one survey entry after 16 September 2021 (N = 32,064)</i> | | |
| | EHR + | EHR – |
| N | 6303 | 25,800 |
| Delayed Recall | | |
| Self-Report + | 6240 (99.0%) | 25,214 (97.7%) |
| Self-Report – | 63 (1.0%) | 586 (2.3%) |

+: Indicates presence of vaccine information; -: Indicates absence of vaccine information.

Table A2. Comparing characteristics of participants with vaccine information in EHR (among participants with information about receipt and date of at least one dose of a COVID-19 vaccine in at least one source, N = 41,484).

| | Without Vaccine in EHR N = 32,972 | With Vaccine in EHR N = 8512 |
|--------------------------------|--------------------------------------|---------------------------------|
| Site | | |
| Atrium | 2689 (8.2%) | 4232 (49.7%) |
| Campbell | 447 (1.4%) | 0 (0.0%) |
| Maryland | 4683 (14.2%) | 0 (0.0%) |
| MedStar | 8551 (25.9%) | 1292 (15.2%) |
| Mississippi | 71 (0.2%) | 272 (3.2%) |
| New Hanover | 369 (1.1%) | 334 (3.9%) |
| Tulane | 166 (0.5%) | 0 (0.0%) |
| Vidant | 752 (2.3%) | 0 (0.0%) |
| Wake Forest | 13,404 (40.6%) | 1495 (17.6%) |
| Wake Med | 1842 (5.6%) | 887 (10.4%) |
| Age (years) | 53.2 ± 14.8 | 51.4 ± 15.1 |
| Sex | | |
| Female | 22,582 (68.5%) | 6179 (72.6%) |
| Male | 10,385 (31.5%) | 2333 (27.4%) |
| Unspecified | 5 (0.0%) | 0 (0.0%) |
| Healthcare Worker | | |
| No | 27,001 (81.9%) | 4433 (52.1%) |
| Yes | 5971 (18.1%) | 4079 (47.9%) |
| Race/Ethnicity | | |
| American Indian/Alaskan Native | 98 (0.3%) | 11 (0.1%) |
| Asian/Pac Islander | 647 (2.0%) | 244 (2.9%) |
| Black/African Am | 2004 (6.1%) | 555 (6.5%) |
| Hispanic/Latino | 834 (2.5%) | 218 (2.6%) |
| Mixed Ethnicity | 445 (1.3%) | 132 (1.6%) |
| Not Specified | 308 (0.9%) | 62 (0.7%) |
| Non-Hispanic White | 28,636 (86.8%) | 7290 (85.6%) |
| Region of Enrollment | | |
| Deep South | 237 (0.7%) | 272 (3.2%) |
| North Atlantic | 13,234 (40.1%) | 1292 (15.2%) |
| South East | 19,501 (59.1%) | 6948 (81.6%) |
| County of Residence | | |
| Rural | 7218 (21.9%) | 1072 (12.6%) |
| Suburban | 7338 (22.2%) | 2093 (24.6%) |
| Urban | 18,415 (55.9%) | 5347 (62.8%) |

$p < 0.001$ for all comparisons. p -values for Pearson's Chi-squared test for categorical variables and Welch's two-sample t -test for normally distributed continuous variables.

Box A1. Vaccine Questions from Immediate Recall.**Daily Questionnaire (Immediate Recall) for Adult participants.**

Please answer the following questions based on what you are experiencing between your last update and now.

- Have you received a vaccine for COVID-19, since your last update?
 - Yes
 - No
 - If YES:
 - Was this part of a clinical trial?
 - Yes
 - No
 - I don't know
 - Which vaccine:
 - I do not know
 - AstraZeneca—AZD1222
 - Johnson & Johnson—Ad26.COV2.S
 - Moderna—mRNA
 - Pfizer—BNT162
 - Other
 - Which dose:
 - Dose 1
 - Dose 2
 - Vaccination date? (If Known)

Box A2. Vaccine Questions from Delayed Recall.**Supplemental Questionnaire (Delayed Recall) for Adult participants**

Questions added to daily survey to replace existing questions:

1. How many doses of the COVID-19 Vaccine have you received in total? *
 - Zero
 - One
 - Two
 - Three
 - Four or more
 - Please enter total number of COVID-19 vaccine doses you have received. [text box]

If a person chooses this is my first dose, two, three, or four or more ask question
2. Tell us about your **first** COVID-19 vaccine dose.
Was this part of a clinical trial? (If Yes, end of questions)
 - Yes
 - No
 - I don't know
3. What was your first dose of the COVID-19 Vaccine? *
 - Moderna
 - Janssen/Johnson & Johnson
 - Pfizer-BioNTech
 - Other/I don't know
4. When did you receive your first COVID-19 Vaccine?
 - Calendar Function

If a person chose two, three, or four or more in Q1 ask question

5. Tell us about your **second** COVID-19 vaccine dose.
What was your second dose of the COVID-19 Vaccine? *
- Moderna
 - Janssen/Johnson & Johnson
 - Pfizer-BioNTech
 - Other/I don't know
6. When did you receive your second COVID-19 Vaccine?
- Calendar Function

If a participant selected "three or four or more" on question 1

7. Tell us about your **third** COVID-19 vaccine dose.
What was your second dose of the COVID-19 Vaccine? *
- Moderna
 - Janssen/Johnson & Johnson
 - Pfizer-BioNTech
 - Other/I don't know
8. When did you receive your third COVID-19 Vaccine?
- Calendar Function
9. Why did you seek an additional dose of COVID-19 Vaccine [select all that applies]
- Additional protection
 - Immunocompromised
 - Greater than 8-months since I received my vaccine
 - Other
 - Text Box for additional response

If a participant selected "one or two" on question 1 on last survey

10. Since your last survey, have you had an additional COVID-19 Vaccine dose? *
- Yes
 - No
11. What was your [dose number] of the COVID-19 Vaccine? *
- Moderna
 - Janssen/Johnson & Johnson
 - Pfizer-BioNTech
 - Other/I don't know
12. When did you receive your COVID-19 Vaccine
- Calendar Function
13. Why did you seek an additional dose of COVID-19 Vaccine [select all that applies]
- Additional protection
 - Immunocompromised
 - Greater than 8-months since I received my vaccine
 - Other
 - Text Box for additional response

Once participant answers Question 1, Question 1–9 should go away. Going forward, participant should be asked Question 10.

We should bring up the next set of vaccine questions (i.e., vaccine 1, 2, or 3) based on the what the participant last completed.

After someone with J&J had at least 2 doses and someone with Moderna/Pfizer had 3 doses, question series goes away.

* Required question

*Authorship: The COVID-19 Research Group (*Site Principal Investigator)*

Wake Forest School of Medicine: Thomas F Wierzba PhD, MPH, MS*, John Walton Sanders, MD, MPH, David Herrington, MD, MHS, Mark A. Espeland, PhD, MA, John Williamson, PharmD, Morgana Mongraw-Chaffin, PhD, MPH, Alain Bertoni, MD, MPH, Martha A. Alexander-Miller, PhD, Paola Castri, MD, PhD, Allison Mathews, PhD, MA,

Iqra Munawar, MS, Austin Lyles Seals, MS, Brian Ostasiewski, Christine Ann Pittman Ballard, MPH, Metin Gurcan, PhD, MS, Alexander Ivanov, MD, Giselle Melendez Zapata, MD, Marlana Westcott, PhD, Karen Blinson, Laura Blinson, Mark Mistysyn, Donna Davis, Lynda Doomy, Perrin Henderson, MS, Alicia Jessup, Kimberly Lane, Beverly Levine, PhD, Jessica McCanless, MS, Sharon McDaniel, Kathryn Melius, MS, Christine O'Neill, Angelina Pack, RN, Ritu Rathee, RN, Scott Rushing, Jennifer Sheets, Sandra Soots, RN, Michele Wall, Samantha Wheeler, John White, Lisa Wilkerson, Rebekah Wilson, Kenneth Wilson, Deb Burcombe, Georgia Saylor, Megan Lunn, Karina Ordonez, Ashley O'Steen, MS, Leigh Wagner.

Atrium Health: Michael S. Runyon MD, MPH*, Lewis H. McCurdy MD*, Michael A. Gibbs, MD, Yhenneko J. Taylor, PhD, Lydia Calamari, MD, Hazel Tapp, PhD, Amina Ahmed, MD, Michael Brennan, DDS, Lindsay Munn, PhD RN, Keerti L. Dantuluri, MD, Timothy Hetherington, MS, Lauren C. Lu, Connell Dunn, Melanie Hogg, MS, CCRA, Andrea Price, Marina Leonidas, Melinda Manning, Whitney Rossman, MS, Frank X. Gohs, MS, Anna Harris, MPH, Jennifer S. Priem, PhD, MA, Pilar Tochiki, Nicole Wellinsky, Crystal Silva, Tom Ludden PhD, Jackeline Hernandez, MD, Kennisha Spencer, Laura McAlister.

MedStar Health Research Institute: William Weintraub MD*, Kristen Miller, DrPH, CPPS*, Chris Washington, Allison Moses, Sarahfaye Dolman, Julissa Zelaya-Portillo, John Erkus, Joseph Blumenthal, Ronald E. Romero Barrientos, Sonita Bennett, Shrenik Shah, Shrey Mathur, Christian Boxley, Paul Kolm, PhD, Ella Franklin, Naheed Ahmed, Moira Larsen.

Tulane: Richard Oberhelman MD*, Joseph Keating PhD*, Patricia Kissinger, PhD, John Schieffelin, MD, Joshua Yukich, PhD, Andrew Beron, MPH, Johanna Teigen, MPH.

University of Maryland School of Medicine: Karen Kotloff MD*, Wilbur H. Chen MD, MS*, DeAnna Friedman-Klabanoff, MD, Andrea A. Berry, MD, Helen Powell, PhD, Lynnee Roane, MS, RN, Reva Datar, MPH, Colleen Reilly.

University of Mississippi: Adolfo Correa MD, PhD*, Bhagyashri Navalkele, MD, Yuan-I Min, PhD, Alexandra Castillo, MPH, Lori Ward, PhD, MS, Robert P. Santos, MD, MSCS, Pramod Anugu, Yan Gao, MPH, Jason Green, Ramona Sandlin, RHIA, Donald Moore, MS, Lemichal Drake, Dorothy Horton, RN, Kendra L. Johnson, MPH, Michael Stover.

Wake Med Health and Hospitals: William H. Lagarde MD*, LaMonica Daniel, BSCR.

New Hanover: Patrick D. Maguire MD*, Charin L. Hanlon, MD, Lynette McFayden, MSN, CCRP, Isaura Rigo, MD, Kelli Hines, BS, Lindsay Smith, BA, Monique Harris, CCRP, Belinda Lissor, AAS, CCRP, Vivian Cook, MA, MPH, Maddy Eversole, BS, Terry Herrin, BS, Dennis Murphy, RN, Lauren Kinney, BS, Polly Diehl, MS, RHIA, Nicholas Abromitis, BS, Tina St. Pierre, BS, Bill Heckman, Denise Evans, Julian March, BA, Ben Whitlock, CPA, MSA, Wendy Moore, BS, AAS, Sarah Arthur, MSW, LCSW, Joseph Conway.

Vidant Health: Thomas R. Gallaher MD*, Mathew Johanson, MHA, CHFP, Sawyer Brown, MHA, Tina Dixon, MPA, Martha Reavis, Shakira Henderson, PhD, DNP, MS, MPH, Michael Zimmer, PhD, Danielle Oliver, Kasheta Jackson, DNP, RN, Monica Menon, MHA, Brandon Bishop, MHA, Rachel Roeth, MHA.

Campbell University School of Osteopathic Medicine: Robin King-Thiele DO*, Terri S. Hamrick PhD*, Abdalla Ihmeidan, MHA, Amy Hinkelman, PhD, Chika Okafor, MD (Cape Fear Valley Medical Center), Regina B. Bray Brown, MD, Amber Brewster, MD, Danius Bouyi, DO, Katrina Lamont, MD, Kazumi Yoshinaga, DO, (Harnett Health System), Poornima Vinod, MD, A. Suman Peela, MD, Giera Denbel, MD, Jason Lo, MD, Mariam Mayet-Khan, DO, Akash Mittal, DO, Reena Motwani, MD, Mohamed Raafat, MD (South-eastern Health System), Evan Schultz, DO, Aderson Joseph, MD, Aalok Parkeh, DO, Dhara Patel, MD, Babar Afridi, DO (Cumberland County Hospital System, Cape Fear Valley).

George Washington University Data Coordinating Center: Diane Uschner PhD*, Sharon L. Edelstein, ScM, Michele Santacatterina, PhD, Greg Strylewicz, PhD, Brian Burke, MS, Mihili Gunaratne, MPH, Meghan Turney, MA, Shirley Qin Zhou, MS, Ashley H Tjaden, MPH, Lida Fette, MS, Asare Buahin, Matthew Bott, Sophia Graziani, Ashvi Soni, MS, Guoqing Diao, PhD, Jone Renteria, MS.

George Washington University Mores Lab: Christopher Mores, PhD, Abigail Porzucek, MS.

Oracle Corporation: Rebecca Laborde, Pranav Acharya.

Sneez LLC: Lucy Guill, MBA, Danielle Lamphier, MBA, Anna Schaefer, MSM, William M. Satterwhite, JD, MD.

Vysnova Partners: Anne McKeague, PhD, Johnathan Ward, MS, Diana P. Naranjo, MA, Nana Darko, MPH, Kimberly Castellon, BS, Ryan Brink, MSCM, Haris Shehzad, MS, Derek Kuprianov, Douglas McGlasson, MBA, Devin Hayes, BS, Sierra Edwards, MS, Stephane Daphnis, MBA, Britnee Todd, BS.

Javara Inc.: Atira Goodwin.

External Advisory Council: Ruth Berkelman, MD, Emory, Kimberly Hanson, MD, U of Utah, Scott Zeger, PhD, Johns Hopkins, Cavan Reilly, PhD, U. of Minnesota, Kathy Edwards, MD, Vanderbilt, Helene Gayle, MD MPH, Chicago Community Trust, Stephen Redd.

References

1. WHO. *Evaluation of COVID-19 Vaccine Effectiveness*; World Health Organization: Geneva, Switzerland, 2021.
2. Bell, S.K.; Delbanco, T.; Elmore, J.G.; Fitzgerald, P.S.; Fossa, A.; Harcourt, K.; Leveille, S.G.; Payne, T.H.; Stametz, R.A.; Walker, J.; et al. Frequency and types of patient-reported errors in electronic health record ambulatory care notes. *JAMA Netw. Open* **2020**, *3*, e205867. [[CrossRef](#)] [[PubMed](#)]
3. Goodyear-Smith, F.; Grant, C.; York, D.; Kenealy, T.; Copp, J.; Petousis-Harris, H.; Turner, N.; Kerse, N. Determining immunisation coverage rates in primary health care practices: A simple goal but a complex task. *Int. J. Med. Inform.* **2008**, *77*, 477–485. [[CrossRef](#)] [[PubMed](#)]
4. Irving, S.A.; Donahue, J.G.; Shay, D.K.; Ellis-Coyle, T.L.; Belongia, E.A. Evaluation of self-reported and registry-based influenza vaccination status in a wisconsin cohort. *Vaccine* **2009**, *27*, 6546–6549. [[CrossRef](#)] [[PubMed](#)]
5. Mac Donald, R.; Baken, L.; Nelson, A.; Nichol, K.L. Validation of self-report of influenza and pneumococcal vaccination status in elderly outpatients. *Am. J. Prev. Med.* **1999**, *16*, 173–177. [[CrossRef](#)]
6. Mangtani, P.; Shah, A.; Roberts, J.A. Validation of influenza and pneumococcal vaccine status in adults based on self-report. *Epidemiol. Infect.* **2007**, *135*, 139–143. [[CrossRef](#)] [[PubMed](#)]
7. Smith, R.; Hubers, J.; Farraye, F.A.; Sampene, E.; Hayney, M.S.; Caldera, F. Accuracy of self-reported vaccination status in a cohort of patients with inflammatory bowel disease. *Dig Dis. Sci.* **2021**, *66*, 2935–2941. [[CrossRef](#)] [[PubMed](#)]
8. Zimmerman, R.K.; Raymund, M.; Janoskym, J.E.; Nowalk, M.P.; Fine, M.J. Sensitivity and specificity of patient self-report of influenza and pneumococcal polysaccharide vaccinations among elderly outpatients in diverse patient care strata. *Vaccine* **2003**, *21*, 1486–1491. [[CrossRef](#)]
9. King, J.P.; McLean, H.Q.; Belongia, E.A. Validation of self-reported influenza vaccination in the current and prior season. *Influenza Other Respir. Viruses* **2018**, *12*, 808–813. [[CrossRef](#)] [[PubMed](#)]
10. Rolnick, S.J.; Parker, E.D.; Nordin, J.D.; Hedblom, B.D.; Wei, F.; Kerby, T.; Jackson, J.M.; Crain, A.L.; Euler, G. Self-report compared to electronic medical record across eight adult vaccines: Do results vary by demographic factors? *Vaccine* **2013**, *31*, 3928–3935. [[CrossRef](#)] [[PubMed](#)]
11. The COVID-19 Community Research Partnership; Sanders, J. The COVID-19 community research partnership: Objectives, study design, baseline recruitment, and retention. *medRxiv* **2022**. [[CrossRef](#)]
12. Jackson, M.L. Use of self-reported vaccination status can bias vaccine effectiveness estimates from test-negative studies. *Vaccine X* **2018**, *1*, 100003. [[CrossRef](#)] [[PubMed](#)]
13. De Smedt, T.; Merrall, E.; Macina, D.; Perez-Vilar, S.; Andrews, N.; Bollaerts, K. Bias due to differential and non-differential disease and exposure misclassification in studies of vaccine effectiveness. *PLoS ONE* **2018**, *13*, e0199180. [[CrossRef](#)]
14. Centers for Disease Control and Prevention. Flu Vaccination Coverage, United States, 2020–2021 Influenza Season. Updated 2021. Available online: <https://www.cdc.gov/flu/fluview/coverage-2021estimates.htm> (accessed on 14 March 2022).
15. Centers for Disease Control and Prevention. COVID-19 Vaccinations in the United States. Updated 2022. Available online: <https://covid.cdc.gov/covid-data-tracker> (accessed on 14 March 2022).
16. Gianfredi, V.; Pennisi, F.; Lume, A.; Ricciardi, G.E.; Minerva, M.; Riccò, M.; Odone, A.; Signorelli, C. Challenges and opportunities of mass vaccination centers in COVID-19 times: A rapid review of literature. *Vaccines* **2021**, *9*, 574. [[CrossRef](#)] [[PubMed](#)]
17. Goralnick, E.; Kaufmann, C.; Gawande, A.A. Mass-vaccination sites—An essential innovation to curb the COVID-19 pandemic. *N. Engl. J. Med.* **2021**, *384*, e67. [[CrossRef](#)] [[PubMed](#)]
18. Sattar, N.S.; Arifuzzaman, S. COVID-19 vaccination awareness and aftermath: Public sentiment analysis on twitter data and vaccinated population prediction in the USA. *Appl. Sci.* **2021**, *11*, 6128. [[CrossRef](#)]