

## Quantifying and adjusting for confounding from health-seeking behaviour and healthcare access in observational research

Graham, S.<sup>1,2\*</sup>, Walker J.L.<sup>1,2,3</sup>, Andrews, N.<sup>2,3</sup>, Hulme, W.J.<sup>4</sup>, Nitsch, D.<sup>1,5,6</sup>, Parker, E.P.K.<sup>1,2†</sup>, McDonald, H.I.<sup>1,2,3,7†</sup>

1. Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom
2. National Institute for Health and Care Research Health Protection Research Unit in Vaccines and Immunisation, London United Kingdom
3. UK Health Security Agency, London, United Kingdom
4. The Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, United Kingdom
5. UK Renal Registry, Bristol, United Kingdom
6. Renal Unit, Royal Free London NHS Foundation Trust, Hertfordshire, United Kingdom
7. Faculty of Science, University of Bath, Bath, United Kingdom

\*Corresponding author:

Sophie Graham ([sophie.graham@evidera.com](mailto:sophie.graham@evidera.com))

†Equal contribution

Name	Institution	Faculty	Country, region and city
Sophie Graham	London School of Hygiene and Tropical Medicine	Epidemiology and Population Health	London, UK
	National Institute for Health and Care Research	Health Protection Research Unit in Vaccines and Immunisation	London, UK
Jemma Walker	London School of Hygiene and Tropical Medicine	Epidemiology and Population Health	London, UK
	National Institute for Health and Care Research	Health Protection Research Unit in Vaccines and Immunisation	London, UK
	UK Health Security Agency	Immunisation and Vaccine Preventable Diseases	London, UK
Nick Andrews	National Institute for Health and Care Research	Health Protection Research Unit in Vaccines and Immunisation	London, UK
	UK Health Security Agency	Immunisation and Vaccine Preventable Diseases	London, UK
Dorothea Nitsch	London School of Hygiene and Tropical Medicine	Epidemiology and Population Health	London, UK
	UK Kidney Association		Bristol, UK
	Royal Free London NHS Foundation Trust	Renal Unit	London, UK
William Hulme	University of Oxford	The Bennett Institute for Applied Data Science, Nuffield Department of Primary Care Health Sciences	Oxford, UK
Edward Parker	London School of Hygiene and Tropical Medicine	Epidemiology and Population Health	London, UK
	National Institute for Health and Care Research	Health Protection Research Unit in Vaccines and Immunisation	London, UK
Helen McDonald	London School of Hygiene and Tropical Medicine	Epidemiology and Population Health	London, UK
	National Institute for Health and Care Research	Health Protection Research Unit in Vaccines and Immunisation	London, UK
	UK Health Security Agency	Immunisation and Vaccine Preventable Diseases	London, UK
	University of Bath	Faculty of Science	Bath, UK

## Abstract

**Objective:** To assess the feasibility and effect of using proxy markers of health-seeking behaviour and healthcare access to quantify and adjust for confounding in observational studies of influenza and COVID-19 vaccine effectiveness (VE).

**Design:** Cohort study for influenza VE in the 2019/2020 influenza season and for early COVID-19 VE (December 2020 – March 2021).

**Setting:** Primary care data pre-linked to secondary care and death data in England.

**Participants:** Individuals aged  $\geq 66$  years on 1 September 2019.

**Interventions:** Vaccination with any influenza vaccination in the 2019/2020 season or with either a BNT162b2 or ChAdOx1-S vaccination from 08/12/2020 to 31/03/2021.

**Main outcome measures:** Influenza or COVID-19 specific infections, hospitalisation and death. VE was estimated with sequential adjustment for demographics, underlying health conditions, and 14 markers reflecting uptake of public health interventions (screenings, vaccinations and NHS health checks), active healthcare access/use (prostate antigen testing, bone density scans, GP practice visits, low value procedures and blood pressure measurements) and lack of access/underuse (hospital visits for ambulatory care sensitive conditions and did not attend primary care visits). Influenza vaccination in the 2019/2020 season was also considered as a negative exposure intervention against the first wave of COVID-19.

**Results:** We included 1,991,284, 1,796,667, and 1,946,943 individuals in the influenza, COVID-19 and negative exposure VE populations, respectively. Vaccinated individuals were more likely to display active health-seeking behaviour, including participation in UK national screening programmes, compared with unvaccinated individuals. In the 2019/2020 influenza season, adjusting for health-seeking markers increased VE against infection from -1.5% (95%CI: -3.2,0.1) to 7.1% (95%CI: 5.4,8.7), but this trend was less apparent for more severe outcomes. For COVID-19 during early vaccine roll out, adjusting for health-seeking markers in addition to demographics and underlying health conditions did not change VE estimates against infection or severe disease (e.g., two doses of BNT162b2 against infection: from 82.8% [95%CI: 78.4,86.3] to 83.1% [95%CI: 78.7,86.5]). Adjusting for health-seeking markers removed bias in the negative exposure analysis of influenza vaccination against SARS-CoV-2 infection (-7.5% [95%CI: -10.6,-4.5] vs -2.1% [95%CI: -6.0,1.7] before vs after adjusting for health-seeking markers).

**Conclusions:** Markers of health-seeking behaviour and healthcare access can be identified in electronic health records, are associated with vaccination uptake, and can be used to quantify and account for confounding in observational studies.

<b>What is already known on this topic</b>
Health-seeking behaviour and healthcare access are recognised confounders in observational studies, but are not directly measurable in electronic health records (EHRs). Previously we systematically identified 14 markers in UK EHRs that reflect different aspects of health-seeking behaviour and healthcare access. We do not know if these markers can be utilised to quantify and account for this type of confounding in observational research using influenza and COVID-19 vaccine effectiveness as examples.
<b>What this study adds</b>
This study demonstrated using the proxy markers that confounding from health-seeking behaviour and healthcare access underestimates influenza VE estimates, but has

negligible impact on COVID-19 VE estimates during early vaccine roll out. We also demonstrated via a negative exposure analysis that residual confounding can be removed by adjusting for these proxy markers.

## Background:

Health-seeking behaviour and healthcare access may be important confounders in observational research. Health-seeking behaviour is defined as seeking care for disease prevention, when asymptomatic or during early symptomatic stages,(1) whereas healthcare access can be defined as the ability to access healthcare services for these purposes.(2) Individuals with active health-seeking behaviour and the ability to easily access healthcare services are generally expected to have favourable clinical outcomes. Previously confounding from health-seeking behaviour and healthcare access has led to overestimates of effectiveness of some preventative therapies in observational research. For example, observational studies of statin use and hip fracture have consistently shown, even in meta-analysed results, to reduce risk by 23%, even though this has not been reflected in clinical trials(3). In cohort studies of influenza vaccine effectiveness (VE), authors have reported reductions in all-cause mortality by 40-50%(4, 5). Findings from these studies have since been disputed as influenza accounts for a maximum of 10% of deaths per year, so even if the vaccination was 100% effective it would only be able to prevent 10% of deaths per year(6) However, authors do not usually systematically attempt to account for this type of confounding, as health-seeking behaviour and healthcare access is typically not directly measurable in routine healthcare data.

Proxy markers identified in electronic health records (EHRs) have been used previously to account for confounding from health-seeking behaviour and healthcare access(7-11). Markers included in these studies vary considerably and therefore the best approach to account for this type of confounding is unclear. We recently identified a systematic set of fourteen markers of health-seeking behaviour and healthcare access in UK EHRs(12) that accounted for a range of determinants from a behavioural model known as the Theory of Planned Behaviour.(13) These markers represented interactions with the healthcare system that are only partially driven by an individual's underlying health need. In the current study, we aimed to assess whether these markers or proxies of health-seeking behaviour and healthcare access in UK EHRs data can be used to quantify and adjust for confounding in observational studies, using seasonal influenza and COVID-19 VE as examples.

## Methods:

### Data sources

We conducted a cohort study using Clinical Practice Research Datalink (CPRD) Aurum(14) data pre-linked to Hospital Episode Statistics (HES) Admitted Patient Care (APC)(15) and Office for National Statistics (ONS) death and socioeconomic status data.(16) CPRD Aurum includes diagnoses, prescriptions, referral and testing information of patients registered to consenting GP practices in the UK that use EMIS® software systems. Medical diagnoses are recorded using SNOMED, Read Coded Clinical Terms version 3 (CTV3), or local EMIS® codes that are each mapped to a unique medcode. Information on prescription medicines is coded using the Dictionary of Medicines and Devices (dm+d) codes that are each mapped to a unique prodcode. CPRD Aurum data was pre-linked to HES APC, which includes all admissions to NHS hospitals in England(15). It includes inpatient hospital admission and discharge dates, diagnoses recorded using International Classification of Diseases 10th (ICD-10) Revision codes(17) and procedures recorded using Operating Procedure Codes Supplement (OPCS) codes(18). The data was also pre-linked to ONS, which includes date and underlying cause of death, recorded using ICD-10 codes, and socioeconomic data based on multiple deprivation index(19) which is based on small area geographical location. Linkage from CPRD Aurum to HES APC and ONS datasets occurs at the patient level and is consented at practice level. At the time of data extraction (May 2022 release), the CPRD

Aurum data included 1,345 general practices for 13,300,067 currently contributing patients (19.8% of the UK population)(20).

### **Study design and population selection**

We created separate cohorts to estimate influenza and COVID-19 VE. In addition, to assess for potential residual confounding, we created a third “negative exposure” cohort. Negative exposures assume there to be no causal mechanism between the negative exposure and outcome, but for the confounding structures to reflect those of the primary exposure(21). In the current study we used seasonal influenza vaccinations in 2019/20 as a negative exposure against COVID-19 infections before COVID-19 vaccinations were available in the UK.

An overview of the study design for each of these three populations can be found in Figure 1 and further details in Supplementary Table 1. We included all individuals aged  $\geq 66$  years on 1 September 2019. This age group is prioritised for vaccination (including for seasonal influenza vaccine and primary COVID-19 vaccination), and is also likely to show distinct patterns of health-seeking behaviour and healthcare access varies (e.g. due to eligibility of various age-specific screening programmes)(22). We required all individuals to have at least one year of registration prior to their index date (Figure 1 and described further below), have a record of ‘acceptable’ quality by CPRD, and be eligible for linkage to HES APC and ONS. We excluded individuals with a death or registration end date before index date. We also excluded those with indeterminate sex due to low numbers. Individuals in the COVID-19 cohort were additionally excluded if they received their first vaccination prior to 8 December 2020 as these were likely to reflect coding errors or trial participants with a different risk profile.

**Figure 1. Study design.** We followed individuals until the earliest of death, transfer out of GP practice, end of data availability (COVID-19), end of influenza season (influenza), or start of COVID-19 vaccination roll out (negative exposure). For the COVID-19 cohort, we censored individuals at first COVID-19 vaccination that was neither BNT162b2 or ChAdOx1, or on receipt of a second heterologous vaccination.

### **Outcomes, exposures, and follow-up**

For all analyses we considered three nested outcomes of increasing severity: infections (based on GP data in CPRD Aurum, hospitalisations in HES APC or deaths in ONS); hospitalisation/deaths; and deaths. All COVID-19 outcomes were required to have a COVID-19 diagnosis code. For influenza we required a diagnosis of acute respiratory infection or influenza like illness (ARI/ILI) using code lists from Davidson et al, 2021(23). For all hospital and death outcomes the diagnosis code was required to be in the primary position. Further details on the study outcome and exposure definitions can be found in Supplementary Table 1.

For influenza a first dose of any seasonal influenza vaccination that was recorded during the 2019/2020 season was included. We identified doses of the COVID-19 vaccines BNT162b2 and ChAdOx1 separately. We also identified second doses for COVID-19, requiring a minimum interval of at least 18 days from first to second dose to reflect Joint Committee on Vaccination and Immunisation dosing schedules.(24) Only prodcodes were used to identify COVID-19 vaccinations, as during the COVID-19 pandemic, COVID-19 vaccinations were automatically recorded in patient GP records. For influenza, we identified vaccinations using both medcodes and prodcodes. Since individuals can have more than one medcode recorded on the same day that can represent conflicting information (e.g., given and not given), an algorithm was used to classify influenza vaccination status (see Supplementary Table 1).

For influenza, we used an index date of 1 September 2019. Individuals were followed up from this date until the earliest of death, transfer out of the practice or start of the COVID-19 pandemic (29 February 2020). For COVID-19, we used an index date of 8 December 2020, when COVID-19 vaccinations were first introduced in the UK. Individuals were followed up from this date until the earliest of death, transfer out of the practice, end of data availability (31 March 2021), first vaccination that was neither BNT162b2 or ChAdOx1, or second heterologous vaccination. For the negative exposure analysis, we used an index date of 1 January 2020 as first SARS-CoV-2 infections were identified in the UK after this date. We only included influenza vaccinations before 31 December 2019 by which time the majority of vaccinations in the UK have been delivered(25). This was to reflect positive health-seeking behaviour/access, and prevent overlap with the outcome period, when vaccination uptake might have been influenced in part by awareness and attitudes towards the emerging COVID-19 pandemic. Individuals were followed-up from this date until the earliest of death, transfer out of the practice, or the day before introduction of COVID-19 vaccinations in the UK (7 December 2020).

### **Sociodemographic variables at index**

At index for each cohort we described the following variables: age (day and month imputed as 01/07 for all individuals as only year of birth is recorded in CPRD), sex, recent infection, local area index of multiple deprivation (IMD), ethnicity, and influenza 'at-risk' conditions(26). Recent infection for SARS-CoV-2 was defined as an infection in the three months pre-index, whereas for influenza this was defined as an influenza infection in the previous influenza season. IMD was identified from the ONS at the patient level, or if missing by the primary care practice. Ethnicity was identified from primary care records as described by Mathur et al.(27). Briefly, the algorithm uses a modal approach with ties resolved by recency. If ethnicity could not be identified in primary care, then ethnicity from HES APC was used. Influenza 'at-risk' groups(26) were identified using primary care records of diagnoses or medications as described previously(23), and grouped into immunosuppressed status or other conditions. Individuals with missing age and sex are not provided by CPRD. We assessed missingness of ethnicity, IMD and region. For all other variables, if codes did not appear in an individual's health record, this was regarded as evidence of absence. Further details on how variables were defined can be found in Supplementary Table 1.

### **Markers of health-seeking behaviour and healthcare access at index**

We used 14 markers of health-seeking behaviour and healthcare access that we identified through a framework based on the updated Theory of Planned Behaviour(13), as described previously.(12) These included markers representing uptake of public health interventions (AAA, breast cancer, cervical cancer and bowel cancer screening; influenza and pneumococcal vaccination and NHS health checks), active healthcare access/use (prostate antigen testing, bone density scans, GP practice visits, low value procedures and blood pressure measurements) and lack of access/underuse (hospital visits for ambulatory care sensitive conditions and did not attend primary care visits). The markers were identified in CPRD Aurum using medcodes or in HES APC using OPCS or ICD-10 codes as described previously.(12) The lookback periods reflect use of these markers in UK clinical practice and are detailed in Supplementary Table 1.

### **Statistical analyses**

We described sociodemographic variables, clinical variables, and markers of health-seeking behaviour at index stratified by final vaccination status. To assess timeliness of vaccination, we calculated median days from index to first vaccination, stratified by marker status and

age categories (since for COVID-19 vaccinations the UK phased approach was based on age) amongst vaccinated individuals. We grouped the older age categories (over 85 years) due to low patient numbers. Outcome rates were represented by vaccination status as number of events divided by total person years. Cox regression models were used to estimate the risk of infection in vaccinated versus unvaccinated individuals. A complete case analysis was conducted (excluding individuals with a missing value for age, sex, ethnicity or IMD). In the influenza and COVID-19 analyses, vaccination status was time-updated, such that all individuals started follow-up unvaccinated at index date and their vaccination status was updated 14 days after a vaccine was administered (to provide time for immune response induction). For COVID-19, we assessed VE (defined as  $[1 - \text{hazard ratio}] \times 100$ ) for BNT162b2 and ChAdOx1 by fitting separate models each restricted to unvaccinated individuals and those who received the vaccine of interest. In the negative exposure analysis, since the vaccination and outcome periods did not overlap, a history of influenza vaccination in the previous season was defined as a binary exposure at index date. We adapted a previous modelling strategy(28) to understand the hierarchical relationships between determinants of vaccine uptake in four iterative steps. First, we fitted baseline models (Model 1) that adjusted for age (as a quadratic polynomial), sex, region and recent infection. In demography-adjusted models (Model 2), we further adjusted for ethnicity and IMD. In comorbidity-adjusted models (Model 3), we further adjusted for immunosuppressive status and other comorbidities. Finally, we fitted fully adjusted models (Model 4) that further adjusted for markers of health-seeking behaviour and healthcare access. For sex-specific markers (cervical cancer screening, breast cancer screening, AAA screening and PSA test), we included an interaction term with sex in the fully adjusted models.

### **Sensitivity analysis**

We conducted a sensitivity analysis to re-run the fully adjusted models fitting age interactions with AAA screening, bowel cancer screening, NHS health checks and ACS conditions. These were selected as we previously found their prevalence to vary markedly by age.(12)

### **Patient and public involvement**

There was no patient or public involvement in the study design or interpretation of results.

### **Ethics and CPRD requirements**

Our study was approved by CPRD's (#21\_000737) and LSHTM's (#28169) independent ethics review committees. To meet CPRD patient confidentiality requirements we redacted counts relating to less than 5 individuals and conducted secondary suppression where necessary.

### **Software, code and reproducibility**

All the analyses were conducted using R versions 4.2.2 to 4.2.4. All programming code from this project can be found in the Github repository: <https://github.com/grahams99/Health-seeking-behaviour>. The code lists and related search terms can be found on LSHTM Data Compass (<https://doi.org/10.17037/DATA.00003684>).

## **Results**

### **Study population**

We included 1,946,943, 1,796,667 and 1,991,284 individuals in the influenza, COVID-19 and negative exposure cohorts, respectively (Figure 2). Compared with individuals who remained unvaccinated throughout enrolment, vaccinated individuals were more likely to be older, of

White ethnicity, and live in less deprived areas. Baseline demographic, underlying health conditions, and health-seeking characteristics are summarised for each cohort in Supplementary Table 3, and stratified by subsequent vaccination status in Table 2.

### **Markers of health-seeking behaviour**

Compared with individuals that remained unvaccinated throughout enrolment, vaccinated individuals had a higher prevalence of all markers of health-seeking behaviour and healthcare access (with the exception of hospital visits for ambulatory care sensitive conditions; Table 2). The differences were greatest for previous vaccinations. In the influenza analysis, 91.2% of vaccinated had an influenza vaccination in the previous season, versus 22.6% in those unvaccinated for influenza vaccination. 72.7% of influenza vaccinated individuals had a pneumococcal vaccination prior to index, compared with only 33.0% in those unvaccinated for influenza. The same was true for the COVID-19 analysis, with 82.0% of ChAdOx1-S vaccinated and 75.9% of BNT162b2 vaccinated individuals having an influenza vaccination in the previous season versus 40.3% in those unvaccinated for any COVID-19 vaccination. 70.9% of individuals vaccinated for ChAdOx1-S and 61.9% of individuals vaccinated for BNT162b2 had a pneumococcal vaccination prior to index, compared with 42.9% in individuals unvaccinated for COVID-19. Among vaccinated individuals, time to vaccination was not strongly associated with health-seeking marker status, with the exception of influenza vaccination in the previous season, which was associated with faster uptake of both COVID-19 and influenza vaccines (Supplementary Figure 1 and 2 with complementary raw data in Supplementary Table 6 and 7).

**Figure 2. Study selection criteria.** Additional details on population selection can be found in Supplementary Table 1. Abbreviations: CPRD: clinical practice research datalink; HES: hospital episode statistics; IMD: index of multiple deprivation; ONS: office for national statistics.



**Table 2. Baseline characteristics stratified by vaccination status at the end of follow-up.**

		Influenza analysis population N=1,796,667		COVID-19 analysis population N=1,796,667			Negative exposure analysis population N=1,946,943	
Variable	Category	Vaccinated N=1,473,955	Unvaccinated N=517,329	ChAdOx1-S N=845,428	BNT162b2 N=811,740	Unvaccinated N=139,499	Vaccinated N=1,437,356	Unvaccinated N=509,587
<b>Age category in years, N (%)</b>	65-69	295,808 (20.1%)	152,255 (29.4%)	196,994 (23.3%)	93,761 (11.6%)	29,203 (20.9%)	288,285 (20.1%)	154,638 (30.3%)
	70-74	409,473 (27.8%)	148,126 (28.6%)	318,400 (37.7%)	185,076 (22.8%)	37,479 (26.9%)	400,663 (27.9%)	149,441 (29.3%)
	75-79	312,214 (21.2%)	89,076 (17.2%)	186,454 (22.1%)	180,327 (22.2%)	24,518 (17.6%)	305,940 (21.3%)	88,128 (17.3%)
	80-84	235,026 (15.9%)	60,466 (11.7%)	65,793 (7.8%)	191,890 (23.6%)	21,093 (15.1%)	229,632 (16.0%)	58,105 (11.4%)
	85-89	142,171 (9.6%)	39,664 (7.7%)	44,040 (5.2%)	110,191 (13.6%)	15,032 (10.8%)	137,602 (9.6%)	36,230 (7.1%)
	90-95	60,924 (4.1%)	20,019 (3.9%)	24,562 (2.9%)	41,348 (5.1%)	8,520 (6.1%)	58,241 (4.1%)	17,037 (3.3%)
	95+	18,339 (1.2%)	7,723 (1.5%)	9,185 (1.1%)	9,147 (1.1%)	3,654 (2.6%)	16,993 (1.2%)	6,008 (1.2%)
<b>Sex, N (%)</b>	Female	795,391 (54.0%)	280,332 (54.2%)	458,466 (54.2%)	442,085 (54.5%)	73,243 (52.5%)	776,347 (54.0%)	276,181 (54.2%)
	Male	678,564 (46.0%)	236,997 (45.8%)	386,962 (45.8%)	369,655 (45.5%)	66,256 (47.5%)	661,009 (46.0%)	233,406 (45.8%)
<b>Ethnicity, N (%)</b>	Asian	49,874 (3.4%)	18,087 (3.5%)	27,362 (3.2%)	26,576 (3.3%)	9,891 (7.1%)	48,257 (3.4%)	19,013 (3.7%)
	Black	21,970 (1.5%)	14,942 (2.9%)	12,665 (1.5%)	9,685 (1.2%)	11,631 (8.3%)	21,023 (1.5%)	15,431 (3.0%)
	Mixed	6,321 (0.4%)	3,751 (0.7%)	3,744 (0.4%)	3,223 (0.4%)	2,242 (1.6%)	6,095 (0.4%)	3,853 (0.8%)
	Other	11,295 (0.8%)	6,416 (1.2%)	6,750 (0.8%)	5,942 (0.7%)	3,841 (2.8%)	10,982 (0.8%)	6,611 (1.3%)
	White	1,331,686 (90.3%)	428,068 (82.7%)	754,194 (89.2%)	734,394 (90.5%)	94,638 (67.8%)	1,299,673 (90.4%)	418,902 (82.2%)
	Missing	52,809 (3.6%)	46,065 (8.9%)	40,713 (4.8%)	31,920 (3.9%)	17,256 (12.4%)	51,326 (3.6%)	45,777 (9.0%)
<b>Region, N (%)</b>	East Midlands	30,432 (2.1%)	[Redacted]	18,636 (2.2%)	14,776 (1.8%)	1,892 (1.4%)	29,726 (2.1%)	[Redacted]
	East of England	70,073 (4.8%)	25,340 (4.9%)	39,048 (4.6%)	37,023 (4.6%)	5,426 (3.9%)	67,180 (4.7%)	21,300 (4.2%)
	London	179,896 (12.2%)	83,929 (16.2%)	86,705 (10.3%)	115,643 (14.2%)	37,022 (26.5%)	174,002 (12.1%)	86,226 (16.9%)
	North East	51,140 (3.5%)	16,138 (3.1%)	30,256 (3.6%)	28,324 (3.5%)	3,873 (2.8%)	50,011 (3.5%)	16,579 (3.3%)
	North West	288,382 (19.6%)	93,210 (18.0%)	165,164 (19.5%)	158,595 (19.5%)	23,774 (17.0%)	282,152 (19.6%)	93,857 (18.4%)
	South East	330,361 (22.4%)	108,046 (20.9%)	192,498 (22.8%)	175,053 (21.6%)	27,873 (20.0%)	323,518 (22.5%)	109,375 (21.5%)

		Influenza analysis population N=1,796,667		COVID-19 analysis population N=1,796,667			Negative exposure analysis population N=1,946,943	
Variable	Category	Vaccinated N=1,473,955	Unvaccinated N=517,329	ChAdOx1-S N=845,428	BNT162b2 N=811,740	Unvaccinated N=139,499	Vaccinated N=1,437,356	Unvaccinated N=509,587
	South West	202,112 (13.7%)	68,372 (13.2%)	126,645 (15.0%)	103,675 (12.8%)	14,764 (10.6%)	196,664 (13.7%)	62,755 (12.3%)
	West Midlands	262,952 (17.8%)	94,411 (18.2%)	151,012 (17.9%)	151,300 (18.6%)	21,235 (15.2%)	256,602 (17.9%)	94,795 (18.6%)
	Yorkshire and The Humber	58,599 (4.0%)	16,206 (3.1%)	35,386 (4.2%)	27,336 (3.4%)	3,630 (2.6%)	57,467 (4.0%)	16,393 (3.2%)
	Unknown	8 (0.0%)	[Redacted]	78 (0.0%)	15 (0.0%)	10 (0.0%)	34 (0.0%)	[Redacted]
<b>IMD, N (%)</b>	1 (least deprived)	389,007 (26.4%)	110,866 (21.4%)	219,539 (26.0%)	213,479 (26.3%)	25,049 (18.0%)	381,652 (26.6%)	110,457 (21.7%)
	2	352,561 (23.9%)	115,220 (22.3%)	198,707 (23.5%)	195,168 (24.0%)	26,339 (18.9%)	344,778 (24.0%)	113,011 (22.2%)
	3	291,822 (19.8%)	107,489 (20.8%)	165,923 (19.6%)	164,259 (20.2%)	27,283 (19.6%)	283,270 (19.7%)	103,275 (20.3%)
	4	245,959 (16.7%)	99,005 (19.1%)	143,497 (17.0%)	136,591 (16.8%)	30,900 (22.2%)	239,223 (16.6%)	97,624 (19.2%)
	5 (most deprived)	194,606 (13.2%)	84,749 (16.4%)	117,762 (13.9%)	102,243 (12.6%)	29,928 (21.5%)	188,433 (13.1%)	85,220 (16.7%)
<b>Influenza 'at-risk' conditions, N (%)</b> <b>Markers of health-seeking behaviour and healthcare access, N (%)</b>	Immunosuppressed status	44,445 (3.0%)	10,708 (2.1%)	20,560 (2.4%)	20,314 (2.5%)	3,181 (2.3%)	51,304 (3.6%)	11,491 (2.3%)
	Other comorbidities*	875,433 (59.4%)	227,851 (44.0%)	455,435 (53.9%)	478,945 (59.0%)	70,209 (50.3%)	859,858 (59.8%)	224,804 (44.1%)
	AAA screen	171,329 (11.6%)	59,759 (11.6%)	76,400 (9.4%)	127,455 (15.1%)	11,087 (7.9%)	167,605 (11.7%)	59,711 (11.7%)
	Bowel screen	1,063,252 (72.1%)	376,160 (72.7%)	556,813 (68.6%)	703,450 (83.2%)	94,562 (67.8%)	1,043,974 (72.6%)	380,264 (74.6%)
	Breast screen	268,370 (18.2%)	77,746 (15.0%)	149,098 (18.4%)	161,443 (19.1%)	16,564 (11.9%)	265,104 (18.4%)	78,616 (15.4%)
	Cervical screen	308,261 (20.9%)	89,042 (17.2%)	171,781 (21.2%)	171,059 (20.2%)	20,103 (14.4%)	301,116 (20.9%)	87,878 (17.2%)
	NHS health checks	278,539 (18.9%)	93,705 (18.1%)	134,883 (16.6%)	170,284 (20.1%)	15,862 (11.4%)	270,812 (18.8%)	92,865 (18.2%)
	Influenza vaccine†	1,343,562 (91.2%)	116,829 (22.6%)	665,364 (82.0%)	641,842 (75.9%)	56,223 (40.3%)	1,314,908 (91.5%)	112,149 (22.0%)
	Pneumococcal vaccine	1,071,867 (72.7%)	170,492 (33.0%)	575,351 (70.9%)	523,544 (61.9%)	59,781 (42.9%)	1,069,249 (74.4%)	164,248 (32.2%)
	ACS hospital care visit	85,734 (5.9%)	43,689 (8.4%)	85,734 (10.6%)	87,757 (10.4%)	17,314 (12.4%)	153,108 (10.7%)	42,755 (8.4%)
	Blood pressure test	1,160,045 (78.7%)	309,961 (59.9%)	718,511 (88.5%)	717,133 (84.8%)	91,327 (65.5%)	1,229,509 (85.5%)	334,066 (65.6%)
	Bone density scan	56,679 (3.8%)	19,011 (3.7%)	56,679 (7.0%)	53,276 (6.3%)	5,282 (3.8%)	87,873 (6.1%)	20,667 (4.1%)
	DNA Primary care visit	459,941 (31.2%)	141,955 (27.4%)	431,120 (53.1%)	432,191 (51.1%)	72,675 (52.1%)	560,407 (39.0%)	171,514 (33.7%)

		Influenza analysis population N=1,796,667		COVID-19 analysis population N=1,796,667			Negative exposure analysis population N=1,946,943	
Variable	Category	Vaccinated N=1,473,955	Unvaccinated N=517,329	ChAdOx1-S N=845,428	BNT162b2 N=811,740	Unvaccinated N=139,499	Vaccinated N=1,437,356	Unvaccinated N=509,587
	GP visits	1,429,058 (97.0%)	415,765 (80.4%)	802,423 (98.9%)	828,713 (98.0%)	110,333 (79.1%)	1,420,864 (98.9%)	426,546 (83.7%)
	Low value procedures	280,546 (19.0%)	78,335 (15.1%)	259,273 (31.9%)	243,222 (28.8%)	34,634 (24.8%)	342,882 (23.9%)	93,058 (18.3%)
	PSA test	285,156 (19.3%)	67,116 (13.0%)	178,655 (22.0%)	168,574 (19.9%)	19,411 (13.9%)	294,032 (20.5%)	69,005 (13.5%)

Abbreviations: AAA: abdominal aortic aneurysm; DNA: did not attend; GP: general practice; IMD: index of multiple deprivation; N: numerator; PSA: prostate specific antigen; VE: vaccine effectiveness.

\*Other comorbidities includes: chronic liver disease, chronic cardiac disease, chronic respiratory disease, asthma, diabetes mellitus, chronic neurological disease, chronic kidney disease, severe obesity, severe mental conditions and severe learning disability. For more information on how these were defined see **Supplementary Table 1**.

†Influenza vaccination that occurred in the influenza season prior to index date. For COVID-19 this was an influenza vaccination that occurred 1 September 2019 – 31 March 2020; for Influenza and Negative Exposure this was an influenza vaccination that occurred 1 September 2018-31 March 2019 and for Negative Exposure.

Notes: for baseline characteristics in overall analysis populations see **Supplementary Table 3**. For each analysis we are comparing individuals with  $\geq 1$  vaccination versus no vaccination throughout follow-up. Cells with <5 individuals are redacted due to CPRD's patient confidentiality requirements and secondary suppression has occurred where necessary. Age was estimated at index date for each cohort and since only year of birth is provided in CPRD, all date of birth were imputed as middle of the year (01/07).

## Vaccine effectiveness estimates

For the influenza analysis population, median (IQR) follow up time was 181 (0) days. Follow-up time after first influenza vaccination was median of 50 (28) days. Unadjusted event rates ranged from 0.84 per 1,000 person-years during unvaccinated time against ARI/ILI-related death to 117.15 per 1,000 person-years after vaccination against influenza infections (Supplementary Table 4). Incremental adjustments across the models led to increased VE estimates. For influenza infections, we observed a negative VE in the minimally adjusted baseline models of -5.5% (95%CI: -7.2,-3.9). Estimated VE increased to -1.5% (95%CI: -3.2,0.1) after adjusting for comorbidities, and to 7.1% (95%CI: 5.4,8.7) after adjusting for all 14 markers of health-seeking behaviour and healthcare access. For severe outcomes, VE estimates were more consistent across models, ranging from 42.5% (95%CI: 32.8 - 50.8) in the baseline model to 47.5% (95%CI: 37.3 - 56.1) for ARI/ILI-related death in the fully adjusted model, although the trend towards increased VE after adjusting for health-seeking behaviour remained (Figure 3 and Supplementary Table 5).

For the COVID-19 analysis, median (IQR) follow-up time was 113 (0) days. Follow-up time on after first BNT162b2 vaccination was a median of 64 (19) days. Unadjusted event rates ranged from 0.54 per 1,000 person-years after two doses of BNT162b2 vaccination against COVID-19-related death to 95.39 per 1,000 person-years during unvaccinated time against COVID-19 infections (Supplementary Table 4). There was very minimal change in VE from the baseline model to the fully adjusted model that included all 14 markers of health-seeking behaviour and healthcare access (e.g. 2-dose VE of 82.7% [95%CI: 78.3,86.2] and 83.1% [95%CI: 78.7,86.5], respectively). This was also the case for more severe outcomes (e.g. 2-dose VE against hospitalisation of 96.2% [95%CI: 93.0,98.0] and 92.3% [95%CI: 93.0,98.0] for baseline and fully adjusted models, respectively). For ChAdOx1, there was very limited follow-up time after two doses; however, we observed consistent 1-dose VE estimates across baseline and incrementally adjusted models (Supplementary Table 5).

For the negative exposure analysis, median (IQR) follow-up time was 341 (0) days. For those with a history of influenza vaccination, follow-up time was also 341 (0) days. Unadjusted event rates ranged from 1.39 per 1,000 person-years for unvaccinated individuals against COVID-19-related death to 13.77 per 1,000 person-years for influenza vaccinated individuals against COVID-19-related infections (Supplementary Table 4). We observed a negative VE for the effect of influenza vaccinations against COVID-19 for all baseline and demography-adjusted models (e.g. -6.4% [95%CI: -15.4,-1.9] and -12% [95%CI: -17.4,-6.9], respectively, for VE against COVID-19-related mortality). For infections, negative VE persisted after adjusting for comorbidities (-7.5% [95%CI: -10.6,-4.5]), but not after adjusting for health-seeking behaviour and healthcare access (-2.1% [95%CI: -6.0,1.7]). For more severe endpoints, adjusting for comorbidities led to VE estimates consistent with a null finding, and this was also the case after additional adjustment for health-seeking behaviour and healthcare access (Figure 3 and Supplementary Table 5).

The sensitivity analyses that additionally included interaction terms between age and age-varying markers did not substantively change VE estimates for any of the analyses (Supplementary Table 5).

**Figure 3. Estimated vaccine effectiveness following sequential confounder adjustment in each study analysis (columns) for each outcome of interest (rows).** COVID-19 estimates are only for BNT162b2 versus unvaccinated as ChAdOx1 follow-up data after 2 doses was limited. Baseline models adjusted for polynomial age, sex, region and recent infection. Demography model further adjusted for ethnicity and IMD. Comorbidity models further adjusted for immunosuppressed status and other comorbidities. Marker models further adjusted for markers of health-seeking behaviour and healthcare access. Abbreviations: Comorb: comorbidities; demog: demography.

## Discussion

### Statement of principal findings

We found that using a range of markers of health-seeking behaviour we were able to address confounding in VE studies of influenza and COVID-19, with consistent findings using a negative exposure. This was assessed using a large cohort of individuals aged  $\geq 66$  years in England and confounding from health-seeking behaviour and healthcare access was adjusted for using proxy markers identified in UK EHRs. We found that influenza and COVID-19 vaccination uptake was higher in those with active health-seeking behaviour and better access to healthcare services. For vaccine effectiveness the impact of health-seeking behaviour and healthcare access varied by context. For influenza in a pre-COVID-19 pandemic context (influenza VE during the 2019/2020 season), we found that VE against infections (including those identified in primary care settings) was underestimated when health-seeking behaviour and healthcare access was not adjusted for. This confounding was less apparent for more severe disease endpoints. For COVID-19 in a pandemic context (during the early stages of COVID-19 vaccine implementation), we found that minimally adjusted models were very similar to more complex models, and additional adjustment for health-seeking behaviour and healthcare access did not substantively change effect estimates. Residual confounding was initially present and successfully removed by adjusting for health-seeking behaviour in a negative control analysis exploring the VE of pre-pandemic influenza vaccination against SARS-CoV-2 infections early in the pandemic.

### Strengths and weaknesses of the study

We used a large cohort of individuals and conducted a harmonised analysis with consistent variable definitions and modelling approaches in both pre- and during the COVID-19 pandemic. Although previous VE analyses have adjusted for variables that aim to capture health-seeking behaviour (e.g. GP visits or SARS-CoV-2 testing behaviour),<sup>(7, 8, 11)</sup> we included a set of proxy markers that were based on a theoretical model published previously<sup>(12)</sup>, which provided a more systematic approach to adjusting for this complex phenomenon. These markers are captured in routinely collected data sources and therefore can be used in other datasets and in other study designs where health-seeking behaviour is a concern. We were also able to identify and quantify residual confounding using a cohort design and negative exposure and demonstrate the impact of adjusting for health-seeking behaviour and healthcare access to address this. We also used nested endpoints which allowed us to assess the implications of confounding on varying outcome severities.

Despite these strengths, we assessed health-seeking behaviour and healthcare access at index date, but this might change over time, especially for the COVID-19 analysis, in which risk perception likely influenced health behaviours. There could potentially be other time-varying confounding<sup>(29)</sup> if for example, non-vaccination leads to infection and therefore temporary ineligibility for vaccination<sup>(30)</sup>. In addition, there may be some selection bias in the negative exposure study as history of influenza vaccination was identified pre-1 September 2019, but COVID-19 outcomes were identified after this date. This could have potentially led to a slight underestimate of COVID-19 infections in those without influenza vaccination.

### Strengths and weaknesses in relation to other studies, discussing important differences in results

VE estimates from the comorbidity adjusted models were similar to previous estimates from other observational studies during the same time period. For influenza, previous test-negative case control studies estimated VE against virology confirmed disease in the

2019/2020 season to be 22.7% (95%CI: -38.5,56.9%),(31) which is broadly consistent with our estimated VE (adjusted for socio-demographic variables and comorbidities i.e., the most comparable) against ARI/ILI hospital/death 24.7% (95%CI: 22.0 - 27.4). For COVID-19, Whitaker et al.(11) conducted both a cohort and test-negative study to estimate effectiveness of COVID-19 vaccinations against polymerase chain reaction confirmed COVID-19 from 8 December 2020 until 16<sup>th</sup> May 2021. They estimated VE amongst individuals aged  $\geq 65$  years after 2 doses of BNT162b2 to be 84.7% (95%CI: 77.7%,89.5) for the cohort and 79.8% (95%CI: 69.0,86.8) in the test-negative design– consistent with our all infection estimate (adjusted for socio-demographic variables and comorbidities) of 82.8% (95%CI: 78.4 - 86.3).

Our results differ to two US Medicare studies that assessed adjusting for proxy markers of health-seeking behaviour and healthcare access on VE estimates(32, 33). Whilst the two US studies both saw a decrease in VE after adjusting for confounding from health-seeking behaviour and healthcare access, we saw an increase in VE for influenza. These differences could be because of the differences in healthcare settings (i.e., nationwide free healthcare versus a premium paid healthcare) or differences in the type of dataset (i.e., EHRs versus insurance claims). It could also be that the US studies used a smaller set of markers than the current study, and therefore it could be that residual confounding remained. One of the US studies(33) used pre-season influenza estimates as a negative control outcome and still found significant residual confounding even after adjustment for health-seeking behaviour and healthcare access as well as other socio-demographic variables and proxies for frailty (32% [95%CI: 30-33%]).

For the negative control analysis we assumed that there was no causal mechanism between influenza vaccination and COVID-19 infection, or that any plausible causal association was minimal in effect. Some studies with non-specific COVID-19 outcomes have shown there to be a minor protective effect of the influenza vaccination against COVID-19 infection. A meta-analysis of 16 observational studies reported pooled odds ratios of influenza vaccination against COVID-19 infection of 0.86 (95%CI: 0.81,0.91)(34). The authors of this study suggested that individuals with influenza vaccinations are likely to have better health-seeking behaviours and healthcare access and may therefore be more compliant with COVID-19 prevention measures. A recent observational study conducted using administrative data in Canada reported reduced COVID-19 infections amongst those with an influenza vaccination, but also reported the same trend for COVID-19 infections amongst those with a previous health examination (adjusted HR:  $\square$ 0.85 [95%CI: 0.78-0.91])(35) and summarised that this provided evidence of confounding from health-seeking behaviour and healthcare access. Our study showed an opposite trend to the Canadian study for our negative exposure. Prior to adjusting for confounding from our proxies of health-seeking behaviour and healthcare access, our negative exposure analysis showed a negative VE. This likely represents those with active health-seeking behaviour being more likely to present to primary care when they contract COVID-19-like symptoms compared to those with poorer health-seeking behaviours. When we adjusted for markers of health-seeking behaviour and healthcare access, we saw a null effect representing that residual confounding had been removed. Although prior influenza vaccination may be relevant as a negative exposure for COVID-19 in the first wave of the pandemic, this may become less appropriate over time due to the strong association between influenza and COVID-19 vaccinations.

**Meaning of the study: possible explanations and implications for clinical and policymakers**

We hope the findings from this study will be useful to other observational researchers. Authors will be able to characterise the health behaviours of their population, to identify the strength and direction of this confounding from health-seeking behaviour and healthcare access on their effect estimates, and to account for any identified confounding using the markers provided. We believe that particularly for seasonal influenza and COVID-19, these markers could be helpful for future researchers to provide more accurate annual VE and cost-effectiveness estimates. Other areas of research whether we believe these markers might be important are for chronic conditions such as chronic kidney disease and diabetes, for which health-seeking behaviour and healthcare access has previously been shown to influence timeliness of seeking care and self-management of these conditions.(36, 37)

Usefulness of these markers is likely to vary by context and study type. For example, they are likely to be more useful for routine rather than pandemic VE estimates. As we saw for the COVID-19 during the pandemic after early vaccine deployment, sequential model adjustments for additional socio-demographic and health-seeking behaviour and healthcare access variables had limited impact on VE estimates. This is likely due to the high-risk perception of the virus and high testing and vaccination capacity during this time which meant that prior health-seeking behaviour and healthcare access was less influential. These markers are likely to be more useful in older cohorts as many of the markers used are relevant to individuals aged  $\geq 65$  years. Marker will have to be adapted when including younger individuals.

The descriptive results of this study are likely to be useful to clinicians and policymakers that are interested in identifying characteristics of individuals that are more likely to take up vaccinations and other nationwide programmes. We showed that individuals that take up UK nationwide screening programmes and NHS health checks are more likely to be those that get vaccinated. We also showed that individuals with previous influenza vaccinations are more likely to take up vaccinations in subsequent seasons. Governments could use this information to target groups of individuals to improve health inequality. Policy makers and clinicians can also use the findings from this study to increase their reassurance that during pandemic contexts, minimally adjusted VE estimates from observational studies, which are rapidly produced to guide decision making, are likely to be a close reflection of final VE estimates.

### **Unanswered questions and future research**

It would also be useful to understand how well these markers perform in different settings, study types and for different research questions, including designs that explicitly account for time-varying confounders(30). The rate of uptake of preventative services, including vaccinations, are likely to vary by country and therefore the usefulness of these markers is likely to differ by settings.

### **Conclusion**

We have identified markers in UK EHRs that can be used to quantify and adjust for confounding from health-seeking behaviour and healthcare access in observational research. Adjusting for health-seeking behaviour had a limited influence on estimates of COVID-19 VE during the pandemic and early stages of the vaccine roll-out. For seasonal influenza VE, severe outcomes were robust to confounding from health-seeking behaviour, but VE against influenza infections identified in primary care were underestimated prior to adjustment for health-seeking behaviour. Residual confounding was also removed as demonstrated in a negative exposure analysis of history of influenza vaccination against COVID-19 infections (pre COVID-19 vaccine roll out).

**Funding:** SG, EPKP, NA, JLW and HIM are funded by the National Institute of Health and Care Research (NIHR) Health Protection Research Unit in Vaccines and Immunisation (grant reference: NIHR200929), a partnership between UK Health Security Agency (UKHSA) and London School of Hygiene and Tropical Medicine. The views expressed are those of the author(s) and not necessarily those of the NIHR, UKHSA or the Department of Health and Social Care.

**Disclaimer:** The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or UKHSA.

**Competing interests:** SG is also a part-time salaried employee of Evidera, which is a business unit of Pharmaceutical Product Development (PPD), part of Thermo Fisher Scientific.

**Data availability statement:** These data were obtained from the Clinical Practice Research Datalink, provided by the UK Medicines and Healthcare products Regulatory Agency. The authors' licence for using these data does not allow sharing of raw data with third parties. Information about access to Clinical Practice Research Datalink data is available here: <https://www.cprd.com/research-applications>. Codelists for this study are available at <https://doi.org/10.17037/DATA.00003684> and code at <https://github.com/grahams99/Health-seeking-behaviour>.



## References

1. Kasl SV, Cobb S. Health behavior, illness behavior, and sick-role behavior. II. Sick-role behavior. *Arch Environ Health*. 1966;12(4):531-41.
2. University of Missouri. Health Care Access [Available from: <https://medicine.missouri.edu/centers-institutes-labs/health-ethics/faq/health-care-access#:~:text=Health%20care%20access%20is%20the.and%20other%20health%20impacting%20conditions>].
3. Toh S, Hernandez-Diaz S. Statins and fracture risk. A systematic review. *Pharmacoepidemiol Drug Saf*. 2007;16(6):627-40.
4. Jefferson T, Rivetti D, Rivetti A, Rudin M, Di Pietrantonj C, Demicheli V. Efficacy and effectiveness of influenza vaccines in elderly people: a systematic review. *Lancet*. 2005;366(9492):1165-74.
5. Nichol KL, Nordin JD, Nelson DB, Mullooly JP, Hak E. Effectiveness of influenza vaccine in the community-dwelling elderly. *N Engl J Med*. 2007;357(14):1373-81.
6. Nelson JC, Jackson ML, Weiss NS, Jackson LA. New strategies are needed to improve the accuracy of influenza vaccine effectiveness estimates among seniors. *J Clin Epidemiol*. 2009;62(7):687-94.
7. Horne EMF, Hulme WJ, Keogh RH, Palmer TM, Williamson EJ, Parker EPK, et al. Waning effectiveness of BNT162b2 and ChAdOx1 covid-19 vaccines over six months since second dose: OpenSAFELY cohort study using linked electronic health records. *BMJ*. 2022;378:e071249.
8. Hulme WJ, Williamson EJ, Green ACA, Bhaskaran K, McDonald HI, Rentsch CT, et al. Comparative effectiveness of ChAdOx1 versus BNT162b2 covid-19 vaccines in health and social care workers in England: cohort study using OpenSAFELY. *BMJ*. 2022;378:e068946.
9. Izurieta HS, Chillarige Y, Kelman J, Wei Y, Lu Y, Xu W, et al. Relative Effectiveness of Influenza Vaccines Among the United States Elderly, 2018-2019. *J Infect Dis*. 2020;222(2):278-87.
10. Izurieta HS, Lu M, Kelman J, Lu Y, Lindaas A, Loc J, et al. Comparative Effectiveness of Influenza Vaccines Among US Medicare Beneficiaries Ages 65 Years and Older During the 2019-2020 Season. *Clin Infect Dis*. 2021;73(11):e4251-e9.
11. Whitaker HJ, Tsang RSM, Byford R, Andrews NJ, Sherlock J, Sebastian Pillai P, et al. Pfizer-BioNTech and Oxford AstraZeneca COVID-19 vaccine effectiveness and immune response amongst individuals in clinical risk groups. *J Infect*. 2022;84(5):675-83.
12. Graham S, Walker JL, Andrews N, Nitsch D, Parker PKE, McDonald HI. Identifying markers of health-seeking behaviour and healthcare access in UK electronic health records. *medRxiv*. 2023:2023.11.08.23298256.
13. Schmid P, Rauber D, Betsch C, Lidolt G, Denker ML. Barriers of Influenza Vaccination Intention and Behavior - A Systematic Review of Influenza Vaccine Hesitancy, 2005 - 2016. *PLoS One*. 2017;12(1):e0170550.

14. Wolf A, Dedman D, Campbell J, Booth H, Lunn D, Chapman J, Myles P. Data resource profile: Clinical Practice Research Datalink (CPRD) Aurum. *Int J Epidemiol*. 2019;48(6):1740-g.
15. Herbert A, Wijlaars L, Zylbersztejn A, Cromwell D, Hardelid P. Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). *Int J Epidemiol*. 2017;46(4):1093-i.
16. Office for National Statistics. Main figures [Available from: <https://www.ons.gov.uk/>].
17. World Health Organisation. ICD-10 Version:2019 2019 [Available from: <https://icd.who.int/browse10/2019/en>].
18. NHS England. OPCS-4 CODE 2021 [Available from: [https://www.datadictionary.nhs.uk/data\\_elements/opcs-4\\_code.html](https://www.datadictionary.nhs.uk/data_elements/opcs-4_code.html)].
19. Clinical Practice Research Datalink. Small area level data based on patient postcode. 2022 [Available from: [https://cprd.com/sites/default/files/2022-02/Documentation\\_SmallAreaData\\_Patient\\_set22\\_v3.2.pdf](https://cprd.com/sites/default/files/2022-02/Documentation_SmallAreaData_Patient_set22_v3.2.pdf)].
20. Clinical Practice Research Datalink. Release Notes: CPRD Aurum May 2022 2023 [Available from: <https://cprd.com/sites/default/files/2022-05/2022-05%20CPRD%20Aurum%20Release%20Notes.pdf>].
21. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383-8.
22. Cowling TE, Ramzan F, Ladbrooke T, Millington H, Majeed A, Gnani S. Referral outcomes of attendances at general practitioner led urgent care centres in London, England: retrospective analysis of hospital administrative data. *Emerg Med J*. 2016;33(3):200-7.
23. Davidson JA, Banerjee A, Smeeth L, McDonald HI, Grint D, Herrett E, et al. Risk of acute respiratory infection and acute cardiovascular events following acute respiratory infection among adults with increased cardiovascular risk in England between 2008 and 2018: a retrospective, population-based cohort study. *Lancet Digit Health*. 2021;3(12):e773-e83.
24. UK Government. Greenbook Chapter 14a [Available from: <https://assets.publishing.service.gov.uk/media/650c0d6afbd7bc0014e54715/Greenbook-chapter-14a-4September2023.pdf>].
25. UK Health Security Agency. Weekly national flu reports: 2019 to 2020 season 2019 [Available from: <https://www.gov.uk/government/statistics/weekly-national-flu-reports-2019-to-2020-season>].
26. UK Government. Greenbook Chapter 19 [Available from: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/931139/Green\\_book\\_chapter\\_19\\_influenza\\_V7\\_OCT\\_2020.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/931139/Green_book_chapter_19_influenza_V7_OCT_2020.pdf)].
27. Mathur R, Bhaskaran K, Chaturvedi N, Leon DA, vanStaa T, Grundy E, Smeeth L. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *J Public Health (Oxf)*. 2014;36(4):684-92.

28. Jain A, Walker JL, Mathur R, Forbes HJ, Langan SM, Smeeth L, et al. Zoster vaccination inequalities: A population based cohort study using linked data from the UK Clinical Practice Research Datalink. *PLoS One*. 2018;13(11):e0207183.
29. Mansournia MA, Etminan M, Danaei G, Kaufman JS, Collins G. Handling time varying confounding in observational research. *BMJ*. 2017;359:j4587.
30. Hulme WJ, Williamson E, Horne EMF, Green A, McDonald HI, Walker AJ, et al. Challenges in Estimating the Effectiveness of COVID-19 Vaccination Using Observational Data. *Ann Intern Med*. 2023;176(5):685-93.
31. Pathirannehelage S, Kumarapeli P, Byford R, Yonova I, Ferreira F, de Lusignan S. Uptake of a Dashboard Designed to Give Realtime Feedback to a Sentinel Network About Key Data Required for Influenza Vaccine Effectiveness Studies. *Stud Health Technol Inform*. 2018;247:161-5.
32. Izurieta HS, Wu X, Lu Y, Chillarige Y, Wernecke M, Lindaas A, et al. Zostavax vaccine effectiveness among US elderly using real-world evidence: Addressing unmeasured confounders by using multiple imputation after linking beneficiary surveys with Medicare claims. *Pharmacoepidemiol Drug Saf*. 2019;28(7):993-1001.
33. Zhang HT, McGrath LJ, Wyss R, Ellis AR, Sturmer T. Controlling confounding by frailty when estimating influenza vaccine effectiveness using predictors of dependency in activities of daily living. *Pharmacoepidemiol Drug Saf*. 2017;26(12):1500-6.
34. Wang R, Liu M, Liu J. The Association between Influenza Vaccination and COVID-19 and Its Outcomes: A Systematic Review and Meta-Analysis of Observational Studies. *Vaccines (Basel)*. 2021;9(5).
35. Hosseini-Moghaddam SM, He S, Calzavara A, Campitelli MA, Kwong JC. Association of Influenza Vaccination With SARS-CoV-2 Infection and Associated Hospitalization and Mortality Among Patients Aged 66 Years or Older. *JAMA Netw Open*. 2022;5(9):e2233730.
36. Wu S, Du S, Feng R, Liu W, Ye W. Behavioral deviations: healthcare-seeking behavior of chronic disease patients with intention to visit primary health care institutions. *BMC Health Serv Res*. 2023;23(1):490.
37. Prabhakar T, Goel MK, Acharya AS. Health-Seeking Behavior and its Determinants for Different Noncommunicable Diseases in Elderly. *Indian J Community Med*. 2023;48(1):161-6.
38. Walker JL, Grint DJ, Strongman H, Eggo RM, Peppas M, Minassian C, et al. UK prevalence of underlying conditions which increase the risk of severe COVID-19 disease: a point prevalence study using electronic health records. *BMC Public Health*. 2021;21(1):484.
39. Wood ME, Chrysanthopoulou S, Nordeng HME, Lapane KL. The Impact of Nondifferential Exposure Misclassification on the Performance of Propensity Scores for Continuous and Binary Outcomes: A Simulation Study. *Med Care*. 2018;56(8):e46-e53.





