Methodological issues for using a common data model (CDM) of COVID-19 vaccine uptake and important adverse events of interest (AEIs)

Methodological issues for using a common data model (CDM) of COVID-19 vaccine uptake and important adverse events of interest (AEIs): the Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) United Kingdom feasibility study.

Gayathri Delanerolle, Robert Williams, Ana Stipancic, Rachel Byford, Anna Forbes, Sneha Anand, Declan Bradley, Ruby Tsang, Siobhán Murphy, Ashley Akbari, Stuart Bedston, Ronan Lyons, Rhiannon Owen, Jillian Beggs, Antony Chuter, Domnique Balharry, Mark Joy, Aziz Sheikh, F.D. Richard Hobbs, Simon de Lusignan

Submitted to: JMIR Formative Research
on: March 08, 2022

Table of Contents

Methodological issues for using a common data model (CDM) of COVID-19 vaccine uptake and important adverse events of interest (AEIs): the Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) United Kingdom feasibility study.

Gayathri Delanerolle[1]; Robert Williams[1] BSc; Ana Stipancic[2] BSc, MSc; Rachel Byford[1] BSc; Anna Forbes[1] MBChB; Sneha Anand[1] BSc, MSc, PhD; Declan Bradley[3] PhD, BSc; Ruby Tsang[1] BSc, MSc, PhD; Siobhán Murphy[3] BSc; Ashley Akbari[4] BSc, PhD; Stuart Bedston[4] BSc, PhD; Ronan Lyons[4] MD, BSc; Rhiannon Owen[4] BSc, PhD; Jillian Beggs[5] BSc, PhD; Antony Chuter[5] MBBS, FRCGP; Domnique Balharry[5] BSc, MSc, PhD; Mark Joy[1] BSc, PhD; Aziz Sheikh[5] MBBS, PhD, FRCGP; F.D. Richard Hobbs[1] MBBS, PhD, FRCGP, MBE; Simon de Lusignan[1] MBBS, MD, MSc, FRCGP

[1]University of Oxford Oxford GB
[2]Royal College of General Practitioners London GB
[3]Centre for Public Health, Queen's University Belfast Belfast GB
[4]Population Data Science, Swansea University Swansea GB
[5]Usher Institute, University of Edinburgh Edingburgh GB

Corresponding Author:
Simon de Lusignan MBBS, MD, MSc, FRCGP
University of Oxford
Eagle house, Walton road
Oxford
GB

Abstract

Background: The Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) UK-wide collaboration was created to monitor vaccine uptake and effectiveness and provide pharmacovigilance using routine clinical and administrative data. To monitor these, pooled analyses may be needed. However, variation in terminologies present a barrier as, England uses the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), while the rest of the UK uses the Readv2 terminology in primary care. The availability of data sources is not uniform across the UK.

Objective: To use the concept mappings in the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to identify common concepts recorded and to report these in a repeated cross-sectional study. We plan to do this for vaccine coverage and two adverse events of interest (AEIs), cerebral venous sinus thrombosis (CVST) and anaphylaxis. We identified concept mappings to SNOMED CT, Readv2, the World Health Organisation's International Classification of Disease version 10 (ICD-10) terminology and the UK's Dictionary of Medicines and Devices (dm+d).

Methods: Exposures and outcomes of interest to DaC-VaP for pharmacovigilance studies were selected. Mappings of these variables to different terminologies used across the UK devolved nations health services were identified from the Observational Health Data Sciences and Informatics (OHDSI) Athena online browser. Lead analysts from each nation then confirm or add to the mappings identified. These mappings will then be used to report AEIs in a common format. We will report rates for windows of 0-2 days and 3-28 days post-vaccine every 28 days.

Results: We list the mappings between Read v2, SNOMED CT, ICD-10 and dm+d. For vaccine exposure, we found clear mapping from OMOP to our clinical terminologies, though dm+d had codes not listed by OMOP at the time of searching. We found a list of CVST and anaphylaxis codes. For CVST we had to use a broader cerebral venous thrombosis conceptual approach to include Read v2. We identified 56 SNOMED clinical terms from which we selected 47, and 15 Read v2 codes. For anaphylaxis, our refined search identified 60 SNOMED codes and 9 from Read v2, from which we selected 10 and 4 clinical terms to include in our repeated cross-sectional studies.

Conclusions: This approach enables the use of mappings to different terminologies within the OMOP CDM without the need to catalogue an entire database. However, Read v2 has less granular concepts than some terminologies such as SNOMED CT. Additionally, the OMOP CDM cannot compensate for limitations in the clinical coding system. Neither Read v2 nor ICD-10 are sufficiently granular to enable CVST to be specifically flagged. Hence, any pooled analysis will have to be at the less specific level of cerebrovascular venous thrombosis. Overall the mappings within this CDM are useful, and our method could be used for rapid collaborations where there are only a limited number of concepts to pool.

(JMIR Preprints 08/03/2022:37821)
DOI: https://doi.org/10.2196/preprints.37821

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Methodological issues for using a common data model (CDM) of COVID-19 vaccine uptake and important adverse events of interest (AEIs): the Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) United Kingdom feasibility study.

Gayathri Delanerolle[1]

https://orcid.org/ 0000-0002-9628-9245
gayathri.delanerolle@phc.ox.ac.uk

Robert Williams[1]

https://orcid.org/0000-0002-1370-8898

robert.williams@phc.ox.ac.uk

Ana Stipancic[1,2]

https://orcid.org/0000-0001-6503-8278

Ana.Stipancic@rcgp.org.uk

Rachel Byford[1]

https://orcid.org/0000-0002-4792-8995

rachel.byford@phc.ox.ac.uk

Anna Forbes[1]

https://orcid.org/ 0000-0001-7860-3782

anna.forbes@phc.ox.ac.uk

Ruby Tsang

https://orcid.org/ 0000-0002-2520-526X
ruby.tsang@phc.ox.ac.uk

Sneha Anand[1]

https://orcid.org/ 0000-0002-0538-3974
sneha.anand@phc.ox.ac.uk

Declan T. Bradley[3,4]

https://orcid.org/0000-0003-1468-1823

Declan.bradley@qub.ac.uk

Siobhán Murphy[3]
https://orcid.org/0000-0001-9765-3873
Siobhan.Murphy@qub.ac.uk

Ashley Akbari[5]
https://orcid.org/0000-0003-0814-0801
a.akbari@swansea.ac.uk

Stuart Bedston[5]
https://orcid.org/0000-0002-5635-5957
stuart.bedston@swansea.ac.uk

Ronan A Lyons[5]
https://orcid.org/0000-0001-5225-000X
r.a.lyons@swansea.ac.uk

Rhiannon K Owen[5]
https://orcid.org/0000-0001-5977-376X
r.k.owen@swansea.ac.uk

Fatemeh Torabi[5]
https://orcid.org/0000-0002-5853-4625
fatemeh.torabi@swansea.ac.uk

Jillian Beggs[6]
https://orcid.org/ 0000-0002-7591-3256
jillianbeggs@hotmail.co.uk

Antony Chuter[6]
https://orcid.org/ 0000-0002-0646-5939
chuter@me.com

Domnique Balharry[6]
https://orcid.org/ 0000-0001-7439-2105
d.balharry@ed.ac.uk

Mark Joy[1]
https://orcid.org/0000-0002-4974-3724
mark.joy@phc.ox.ac.uk

Aziz Sheikh[6]
https://orcid.org/0000-0001-7022-3056
Aziz.Sheikh@ed.ac.uk

F.D. Richard Hobbs[1]
https://orcid.org/0000-0001-7976-7172
richard.hobbs@phc.ox.ac.uk

Simon de Lusignan[1,2]
https://orcid.org/0000-0002-8553-2641
simon.delusignan@phc.ox.ac.uk

Affiliations:
1 Nuffield Department of Primary Care Health Sciences, University of Oxford
2 Royal College of General Practitioners Research Surveillance Centre
3 Centre for Public Health, Queen's University Belfast, Belfast, UK
4 Public Health Agency, Belfast, UK
5 Population Data Science, Swansea University
6 Usher Institute, University of Edinburgh

Abstract

Background: The Data and Connectivity COVID-19 Vaccines Pharmacovigilance (DaC-VaP) UK-wide collaboration was created to monitor vaccine uptake and effectiveness and provide pharmacovigilance using routine clinical and administrative data. To monitor these, pooled analyses may be needed. However, variation in terminologies present a barrier as, England uses the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT), while the rest of the UK uses the Readv2 terminology in primary care. The availability of data sources is not uniform across the UK.

Objective: To use the concept mappings in the Observational Medical Outcomes Partnership (OMOP) common data model (CDM) to identify common concepts recorded and to report these in a repeated cross-sectional study. We plan to do this for vaccine coverage and two adverse events of interest (AEIs), cerebral venous sinus thrombosis (CVST) and anaphylaxis. We identified concept mappings to SNOMED CT, Readv2, the World Health Organisation's International Classification of Disease version 10 (ICD-10) terminology and the UK's Dictionary of Medicines and Devices (dm+d).

Methods: Exposures and outcomes of interest to DaC-VaP for pharmacovigilance studies were selected. Mappings of these variables to different terminologies used across the UK devolved nations health services were identified from the Observational Health Data Sciences and Informatics (OHDSI) Athena online browser. Lead analysts from each nation then confirm or add to the mappings identified. These mappings will then be used to report AEIs in a common format. We will report rates for windows of 0-2 days and 3-28 days post-vaccine every 28 days.

Results: We list the mappings between Read v2, SNOMED CT, ICD-10 and dm+d. For vaccine exposure, we found clear mapping from OMOP to our clinical terminologies, though dm+d had codes not listed by OMOP at the time of searching. We found a list of CVST and anaphylaxis codes. For CVST we had to use a broader cerebral venous thrombosis conceptual approach to include Read v2. We identified 56 SNOMED clinical terms from which we selected 47, and 15 Read v2 codes. For anaphylaxis, our refined search identified 60 SNOMED codes and 9 from Read v2, from which we selected 10 and 4 clinical terms to include in our repeated cross-sectional studies.

Discussion: This approach enables the use of mappings to different terminologies within the OMOP CDM without the need to catalogue an entire database. However, Read v2 has less granular concepts than some terminologies such as SNOMED CT. Additionally, the OMOP CDM cannot compensate for limitations in the clinical coding system. Neither Read v2 nor ICD-10 are sufficiently granular to enable CVST to be specifically flagged. Hence, any pooled analysis will have to be at the less specific level of cerebrovascular venous thrombosis. Overall the mappings within this CDM are useful, and our method could be used for rapid collaborations where there are only a limited number of concepts to pool.

450/450

Key Words:

Systematized Nomenclature of Medicine, COVID-19 Vaccines, COVID-19, Sinus Thrombosis, Intracranial, Anaphylaxis, computerized, Medical Record Systems

Introduction

Coronavirus 2019 (COVID-19) vaccination is the best option for controlling the current pandemic, with data about uptake and pharmacovigilance therefore essential for monitoring its progress. Since COVID-19 was first identified in Wuhan, China, at the end of 2019, the virus has spread globally, with more than 190 million confirmed cases, and over 5.9 million COVID-19 related deaths as of the 28th Feburary 2022 (1, 2). Globally, most healthcare systems have opted for a vaccination strategy to protect public health by reducing the incidence, but most importantly serious outcomes leading to hospitalisation and death. Five vaccines have been approved for use in the UK by the Medicines and Healthcare products Regulatory Agency (MHRA). The first two, Pfizer-BioNTech and Oxford-AstraZeneca, have been used since December 2020 and January 2021 respectively (3, 4). Real-world data suggest that these vaccines are effective (5, 6) for preventing severe disease and death. However, there has been concern about the risk of adverse effects such as thrombotic thrombocytopenia and anaphylaxis (7, 8). It is important to be able to monitor these at scale to give power to detect potential associations with rare AEIs.

Medical record systems enable information flow beyond oragnisational boundaries. General practices with their own IT systems that record millions of patient interactions daily. A challenge for this partnership is the heterogeneity of routine primary care data due to variation in the clinical terminologies used across the four UK nations. The Data and Connectivity: COVID-19 Vaccines Pharmacovigilance (DaC-VaP) collaboration was formed to explore vaccine effectiveness, uptake and safety across the UK. England has transitioned to the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) and has not updated Read version 2 since April 2016 (9). However, the devolved nations (Scotland, Wales and Northern Ireland) remain using the Read terminology. In addition, the levels of granularity and hierarchy of the two are incompatible, making comparison of the results of any analysis a challenge. Primary care data have not been included in the Northern Ireland component of the study because systems to make anonymised primary care data available for research are under development.

The use of a common data model (CDM) could provide a solution faced by many looking to aggregate data from different sources (10,11). CDMs use logic and semantics to standardise data and enable data from different sources to be used in pooled analyses. The use of CDMs are common within clinical research, and at present, three are cited in many studies; (1) Observational Medical Outcomes Partnership (OMOP); (2) US Food and Drug Administration (FDA) Sentinel CDM; and (3) Patient Outcomes Research Institute CDM (PCORnet). (12). The OMOP CDM, the most cited of the three, OMOP enables the transformation of data from diverse observational databases into a common format using a standardised vocabulary (13). The OMOP CDM includes different data domains required for observational studies, including demographics, vaccine exposure and adverse events of interest (AEIs) relevant to this study. (14)

Other groups including the National COVID Cohort Collaborative (N3C) in the USA have faced challenges in how to achieve harmonisation between data sources, whilst this was customised and drew together data from different sources and CDMs, N3C also extensively used OMOP. (24) We carried out this study therefore to test the feasibility of using the OMOP CDM for comparisons of vaccine uptake and AEIs across the four UK nations.
Our primary aim is to assess the feasibility of using the OMOP CDM to report the incidence of exemplar AEIs following COVID-19 vaccination across the DaC-VaP collaboration and report these as repeated cross-sectional analyses. The objectives of the study include the following;

1. To test the validity of the mappings within the OMOP Athena browser to our exemplar AEIs.
2. To report a vaccine uptake rate across the UK wide stratified by age group, sex, vaccine type,

and ethnicity; with the goal of reporting a UK-wide vaccination uptake rate. We will differentiate people who have had their first and second dose.

3. To report the rates for England, Scotland, Wales, and Northern Ireland and overall for the two exemplar AEIs – cerebral venous sinus thrombosis (CVST) and anaphylaxis.

Method

Overview

We used the OMOP Athena online browser to identify mappings to SNOMED CT, the Read terminology and ICD-10. We will report vaccine exposure using SNOMED CT for vaccine administration and dm+d codes for vaccine prescriptions. Each national team validated these mappings and any differences between the terms they would use to represent each concept discussed with a decision made by consensus. We then used to create monthly reports of vaccine coverage and AEIs. We elected to use cerebral venous sinus thrombosis (CVST) and anaphylaxis as demonstration AEIs (15).

Settings

The data were drawn from the data sources of the four DaC-VaP partners.

English data:

The data from England are from the Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), one of Europe's oldest surveillance networks. It is now in its 53rd season of operation, working alongside Public Health England (PHE) (16). The RSC has been involved in monitoring studies of influenza vaccine safety. This network is active in COVID-19 research, including the PRINCIPLE trial (17). The RSC is the surveillance platform of the Oxford RCGP clinical informatics digital hub (ORCHID).(18, 19)

Scottish data:

The EAVE II data of 5.4 million people registered in general practices in Scotland tracks COVID-19 within the Scottish population. This effort has led to impactful findings used by the Scottish and UK governments to respond to the COVID-19 pandemic.

Welsh data:

The Secure Anonymised Information Linkage (SAIL) Databank is a trusted research environment (TRE), which includes primary care general practitioner (GP) records, secondary care hospital and emergency services data, along with a range of administrative, governmental, education, social care and specialist audit, register and services data. This data is also used by the National Institute for Health and Care Excellence (NICE) to shape policies that cover England and Wales. SAIL is powered by the Secure e-Research Platform (SeRP).

Northern Ireland data:

Data are accessed through the Business Services Organisation Honest Broker Service, which also uses SeRP. It has available GP registration (but not primary care clinical records), the enhanced prescribing database, emergency department attendances, hospital admissions, COVID-19 testing and the vaccine management system.

Study Design

Repeated cross-sectional report of the incidence of the AEIs in the vaccinated population in a single time interval post-vaccination.

Phase 1, validation of the OMOP mapping and creating searches for the distributed analyses:

We will search the OMOP CDM for the concepts of interest using the Athena browser. These are demographic details, COVID-19 vaccine uptake, and the AEIs. The demographic and socioeconomic status (SES) data of interest are: age and sex, SES divided into quintiles (quintile 5 being the most deprived). We will also collect data about obesity (defined as latest body mass index ≥30, or coded as obesity), smoking status (current, ex- or never smoked).

We will compare the linkages flagged by the OMOP Athena browser with those currently used across the four nations. We will report any differences and achieve a consensus as to which terms/codes will be used in each nation.

OMOP also maps to the Medical disctictionarymaps to the Medical Dictionary for Regulatory Activities (MedDRA), and if the method in this protocol became established, considerations for enabling reporting of pharmacovigilance study findings mapped to MedDRA.

Each DaC-VaP partner nation will restrict their Athena search using the “VOCAB” (vocabulary tool) to SNOMED CT or Read as relevant, and dm+d.

The medication dictionary, dm+d, is made up of a hierarchy of generic terms (termed “virtual”) and real prescribable items (termed “actual”). The dm+d use-case for the COVID-19 vaccine is set out below:

- Virtual therapeutic moiety (VTM) – top of the hierarchy, in this use case, this is COVID-19 vaccine.
- Virtual medicinal product (VMP) - This is the next level of notional product and allows vaccine types, in our use case messenger ribonucleic acid (mRNA) vaccines and their manufacturer to be distinguished from recombinant vaccines.
- Virtual medicinal product pack (VMPP) – this is the notional product pack for the medical product above, for example a six-dose multi-dose vial.
- Actual medicinal product packs (AMPP) – are the medicinal product packs distributed. and finally,
- Actual medicinal products (AMP) – are the medicinal products prescribed to an individual.

The analyst team from each nation will report if they included all the terms identified from their search of OMOP for mapping to their terminology using ATHENA and if they have added others they routinely use.

Phase 2, monthly reports, and aggregation of results

We will run these searches monthly to produce a monthly output of vaccine coverage by demographic group and report the incidence of our AEIs.

Vaccine uptake will be reported as the percentage of adults vaccinated per nation and stratified by age, sex, smoking status, and obesity. We will report two exemplars of AEIs following vaccination using two-time windows (0-2 days and 3-28 days).

Cross-sections, Exposures, and Outcomes

DaC-VaP partners will run cross-sectional studies for the previous 28 days.

The period for the first search will be from 8th December 2020 (first dose of Pfizer vaccine given in the UK), the second on 5th January. These will run in 28 day intervals (02/02/2021, 02/03/2021, 30/03/2021, 27/04/2021 through to 17th August 2021.

The cross-sections will include all individuals registered with general practices on the date of vaccination and remain registered for 28 days. The outputs will be reported in the following age-bands: <16 years old, 16 to 39 years, 40 to 64 years, and 65 years and older. Mortality in the post-vaccination period will also be reported for those with AEIs.

We will report by vaccine brand, including reporting unknown vaccines. We will presume that unknown vaccine brand for December 2020 was Pfizer-BioNTech, as the Oxford-Astra Zeneca was unavailable until January 2021 and other vaccine types later.

We aim to include a statistical a reporting ration and a disproportional analysis metrics, a proportional reporting ratio and a reporting odds ratio (ROR). We would use a Bayesian method which provides a framework to combine prior information/knowledge and data to account for conceptually transparency. Our aim would be to use $IC = \log_2 \frac{(Observed + ½)}{(Expected + ½)}$

Ethical Considerations

The DaC-VaP collaborators have individual ethical controls of their data. No data will be reported that might risk identifying individuals. Where less than five individuals are in a group, this will be reported as <5. This exercise aims to demonstrate the potential of the DaC-VaP collaboration to report outcomes of interest.

English data:

University of Oxford is compliant with the General Data Protection Regulation and the NHS Digital Data Security and Protection Policy (20). This is an approved study (Integrated Research Application (IRAS) ID 301740, Health Research Authority (HRA) Research Ethics Committee 21/HRA/2786). ORCHID meets NHS Digital's Data Security and Protection Toolkit requirements. (21).

Scottish data:

Ethical permission for this study was granted from South-East Scotland Research Ethics Committee 314 02 [12/SS/0201]. The Public Benefit and Privacy Panel Committee of Public Health Scotland 315 approved the linkage and analysis of the de-identified datasets for this project [1920-0279].

Welsh data:

This study makes use of anonymised data held in the Secure Anonymised Information Linkage (SAIL) Databank. This work uses data provided by patients and collected by the NHS as part of their care and support. All research conducted has been completed under the permission and approval of the SAIL independent Information Governance Review Panel (IGRP) project number 0911.

Northern Ireland data:

Nothern Irish data was accessed from the Business Services Organisation Honest Broker Service (HBS) who provided de-identified linked data via SeRP. All research conducted has been approved by the HBS Governance Board (HBSGB) project number 064.

Results

Study concepts within the OMOP CDM

We initially report whether the data items / clinical concepts required for the study exist within the OMOP CDM and whether there are mappings to SNOMED CT, Read or dm+d (Table 1).

Data category	Data item	OMOP Y/N	Data source
Demographic	Gender	Y	Standardised sex (gender) codes are used in OMOP CDM mapping. Date of birth and age concepts also exist.
	Age-band	Y	Date of birth and age concepts exist in OMOP.
Socioeconomic status (SES) quintile	e.g. Index of multiple deprivation (IMD), in England, Scottish Index of Multiple Deprivation (SIMD), Welsh Index of Multiple Deprivation (WIMD), and the Northern Ireland Multiple Deprivation Measure (NIMDM)	N	Doesn't exist in OMOP CDM. Can be introduced as a custom mapping in all UK databases within OMOP. To be harmonised across the DaC-VaP data partners (Quintile 5 most deprived, quintile 1 least deprived).
Other characteristics	Body Mass Index (BMI) >30 / Obesity	Y	Will be found in the Measurement table, or from a diagnosis of obesity.
	Smoking status	Y	Will be found in the Observation table.
Vaccinations	Vaccine type	Y	Will be found in the Drug, Procedure and Event Tables For England, the source codes are dm+d, or SNOMED.
	Vaccine dose	Y	In vaccine administration.
	Vaccination date	Y	Date of Event where the event is COVID-19 vaccination.
Exemplar AEIs	A) CVST B) Anaphylaxis	Y Y	Will be found in the Condition table. Mapped to SNOMED or Read. We will not include medications in this feasibility study.

Table 1: **The variables included in the common data model (CDM) conceptual mapping exercise, with counts subsequently reported monthly.**

We then reported the components by terminology (Table 2). Age does not map to the Read terminology, but this is of no practical significance. We noted that socioeconomic status (SES) only exists as a generic concept in OMOP, and that a custom mapping would be required. There was no mapping of vaccine dose (first or second) to Read. However, like age, this would not be practically important as these data are well ordered in the DaC-VaP collaborators data sources.

Primary term	OMOP Athena	ICD-10	Dm+D	SNOMEDCT	Read (v2)

	concept ID				
Cerebral venous sinus thrombosis (non-standard to standard map)	10083037	I63.6, I67.6, U07.7 (vaccine caused adverse effects) and P3.344 (CVST in hospitalised adults)	N/A	4102202	N/A
Cerebral venous sinus thrombosis (standard to non-standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	4102202	N/A
Anaphylaxis (localised)	4034658	T78.2 (anaphylactic shock unspecified)	N/A	40316757 (systemic); 42536383 (anaphylactic shock); 4294049 (sudden onset); 2084167 (allergic); 4084167 (acute allergic reaction); 441202 (non standard to standard OMOP map); 441202; 40640468 (generalised anaphylaxis)	N/A
Anaphylaxis	441202	45537000 (anaphylactic shock unspecified)	N/A	40316757 (systemic); 40640468 (generalised anaphylaxis)	N/A
Anaphylaxis (Anaphylactic shock, due to adverse effect of correct medicinal substance properly administered)	45376003	45537000 (anaphylactic shock unspecified), 19746	N/A	4254051 (drug or medicament), 441297 (adverse reaction to drug)	N/A
Anaphylaxis (Drug-induced anaphylaxis)	241937000	45537000 (anaphylactic shock unspecified)	N/A	46274027 4084168 (non-standard OMOP)	N/A
Anaphylaxis (procedure)	42537947	45537000 (anaphylactic shock unspecified)	N/A	44807057 (anaphylaxis care), 4021200 (care of patient states), 42537947 (non standard to standard OMOP map), 44807057 (standard to non-standard OMOP map)	N/A
Anaphylaxis (Anaphylaxis due to substance)	4221182	45537000 (anaphylactic shock unspecified)	N/A	4022675 (substance), 4294049 (sudden onset), 441202 (anaphylaxis), 4221182 (non-standard to standard OMOP map), 4083868 (standard to non-standard OMOP map)	N/A

Table 2: Study concepts identified within OMOP, ICD10, and any mapping to SNOMED CT, Read, dm+d

Of most importance are the AEIs. For CVST the specific concept exists within OMOP, it also exists in SNOMED CT. SNOMED concept IDs 95455008 and19522900 from CVST (concept ID 195229008). For anaphylaxis, SNOMED and the Athena browser showed 130 and 161 items,

respectively. SNOMED CT and OMOP has 16 and 15 items, respectively. Read v2 codes are generic for CVST and Anaphylaxis. Of these, those relevant to vaccination is shown in Table 3, 4, 5 and 6 England, Wales, Scotland and Nothern Ireland.

Table 3: Study concepts identified and within OMOP and any mapping to ICD 10, dm+dx, SNOMED CT or Read v2, in England

Primary term	OMOP Athena concept ID	ICD-10	dm+d	SNOMED CT	Read v2
Cerebral venous sinus thrombosis (non-standard to standard map)	10083037	I63.6, I67.6, U07.7 (vaccine caused adverse effects) and P3.344 (CVST in hospitalised adults)	N/A	4102202	N/A
Cerebral venous sinus thrombosis (standard to non-standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	4102202	N/A
Anaphylaxis (localised)	4034658	T78.2 (anaphylactic shock unspecified)	N/A	40316757 (systemic); 42536383 (anaphylactic shock); 4294049 (sudden onset); 2084167 (allergic); 4084167 (acute allergic reaction); 441202 (non standard to standard OMOP map); 441202; 40640468 (generalised anaphylaxis)	N/A
Anaphylaxis	441202	45537000 (anaphylactic shock unspecified)	N/A	40316757 (systemic); 40640468 (generalised anaphylaxis)	N/A
Anaphylaxis (Anaphylactic shock, due to adverse effect of correct medicinal substance properly administered)	45376003	45537000 (anaphylactic shock unspecified), 19746	N/A	4254051 (drug or medicament), 441297 (adverse reaction to drug)	N/A
Anaphylaxis (Drug-induced anaphylaxis)	241937000	45537000 (anaphylactic shock unspecified)	N/A	46274027 4084168 (non-standard OMOP)	N/A
Anaphylaxis (procedure)	42537947	45537000 (anaphylactic shock unspecified)	N/A	44807057 (anaphylaxis care), 4021200 (care of patient states), 42537947 (non standard to standard OMOP map), 44807057 (standard to non-standard OMOP map)	N/A
Anaphylaxis (Anaphylaxis due to substance)	4221182	45537000 (anaphylactic	N/A	4022675 (substance),	N/A

		shock unspecified)		4294049 (sudden onset), 441202 (anaphylaxis), 4221182 (non-standard to standard OMOP map), 4083868 (standard to non-standard OMOP map)	

Table 4: Study concepts identified and within OMOP and any mapping to ICD 10, dm+dx, SNOMED CT or Read v2, in Scotland

Primary term	OMOP Athena concept ID	ICD-10	dm+d	SNOMED CT	Read v2
Cerebral venous sinus thrombosis (non-standard to standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects) and P3.344 (CVST in hospitalised adults)	N/A	N/A	N/A
Cerebral venous sinus thrombosis (standard to non-standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	N/A
Cerebral vein thrombosis	45446702	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	G67A
Thrombosis of central nervous system venous sinus NOS	3534267	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	F051z
Thrombophlebitis of central nervous system venous sinuses	4100223	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	F053
Nonpyogenic venous sinus thrombosis	45456755	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	G676
Anaphylaxis (localised)	4034658	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis	441202	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (Anaphylactic shock, due to adverse effect of correct medicinal substance properly administered)	45376003	45537000 (anaphylactic shock unspecified), 19746	N/A	N/A	N/A
Anaphylaxis (Drug-induced anaphylaxis)	241937000	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (procedure)	42537947	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (Anaphylaxis due to substance)	4221182	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A

Table 5: Study concepts identified and within OMOP and any any mapping to ICD 10, dm+dx, SNOMED CT or Read v2, in Wales

Primary term	OMOP Athena concept ID	ICD-10	dm+d	SNOMED CT	Read v2
Cerebral venous sinus thrombosis (non-standard to standard map)	10083037	I63.6, I67.6, U07.7 (vaccine caused adverse effects) and P3.344 (CVST in hospitalised adults)	N/A	N/A	N/A
Cerebral venous sinus thrombosis (standard to non-standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects)	N/A	N/A	N/A
Anaphylaxis (localised)	4034658	T78.2 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis	441202	45537000 (anaphylactic shock unspecified)	N/A	N/A	SN50.11
Anaphylaxis (Anaphylactic shock, due to adverse effect of correct medicinal substance properly administered)	45376003	45537000 (anaphylactic shock unspecified), 19746	N/A	N/A	SN50110
Anaphylaxis (Drug-induced anaphylaxis)	241937000	45537000 (anaphylactic shock unspecified)	N/A	N/A	SN50.00, 14M5.00
Anaphylaxis (procedure)	42537947	45537000 (anaphylactic shock unspecified)	N/A	N/A	SN50.11, SN50.00, 14M5.00
Anaphylaxis (Anaphylaxis due to substance)	4221182	45537000 (anaphylactic shock unspecified)	N/A	N/A	SN50.11, SN50.00, 14M5.00

Table 6: Study concepts identified and within OMOP and any mapping to SNOMED CT, Read v2, dm+d or ICD-10 in Northern Ireland

Primary term	OMOP Athena concept ID	ICD-10	dm+d	SNOMED CT	Read v2
Cerebral venous sinus thrombosis (non-standard to standard map)	10083037	I63.6, 167.6, UO7.7 (vaccine caused adverse effects) and P3.344 (CVST in hospitalised adults)	N/A	N/A	N/A
Cerebral venous sinus thrombosis	10083037	I63.6, 167.6, UO7.7 (vaccine caused	N/A	N/A	N/A

(standard to non-standard map)		adverse effects)			
Anaphylaxis (localised)	4034658	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis	441202	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (Anaphylactic shock, due to adverse effect of correct medicinal substance properly administered)	45376003	45537000 (anaphylactic shock unspecified), 19746	N/A	N/A	N/A
Anaphylaxis (Drug-induced anaphylaxis)	241937000	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (procedure)	42537947	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A
Anaphylaxis (Anaphylaxis due to substance)	4221182	45537000 (anaphylactic shock unspecified)	N/A	N/A	N/A

Vaccine exposure:

COVID-19 vaccine exposure was well recorded with dm+d and SNOMED CT. The vaccine unsurprisingly was listed as a VTM at the top of the drug dictionary hierarchy, with VMPs created for each vaccine type. There were virtual, actual packs and products to match the vaccines available. Additional administration and vaccine type clinical terms were also within SNOMED (Table 7). Finally, we found a small number of vaccine administration codes within dm+d that were not mapped to OMOP.

Vaccine Brand/Generic/Admin	**Number of dm+d or SNOMED CT terms**						
	Admin	**VTM**	**VMP**	**VMPP**	**AMPP**	**AMP**	**Ingredient**
Generic COVID		1					
Generic mRNA	3						
Generic recombinant							3
Vaccine administration	1						
Covid-19 Vaccine administration	6						
Covid-19 1st dose vaccine administration	2						
Covid-19 2nd dose vaccine administration	2						
AstraZeneca (AZ)			1	4	4	1	
Moderna			1	2	2	1	
Pfizer-BioNTech (PB)			1	2	2	1	

Table 7 Covid-19 Vaccine Concepts

Discussion

Principle findings

This protocol shows a mapping method to identify codes relevant to CVST and Anaphylaxis using the OMOP CDM to link common concepts required for COVID-19 vaccine pharmacovigilance to different terminologies relevant to the UK. All of our predefined concepts were represented in the OMOP CDM. However, some like socioeconomic status, did not have specific mappings thus, custom mappings would need development. We noted local codes and curation of variables may be used to enable specificity where the concepts are less granular, especially for CVST.

Comparison with prior work

The OMOP common data model may be sub-optimal to overcome the limitation in the granularity of the coding systems used for AEIs. As well as being less granular, the Read terminology hasn't been updated formally since April 2016, so local adaptions have been undertaken in the developed UK nations to enable new conditions and treatments such as COVID-19 and vaccination to be recorded.

Conventionally, CDMs such as OMOP are used by each database, mapping their data and querying it using the script created by one of the teams. The cataloguing is carried out using applications such as White Rabbit and Rabbit in a Hat (Figure 1).

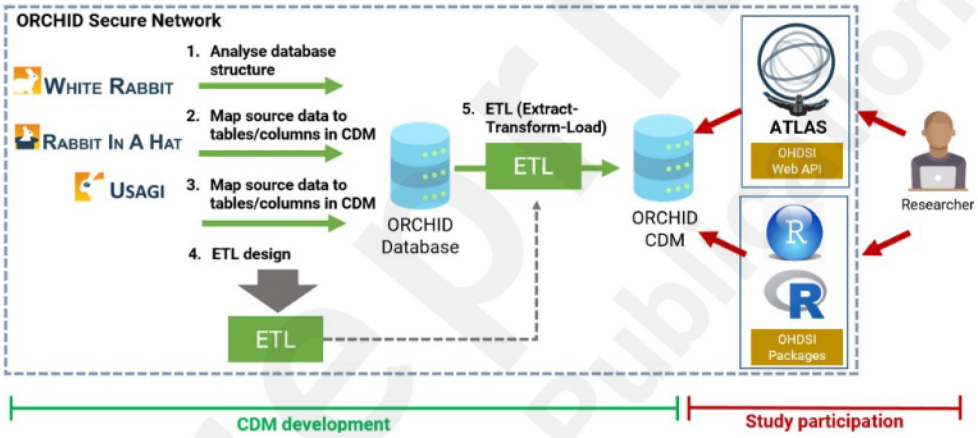

Figure 1 **Formal mapping of a database (in this case ORCHID) to OMOP-CDM. We used three steps to develop the Extract, Transform and Load (ETL). We tested this specifically in relation to the SQL environment of the ORCHID database.**

Limitations

The OMOP CDM provides a framework for capturing patient demographic and socioeconomic characteristics, varying vaccine exposure (22), and AEI data. Although others could replicate our approach, avoiding the need to map whole databases, initial findings reflect the relative size of the terminologies we use, and their granularity needs improvement. Therefore, to conduct future research, concepts will require localisation to evaluate COVID-19 vaccination associated with CVST and Anaphylaxis. We selected dm+d rather than the better known British National Formulary (BNF) although the latter is mapped to SNOMED CT. Its limitations are only lists prescribable drugs. Its chapter headings change from time to time, and is not mapped to the Athen OMOP hierarchy. We considered Medical disctionary for regulatory activities (MedDRA) with clincically-validated medical terminologies for clinical conditions, medical devices and medicines that is commonly used to share AEIs mainly with regulators. MeDRA is primaily used to report pharmacovigilance using acute care and clinical trial data. This approach cannot be directly used within primary care but will be considered for future studies.

Conclusions:

Concept mapping to a large number of terminologies, such as within OMOP and its Athena browser, are usable and valuable for those conducting studies that draw together heterogeneous data to perform pooled analyses. Comprehensive mappings have to set a level of granularity that may be more or less specific than the terminologies they map to. Clinical variable curation at a local database level would prove useful to address issues around granularity. This would allow local expert refinement of the mappings that could be used by others looking to do a limited pooled analysis of a small number of clinical concepts. The interconnectivity of the pooled analysis may also support the MHRA's Sponstaneous Report System used for optimising patient safety.

Authors' contributions:

SdeL conceived the approach in collaboration with FDRH and AS. SdeL, GD, and AS drafted versions of the protocol with input from all authors who have read and approved the paper.

Acknowledgments

Patients registered with practice members of the Oxford-RCGP RSC who allow their pseudonymised data to be shared for this research. EMIS, TPP, InPractice Systems and Wellbeing for cooperation to facilitate data extraction The RSC is principally funded by the UK Health Security Agency.

We would also like to acknowledge all data providers who make anonymised data available for research. We wish to acknowledge the collaborative partnership that enabled acquisition and access to the de-identified data, which led to this output. The collaboration was led by the Swansea University Health Data Research UK team under the direction of the Welsh Government Technical Advisory Cell (TAC) and includes the following groups and organisations: the SAIL Databank, Administrative Data Research (ADR) Wales, Digital Health and Care Wales (DHCW), Public Health Wales, NHS Shared Services Partnership (NWSSP) and the Welsh Ambulance Service Trust (WAST).

Competing interests:

SdeL reports that through his University he has had grants, from AstraZeneca, GSK, Sanofi, Seqirus and Takeda for vaccine-related research and membership of advisory boards for AstraZeneca, Sanofi and Seqirus. FDRH acknowledges part support as Director of the NIHR Applied Research Collaboration (ARC) Oxford Thames Valley, and Theme Lead of the NIHR OUH BRC. FDRH has also received occasional fees or expenses for speaking or consultancy from AZ, BI, Bayer, BMS/Pfizer, and Novartis. All other authors declare no conflicts of interest.

Abbreviations

AEI – Adverse Events of Interests (Medical conditions arising after the administration of a vaccine, not necessarily causally linked).

AMP – Actual Medicinal Product – Vaccine brand (Pfizer-BioNTech, Moderna)
AMPP – Actual Medicinal Product Pack – Distribution pack of a branded medical product

ATHENA – Automated Terminology Harmonization, Extraction and Normalization for Analytics – A repository of all the latest OMOP CDM Vocabularies and mappings are hosted and can be searched and downloaded.

ATLAS – is an open-source application developed as a part of OHDSI intended to provide a unified interface for access to patient-level data and analytics.

BMI – The body mass index (BMI) is a measure that uses your height and weight to work out if your weight is healthy.

CDM – Common Data Model.

CMR – Computerised Medical Records.

CO-CONNECT – COvid - Curated and Open aNalysis aNd rEsearCh platform, https://co-connect.ac.uk.

CPRD – Clinical Practice Research Datalink (multiple GP practices data).

CPT4 – Current Procedural Terminology Code Medical Billing is a numeric code that describes the service performed by the physician.

CTv3 – (Read) Clinical Terminology version 3.

CVST – Cerebral Venous Sinus Thrombosis.

DaC-VaP –National Core Studies - Data and Connectivity: COVID-19 Vaccines Pharmacovigilance. Pharmacovigilance is a word used to describe the science and activities relating to the detection, assessment, understanding and prevention of any side effects of a vaccine or drug.

dm+d – Dictionary of Medicines and Devices. A dictionary of descriptions and codes which represent medicines and devices in use across the NHS.

EAVE II – Early Pandemic Evaluation and Enhanced Surveillance of COVID-19 Study, a cooperation between Public Health Scotland and University of Edinburgh.

FDA – US Food and Drug Administration.

FDA Sentinel – Common Data Model used by the FDA. The Sentinel Operations Center (SOC) coordinates the network of Sentinel Data Partners and leads development of the Sentinel Common Data Model (SCDM), a standard data structure that allows Data Partners to quickly execute distributed programs against local data.

GitHub – a provider of Internet hosting for software development and version control using Git, where GIT is software for tracking changes in any set of files, usually used for coordinating work among programmers collaboratively developing source code during software development. Its goals include speed, data integrity, and support for distributed, non-linear workflows.

IMD – Index of Multiple Deprivation – Set of relative measures of deprivation for small areas (Lower-layer Super Output Areas*) across England, based on seven domains of deprivation.

MHRA – Medicines and Healthcare products Regulatory Agency.

OHDSI – The Observational Health Data Sciences and Informatics (or OHDSI, pronounced

"Odyssey") program is a multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics.

ONS – Office for National Statistics.

ORCHID – Secure data processing environment for the Oxford-Royal College of GPs Clinical Informatics Digital Hub.

OMOP – Observational Medical Outcomes Partnership, used to be a partnership project, but now it only designates a type of a common data model for Real World Data/Evidence in clinical research: https://ohdsi.github.io/TheBookOfOhdsi/.

PCORI – Patient Centred Outcomes Research Institute.

PCORnet – A Common Data Model created by PCORI to facilitate the sharing of information across Its wider network. The PCORnet Common Data Model is based on the Mini-Sentinel Common Data Model (FDA Sentinel).

PHE – Public Health England.

PRINCIPLE – Platform Randomised trial of treatmeNts in the Community for epidemic and Pandemic iLInEsses. A nationwide clinical study from the University of Oxford to find COVID-19 treatments for recovery at home (includes various medicine trials).

RCGP – Royal College of General Practitioners

Read v2 – Clinical Terminology System (5-byte version) version 2 which was developed by Dr James Read.

RSC – Research and Surveillance Centre, part of RCGP.

RWD – Real World Data, in medicine, is data derived from any number of sources that are associated with outcomes in a heterogeneous patient population in real-world settings, such as patient surveys, clinical trials, and observational cohort studies.

RxNorm – A US-specific terminology in medicine that contains all medications available on the US market. It can also be used in personal health records applications. RxNorm is part of Unified Medical Language System terminology and is maintained by the United States National Library of Medicine.

SAIL Databank – SAIL stands for Secure Anonymised Information Linkage. SAIL Databank is a Wales-wide research resource focused on improving health, well-being, and services.

SeRP – Secure eResearch Platform.SeRP is the technology platform and services that enabled the SAIL Databank and other trusted research environments and platforms in the UK and across the world.

Semantic – in case of OMOP CDM having one standard reference vocabulary per domain, so that everyone is “speaking the same language”.

SNOMED CT – Combination of Systematized Nomenclature of Medicine Reference Terminology (US) and Clinical Terms Version 3 (UK)STROBE – a methodological tool for STrengthening the Reporting of OBservational studies in Epidemiology.

TRE – Trusted Research Environment.

VMP – Virtual Medicinal Product – COVID-19 Vaccine type (mRNA or vector, etc).

VMPP – Virtual Medicinal Product Pack – COVID-19 Vaccine type + dosage.

VTM – Virtual Therapeutic Moiety – Type of therapy indicator – e.g. COVID-19 Vaccine.

References

1. World Health Organization. WHO Coronavirus (COVID-19) Dashboard [cited 2021 21st July]. Available from: https://covid19.who.int/?adgroupsurvey={adgroupsurvey}&gclid=CjwKCAjwi9-HBhACEiwAPzUhHNJHRiN1x-9-iBCLShC39mkfDoQ_3a1ricqR66a_2xrfKO7kTpGfvBoCCtsQAvD_BwE.

2. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.

3. Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet. 2021;397(10269):99-111.

4. Walsh EE, Frenck RW, Jr., Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. Safety and Immunogenicity of Two RNA-Based Covid-19 Vaccine Candidates. N Engl J Med. 2020;383(25):2439-50.

5. Lopez Bernal J, Andrews N, Gower C, Robertson C, Stowe J, Tessier E, et al. Effectiveness of the Pfizer-BioNTech and Oxford-AstraZeneca vaccines on covid-19 related symptoms, hospital admissions, and mortality in older adults in England: test negative case-control study. Bmj. 2021;373:n1088.

6. Vasileiou E, Simpson CR, Shi T, Kerr S, Agrawal U, Akbari A, et al. Interim findings from first-dose mass COVID-19 vaccination roll-out and COVID-19 hospital admissions in Scotland: a national prospective cohort study. Lancet. 2021;397(10285):1646-57.

7. Greinacher A, Thiele T, Warkentin TE, Weisser K, Kyrle PA, Eichinger S. Thrombotic Thrombocytopenia after ChAdOx1 nCov-19 Vaccination. N Engl J Med. 2021;384(22):2092-101.

8. Simpson CR, Shi T, Vasileiou E, Katikireddi SV, Kerr S, Moore E, et al. First-dose ChAdOx1 and BNT162b2 COVID-19 vaccines and thrombocytopenic, thromboembolic and hemorrhagic events in Scotland. Nat Med. 2021;27(7):1290-7.

9. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. Inform Prim Care. 2005;13(1):65-70.

10. Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med. 2014;275(6):551-61

11. Li X, Ostropolets A, Makadia R, Shaoibi A, Rao G, Sena AG, et al. Characterizing the incidence of adverse events of special interest for COVID-19 vaccines across eight countries: a multinational network cohort study. medRxiv. 2021.

12. Liyanage H, Liaw ST, Jonnagaddala J, Hinton W, de Lusignan S. Common Data Models (CDMs) to Enhance International Big Data Analytics: A Diabetes Use Case to Compare Three CDMs. Stud Health Technol Inform. 2018;255:60-4.

13. Voss EA, Makadia R, Matcho A, Ma Q, Knoll C, Schuemie M, et al. Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. J Am Med Inform Assoc. 2015;22(3):553-64.

14. Matcho A, Ryan P, Fife D, Reich C. Fidelity assessment of a clinical practice research datalink conversion to the OMOP common data model. Drug Saf. 2014;37(11):945-59.

15. Rüggeberg JU, Gold MS, Bayas JM, Blum MD, Bonhoeffer J, Friedlander S, et al.

Anaphylaxis: case definition and guidelines for data collection, analysis, and presentation of immunization safety data. Vaccine. 2007;25(31):5675-84.

16. de Lusignan S, Lopez Bernal J, Byford R, Amirthalingam G, Ferreira F, Akinyemi O, et al. Influenza and Respiratory Virus Surveillance, Vaccine Uptake, and Effectiveness at a Time of Cocirculating COVID-19: Protocol for the English Primary Care Sentinel System for 2020-2021. JMIR Public Health Surveill. 2021;7(2):e24341.

17. PRINCIPLE Trial Collaborative Group. Azithromycin for community treatment of suspected COVID-19 in people at increased risk of an adverse clinical course in the UK (PRINCIPLE): a randomised, controlled, open-label, adaptive platform trial. Lancet. 2021;397(10279):1063-74.

18. de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, et al. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract. 2017;67(663):440-1.

19. de Lusignan S, Jones N, Dorward J, Byford R, Liyanage H, Briggs J, et al. The Oxford Royal College of General Practitioners Clinical Informatics Digital Hub: Protocol to Develop Extended COVID-19 Surveillance and Trial Platforms. JMIR Public Health Surveill. 2020;6(3):e19773.

20. NHS Digital. Data security and information governance [cited 2021 21st July]. Available from: https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance.

21. Information Security [cited 2021 21st July]. Available from: https://www.infosec.ox.ac.uk/apply-to-be-made-part-of-university-of-oxford-dspt#/.

22. Duarte-Salles T, Prieto-Alhambra D. Heterologous vaccine regimens against COVID-19. Lancet. 2021;398(10295):94-5.

23. Lyons J, Akbari A, Agrawal U, Harper G, Azcoaga-Lorenzo A, Bailey R, Rafferty J,Watkins A, Fry R, McCowan C, Dezateux C, et al. Protocol for the development of the Wales Multi-morbidity e-Cohort (WMC): data sources and methods to construct a population-based research platform to investigate multi-morbidity. BMJ Open 2021;11

24. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, Payne PRO, Pfaff ER, Robinson PN, Saltz JH, Spratt H, Suver C, Wilbanks J, Wilcox AB, Williams AE, Wu C, Blacketer C, Bradford RL, Cimino JJ, Clark M, Colmenares EW, Francis PA, Gabriel D, Graves A, Hemadri R, Hong SS, Hripscak G, Jiao D, Klann JG, Kostka K, Lee AM, Lehmann HP, Lingrey L, Miller RT, Morris M, Murphy SN, Natarajan K, Palchuk MB, Sheikh U, Solbrig H, Visweswaran S, Walden A, Walters KM, Weber GM, Zhang XT, Zhu RL, Amor B, Girvin AT, Manna A, Qureshi N, Kurilla MG, Michael SG, Portilla LM, Rutter JL, Austin CP, Gersing KR; N3C Consortium. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc. 2021 Mar 1;28(3):427-443. doi: 10.1093/jamia/ocaa196

Appendix 1: Mapping of COVID-19 vaccination clinical terms to OMOP.

We list the dm+d COVID-19 vaccine codes linked to the OMOP CDM (Table A2.1), as well as SNOMED CT codes (Table A2.2) used to identify people exposed to COVID-19 vaccines. Finally, we list three vaccine codes in dm+d not listed in the OMOP Athena browser and therefore not linked to an OMOP concept (Table A2.3).

Vaccine Brand	OMOP CDM conceptID	dm+d code	COVID-19 Vaccine Name	Class
Generic	35895098	39330711000001103	COVID-19 vaccine	VTM
Generic	36122818	OMOP5047743	COVID-19 vaccine, recombinant, full-length nanoparticle spike (S) protein, adjuvanted with Matrix-M	Ingredient
Generic	36122823	OMOP5047742	COVID-19 vaccine, whole virus, inactivated, adjuvanted with Alum and CpG 1018	Ingredient
Generic	36122819	OMOP5047744	COVID-19 vaccine, recombinant, plant-derived Virus-Like Particle (VLP) spike (S) protein, adjuvanted with AS03	Ingredient
AZ	35891649	39116211000001106	Generic COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials	VMP
AZ	35895194	39301011000001100	Generic COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials 80 dose	VMPP
AZ	35895193	39301111000001104	Generic COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials 100 dose	VMPP
AZ	35891433	39114711000001108	Generic COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials 8 dose	VMPP
AZ	35891890	39114811000001100	Generic COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials 10 dose	VMPP
AZ	35895100	39301211000001105	COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials (AstraZeneca UK Ltd) 80 dose 10 x 8 dose vials	AMPP
AZ	35895099	39301311000001102	COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials (AstraZeneca UK Ltd) 100 dose 10 x 10 dose vials	AMPP
AZ	35891906	39115011000001105	COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials (AstraZeneca UK Ltd) 8 dose	AMPP
AZ	35891864	39115111000001106	COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials (AstraZeneca UK Ltd) 10 dose	AMPP
AZ	35891522	39114911000001105	COVID-19 Vaccine AstraZeneca (ChAdOx1 S [recombinant]) 5x10,000,000,000 viral particles/0.5ml dose solution for injection multidose vials	AMP
PB	35891484	39116111000001100	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech)	VMP

PB	35895192	39214411000001100	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) 1170 dose	VMPP
PB	35891695	39115311000001108	Generic COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) 6 dose	VMPP
PB	35891603	39115711000001107	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) (Pfizer-BioNTech) 6 dose	AMPP
PB	35895097	39214511000001101	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech) (Pfizer-BioNTech) 1170 dose 195 x 6 dose vials	AMPP
PB	35891709	39115611000001103	COVID-19 mRNA Vaccine BNT162b2 30micrograms/0.3ml dose concentrate for suspension for injection multidose vials (Pfizer-BioNTech)	AMP
MOD	35895190	39326811000001106	Generic COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials	VMP
MOD	35895191	39326611000001107	Generic COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials 100 dose	VMPP
MOD	36123045	39375311000001106	Generic COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials 10 dose	VMPP
MOD	36122810	39375411000001104	COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials (Moderna, Inc) 10 dose	AMPP
MOD	35895096	39327011000001102	COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials (Moderna, Inc) 100 dose 10 x 10 dose vials	AMPP
MOD	35895095	39326911000001101	COVID-19 mRNA (nucleoside modified) Vaccine Moderna 0.1mg/0.5mL dose dispersion for injection multidose vials	AMP