

Detection of vaccine-induced thrombotic thrombocytopenia with regional- and hospitallevel data: A pilot and feasibility study

 \bullet

 \bullet

 \bullet

Exploring Regional Linked Health Data Capability for Research

March 2023

Contents

Executive Summary

- There are no coordinated UK data feeds that provide detailed and near-real time information on hospital admissions or hospital activity to enable research across the UK. **Currently, nationally available acute admissions data are only accessible after discharge, some with up to a 6-week time delay. The available data lack granularity. These gaps in the UK health data research ecosystem need to be addressed.**
- Health Data Research UK (HDRUK) convened a collaborative group **("The Regional Linked Health Data Group")** under the leadership of Luke Readman, Director of Digital Transformation for NHS London, to curate regional rapid near-real time hospital admissions data feeds from **five local health systems across England and Scotland.**
- **Using identification of thrombotic adverse events associated with COVID-19 vaccination as a use case, the group identified a requirement for near-real time hospital admissions data feeds and curation of a data set across the five UK local health systems**. This work led to insights into methods for scale up for different use cases. Challenges identified were variation the availability of similar data sources across local health systems, variable and complex governance processes, and the person-resources needed at Trust and Integrated Care System (ICS) level to implement these data feeds within current infrastructure.

Recommendations

- More work is needed to improve coding. This would improve the identification of important diagnoses with greater fidelity at hospital admission, during hospital stay, and at discharge. **We recommend identifying and addressing the barriers that limit the use of the NHS-mandated¹ SNOMED coding system at the point of care.**
- Although the 'pull' factors for better coding in secondary care are strong, including use cases in health care quality improvement, automated decision systems and personalising information to patients, these are unlikely to lead to substantial change. **We recommend that point of care coding with SNOMED should be mandated, starting with digitally mature integrated care systems.**
- **We recommend that regional secure data environments (SDEs) have seamless and streamlined access to the data arising from or pertinent to their local health systems that is collated and curated nationally.**
- Governance processes differ in speed and scope between local health systems, adding delay to any project working across local health systems. We recommend that national and regional SDEs should agree common processes and trust approvals granted in one SDE across all SDEs. **We recommend a national approval process modelled on current ethics approvals, with time targets from application to approval, similar to the Integrated Research Application System.**
- • We should leverage regional expertise embedded within each ICS. **We recommend that each SDE should be supported to train and support talented analysts from across the data ecosystems within their region.**

¹ [SCCI0034: SNOMED CT -](https://gbr01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdigital.nhs.uk%2Fdata-and-information%2Finformation-standards%2Finformation-standards-and-data-collections-including-extractions%2Fpublications-and-notifications%2Fstandards-and-collections%2Fscci0034-snomed-ct&data=05%7C01%7CLara.Edwards%40hdruk.ac.uk%7Cd32f4a71f2204adf68fe08db3a5f2f10%7C844cacb1702742639d8a18faa5bf0839%7C1%7C0%7C638167952846865635%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=z64Tm4ToREsPeFYzKx%2BjgOqyIpXebAT1fQEVeNjBV0o%3D&reserved=0) NHS Digital

Authors

William Whiteley¹, Sophie Williams², Polchronis Kemos², Charles Gutteridge², Christopher Cheyne³, Iain Buchan³Gary Leeming³, Marta García-Fiñana³, Simon Weston Smith⁴, Anda Bularga⁵, Sophie McCall⁵

Nick Mills⁵, Ben Glampson⁶, Kavitha Saravanakumar⁶, Erik Mayer⁶, Suzy Gallier⁷, Sophie Morris⁸, Lara Edwards⁹, Luke Readman⁹, Cathie Sudlow⁹

- 1. Centre for Clinical Brain Sciences, University of Edinburgh
- 2. Barts Health NHS Trust, London
- 3. Institute of Population Health, University of Liverpool, University of Liverpool
- 4. Graphnet Ltd
- 5. DataLoch, University of Edinburgh
- 6. iCARE, Imperial College NHS Trust
- 7. PIONEER, University of Birmingham and University Hospitals Birmingham NHS Foundation Trust
- 8. IQVIA (formerly HDRUK)
- 9. Health Data Research UK

Abbreviations

Introduction

Population-scale epidemiological studies often use data sent by hospitals to national NHS organisations such as NHS England. These data record information (including disease diagnoses) from hospital admissions. However, these data take time to process before being available for healthcare planning, quality improvement, and research. During the COVID-19 pandemic, hospital admission data were very important; for example, they were used to make estimates of the incidence of COVID vaccination complications [1]. However, analyses using national data [2, 3] to determine association between COVID-19 vaccines and thrombotic adverse events were limited by the time lag in acute admissions data and lack of linked laboratory and radiology data flows. Although they have the substantial advantage of whole population scale, national hospital admissions data have limited depth of disease information; have variable and sometimes uncertain accuracy depending on the disease in question, e.g. under-detection of many diseases; have a lag time of weeks between data collection and data availability for analysis; and are usually not available before patients are discharged from hospital. These limitations can constrain the accuracy and speed of research for a range of public health and healthcare questions.

Data available within hospital electronic health record (EHR) systems, including data created in real time by clinicians and in laboratory and radiology systems, contain rich and varied information about disease. Rapid access to real time accurate coding of clinical data with widely used systems such as SNOMED CT, would allow timely ascertainment of new and existing diseases at larger scale and lower costs than current methods (e.g. clinical reporting systems which rely on clinician enthusiasm and time).

In this study, we tested whether we could detect the clinical syndrome of COVID-19 vaccine induced thrombotic thrombocytopenia (VITT) with greater accuracy and speed from hospital EHR data than from nationally collated datasets.

The [Data and Connectivity National Core Study,](https://www.hdruk.ac.uk/covid-19-data-and-connectivity/) led by HDRUK in partnership with the Office for National Statistics, convened a collaborative group (The Regional Linked Health Data Group), representing four local health systems across England and one in Scotland, to curate regional rapid near-real time hospital admissions data feeds. This report details the findings of Phase 1 of this work which focused on using a driver use case to explore data flows, capabilities and variances in governance processes across different local health systems. The first use case implemented was conducted as part of the NIHR-funded Thrombosis and Thrombocytopenia Consortium research programme and focused on safety of the COVID-19 vaccination programme, in particular on the ascertainment of the rare adverse event, vaccine-induced immune thrombocytopenia and thrombosis (VITT) [4].

In addition, views from the public were sought on public perception of access to and use of regional data for research and surveillance by researchers from across the UK [5]. Finally, key recommendations to enable more streamlined access to rapid, granular regional linked health data for priority research are provided and suggested next steps for potential scale up of this work are outlined.

Objectives

We aimed to test the feasibility of detecting VITT with automated algorithmic methods deployed within EHR in hospitals in near real time.

We intended to measure feasibility by:

- Determining the number and positive predictive value (PPV) of VITT cases identified with automated algorithmic methods in EHR compared with clinicians.
- Estimating the time potentially saved by using hospital EHR systems rather than nationally collated data sources, in order to assess the potential of such systems to reduce time to ascertainment of important diseases (in this case VITT).

Data Sources

We planned to perform analyses in five regional acute hospital-based secure data environments (SDEs): Barts Health NHS Trust (northeast London) [6], Combined Intelligence for Population Health Action (CIPHA) (northwest England) [7], iCare/WSIC (northwest London) [8], PIONEER Data Hub (Birmingham) [9] and DataLoch (southeast Scotland) [10].

Characteristics of each region, with respect to the data available and the processes in place to manage data access requests and information governance are detailed below:

Table 1: Characteristics of regional data sources

Analytical Report

Data Sources Analysed

PIONEER Data Hub – request for national data via NHS Digital (NHS England)

- PIONEER Data Hub does not currently have linkage to primary care data and so needed to submit a form to the Data Access Request Service (DARS) to NHS Digital (now NHS England) to obtain COVID 19 Vaccination data and Adverse Reactions for a specific cohort of patients from within PIONEER NHS Trust providers. All other data required for the analysis (ED admissions, inpatient care, discharge diagnosis, laboratory data and clinical review) were available in PIONEER for the specific cohort of patients from the outset of the study.
- PIONEER commenced the process to obtain a linked extract of the relevant vaccination/adverse events data in September 2021. This was finally approved in March 2023, at which point it was too late to complete the analysis using PIONEER data as the project had concluded.
- **Figure 1** below summarises the process and timeline. Key issues were:
	- o The burden of the DARS process is not proportionate for an NHS Trust to access data for their own population. In PIONEER's case, this was particularly burdensome, given that data was to be provided to

the PIONEER Data Hub only (run by the UH Birmingham NHS Trust) and made available to research users via the secure PIONEER Trusted Research Environment.

- \circ There were significant periods of time with no progress or communication at all from NHSD/E to the PIONEER team.
- o Some of the clarification and requests submitted to the PIONEER team between submitting the DARS request and submission of the request to Independent Group Advising on the Release of Data (IGARD) for review were repetitive and unnecessary, delaying the process even further. For example, a requirement to spell out in full all acronyms throughout the form even though all acronyms had been spelt out at the beginning of the document.
- \circ Although this work was conducted under the auspices of the [National Core Studies](https://www.gov.uk/guidance/national-core-studies-programme) i.e. urgent, high priority COVID-19 research to inform the national pandemic response – this did not result in a rapid, streamlined and prioritised approval process.
- This experience informed some of the key recommendations for this work **we recommend that each ICS receives timely data from the national datasets to which they contribute** to carry out high priority research studies of regional and national relevance.

Figure 1: Timeline of NHS Digital (now NHS England) DARS request submitted by PIONEER Data Hub

Data sources included in analysis were therefore as shown in **Table 2**.

Table 2: Data sources analysed

Methods (see Appendix 1: Protocol)

In each of the local health systems, the intention was to:

• Define the population of interest as people arriving in the emergency department (ED) from 5 to 30 days after vaccination and between 8th December 2020 up to 1st May 2022, although dates analysed varied (**Table 2**). People had to have vaccination data available, which was usually only for people registered at a GP providing data to the SDE.

² Within hospital data SNOMED coded data not available in iCARE/WSIC or CIPHA

³ ECDS only available from April 2021

- Classify people who arrived at ED as definite, probable, or possible VITT using three algorithms:
	- o *SNOMED + laboratory data*: using information collected on vaccination, laboratory systems and SNOMED-coded emergency care data to identify thrombotic events. Barts Health supplemented this data with SNOMED-coded in-hospital data.
	- o *ICD-10 + laboratory data*, where SNOMED ED data was not available, using information collected on vaccination, laboratory systems and ICD10-coded discharge data to identify thrombotic events
	- o *ICD-10 only*: Classify people who arrived in ED as definite, probable, possible, or not VITT based on an algorithm that used ICD-10 data available at discharge and vaccination data, without using laboratory data (to emulate data returned to nationally available datasets). (See protocol, Appendix 1)
- Present a selection of algorithmically-detected VITT cases (definite, probable and possible) to clinicians, who reviewed the patients' medical records and reached a clinical diagnosis of definite, probable or possible VITT. Where the case was not thought to be definite VITT, an alternative diagnosis was suggested.
- Calculate the PPV of the algorithm, i.e. the proportion of cases identified by the algorithm who had a clinical diagnosis of VITT with its 95% confidence interval.
- Calculate the number of days between the record of a code in an in-hospital system (e.g. SNOMED-coded emergency department data) and reporting of ICD-10 coded hospital discharge data to NHS Digital. Because we were unable to determine when data arrived at NHS Digital, we estimated when ICD-10 was coded in each SDE.

We use the term "algorithmically-definite" for people who reached all of the criteria in Table 3 for definite VITT.

We use the term "clinically confirmed" for people identified by algorithm where the clinical notes were reviewed by a clinician and the diagnostic certainty of VITT was defined as definite, probable or possible.

It was not possible to aggregate results between local health systems, because the available data sources and time periods of analyses were different.

The British Society of Haematologists (BSH) surveyed cases of VITT reported to them from haematology colleagues based in hospitals all over the UK during 2021. The BSH shared the numbers of VITT cases reported and their reporting hospital for the period March 22 and June 6, 2021. However, we did not obtain the identities of the BSH reported VITT cases and so could not assess whether the same, different or overlapping sets of patients with definite, probable or possible VITT were picked up by the BSH reporting system and our algorithms.

Results

Data were available from four regional acute hospital SDEs, covering populations of between 0.9 and 2.7 million people. The number of ED admissions per site for the analysed periods was between 128,000 and 195,612 people. The British Society of Haematologists had cases of VITT reported to them from three of the four sites (**Table 4**). Across the four SDEs, we identified a total of six cases of clinically definite VITT (**Table 5**).

SNOMED algorithm + laboratory data (Table 6, Table 7)

It was possible to implement a SNOMED algorithm to identify thrombotic phenotypes linked with laboratory and vaccine data in three SDEs (Barts Health, CIPHA, and WSIC). In two of these (Barts Health and WSIC), clinical validation was possible.

Using the SNOMED algorithm, Barts Health identified one case of clinically definite VITT during the study period, and one case of algorithm-definite VITT outside the study period, which was not clinically reviewed. A further four algorithm-definite cases, who were known to Barts Health haematologists (and hence likely clinically definite), were in patients not registered with a GP in north east London and would have been identified had vaccination data been available for them. The SNOMED algorithm identified two cases of clinically definite VITT in iCARE/WSIC and no cases of clinically definite VITT in CIPHA.

In Bart's Health, because of the large number of algorithmically-defined possible VITT cases a randomly selected subset (n=60) was presented to clinical validators. Clinical validators had modest inter-person agreement (Randolph's K = 0.45).

The PPV of a SNOMED algorithmically-defined 'definite VITT' for clinically-confirmed 'definite VITT' was 1.0 (95% CI: 0.05–1.00; 1/1) in Barts Health. The PPV of algorithmically-defined 'definite or probable VITT' for clinically-confirmed 'definite or probable VITT' was 0.09 (95% CI: 0.00–0.43. 1/11) in Barts Health, and 0.83 (95% CI: 0.36–0.99, 5/6) in iCARE/WSIC. The PPV of algorithmically-defined 'definite or probable or possible' VITT for clinically-confirmed 'definite or probable or possible' VITT was 0.31 (95% CI: 0.21-0.43; 22/71) in Barts Health, and 0.94 (95% CI: 0.81-0.98; 44/47) in iCARE/WSIC.

Across SDEs, of 17 cases of SNOMED algorithmically-defined 'definite or probable VITT' cases, six were clinically confirmed definite or probable VITT cases (35%).

ICD-10 algorithm + laboratory data (Table 6, Table 7)

It was possible to implement an ICD-10 algorithm to identify thrombotic phenotypes linked with laboratory and vaccine data in three SDEs (Dataloch, CIPHA, and iCARE/WSIC). In two of these (DataLoch and iCARE/WSIC), clinical validation was possible. With this algorithm, DataLoch identified two cases of clinically definite VITT, and iCARE/WSIC and CIPHA identified no clinically definite cases.

The PPV of the ICD-10 algorithm for 'definite VITT' was 1.00 (95% CI: 0.20–1.00; 2/2) in DataLoch. No algorithmicallydefined 'definite VITT' cases were identified in iCARE/WSIC. The PPV of algorithmically-defined 'definite or probable VITT' for clinically-confirmed 'definite or probable' VITT was 0.89 (95% CI 0.51-0.99; 8/9) in DataLoch and 1.00 (95% CI: 0.05–1.00; 1/1) in iCARE/WSIC. The PPV of algorithmically-defined 'definite or probable or possible VITT' for

clinically-confirmed' definite or probable or possible VITT' was 0.97 (95% CI: 0.84-1.00; 36/37) in DataLoch and 0.64 (95% CI: 0.32–0.88; 7/11) in iCARE/WSIC.

Across SDEs, of ten cases of algorithmically-defined 'definite or probable VITT' cases, nine were clinically confirmed definite or probable VITT cases (90%).

ICD-10 alone algorithm (Table 6, Table 7)

An ICD-10 alone algorithm was implemented in three SDEs (Barts Health, CIPHA, and DataLoch). With this algorithm, DataLoch found two cases of clinically definite VITT, Bart's Health one case, and CIPHA no cases. In Bart's health, this algorithm was used only in the population who had had clinical validation of the SNOMED algorithm (n=71).

In Bart's Health, the PPVs of algorithms were: for ICD-10 algorithmically-defined definite VITT for clinically-confirmed definite VITT, 0.50 (95% CI: 0.09–0.91; 1/2); algorithmically-defined 'definite or probable VITT' for clinically-confirmed 'definite or probable VITT', 0.33 (95% CI: 0.02–0.87; 1/3); algorithmically-defined 'definite or probable or possible' VITT for clinically-confirmed 'definite or probable or possible VITT, 0.33 (95% CI: 0.15–0.77; 3/9).

In DataLoch, ICD-10 codes for VITT had a PPV for: clinically-confirmed definite cases of 0.18 (95%CI 0.03, 0.52); 2/11, clinically-confirmed 'definite or probable VITT' of 0.45 (95%CI 0.18, 0.75; 5/11) and clinically-confirmed 'definite or probable or possible VITT' of 0.81 (95%CI 0.48, 0.97; 9/11).

Agreement between algorithms (Table 8, Table 9)

It was possible to directly compare the ICD-10 + laboratory algorithm and ICD-10 alone algorithm in DataLoch. Of the two ICD-10 + laboratory algorithm definite cases, one was identified with an ICD-10 code for VITT.

It was possible to compare the SNOMED + laboratory data algorithm and the ICD-10 alone algorithm in Barts Health. Of the two SNOMED + laboratory data algorithm definite cases, one was identified with the ICD-10 alone algorithm; of the 11 ICD-10 alone identified cases, 4 were identified as definite or probable by the ICD-10 + laboratory algorithm.

There was no consistent pattern in alternative diagnosis between sites. (**Table 10, Table 11**)

Delay to data provision

Two sites were able to estimate the delay from the data of data collection to the date of sending ICD-coded discharge data to nationally available datasets, although were not able to calculate this delay specifically for individuals included in the study. In CIPHA, ICD-10 coded discharge data is available seven days after the end of the month in which the data is coded. The mean delay from admission during the period of the present study was 31.4 days with a minimum and maximum delay of 6 and 175 days, respectively. In Barts Health, 95% of codes are submitted by 1 month after discharge and 100% of codes are submitted by 2 months after discharge.

Table 4. Characteristics of sites

Table 5. Clinically definite cases in each SDE by algorithm identification method (including all algorithm-defined definite, probable or possible cases)

⁴ Emergency departments & mono speciality A&E

⁵ covers 2.7 million people and multiple trusts; only data from St Helens & Knowsley was available for analysis.

⁶ The British Society of Haematology was notified by haematologists from across the UK.

 $⁷$ 1 fitting study inclusion criteria.</sup>

Table 6. The number of cases of definite, probable and possible VITT identified in each site using an algorithm of SNOMED+ laboratory data; ICD-10 + laboratory data; or ICD-10 alone.

⁸ vaccinated against COVID-19 before 31.12.21 and attended hospitals on or after 01.04. 2021

⁹ vaccinated against COVID-19 before 31 March 2021 and admitted on or before 31 March 2021

¹⁰ 2 out of the 6 SNOMED CT algorithm-identified 'probable' VITT cases subsequently confirmed as 'definite' at clinical validation (see Table 5)

¹¹ Random selection of possible cases

For clinician review, *probable VITT*: VITT most likely diagnosis but other diagnosis possible; *possible VITT*: VITT possible but other diagnoses more likely. TP: true positive

Table 7. Positive predictive value (PPV) of SNOMED+ laboratory data; ICD-10 + laboratory data; or ICD-10 alone, versus different threshold of expert VITT.

Table 8. Agreement between' ICD-10 alone' ' and 'ICD-10 and laboratory' algorithms for VITT in DataLoch

Table 9. Agreement between ICD-10 alone algorithm for VITT and ICD-10 and laboratory algorithm in Barts Health

Table 10. Of cases identified by the algorithm using the SNOMED coded phenotypes with probable or possible VITT, alternative diagnosis given by clinicians reviewing data in iCARE/WSIC

Table 11. Of cases identified by the algorithm using the SNOMED coded phenotypes with probable or possible VITT, alternative diagnosis given by clinicians reviewing data in Barts Health

Discussion

From data sources covering a total population of around 8.5 million people, we were able to identify three clinicallydefinite cases of VITT 5 to 30 days after COVID vaccination with a SNOMED and laboratory data algorithm, and three clinically-definite cases of VITT with an ICD-10 code only algorithm that emulates data that would have been available in a national NHS datasets.

An algorithmic definition of 'definite, probable or possible VITT' using data available in near real time from point of care coding in hospitals (i.e., SNOMED ECDS +/- SNOMED in hospital + laboratory + vaccination data) had either modest (in Barts Health) or excellent (in WSIC/iCARE) precision for a clinician diagnosis of 'definite or probable or possible VITT'

Hence, we have demonstrated, albeit at relatively small scale and only across four SDEs, that for some conditions (in this case the new, rare condition of VITT), it is possible to establish automated alerts of potential cases for clinicians to follow up on to confirm clinical diagnoses that might otherwise be delayed or missed if relying on nationally available data. In the case of VITT, the data we have generated to date from the SDEs able to support clinical validation suggest that many cases of algorithmically-detected 'definite or probable' VITT were deemed clinically 'definite or probable'. What is less certain is: how many cases might be missed by this near real time approach (i.e. the sensitivity of the detection method), even where point of care SNOMED coding is available from in-hospital as well as ED data feeds (true for only one of our sites, Barts Health); how many additional cases might emerge after hospital discharges have been ICD-10-coded; and whether further true cases might be detected through manual reporting mechanisms such as those adopted by the British Society of Haematology. Without triangulation and comparison across all these different mechanisms for case detection, which were not possible during this study, these remain unanswered questions. A further important point is that many acute hospitals across the UK have not yet implemented point of care SNOMED coding in clinical practice and within their EHR systems, so currently this type of approach is only possible for subset of hospitals/local health systems.

It is difficult to recommend a universal strategy to identify as many clinically definite cases as possible with as little clinical validation time as possible, because this study is limited by the small number of cases identified, and the apparent difference (albeit with wide bounds of uncertainty) of algorithm performance between SDEs. However, the

limited available information would suggest that to identify as many definite cases as possible with a yield of around 10% after clinical review, a strategy using an algorithm to identify definite or probable VITT could be adopted.

The mean delay between discharge and provision of data to nationally available datasets in the sites, where it was assessed, was about a month, although in some cases it was considerably longer (up to 175 days). This does not take into account the delay between admission and discharge, which can be up to several months.

Our analyses had a number of limitations:

- We were unable to link people with VITT identified by the BSH to either data collected nationally by NHS Digital or data in our four SDEs, and so we could not assess how the larger number of cases of definite VITT identified by the BSH (n=220) had been coded across the country and whether they would have been detected by our data-driven algorithms.
- We were not able to identify anti-platelet factor 4 in laboratory records in some sites.
- Each site had a slightly different data governance model, which took considerable time to navigate. One region (PIONEER, Birmingham) had to obtain vaccine data from NHS Digital, a process that took over a year; as a result, data from that region could not be included in this report, as it was not available when final analyses for this report were run.
- There was variation in the availability of different data types between sites; only one site (Barts Health) could make inpatient SNOMED data available, although it is collected in many hospitals. Widespread use of SNOMED CT coding very early in the patient journey (through ECDS) is available nationally through NHS England, although it is unclear whether potentially more accurate inpatient coding is - or could be - available to SDEs or national bodies. This is important, because the only way that any analytic system could give results in real time is if diagnoses are coded in some form in real time.
- Sharing analytic code directly between sites was challenging, because different sites had different preferred or available analysis coding languages and different ways of working. Translation was possible, but time consuming.
- There was some drift in the interpretation and application of the pre-specified protocol between sites, largely due to variation in availability of data between sites.
- We were unable to calculate disease incidence, because of the difficulties of defining a population at risk based on acute hospital rather than population-based data systems.

Unfortunately, our current assessment is that near real time ascertainment of many diseases of public health importance (for example in a resurgent or new pandemic) using data from across multiple acute hospital sites' SDEs is not yet feasible at scale although it may be possible in some digitally mature regional SDEs. Further developments in clinical coding practices, digital maturity, inter-regional interoperability and streamlined data access processes are required to enable scalable analyses across local health systems that are responsive to the needs of population health.

Recommendations

Given the problems and limitations we have identified in this project, we make the following recommendations to improve working across regional SDEs:

- More work is needed to improve coding, to identify important diseases with greater fidelity at hospital admission, during hospital stay, and at discharge. **We recommend identifying and addressing the barriers to using the NHSmandated¹² SNOMED coding system much more widely at the point of care.**
- Although the 'pull' factors for better coding in secondary care are strong, including use cases in health care quality improvement, automated decision systems and personalising information to patients, these are unlikely to lead to substantial change. We **recommend that point of care coding with SNOMED should be mandated, starting with digitally mature integrated care systems.**
- It is time consuming for regional data environments to apply for and obtain regional data curated by national data custodians. **We recommend that regional SDEs should have seamless and streamlined access to the data arising from or pertinent to their local health systems that is collated and curated nationally.**
- Governance processes differ in speed and scope between local health systems, adding delay to any project working across SDEs. We recommend that national and regional SDEs should agree common processes and trust approvals granted in one SDE across all SDEs. **We recommend a national approval process modelled on current ethics approvals, with time targets from application to approval, similar to the Integrated Research Application System.**
- We should leverage regional expertise embedded within each Integrated Care System e.g build relationships between population health groups, data groups, research and standardize ways of working etc. **We recommend that each SDE should be supported to train and support talented analysts from across the data ecosystems within their region and nationally**

¹² [SCCI0034: SNOMED CT -](https://gbr01.safelinks.protection.outlook.com/?url=https%3A%2F%2Fdigital.nhs.uk%2Fdata-and-information%2Finformation-standards%2Finformation-standards-and-data-collections-including-extractions%2Fpublications-and-notifications%2Fstandards-and-collections%2Fscci0034-snomed-ct&data=05%7C01%7CLara.Edwards%40hdruk.ac.uk%7Cd32f4a71f2204adf68fe08db3a5f2f10%7C844cacb1702742639d8a18faa5bf0839%7C1%7C0%7C638167952846865635%7CUnknown%7CTWFpbGZsb3d8eyJWIjoiMC4wLjAwMDAiLCJQIjoiV2luMzIiLCJBTiI6Ik1haWwiLCJXVCI6Mn0%3D%7C3000%7C%7C%7C&sdata=z64Tm4ToREsPeFYzKx%2BjgOqyIpXebAT1fQEVeNjBV0o%3D&reserved=0) NHS Digital

References

- 1. Whiteley WN, Ip S, Cooper JA, Bolton T, Keene S, Walker V et al. Association of COVID-19 vaccines ChAdOx1 and BNT162b2 with major venous, arterial, or thrombocytopenic events: A population-based cohort study of 46 million adults in England. *PLoS Med.* 2022;19.<https://doi.org/10.1371/journal.pmed.1003926>
- 2. 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. *Nature Medicine,* 27, July 2021 1290-1297. <https://doi.org/10.1038/s41591-021-01408-4>
- 3. Hippisley-Cox J, Patone M, Mei XW, Saatci D, Dixon S, Khunti K et al. Risk of thrombocytopenia and thromboembolism after covid-19 vaccination and SARS-CoV-2 positive testing: self-controlled case series study. BMJ 2021;374:n193[1 https://doi.org/10.1136/bmj.n1931](https://doi.org/10.1136/bmj.n1931)
- 4. Pavord S, Scully M, Hunt BJ, Lester W, Bagot C, Craven B, Rampotas A, Ambler G, Makris M. Clinical Features of Vaccine-Induced Immune Thrombocytopenia and Thrombosis. *N. Engl. J. Med.* 2021;385:1680–1689. <https://doi.org/10.1056/NEJMoa2109908>
- 5. HDRUK Patient and Public Involvement and Engagement Report: Exploring public perspectives on enabling access to regional health data to support COVID-19 Vaccine safety research. February 2022. [https://www.hdruk.ac.uk/wp-content/uploads/2022/02/Vaccines-and-Acute-Admisison-PPIE-Workshop-](https://www.hdruk.ac.uk/wp-content/uploads/2022/02/Vaccines-and-Acute-Admisison-PPIE-Workshop-Report.pdf)[Report.pdf](https://www.hdruk.ac.uk/wp-content/uploads/2022/02/Vaccines-and-Acute-Admisison-PPIE-Workshop-Report.pdf)
- 6. Discovery Data Services/Barts Health NHS Trust website: <https://www.discoverydataservice.org/Content/Home.htm> (accessed 6th January 2023)
- 7. Combined Intelligence for Population Health Action (CIPHA) website[: https://www.cipha.nhs.uk/](https://www.cipha.nhs.uk/) (accessed 6th January 2023)
- 8. iCare/WSIC website:<https://imperialbrc.nihr.ac.uk/facilities/icare/> (accessed 6th January 2023)
- 9. PIONEER Data Hub website:<https://www.pioneerdatahub.co.uk/>(accessed 6th January 2023)
- 10. DataLoch website:<https://dataloch.org/>(accessed 6th January)

Appendix 1

PROTOCOL 18th May 2022

Rapid ascertainment of COVID vaccination complications with hospital admissions data: pilot project

AIM

To determine the feasibility of rapid ascertainment of COVID vaccine related complications from data obtained during hospital admission compared with hospital discharge.

RESEARCH QUESTIONS

- 1. What is the agreement between vaccine-related complications ascertained with hospital data (ED, pathology and inpatient data), data provided by NHS Digital (data at discharge) and clinician validation?
- 2. What is the average difference in time from vaccination to the identification of a complication using data from ECDS (or data from hospital admission where available) versus using hospital discharge data only?

STUDY DESIGN

Inter-observer agreement between different data sources and estimation of case numbers

DATA SOURCES

- CIPHA (lab data from Liverpool University Hospitals NHS Trust) available in SNOMED
	- o SNOMED records from ED
	- o Platelet and fibrin degradation products or D-Dimer
	- o ICD-10 codes at discharge
- PIONEER
	- o SNOMED records from ED and inpatient stay
	- o Platelet and fibrin degradation products or D-Dimer
	- o ICD-10 codes at discharge
- **DISCOVERY**
	- o SNOMED records from ED
	- o SNOMED from within hospital stay
	- o Platelet and fibrin degradation products or D-Dimer
	- o ICD-10 codes at discharge
- **DATALOCH**
	- o Emergency department biobank (ICD-10 codes) admission diagnosis
	- o Platelet and fibrin degradation products or D-Dimer
	- o Admission and in-patient radiology reports
	- o ICD-10 codes at discharge

POPULATION OF INTEREST

Time period: from 8/12/2020 to data up to 1/5/2022

1. Admitted to hospital and discharged during time period

2. Received any one of the COVID vaccines 90 days prior to admission

OBJECTIVES

- **1. Define phenotypes of interest using within hospital and hospital discharge data, and calculate numbers in each data source**
- **2. Measure agreement between different sources**
- **3. Calculate delay in detecting VITT and other phenotypes**

BACKGROUND

As of August 2021, the COVID vaccines available in the UK are made by AstraZeneca, Pfizer, and Moderna. The side effects that have been reported to the MHRA through the Yellow Card scheme are¹³:

- Oxford AstraZeneca anaphylaxis, vaccine induced immune thrombotic thrombocytopenia (VITT), Guillain Barre syndrome/Miller Fisher syndrome, Bell's palsy, capillary leak syndrome, menstrual disorders and unexpected vaginal bleeding;
- Pfizer: myocarditis and pericarditis, Guillain-Barre syndrome, swelling of vaccinated limb;
- Moderna: myocarditis and pericarditis, delayed hypersensitivity reactions, Guillain-Barre syndrome, facial swelling in people with dermal fillers.

For the purpose of these analyses, we will concentrate on the syndrome of VITT, and - because they have few codes associated with them, Guillain-Barre syndrome, Bell's palsy and myocarditis/pericarditis. Clearly there other potential side effects are of importance but their pathway through the emergency department is less clear (menstrual bleeding or monitoring of miscarriage) or they are more difficult to identify from codes (facial swelling in people with dermal fillers, capillary leak syndrome).

We propose to define phenotypes as ascertained:

- 'prior to discharge': e ents recorded in hospital-based ED or medical admission records, with the date first ascertained as first date a phenotype can be assigned, and date of onset as date of admission or ED attendance
- 'after discharge': events recorded at hospital discharge and returned to NHS Digital for inclusion in HES/SUS statistics, with date ascertained as 'date of discharge' or 'date returned to NHS Digital' and 'date of onset' as date of admission or ED attendance

We will use the code lists in the appendix to define the conditions of interest.

ANALYSIS 1: to measure agreement between different sources and expert opinion

- 1. **Define phenotypes of interest:** Within each data source, define phenotypes of interest from within hospital data, hospital discharge data and expert review of medical records (see appendix for coded definitions using within hospital and at discharge data) both 'prior to discharge' and 'after discharge'.
- 2. **Present cases to experts:** We will present all definite and probable cases, and a random subset of possible cases to the experts, defined both 'prior to discharge' and 'after discharge'. Using a case report form based on standard diagnostic criteria,¹⁴ experts will define each case as fitting the case definition or not. Where more

¹³ https://www.gov.uk/government/publications/coronavirus-covid-19-vaccine-adverse-reactions/coronavirus-vaccine-summary-of-yellow-card-reporting

¹⁴ DOI: 10.1056/NEJMoa2109908

than one expert is available, depending on expert time, at least 15% subset of cases will be double annotated to measure agreement between experts. Experts should mark each case as:

In case of b,c,or d give alternative diagnosis:

3. *Expert panel agreement in the diagnosis of each phenotype with (a) within hospital records and (b) hospital discharge records.*

We will calculate the positive predictive value (PPV) (=precision), or the proportion of cases identified by each method where an expert agrees with the diagnosis of VITT, using the expert as the gold standard for a VITT diagnosis.

We will calculate 6 PPVs: definite or probable VITT, and definite or probable or possible VITT ascertained in 'prior to discharge'; 'after discharge' ; and 'prior to discharge' OR 'after discharge'

Calculate PPV for ascertainment as a/a+b+c

Calculate agreement (with Cohen kappa) between 'prior to discharge' and 'after discharge'.

ANALYSIS 2: Delay in detecting phenotypes.

To calculate the number of cases of each phenotype using within hospital data, and – separately - at discharge data, and where recorded in both datasets, the number of days between the records.

PHENOTYPE DEFINITION AND CODE LISTS

We will study 4 phenotypes defined in and out of hospital

- 1. VITT
- 2. Myocarditis/pericarditis
- 3. Bell's palsy
- 4. Guillain-Barre syndrome

For each phenotype, we will have a definition using within-hospital data and a definition using hospital-discharge data. The general principle is that the within-hospital phenotypes are defined using the first recorded SNOMED code, and the hospital-discharge are defined using an ICD-10 code at discharge.

VITT¹⁵

Definition using within hospital data

Onset time: defined as days from vaccination to date of admission

Thrombosis, defined as one or more of the following during admission with SNOMED (or other available system) in ECDS or in hospital SNOMED or other coding (see TABLE 1)

¹⁵ DOI: 10.1056/NEJMoa2109908

- Transient ischaemic attack
- Arterial embolus and thrombosis
- Cerebral haemorrhage
- Acute non-ST segment MI
- Subarachnoid haemorrhage
- Intracranial venous thrombosis
- Central retinal vein occlusion
- Cerebrovascular accident
- Upper limb ischaemia
- Lower limb ischaemia
- Amaurosis fugax
- Arterial embolus and thrombosis
- Central retinal artery occlusion
- Deep vein thrombosis
- Pulmonary embolism

Platelet count: defined as the first count below threshold during admission

D-dimer: first count above threshold of interest

Anti-APF-4 antibodies – defined as present or absent at any time

Definition using ICD-10 hospital discharge data

Cases to be defined as 'possible' if:

• *Onset time*: defined as days from vaccination to date of admission <30 days

AND ONE OF:

- *Thrombosis*, defined as one or more venous or arterial thrombosis with ICD-10 at discharge any position (TABLE 2) OR
- *Thrombocytopenia*: defined as thrombocytopenia of any cause in any position at hospital discharge TABLE 2)

Cases to be defined as 'probable' if all of the following are present:

Onset time: defined as days from vaccination to date of admission <30 days

Thrombosis, defined as one or more venous or arterial thrombosis with ICD-10 at discharge any position (TABLE 2)

Thrombocytopenia: defined as thrombocytopenia of any cause in any position at hospital discharge TABLE 2)

Cases to be defined as 'definite' if all of the following are present:

Onset time: defined as days from vaccination to date of admission <30 days

COVID vaccine complication in same admission, defined as presence of U07.7

Thrombosis, defined as one or more venous or arterial thrombosis with ICD-10 at discharge any position (TABLE 2)

Thrombocytopenia: defined as thrombocytopenia of any cause in any position at hospital discharge TABLE 2)

GUILLAINE-BARRE SYNDROME

- o Any Guillain Barre code during admission, with date of onset defined as date of admission/ED attendance (TABLE 1)
- o Days of onset from vaccination calculated
- o Within-hospital phenotypes defined with first recorded in hospital code; hospital-discharge phenotype defined with ICD-10 code at discharge. (TABLE 2)

BELL'S PALSY

- \circ Any Bell's palsy code during admission, with date of onset defined as date of admission/ED attendance (TABLE 1)
- o Days of onset from vaccination calculated
- o Within-hospital phenotypes defined with first recorded in hospital code; hospital-discharge phenotype defined with ICD-10 code at discharge (TABLE 2)

MYOCARDITIS/PERICARDITIS

- o Any myocarditis/pericarditis code, with date of onset defined as date of admission/ED attendance (TABLE 1)
- o Within-hospital phenotypes defined with first recorded in hospital code; hospital-discharge phenotype defined with ICD-10 code at discharge. (TABLE 2)
- o Days of onset from vaccination calculated

TABLE 1 SNOMED codes in ECDS

Thrombosis

Myocarditis

TABLE 2

ICD-10 codes for thrombotic events

Guillain Barre syndrome

Guillain Barre syndrome

G61 Guillain Barre syndrome