Informing better use of Welsh hospital admissions data in the SAIL Databank:

A review of the clinical coding completeness and lag in the Patient Episode Database for Wales (PEDW)

Authors: Victoria Best, Ashley Akbari, Stuart Bedston, Gareth Davies, Emily Lowthian, Jane Lyons, Fatemeh Torabi, Ronan A Lyons

Date: January 2022

Key Findings

- Prior to April 2020 the daily average number of episodes was 3,167, which subsequently reduced between April 2020 and July 2021 to a daily average of 2,292 (-28%);

- An ICD-10 completion rate of 96% was achieved 6 months after the latest data extraction;

- An OPCS-4 coding completion rate of 60% was achieved 6 months of the latest data extraction;

- All ICD-10 and OPCS-4 chapters demonstrate some retrospective coding, but there is some variation between chapters in terms of the extent.

Background

Evaluating the impact of COVID-19 on clinical coding completeness and lag within hospital admissions data

"To free up enough capacity to deal with the initial peak of the pandemic, the NHS was forced to shut down or significantly reduce many areas of non-COVID-19 care during April, May and June 2020. This, combined with fewer patients seeking care during lockdown, means that there has been a significant drop in elective procedures, urgent cancer referral, first cancer treatments and outpatient appointments" (BMA, 2020). This will impact not only the overall number of episodes recorded, but the nature of those episodes, as elective procedures are likely to have been more heavily impacted than urgent Emergency Department (ED) care.

COVID-19 and clinical coding

Emergency ICD-10 codes for recording COVID-19 admissions were activated by World Health Organisation (WHO) in February 2020 (WHO, 2021) but these weren't made available in the UK until 26th March 2020 (NHS Digital, 2020). There was a requirement for any episodes prior to this date to be retrospectively coded but it is not possible to know how completely this was carried out both across Wales, and within specific health boards or organisations by clinical coders.

Between the introduction and availability of the COVID-19 ICD-10 codes in late March and 30th June 2020, NHS Digital advised that all episodes which included COVID-19 were required to have COVID-19 recorded as their primary diagnosis (NHS Digital, 2021). From 1st July 2020, this process was changed to COVID-19 being recorded as the secondary diagnosis (unless it was in fact the primary). Any analyses investigating only primary ICD-10 codes are therefore likely to be impacted by the standards changes, not only when investigating COVID-19, but other diagnoses where COVID-19 was present during the episode.

Clinical coding targets:

The target of 95% completion for clinical codes was retired from the NHS delivery framework for 20/21, meaning that the target was removed from 1st April 2020 (Welsh Government, 2020). This may have further reduced the completion rates for all clinical coding. Despite this change, it is likely that coding relating to COVID-19 episodes will have remained high due to the severity and impact of COVID-19 on patient care and national reporting. This may skew the picture when comparing COVID-19 coding completeness and episodes against non-COVID-19 coding completeness and episodes. It is not known whether certain conditions and their respective codes or coding chapters were and are being prioritised, and if this is being completed systematically across Wales or at the discretion of clinical coders in each health board/organisation.

How are hospital admissions recorded?

Health Boards download standardised data from all hospital Patient Administration Systems (PAS) and data is received for Welsh patients treated in England through the NHS switching service (ESCR, 2013). These data for all the health boards are then made available to the Secure Anonymised Information Linkage (SAIL) Databank from Digital Health and Care Wales (DHCW) via the standard governance, anonymisation and linkage processes established for over 10 years.

The Patient Episode Database for Wales (PEDW), which is available in SAIL, holds hospital data for inpatient and day cases within NHS Wales as well as for Welsh residents treated in English Trusts since 1991 (DHCW, 2021). Following the discharge of a patient, clinical coders translate the handwritten patient notes into ICD-10 and OPCS-4 codes (HDR UK, 2020).

What is PEDW used for?

The data held within PEDW can "provide information regarding both health service utilisation and also the incidence and prevalence of disease" (HDR UK, 2020). PEDW is used by multiple projects and users within the SAIL Databank for both COVID-19 and non-COVID-19 research and intelligence. A full list of projects accessing SAIL and PEDW can be seen on the SAIL website: https://saildatabank.com/saildata/projects-using-sail/.

PEDW data contained can be linked anonymously and securely following appropriate governance and approvals being received by a project/users, and linked to other data sources to provide greater insight into the patient pathway and longitudinal data analysis at a population-scale.

PEDW structure and content

PEDW holds individual-level population-scale records based on finished consultant episodes, defined as periods of continuous care during which an admitted patient is in the care of one consultant within one NHS provider. If there is a change in consultant or NHS provider then a new consultant episode is initiated (DHCW, 2021). Unfinished consultant episodes are also available but are subject to change pending the episode being

finished and updated in a future extract refresh from the health board systems to DHCW and SAIL.

Multiple diagnostic and operative procedure codes can be recorded within each episode. Within PEDW ICD-10 diagnostic and OPCS-4 operational procedure codes are used (ESCR 2013, p.25).

There are multiple possible ways to count PEDW data, including by super spells (also known as continuous inpatient spells (CIPs) in other UK nations), spells, episodes, diagnoses or procedures.

There is some variation in coding practices, particularly in relation to secondary diagnoses, which can complicate regional comparison. The Corporate Health Improvement Programme implemented by DHCW is aiming to increase consistency in this area (HDR UK, 2020).

What did we do?

Our main aim was to understand what impact COVID-19 has had on clinical coding completeness, as well as any impact lag on specific codes or coding chapters. Doing this will improve the understanding and use of PEDW for all those looking to use it in their research.

We looked to answer the following questions:

1. Has all clinical coding completeness changed since the COVID-19 outbreak?
2. Has clinical coding completeness changed in only certain coding chapters?
3. Is there any interpretable or consistent lag within certain chapters?

Using ICD-10 and OPCS-4 codes, clinical coding of records within PEDW were investigated by comparing records prior to, and post the start of the COVID-19 pandemic. We defined post COVID-19 pandemic as 1st April 2020 onwards.
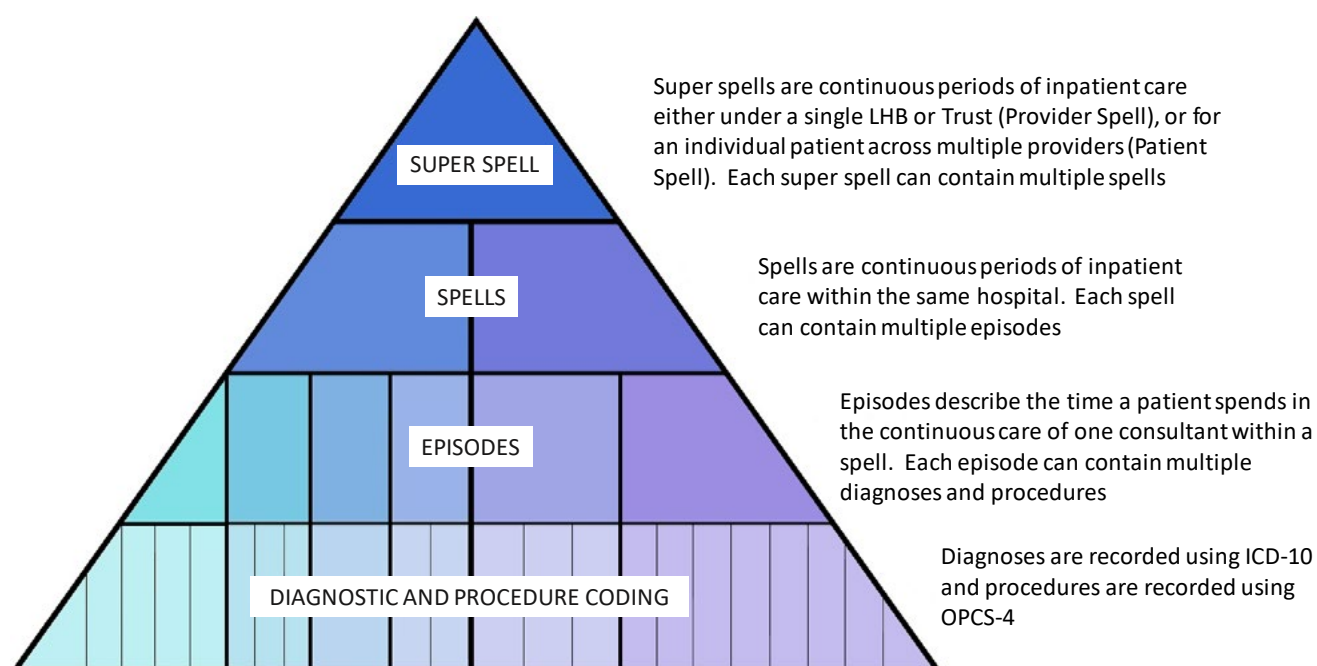

Super spells are continuous periods of inpatient care either under a single LHB or Trust (Provider Spell), or for an individual patient across multiple providers (Patient Spell). Each super spell can contain multiple spells

Spells are continuous periods of inpatient care within the same hospital. Each spell can contain multiple episodes

Episodes describe the time a patient spends in the continuous care of one consultant within a spell. Each episode can contain multiple diagnoses and procedures

Diagnoses are recorded using ICD-10 and procedures are recorded using OPCS-4

SUPER SPELL

SPELLS

EPISODES

DIAGNOSTIC AND PROCEDURE CODING

Figure 1: Visual representation of the relationship between PEDW Super spells, Spells, Episodes and clinical diagnosis and procedure codes.

Our analysis incorporated all finished consultant episodes (FCEs) ending between April 2016 and July 2021. Unless otherwise stated, results shown relate to the number of distinct FCEs and coding refers to the primary ICD-10 or OPCS-4 code for each episode. Note the primary ICD-10 and OPCS-4 code is denoted as the first code recorded on the episode i.e. position one or code one, where the first position was NULL we did not check the second position. It should be noted that a single hospital spell may contain multiple episodes with different primary diagnosis and operational coding.

The following data inclusion criteria were applied (see definitions section for further information):

- Episodes with sufficiently robust data linkage (i.e. NHS number passed check digit test; Surname, First Name, Postcode and Sex matched exactly to Welsh Demographic Service Dataset (WDSD); or fuzzy matching probability >= 0.9).

- Episodes whose primary diagnosis and operational procedure code were not listed as sensitive, as defined within the currently known sensitive ICD-10 Codes in the SAIL concept library https://conceptlibrary.saildatabank.com/concepts/689/detail/ and OPCS-4 Codes lists https://conceptlibrary.saildatabank.com/concepts/691/detail/.

- Episodes where the care provider was either one of the seven Welsh Health Boards or Velindre NHS Trust.

- Episodes where the mortality data available in SAIL from the Annual District Death Extract (ADDE) based on Office for National Statistics (ONS) mortality register, indicated a date of episode start date prior to date of death.

Unfinished consultant episodes are not included in this study because the records are subject to further change and therefore potentially incomplete. In total there were 8,051 unfinished episodes available, with the earliest from 2001. During the analysis period from April 2016, there were a total of 7,995 unfinished episodes, representing 0.1% of episodes.

Central seven day rolling averages were plotted to smooth daily variation which might be seen for example during weekends or on public holidays. For this an average was calculated each day based on the value for that day, the three preceding and three subsequent days.

What did we find?

Total Episodes

The number of episodes recorded in PEDW per month has seen a consistent reduction following the COVID-19 outbreak, with the lowest number of episodes occurring on 13th April 2020.

Prior to April 2020 the daily average number of episodes was 3,167, which subsequently reduced between April 2020 and July 2021 to a daily average of 2,292 (-28%).
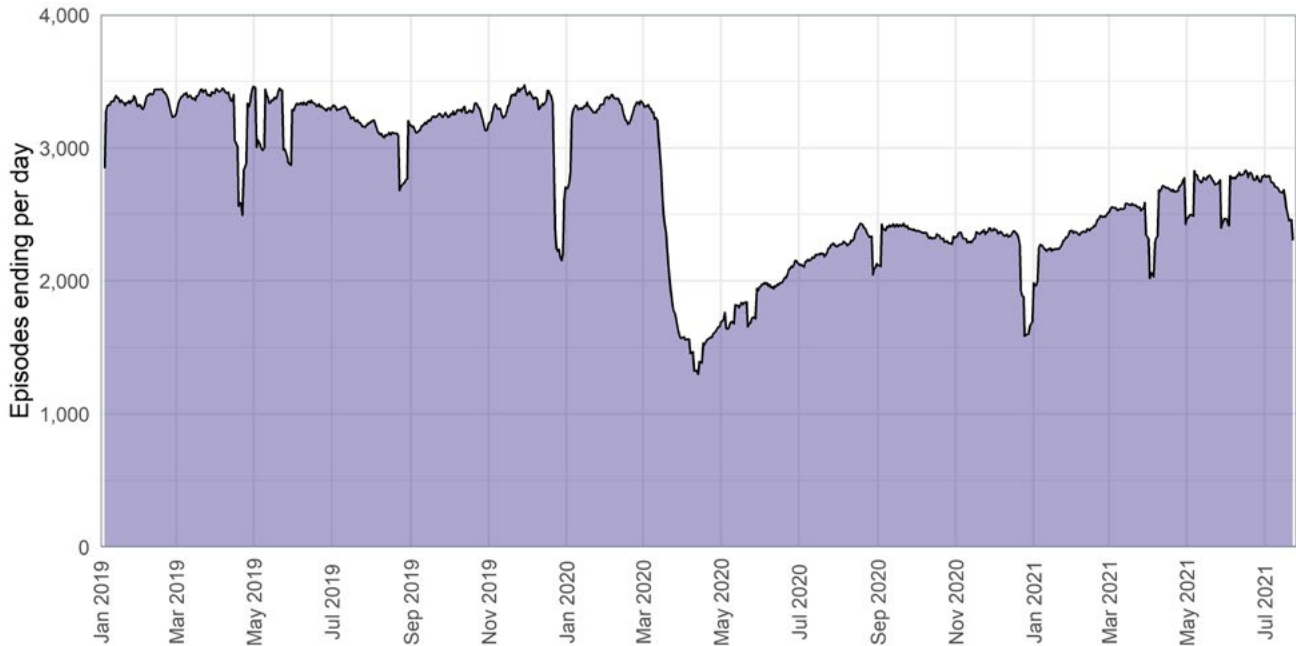

Figure 2: Total number of Patient Episodes between January 2019 and Jul 2021 (7 day rolling average).

The majority of FCE's are entered within their respective monthly reporting period. However, some FCEs are not available in PEDW until subsequent monthly extracts, if data is entered or confirmed retrospectively. Note this refers to the initial entry of the administrative record for the FCEs which may be partially completed and not yet contain clinical coding.

28%
reduction in episodes

Since April 2020
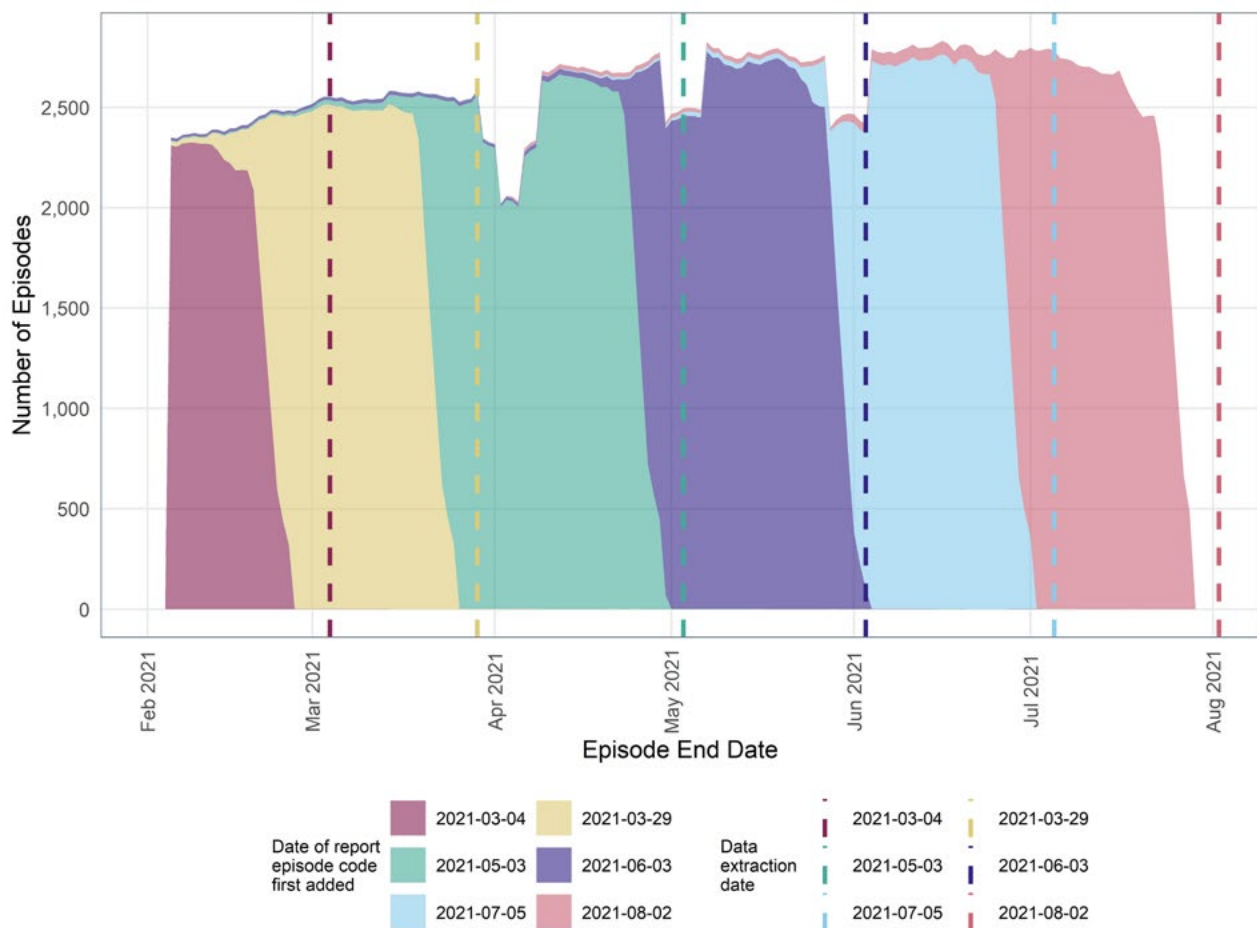
Figure 3: Count of episodes by date added to PEDW (7 day rolling average).

ICD-10 coding completion

The proportion of un-coded or incomplete clinical coded episodes increased from 5% of recorded episodes prior to April 2020, to 11% from April 2020 onwards. However, this is heavily skewed by a low coding rate in the recent months, excluding the latest 6 months results in an un-coded rate of just 3% which is lower than the pre-COVID average.
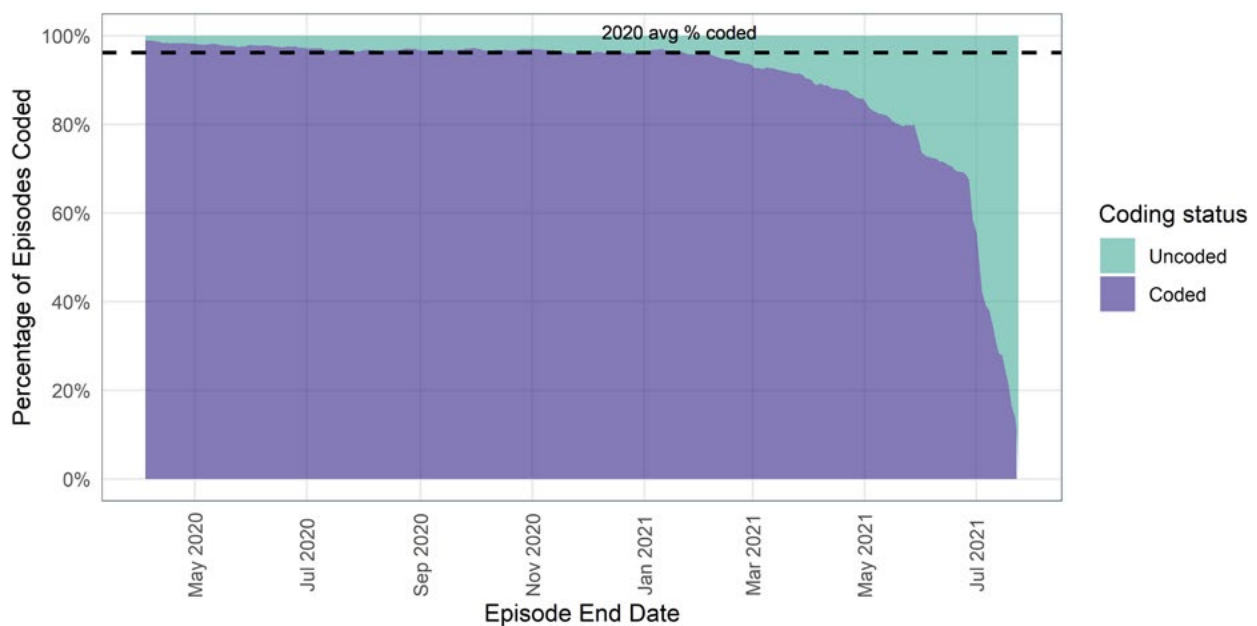
Figure 4: Percentage of episodes with a primary ICD-10 code recorded (7 day rolling average).

Table 1: Percentage of Episodes with Primary ICD-10 Coding by Month of Episode End.

Time Period	Episode End Date	Percentage Coded (7 day rolling average)
Latest Episode End	27/07/2021	N/A
1 Week Prior	20/07/2021	17%
2 Weeks Prior	13/07/2021	29%
3 Weeks Prior	06/07/2021	39%
1 Month Prior	27/06/2021	67%
2 Months Prior	27/05/2021	80%
3 Months Prior	27/04/2021	86%
4 Months Prior	27/03/2021	91%
5 Months Prior	27/02/2021	94%
6 Months Prior	27/01/2021	96%

As seen in Table 1, a primary ICD-10 code had been recorded for 80% of episodes, after a lag of two months from the episodes ending. A completion rate of 96% was achieved after a lag of 6 months. For comparison with previous years, 96% of episodes had an ICD-10 code recorded in 2020 and 95% in 2019.
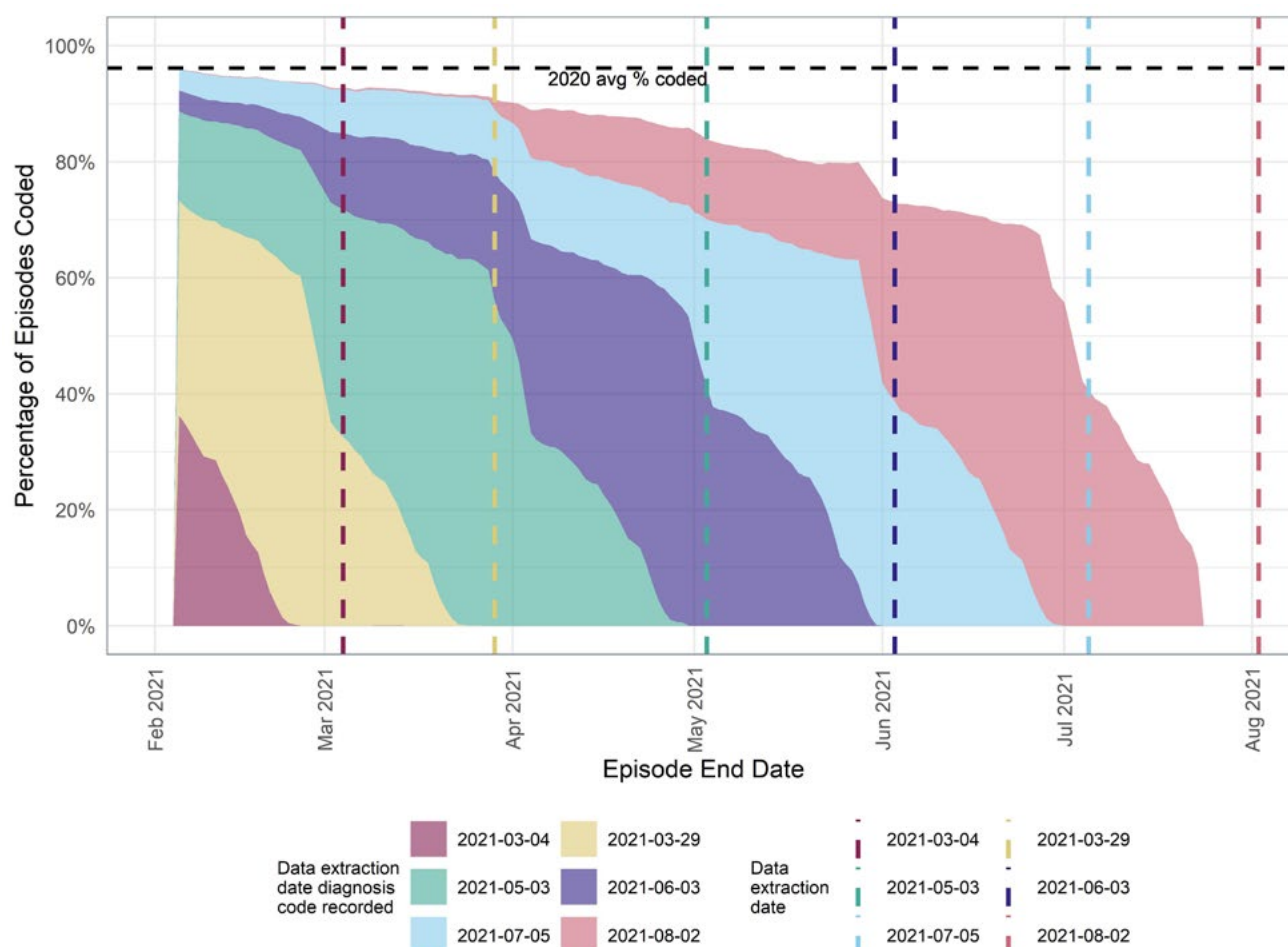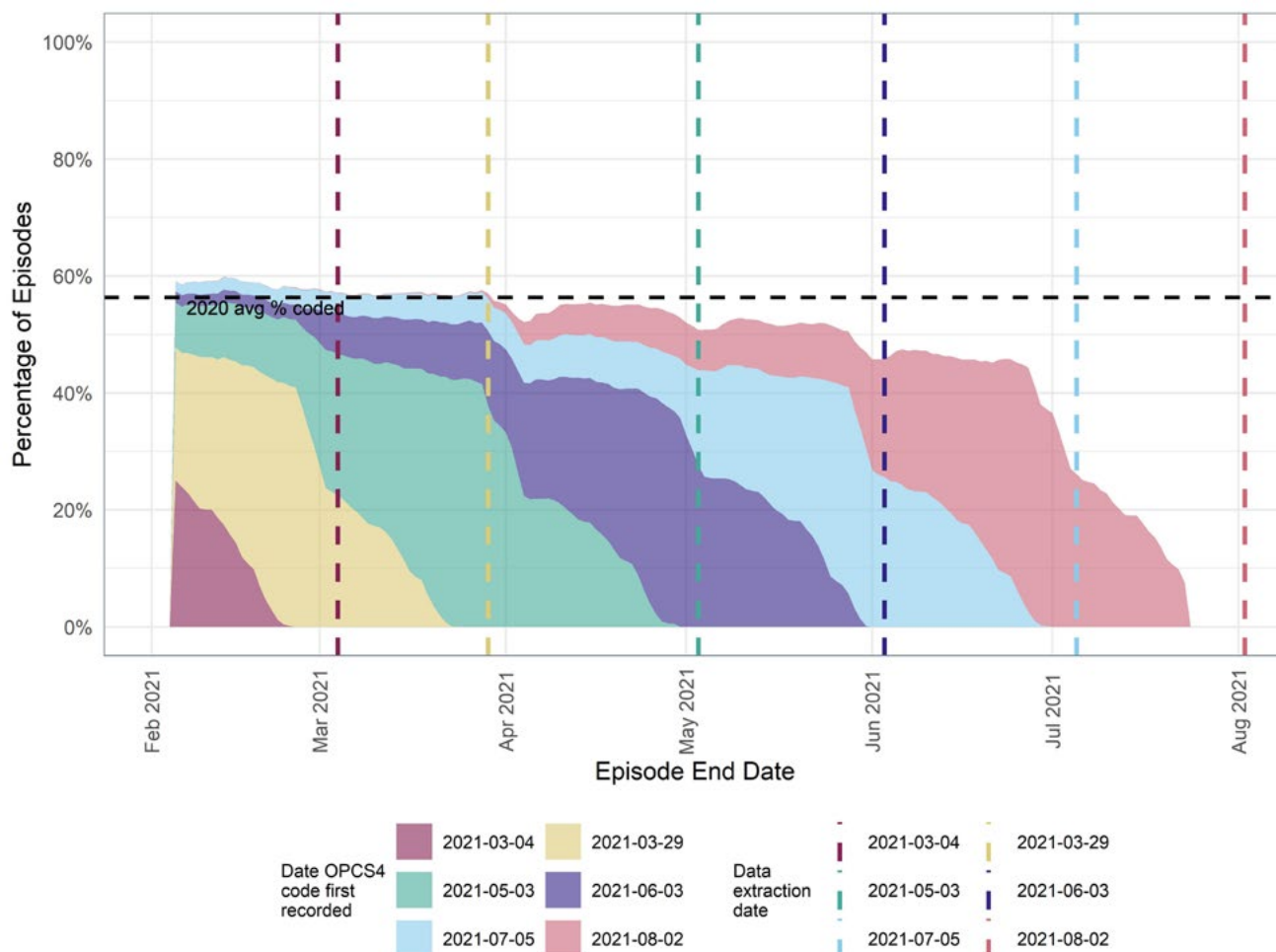

Figure 5: Percentage of episodes with a primary ICD-10 code recorded by month coding was first added to PEDW (7 day rolling average).

The below graphs show that while all chapters demonstrate some retrospective ICD-10 coding, there is some variation between chapters in the extent. Note that chapters with small counts are excluded to prevent disclosure.
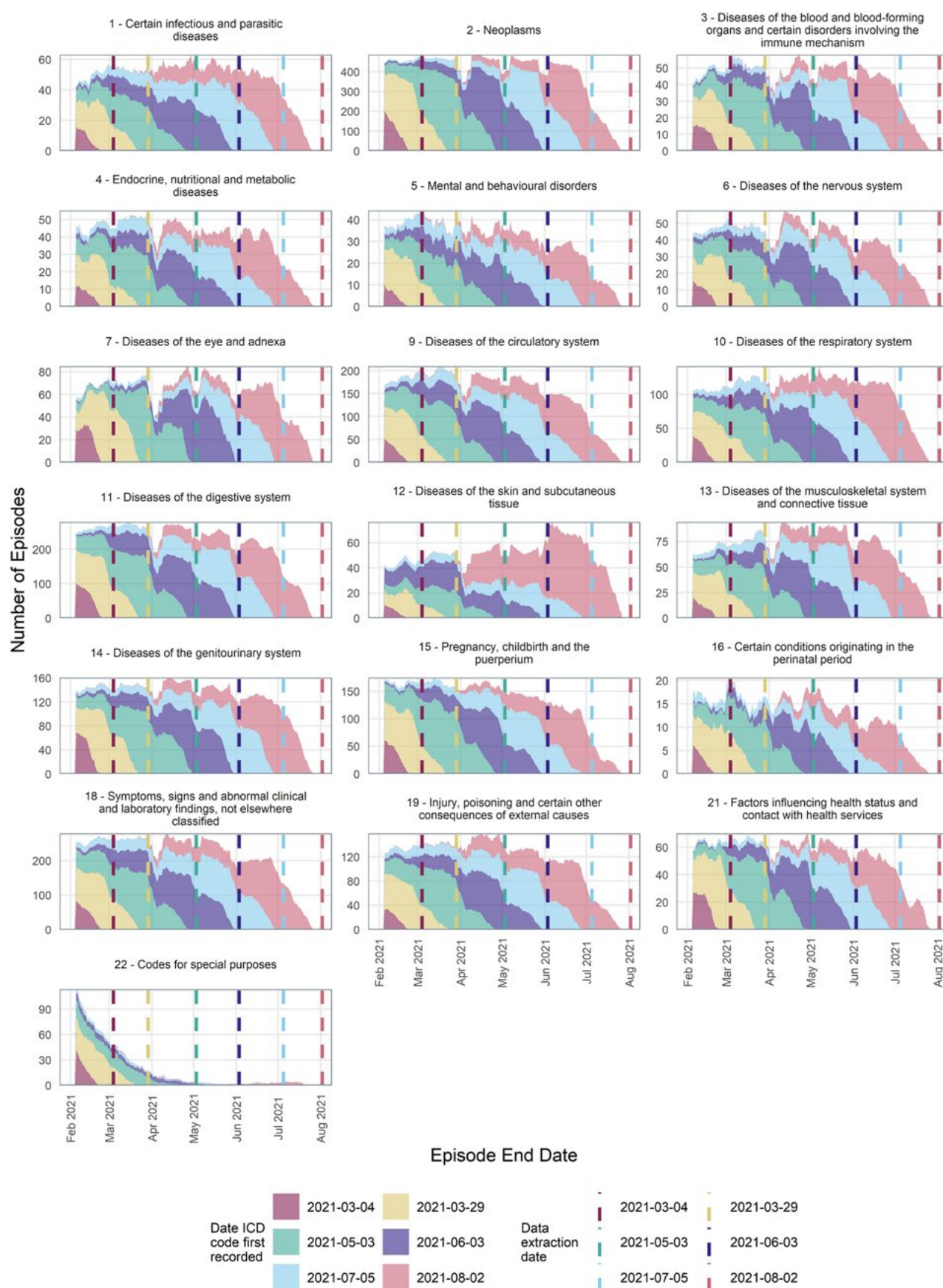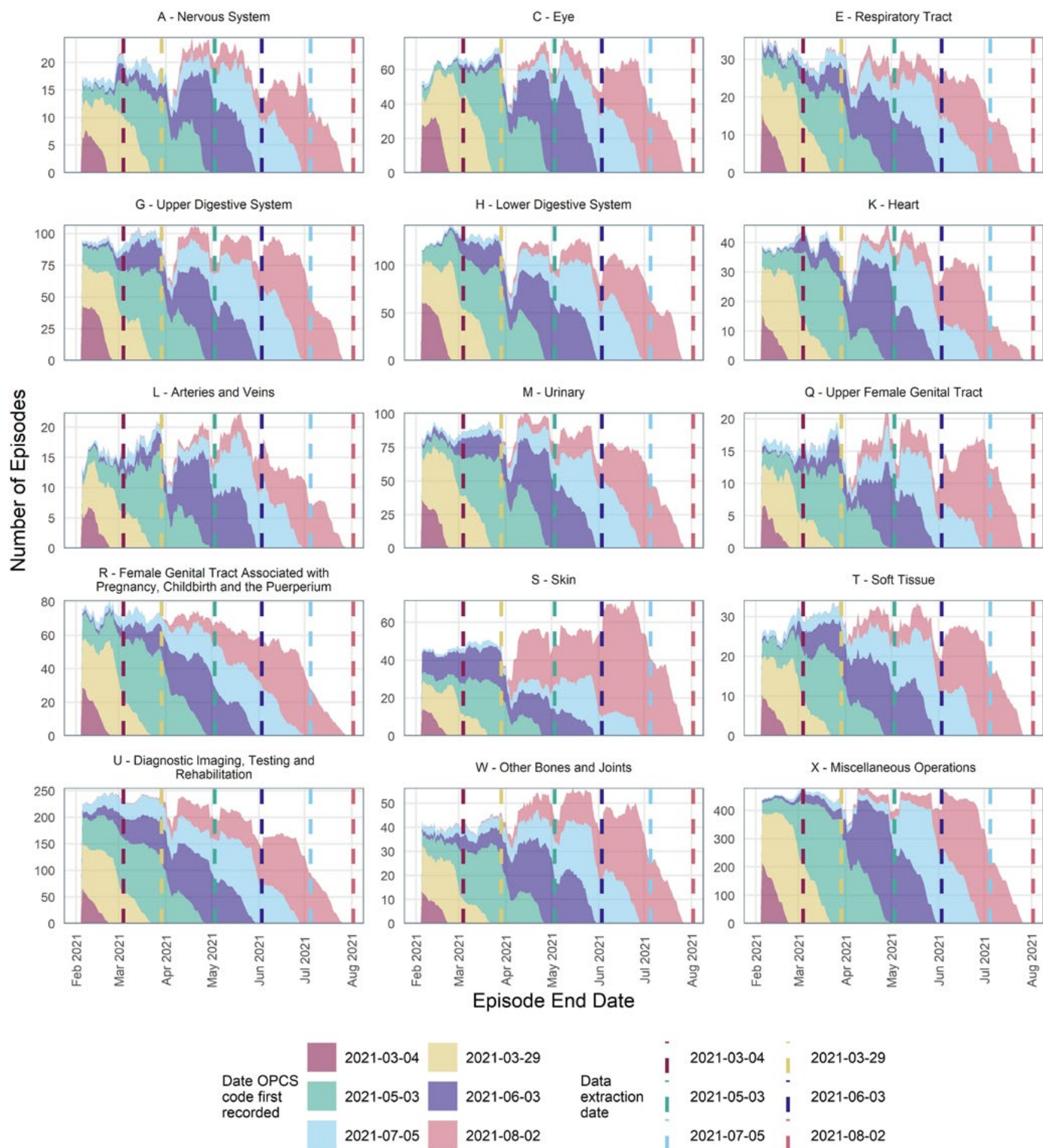

Figure 6: Count of episodes with a primary ICD-10 code recorded by month coding was first added to PEDW (7 day rolling average) and ICD-10 Chapter. Note that the y-axis is adjusted to improve observable detail for each Chapter

53,616 episodes had a change to their ICD-10 diagnosis code since the previous data extract (including added, removed or changed from one ICD-10 code to another), 100% of which had an episode end date within six months of the previous data extracted 4th July 2021.
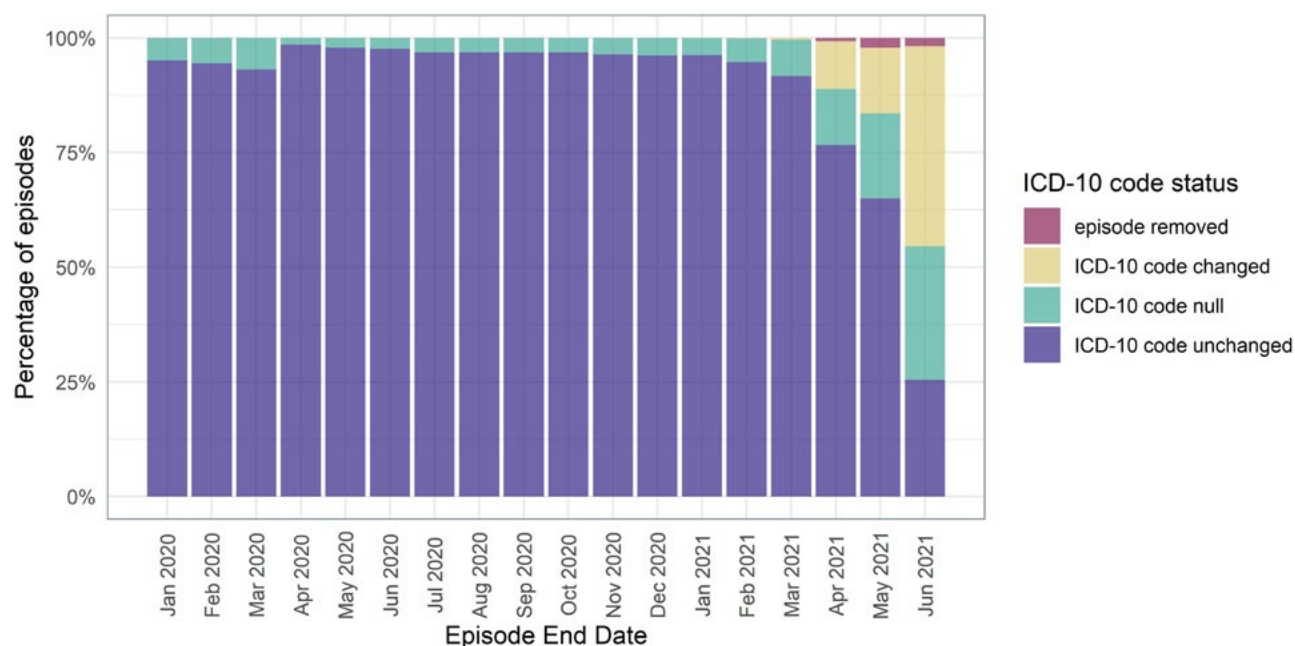
Figure 7: Changes to primary ICD-10 codes since the previous data extracted.

OPCS-4 Coding Completion

OPCS-4 completion is not expected to be 100% because not all episodes require procedural interventions, therefore an episode can have a complete record without having an OPCS-4 code recorded.

Table 2: Percentage of Episodes with Primary OPCS-4 Coding by Month of Episode End.

Time Period	Episode End Date	Percentage Coded (7 day rolling average)
Latest Episode End	27/07/2021	N/A
1 Week Prior	20/07/2021	12%
2 Weeks Prior	13/07/2021	19%
3 Weeks Prior	06/07/2021	25%
1 Month Prior	27/06/2021	44%
2 Months Prior	27/05/2021	51%
3 Months Prior	27/04/2021	54%
4 Months Prior	27/03/2021	57%
5 Months Prior	27/02/2021	58%
6 Months Prior	27/01/2021	60%

Data entry of primary OPCS-4 codes appears to be more timely than that of ICD-10 codes. Using the 2019 and 2020 coding completion average of 56% as a benchmark, shows that this percentage is usually reached within three to four months of the latest data extraction.

Figure 8: Percentage of episodes with a primary OPCS-4 code recorded by month coding was first added to PEDW (7 day rolling average).

The below graphs show that while all chapters demonstrate some retrospective OPCS-4 coding, there is some variation between chapters in the extent. Note that chapters with small counts are excluded to prevent disclosure.

Figure 9: Count of episodes with a primary OPCS-4 code recorded by month coding was first added to PEDW (7 day rolling average) and OPCS-4 Chapter. Note that the y-axis is adjusted to improve observable detail for each Chapter.

Why it matters

These findings indicate that there is a significant lag in clinical coding which any people using PEDW should be aware of when making use of PEDW for research and intelligence. This is important because it indicates that partial episode information is initially entered into PEDW and the records are subsequently updated with clinical coding information. Therefore caution must be exercised when using recent extracts of the dataset for analysis to take into consideration time lag for ICD-10 and OPCS-4 coding.

This information can be used to assist researchers, data scientists, business analysts, intelligence officers and anyone else who would wish to use PEDW to generate findings in deciding an appropriate time frame between PEDW extracts becoming available and using it for research purposes.

What next?

A fuller report is available, maintained and updated on a monthly basis to monitor the PEDW coding completeness and lag, and contains further information and breakdowns, as well as more technical details. This is available outside SAIL on the Swansea University Medical School GitHub at: https://github.com/SwanseaUniversityMedical/DQ_Report_PEDW_Clinical_Coding.

This is also available in SAIL on GitLab with greater level of detail not permitted to be extracted out of SAIL through the disclosure control process. Further information about clinical coding completeness and lag by ICD-10 and OPCS-4 chapters is contained, to assist researchers in understanding the potential impact on their specific area of research.

Availability of data and materials

The data used in this study are available in the SAIL Databank at Swansea University, Swansea, UK, but as restrictions apply they are not publicly available. All proposals to use SAIL data are subject to review by an independent Information Governance Review Panel (IGRP). Before any data can be accessed, approval must be given by the IGRP. The IGRP gives careful consideration to each project to ensure proper and appropriate use of SAIL data. When access has been granted, it is gained through a privacy protecting safe haven and remote access system referred to as the SAIL Gateway. SAIL has established an application process to be followed by anyone who would like to access data via SAIL at: https://www.saildatabank.com/application-process
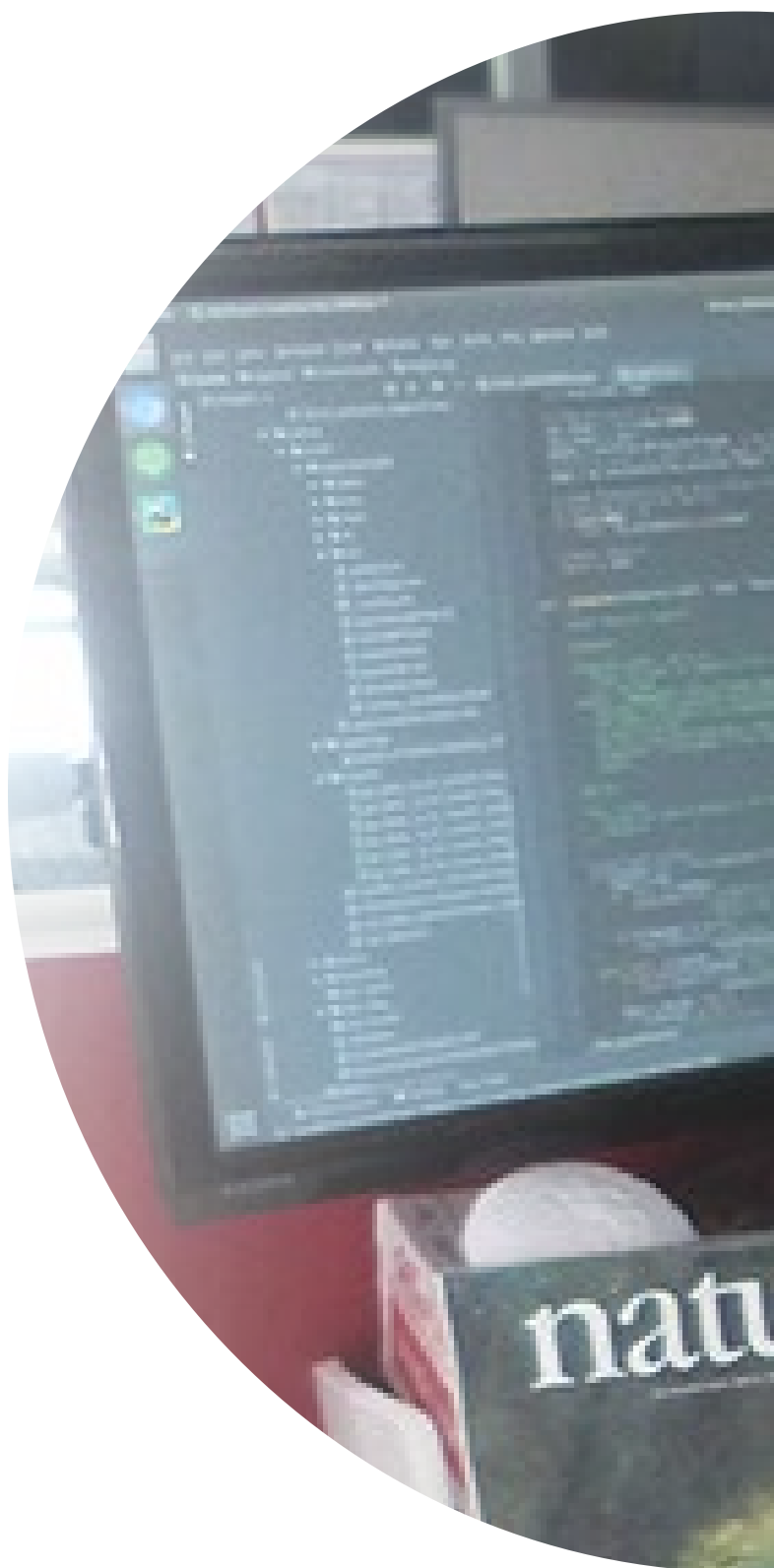
Definitions

Anonymised linking field status
The ALF status indicates the quality of the data linkage, for this report the following statuses were included:

- 1 – NHS number passes check digit test;
- 4 – Surname, First Name, Postcode and Sex match exactly to Welsh Demographic Service Dataset (WDSD);
- 39 – Fuzzy matching Probability >= 0.9.

Clinical coding
"Clinical coding is the translation of medical terminology that describes a patient's complaint, problem, diagnosis, treatment or other reason for seeking medical attention into codes that can then be easily tabulated, aggregated and sorted for statistical analysis in an efficient and meaningful manner." (Clinical Classifications Service, 2017)

ICD-10
The International Statistical Classification of Diseases and Related Health Problems "is used to translate diagnoses of diseases and other health problems from words into an alphanumeric code, which permits easy storage, retrieval and analysis of the data" (World Health Organisation, 2016). ICD-10 is the 10th revision and was the latest version at the time of publication, ICD-11 is due to come into effect on 1st January 2022.

OPCS-4
"The OPCS Classification of Interventions and Procedures (OPCS-4) is a statistical classification of interventions and procedures undertaken in the National Health Service (NHS) reflecting current clinical practice" (Clinical Classifications Service, 2017).

ICD-10 and OPCS-4 coded episodes
Episodes with a valid primary ICD-10 or OPCS-4 code respectively is recorded within the PEDW episodes data view, and linkable to the PEDW Spell data view.

ICD-10 and OPCS-4 un-coded episodes
Episodes where the primary ICD-10 or OPCS-4 code respectively is null within the PEDW episodes data views, and linkable to the PEDW Spell data view.

Sensitive codes
Episodes containing ICD-10 and OPCS-4 codes relating to sensitive topics have been excluded from the data. The following are currently deemed as sensitive topics:

- Gender Reassignment;
- Human immunodeficiency virus (HIV) and sexually transmitted diseases (STDs);
- Miscarriages and terminations;
- Fertility Treatments.

Unfinished episodes
Episodes where there is no episode end date currently available/recorded

References

Audit Wales (2020). Cracking the Code, Management of Clinical Coding Across Wales (Sep 2020). Available at: https://www.audit.wales/sites/default/les/clinical_coding_eng_10.pdf. (Accessed: 11/05/21)

BMA (2020), The Hidden Impact of Covid. Available at: https://www.bma.org.uk/media/2840/the-hidden-impact-of-covid_web-pdf.pdf. (Accessed 11th May 2021)

Clinical Classifications Service (2017). National Clinical Coding Standards OPCS-4 (2017): Accurate data for quality information. Available at https://hscic.kahootz.com/gf2.ti/f/762498/27837541.1/PDF/-/NCCSOPCS42017.pdf (Accessed: 13th July 2021)

DHCW (Digital Health and Care Wales) (2021). PEDW Data Online. Available at https://nwis.nhs.wales/information-services/welsh-data-hub/pedw-data-online/ (Accessed: 20th July 2021)

ESRC (Economic and Social Research Council) (2013). Data Linking Demonstration Project - Journey Mapping for Patients with Multiple Chronic Health Conditions. Available at https://gov.wales/sites/default/files/statistics-and-research/2019-01/data-linking-demonstration-project-journey-mapping-patients-with-multiple-chronic-health-conditions_0.pdf (Accessed: 21st July 2021)

HDR UK (Health Data Research UK) (2020). Patient Episode Dataset for Wales. Available at: https://web.www.healthdatagateway.org/search?search=&tab=Datasets. Accessed: 21st July 2021.

HSCIC (last updated April 2021), Covid-19 National Clinical Standard and Guidance. Available at: hscic.kahootz.com/connect.ti/t_c_home/view?objectID=19165328. (Accessed: 11th May 2021)

NHS Digital (2020). COVID-19 National Clinical Coding Standards and Guidance. Available at https://hscic.kahootz.com/t_c_home/view?objectID=19165328 (Accessed: 9th August 2021)

NHS Wales (2021). Patient Episode Database for Wales (PEDW). Available at https://www.wales.nhs.uk/document/176173 (Accessed: 13th July 2021)

Welsh Government (2020). NHS Delivery Framework 2020-2021. Available at: http://www.wales.nhs.uk/sitesplus/documents/862/2020-21%20Delivery%20Framework.pdf. (Accessed 9th August2021)

World Health Organisation (2016). International statistical classification of diseases and related health problems 10th revision Volume 2 Instruction manual Fifth edition. Available at https://icd.who.int/browse10/Content/statichtml/ICD10Volume2_en_2019.pdf (Accessed: 13th July 2021)

World Health Organisation (2021). Coronavirus symptoms. Available at: who.int/healthtopics/coronavisurs#tab=tab_3. (Accessed 11th May 2021)

World Health Organisation (2021). Emergency use ICD codes for COVID-19 disease outbreak. Available at: https://www.who.int/standards/classifications/classification-of-diseases/emergency-use-icd-codes-for-covid-19-disease-outbreak. (Accessed 13th September 2021)

Acknowledgements

Funding

One Wales

This study is a part of the One Wales response to COVID-19 which brings together cross-institutional teams of experts from across Wales to provide timely evidence to inform policy and practice to tackle the epidemic and its impact in the UK.

Bringing together colleagues from within the Population Data Science group and across Wales including HDR UK, ADR Wales, SAIL Databank, Adolescent Mental Health Data Platform, BREATHE, Welsh Government, Public Health Wales and Digital Health & Care Wales has resulted in an agile and responsive approach to tackling data analysis and intelligence generation based on both the constant and newly developing priorities for tackling COVID-19 in Wales.

The One Wales team will continue to work together to identify gaps in knowledge and streamline efforts to deliver vital intelligence to help policymakers understand and plan around the issue of COVID-19 in Wales and across the UK.

ADR Wales Partners