English Prescribing Dataset Release Guidance

Version: v003

Contents

Document Release Note

Document Name: English Prescribing Dataset Release Guidance

Document Details Name	Version Number	Description
English Prescribing Dataset Release Guidance	v003	This document describes the content and some instruction on how to use the English Prescribing Dataset that is released using the NHSBSA Open Data Portal.

Revision Details Revision Number	Revision Date	Revision Description	Page Number	Previous Page Number	Action Taken	Addenda/ New Page
003	[TBD]	Update guidance regarding addition of SNOMED codes to data			Guidance on the addition of SNOMED data, the implications for users and how to recreate previous versions of EPD data	
002	10/06/2020	Update guidance regarding transformatio n of EPD to	6-7		Guidance on how to manipulate EPD data to recreate PLP and DPI datasets	
002	10/06/2020	Change document title remove reference to Statistics Team	Cover page		Removed 'Statistics at NHSBSA'. Inserted Version number on cover page	
002	10/06/2020	Change contact details for NHSBSA	10		Written address updated to reflect Information Services Team who are data owners.	
002	18/06/2020	Added details of NHS re-organisation.	6		Added information and supporting	

					documentation for NHS re-organisation in April 2020.	

About this document

NHSBSA publishes detailed information on prescribing by each practice in England. This document describes the content of the main data release and provides instructions for users of data that was previously published by NHS BSA and NHS Digital.

English Prescribing Dataset

Summary

The English Prescribing dataset contains detailed information on prescriptions issued in England that have been dispensed in England, Wales, Scotland, Guernsey, Alderney, Jersey, and the Isle of Man.

The dataset combines elements of the [Detailed Prescribing Information](#) (DPI) data previously released by NHS Business Services Authority (NHSBSA) via the Information Services Portal (ISP), and the [Practice Level Prescribing in England](#) (PLP) data released by NHS Digital via their website. It is intended to replace both of those sources.

These datasets have been brought together to provide end users with a single comprehensive, consistent, and accessible source of prescribing information.

Contents of the dataset

The data summarises 'prescription items', sometimes called 'items'. This refers to an instance of a type of drug, appliance, dressing or reagent that is requested by a prescription form. For example, a prescription form may list two items as so:

- 'Warfarin 1mg tablets 84 tablet'

- 'Warfarin 3mg tablets 56 tablet'

Such a form would include two items; the quantity of the first item is 84, and the quantity of the second item is 56.

The data includes figures on the number and quantity of prescription items that are dispensed each month, and information relating to costs for General Practitioner Practices and Cost Centres in England.

The data relates to items that have been prescribed in England and dispensed in England, Scotland, Wales, Isle of Man, Jersey, and Guernsey. It does not include items that have been dispensed in England but prescribed in another country.

It includes for each General Practitioner (GP) Practice or Cost Centre:

- All prescribed and dispensed medicines (by British National Formulatory (BNF) Chapter, Chemical Substance and Presentation)

- All prescribed and dispensed dressings and appliances (by BNF Chapter, Section and Presentation)

- **Items** – the total number of times that the medicine, dressing or appliance appeared on prescription forms that were prescribed and dispensed

- **Quantity** – The quantity of a medicine, dressing or appliance for which an individual item was prescribed and dispensed, for each BNF Presentation. This represents a pseudo pack size, to illustrate the typical range of prescribed quantities of a given presentation

- **Total Quantity** – the total quantity of a medicine, dressing or appliance prescribed and dispensed for each BNF Presentation. This is calculated by multiplying Quantity by Items

- The **'Net Ingredient Cost (NIC)'** of the items normally based on the price given in the Drug Tariff or published wholesale prices

- The '**Actual Cost'** that accounts for the national average discount and some payments to dispensers

- **ADQ Usage** - Average Daily Quantity (ADQ) is the typical daily dose of a medication, prescribed to adult patients by GP Practices

- **Practice Name** and **Address** details, where known; and

- A column marked '**Unidentified'** which can be used to include or exclude information from prescriptions forms for which a practice could not be identified (see below for more information)

- The **SNOMED** code relating to the BNF presentation.

This data is sourced from the NHSBSA Information Services Data Warehouse.

Exclusions

The data excludes:

- Data relating to prescriptions issued in Wales, Scotland, Guernsey, Alderney, Jersey, and the Isle of Man

- Any patient identifiable data

- Items not dispensed, disallowed and those returned to the contractor for further clarification

- Prescriptions prescribed and dispensed in Prisons, Hospitals and Private prescriptions

- Items prescribed but not presented for dispensing or not submitted to NHS Prescription Services by the dispenser.

NHS Re-organisation April 2020

From the 1st of April 2020 a new organisation structure was implemented by the NHSBSA as agreed by NHS England. Reporting against this new structure will commence from April 2020 dispensing data. Data prior to this will remain in the old structure and will not be updated.

Summary of April 2020 structure and changes for EPD:

- 7 regions replacing 4 regions

- 42 STPs replacing 27 Area Teams & 14 Local Offices

You can find further information about these changes and how they impact our data at https://www.nhsbsa.nhs.uk/nhs-structure-re-organisation-implementation.

You can also find further information on the structural changes at NHS Digital's change summary – 2020 STP, CCG and commissioning hub reconfiguration page. NHS Digital have also produced an April 2020 reconfiguration mapping file.

Unidentified Practice Data

This dataset includes some records where the prescription form cannot be linked to a practice or cost centre.

Sometimes the NHSBSA is sent a prescription form that has incomplete or missing information about the practice.

In the dataset that information is marked with '**Unidentified Doctors'** as the Practice Name. These cases typically account for less than 0.1% of all prescribed items. *

Unidentified Doctors data can be put in two groups:

1. where the Clinical Commissioning Group (CCG) can be identified; these are shown with a Practice Code and Address relating to the CCG, and will hold values for CCG, Area Team and Region; and

2. where the CCG cannot be identified. These are shown with no Practice Code or Address, and all other organisation levels will be marked as 'Unidentified'.

Records that fall into the second category are referred to by NHSBSA as **'National Unidentified'**, which is defined as follows:

> *National unidentified totals include data for prescriptions that could not be allocated to a practice or their Commissioner / Provider Organisation. These totals may include prescriptions written by prescribers who have not been linked to a practice for more than six months. The reason for this six month 'rule' is that prescriptions are only valid for six months.*

For reporting purposes, the 'Unidentified' column can be used to restrict the dataset to return only unidentified practice records; adding a further filter to the region name column will allow the inclusion or exclusion of national unidentified records.

* 0.05%, based on analysis of March 2019 prescribing data.

Using large data files

Each release of EPD typically contains in the region of 17 million rows of data.

Loading data files of these sizes poses a challenge for spreadsheet packages such as Microsoft Excel and for computers with little available memory. Microsoft Excel's Power Pivot sometimes reports a size problem for files larger than 4GB, or 2GB if using the 32-bit version.

We recommend using purpose built statistical or database software to process these files. For example, the freely available 'R', Python, MySQL, MariaDB, PostgreSQL programs; many other free, open source, and commercial alternatives are also available. Another option is to use 'csv-splitter' utilities to separate larger files into smaller parts that can be loaded separately into downstream applications.

It may also be possible to use Microsoft Power Query to import data into Microsoft Excel or to import the csv as a source for a Microsoft Excel Pivot Table; however, there may be limitations for any of these solutions, which may prove unreliable or perform poorly.

Finally, if you know in advance that you only need a subset of the data, you can use the Data Explorer within the Open Data Portal to "Add Filters" before you download the filtered results via "Data explorer options" this may result in a smaller file that is easier to handle for applications or systems with resource limitations.

If using the Data Explorer to download a subset of EPD data, this will be delivered in a .GZIP format. GZIP archive files cannot be extracted using the Windows compressed file system, so an external solution must be installed in order to extract the data. An example of such a solution is 7-zip, a free, opens source software, although other programs are available.

For users of previous datasets (PLP and DPI)

The new dataset combines elements of the existing PLP data - from NHS Digital - and DPI datasets - from the NHS BSA Information Services Portal.

These datasets have been brought together to provide end users with a single, comprehensive, consistent and accessible source of prescribing information.

While the new dataset will bring about benefits, we recognise that existing users will have based reporting on the existing versions of the data or use one or both datasets in downstream applications.

For consistency with previous reporting, the following steps can be taken to map the new DPI data back to the existing PLP and DPI datasets.

For Practice Level Prescribing Users (PLP)

The existing PLP dataset does not include data for unidentified practices. These are included in the new DPI data. Filter out any rows with a value of **TRUE** in the **Unidentified** column to remove these practices, then remove the **Unidentified** column. The current PLP also contains the fields:

- SHA
- PCT
- PRACTICE
- BNF CODE
- BNF NAME
- ITEMS
- NIC
- ACT COST
- QUANTITY
- PERIOD

The corresponding fields in the new EPD data set are:

- STP_CODE
- PCO_CODE
- BNF_CODE
- BNF_DESCRIPTION
- ITEMS
- NIC
- ACTUAL_COST
- TOTAL_QUANTITY
- YEAR_MONTH

As the data in the new EPD file is disaggregated across prescribed quantity and SNOMED Code, in order to get to the format of the original PLP file you must group the data by STP_CODE, PCO_CODE, BNF_CODE, BNF_DESCRIPTION, and YEAR_MONTH. Then sum ITEMS, NIC, ACTUAL_COST and TOTAL_QUANTITY.

For Detailed Prescribing Information Users (DPI)

The previous DPI dataset does not include all the columns that feature in the new DPI dataset. **Exclude** the following columns to get the DPI column format:

- ADDRESS_1

- ADDRESS_2

- ADDRESS_3

- ADDRESS_4

- POSTCODE

- BNF_CHEMICAL_SUBSTANCE

- BNF_CHAPTER_PLUS_CODE

- SNOMED_CODE

- TOTAL_QUANTITY

- UNIDENTIFIED

- YEAR_MONTH

The new dataset includes Unidentified Regions (and any Area Teams, CCGs or Practices that sit within them). **Filter out** all rows with **'Unidentified'** in the **'Region Name'** column to get the same rows as the DPI format.

SNOMED CT Data

Commencing [month], a column has been added to the EPD data to hold the SNOMED code for each BNF presentation.

SNOMED CT is an internationally recognised coding system for electronic health records, which enables information to be shared between clinical systems in an easy, secure and accurate manner.

SNOMED CT contains clinical terms for all aspects of NHS activity, from procedures and symptoms, through to clinical measurements, diagnoses and – as contained in the EPD data release – medications.

Implications and changes to previous EPD structure

The introduction of SNOMED data in EPD has resulted in some changes to the structure, level of detail, and file size of the release, compared to previous releases.

Adding the SNOMED code has necessitated the creation of a new column in the data structure – SNOMED_CODE. This column is formatted as an integer, with around 66% of codes having a precision of 17 digits, and another 32% 16 digits (a very small number of codes have shorter lengths).

Some BNF presentation codes are associated with more than one SNOMED code, so the introduction of SNOMED CT data has increased the granularity of the release, increasing the number of rows (resulting in a larger file size), and creating a potential need to group queries based on BNF code in order to remove any multiple rows.

In the November 2020 dataset, the introduction of SNOMED data has caused the row count in the file to increase by 342,858 rows in total.

An example of the change in row structure, for Warfarin 5mg tablets in November 2020, is as follows:

Original Dataset

BNF Code	BNF Description	SNOMED CT Code	Items
0208020V0AAADAD	Warfarin 5mg tablets	n/a	62,742

New Dataset

BNF Code	BNF Description	SNOMED CT Code	Items
0208020V0AAADAD	Warfarin 5mg tablets	1170211000001106	62,711
0208020V0AAADAD	Warfarin 5mg tablets	9797611000001104	12
0208020V0AAADAD	Warfarin 5mg tablets	1530711000001105	10
0208020V0AAADAD	Warfarin 5mg tablets	10418711000001100	3
0208020V0AAADAD	Warfarin 5mg tablets	1530511000001100	3
0208020V0AAADAD	Warfarin 5mg tablets	1530611000001101	2
0208020V0AAADAD	Warfarin 5mg tablets	28989411000001100	1

An existing query that returns all rows associated with Warfarin 5mg tablets will return more rows in the new dataset than before. The total number of items is the same across all rows.

In order to retain the data structure in earlier versions of the EPD data, do not include SNOMED_CODE in the query definition, and group by a higher element in the data structure, such as BNF_CODE, PRACTICE_CODE, PCO_CODE, etc.

Variation in Totals

Based on testing of data from November 2020, the totals for Items, Total Quantity, and Net Ingredient Cost are unchanged by the introduction of SNOMED data. There are variations in the totals for ADQ Usage and Actual Cost, which have been introduced by the rounding used to arrive at the totals for each row.

For November 2020, these differences amount to a reduction of 0.00067 for ADQ Usage, and 0.02305p for Actual Cost, at a national level (0.000000000032% and 0.0000000031%, respectively).

Contact us

Feedback is important to us; we welcome any questions and comments relating to this document.

Please quote 'English Prescribing Dataset Release Guidance' in the subject title of any correspondence.

You can contact us by:

Email: DataServicesSupport@nhsbsa.nhs.uk

You can also write to us at:

Information Services
NHS Business Services Authority
Stella House
Goldcrest Way
Newburn Riverside
Newcastle upon Tyne
NE15 8NY.

END.