



Hospital Episode Statistics (HES) Outpatient Care and CPRD primary care data Documentation (November 2024)

Version: 2.1

Date: 29 October 2024

Documentation Control Sheet

Over time, it may be necessary to issue amendments or clarifications to parts of this document. This form must be updated whenever changes are made.

Version	Affected Areas Summary of Change	Prepared By	Reviewed By
1.0	Initial	Tarita Murray-Thomas	Helen Strongman
1.1	Formatted	Grant Lee	Wilhelmine Meeraus
1.2	Modified	Wilhelmine Meeraus	Tarita Murray-Thomas
1.3	Modified	Wilhelmine Meeraus	Tarita Murray-Thomas, Arlene Gallagher
1.4	Modified	Shivani Padmanabhan	Tarita Murray-Thomas
1.5	Modified	Tarita Murray-Thomas	Shivani Padmanabhan
1.6	Modified	Tarita Murray-Thomas	Sonia Coton
1.7	Modified	Sonia Coton	Tarita Murray-Thomas
1.8	Modified	Tarita Murray-Thomas	Susan Hodgson
1.9	Modified	Tarita Murray-Thomas	Sonia Coton, Rebecca Ghosh
2.0	Modified	Helen Booth	Mia Harley
2.1	Modified	Alice Hinchliffe	Jenny Campbell

Summary of Changes

Version 1.1

- Formatted with new agency branding and updated document title
- Included version of HES on front page

Version 1.2

- Updated document version number, date, HES set, and end of the HES OP coverage date
- Clarified information relating to the 'match_rank' variable under 'HES Outpatient data and GOLD'
- Included new section under 'Data structure and formatting' to record changes introduced in set 11
- Updated the information relating to ethnicity under 'Known Issues'.

Version 1.3

- Updated document version number, date, HES set, and end of the HES OP coverage date
- Added information on the coverage period of HES OP data by HES
- Added table of proportion of patients linked by match_rank
- Added details about availability of records with match_rank values of 6 to 8
- Added details about availability of records with multiple HESIDs
- Added changes introduced in set 12
- Added information under 'Known issues' relating to provisional HES data
- Added the data dictionary to the end of the document

Version 1.4

- Updated document version number, date, HES set, and end of the HES OP coverage date
- Added explanation of changed definition of the derived ethnicity variable
- Added changes introduced in set 13
- Updated references to reflect change of name from HSCIC to NHS Digital

Version 1.5

- Updated document version number, date, HES set, and end of the HES OP coverage date

Version 1.6

- Updated document version number, date, HES set, and end of the HES OP coverage date

Version 1.7

- Updated document version number, date, HES set and end of the HES OP coverage date
- Updated to include CPRD Aurum

Version 1.8

- Updated document version number, date, HES set and end of the HES OP coverage date

Version 1.9

- Updated document version number, date, HES set and end of the HES OP coverage date
- Updated NIHR logo
- Specified primary keys in each data table

Version 2.0

- Updated document version number, date, HES set and end of the HES OP coverage date; added DOIs

Version 2.1

- Updated document version number, date, HES set and end of the HES OP coverage date; added DOIs
- Removed references to hes_id
- Added operations table

HES Outpatient Care (OP) data linked to CPRD primary care data

This document provides an overview of HES Outpatient (HES OP) data, and the available subset that is linked to CPRD GOLD and CPRD Aurum.

What are the HES Outpatient data?

HES OP data are a collection of individual records of outpatient appointments occurring in England only. It includes information on the type of outpatient consultation appointment dates, the main specialty and treatment specialty under which the patient was treated, referral source, waiting times, clinical diagnosis and procedures performed.

HES OP data were collected by the NHS for the first time in 2003-04. NHS Digital (formerly known as the Health and Social Care Information Centre) first released HES OP data on an 'experimental' basis at the end of July 2006. Since then, the experimental label has been removed and in 2008 the outpatient published tables were accredited as a National Statistic. There may be potential data quality issues in the early periods when the data were collected experimentally.

HES OP data can be used to support health resource utilisation studies, clarify clinical health care pathways, and enable variations in the uptake of services to be evaluated, for example by gender and age. Before requesting HES OP data, users are encouraged to familiarise themselves with the content of HES OP data. Details on the fields available can be found at: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary>. Details of HES Outpatient activity statistics can be found at: <https://digital.nhs.uk/data-and-information/publications/statistical/hospital-episode-statistics-for-admitted-patient-care-outpatient-and-accident-and-emergency-data>

Accessing HES Outpatient Care data linked to CPRD GOLD and CPRD Aurum

HES OP data can only be accessed as part of a data extract linked to CPRD primary care data (CPRD GOLD or CPRD Aurum). Access is provided by CPRD for a fee subject to protocol approval.

Not all patients in CPRD GOLD or CPRD Aurum are eligible to be linked to HES, for example, due to the region in which they reside (outside England), or the lack of a valid NHS identifier. Source files (linkage_eligibility.txt) are provided to allow researchers to identify the subset of patients who are eligible to have linked HES data.

Linkage coverage period

The latest release of HES OP data linked to CPRD primary care data (November 2024) covers the period **April 2003 – March 2023**.

DOI

Please cite in any publications using these data:

CPRD GOLD HES OP November 2024 <https://doi.org/10.48329/zamg-gb30>

CPRD Aurum HES OP November 2024 - <https://doi.org/10.48329/k719-t939>

Linkage

Linkage of HES OP to CPRD primary care data is carried out by our trusted third-party NHS England. Linkage is done using NHS England's Master Person Service (MPS) which replaces the eight-step deterministic algorithm used previously. The Master Person Service extracts demographic information from the Personal Demographics Service database (PDS) held by NHS England, which has approximately 80 million patient records. CPRD patients are matched to PDS patients on NHS number and date of birth. If patients cannot be matched to the PDS database, then the MPS process attempts to ascertain a match based on date of birth and either a local patient identifier or postcode to records in the MPS record bucket (~6% of linked patients in CPRD GOLD and ~2% of linked patients in CPRD Aurum were matched using this secondary method) Each matched patient is assigned a CPRD specific MPS identifier (CPRD MPSid). A list of matched patients' CPRD MPSid's together with their data is provided to CPRD. Further information on how the Master Person Service works can be found in NHS England's Person ID Handbook (1).

Data structure and formatting

HES OP data provided by CPRD represents only a subset of the variables that are collected in the National HES OP dataset provided by NHS Digital. Fields such as organisation fields which may lead to the potential re-identification of patients or practices are not collected by the CPRD and/or not supplied to users.

The data are arranged into files relating to appointments and contact in the outpatient clinic setting. Patients may see the same consultant, or members of his/her team, on more than one occasion during a single year. They may attend more than one consultant clinic in the same or different specialties, or in different providers, for the same or different conditions. Each record represents a single outpatient attendance at a consultant or allied health clinic at a single hospital provider. The patient identifier (**patid**) may be used to link together outpatient attendance records for a single patient with CPRD HES admitted patient care and/or A&E attendance records.

For each patient cohort, HES OP data will be provided as separate text tab delimited files. Files can be imported into statistical software such as Stata or SAS, or into data management packages such as Microsoft Access, for further data processing and analysis.

Changes introduced in HES OP sets

set 11

- CPRD introduced the HES patient identifier (**gen_hesid**) in OP data. This is unique across all CPRD linked HES datasets including HES admitted patient care (APC), HES OP and HES accident and emergency (A&E) data.
- The attendkey variable has been altered so that this is now unique (by patient identifier) across all HES OP data.
- CPRD provided a derived ethnicity variable (**gen_ethnicity**) which is the most commonly recorded ethnicity for each patient among all HES data including HES APC, HES OP and HES A&E. The ethnicity recorded at the outpatient appointment (**ethnos**) remains unchanged.

set 12

Licensing obligations require that no attempts are made to re-identify patients in CPRD datasets. The attendkey variable has been encoded by the CPRD to minimise the risk of breaching licensing conditions through linkage of these data to other HES data sources containing patient identifiable information. What this means is that from set 12, the attendkey variable is different from that of previous sets and will differ in each future release of HES OP linkage sets.

set 13

The definition of the derived ethnicity variable (gen_ethnicity) in the patient file has been changed so that ethnicity is specified where at least one episode has a specific ethnicity recorded but the majority of values are “unknown”.

November 2024

The linkage algorithm method has been replaced by MPSid, and the variable match_rank is no longer provided. The cprd_mpsid variable replaces the hesid variable which was available in previous versions of HES OP. The Clinical table has now been split into Clinical and Operations tables.

Known issues

Some fields, which are of great potential interest, were not designated as mandatory when the dataset was originally designed. This has led to apparently low coding levels over the years. These fields include:

- **Ethnic Group** (low levels of coding): Recording of ethnicity is lower in the earlier years of collection and better in the later years.
- **Diagnosis** (less than 5% of all attendances): Unlike HES admitted patient care data, it is not mandatory to record diagnostic information in HES OP using ICD-10 codes. For this reason, diagnostic information is not widely available in this data offering. HES OP data will therefore have limited use in epidemiology research where clinical diagnostic data is needed.
- **Procedure** (about 5-15% of all attendances): Procedures undertaken in the outpatient setting can be identified in HES OP data. However, the completeness of recording of these data for select procedures is unknown.
- There are minor problems with national data coverage reported with each annual dataset.

Considering the limitations listed above, a conservative approach would be to use HES OP data from 2007 onwards.

Provisional HES OP data are monthly publications of HES data. These data may be incomplete or contain errors for which no adjustments have yet been made by HES. Counts produced from provisional data are likely to be lower than those generated for the same period in the final dataset. It is also probable that clinical data are not complete, which may affect the last two months of any given period. There may also be errors due to coding inconsistencies that have not yet been investigated and corrected. At the end of the fiscal year there is a "month 13" annual refresh which corrects known data quality issues prior to locking the annual published data.

Look-up files

Lookup files relating to the use of HES OP data will not be provided by the CPRD. These can be obtained online from NHS Digital using this link <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics/hospital-episode-statistics-data-dictionary>

HES OP: Data dictionary

1. Patient (hesop_patient.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique key given to a patient in CPRD GOLD or CPRD Aurum [primary key]	INTEGER	20
pracid	Encrypted unique key given to a practice in CPRD GOLD or CPRD Aurum	INTEGER	8
cprd_mpsid		INTEGER	12
gen_ethnicity ¹	Patient's ethnicity derived from all HES data (including HES outpatient, HES admitted patient care and HES A&E)	CHAR	8

2. System/Patient pathway (hesop_patient_pathway_XXX.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique key given to a patient in CPRD GOLD or CPRD Aurum [primary key, in combination with attendkey]	INTEGER	20
attendkey ¹	Record identifier (unique in combination with patid) [primary key, in combination with patid]	INTEGER	12
subdate	Submission date	CHAR	yyyy/mm/dd

¹ This variable has been altered so it is now unique (by patid) within and across all HES years.

3. Appointment (hesop_appointment_XXX.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique key given to a patient in CPRD GOLD or CPRD Aurum [primary key, in combination with attendkey]	INTEGER	20
attendkey ³	Record identifier (unique in combination with patid) [primary key, in combination with patid]	INTEGER	20
ethnos	Ethnic category as recorded at appointment	CHAR	9
admindcat	Administrative category 1=NHS, 2=Private	INTEGER	8
apptdate	Appointment date	CHAR	yyyy/mm/dd
apptage	Age on day of appointment	INTEGER	8
atentype	Attendance type	INTEGER	8
attended	Attended or did not attend	INTEGER	8
dnadate	Last DNA or patient cancelled date	CHAR	yyyy/mm/dd
firstatt	First attendance	CHAR	9
outcome	Outcome of attendance	INTEGER	8
priority	Priority type	INTEGER	8
refsourc	Source of referral	INTEGER	8
reqdate	Referral request received date	CHAR	yyyy/mm/dd
servtype	Service type requested	INTEGER	8
stafftyp	Medical staff type seeing patient	INTEGER	8
wait_ind	Waiting calculation indicator	INTEGER	8
waiting	Days waiting	INTEGER	8

³ This variable has been altered so it is now unique (by patid) within and across all HES years.

4. Clinical (hesop_clinical_XXX.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique key given to a patient in CPRD GOLD or CPRD Aurum [primary key, in combination with attendkey]	INTEGER	20
attendkey ³	Record identifier (unique in combination with patid) [primary key, in combination with patid]	INTEGER	20
diagnosis		CHAR	9
icdx		CHAR	9
icd		CHAR	9
diag_order		INTEGER	8
tretspef	Treatment speciality	CHAR	9
mainspef	Main speciality	CHAR	9

5. Operations (hesop_operation_XXX.txt)

<i>Column name</i>	<i>Description</i>	<i>Type</i>	<i>Format</i>
patid	Encrypted unique key given to a patient in CPRD GOLD or CPRD Aurum [primary key, in combination with attendkey]	INTEGER	20
attendkey ³	Record identifier (unique in combination with patid) [primary key, in combination with patid]	INTEGER	20
operation		CHAR	9
opcs		CHAR	9
opertn_order		INTEGER	8
operstat		CHAR	9
tretspef	Treatment speciality	CHAR	9
mainspef	Main speciality	CHAR	9

Reference

1. Data Science Team NHS England (2024). The Person_ID Handbook. A user guide to Person_ID and its creation via the Master Person Service, with applications to HES data sets. [Last accessed: 11 September 2024] Available here: [The Person_ID Handbook - NHS England Digital](#)