

# HES - DID Linkage Report

Provisional Summary Statistics  
England, April 2016-May 2016, experimental statistics

Published 06 October 2016

**Following the NHS Digital 'Consultation on Statistics 2016-17 to 2018-19', we are stopping production of this monthly HES-DID Linkage publication series. This is the last release but linkage data will continue to be made available to extract customers. In addition, key figures about each month's linkage, which are currently made available in this publication, will instead be made available on the NHS Digital website each month.**

## **This is an Experimental Statistics publication**



**This document is published by NHS Digital,  
part of the Government Statistical Service**

Experimental statistics are official statistics which are published in order to involve users and stakeholders in their development and as a means to build in quality at an early stage. It is important that users understand that limitations may apply to the interpretation of this data. More details are given in the report.

All official statistics should comply with the UK Statistics Authority's Code of Practice for Official Statistics which promotes the production and dissemination of official statistics that inform decision making.

Find out more about the Code of Practice for Official Statistics at [www.statisticsauthority.gov.uk/assessment/code-of-practice](http://www.statisticsauthority.gov.uk/assessment/code-of-practice)

Find out more about Experimental Statistics at [https://gss.civilservice.gov.uk/wp-content/uploads/2016/02/Guidance-on-Experimental-Statistics\\_1.0.pdf](https://gss.civilservice.gov.uk/wp-content/uploads/2016/02/Guidance-on-Experimental-Statistics_1.0.pdf)

ISBN: 978-1-78386-830-8

This report will be of interest to people wishing to make use of the HES and DID linked datasets. This will include commissioners, policy makers, NHS organisations, benchmarking organisations, academics and researchers.

## Contents

---

|   |          |
|---|----------|
| <b>Executive Summary</b>                | <b>4</b> |
| <b>Findings</b>                         | <b>4</b> |
| <b>Understanding the data</b>           | <b>6</b> |
| <hr/>                                   |          |
| HES (Hospital Episode Statistics)       | 6        |
| What is Hospital Episode Statistics?    | 6        |
| DID (Diagnostic Imaging Dataset)        | 7        |
| What is the Diagnostic Imaging Dataset? | 7        |
| HES-DID Linkage                         | 8        |
| <hr/>                                   |          |
| <b>Linkage methodology</b>              | <b>9</b> |
| <hr/>                                   |          |
| Duplicate matches                       | 10       |
| Excluded data                           | 11       |

## Executive Summary

In the period April 2016-May 2016:

- **6.9 million**<sup>1</sup> imaging procedures were recorded in the Diagnostic Imaging Dataset as having taken place. It was possible to link **5.7 million** of these records to individuals in the Hospital Episode Statistics data set.
- **1.1 million** patients<sup>2</sup> who were admitted to hospital at least once were identified as having at least one imaging procedure during the same period.
- **1.3 million** patients who had at least one accident and emergency attendance were identified as having at least one imaging procedure during the same period.
- **2.5 million** patients who had at least one outpatient appointment were identified as having at least one imaging procedure during the same period.

## Findings

This is the latest statistical publication of linked HES (Hospital Episode Statistics) and DID (Diagnostic Imaging Dataset) data held by NHS Digital. The HES-DID linkage provides the ability to undertake national (within England) analysis along acute patient pathways to understand typical imaging requirements for given procedures, and/or the outcomes after particular imaging has been undertaken, thereby enabling a much deeper understanding of outcomes of imaging and to allow assessment of variation in practice.

This publication aims to highlight to users the availability of this updated linkage and provide users of the data with some standard information to assess their analysis approach against.

The two data sets have been linked using specific patient identifiers collected in HES and DID. The linkage allows the data sets to be linked from April 2012 when the DID data was first collected; however this report focuses on patients who were present in either data set for the period April 2016-May 2016 only. For DID this is provisional 2016/17 data. For HES this is provisional 2016/17 data.

The linkage used for this publication was created on 7th September 2016 and released together with this publication on 06 October 2016.

<sup>1</sup> This figure excludes procedures specifically relating to fertility which are deemed sensitive. See the linkage Methodology section for further details on excluded imaging codes. This figure may also be slightly higher than figures published in the NHS England report on the DID data set as it includes all recorded activity, even when the modality (type of imaging) is unknown.

<sup>2</sup> For the purposes of this publication HES ID has been used to identify individuals. Counting unique HES IDs recorded in HES provides an approximation of the number of people accessing hospital services. In some circumstances the same HES ID may have been assigned to different individuals and in some circumstances the same individual may be represented by more than one HES ID.

**Table 1: Number of DID records in the period April 2016-May 2016 which can be linked to an individual in HES**

| <b>Number of Records in DID</b> | <b>Number of records in DID that can be linked to an individual in HES (provisional)</b> |
|---------------------------------|--|
| <b>6,850,158</b>                | <b>5,719,045</b>   |

Table 1 shows that a high proportion of individuals who accessed diagnostic imaging during April 2016-May 2016 can be identified as an individual who also accessed hospital services at least once during the same period. This use of the linkage does not map imaging procedure records to the corresponding HES episode or attendance record. It simply identifies the presence of an individual in both data sets during the period of the data analysed.

In the HES data sets the HES ID can be used to identify unique individuals. The HES ID is created using the patient identifiers in the HES index. Full details of the methodology used to derive the HES ID and its limitations can be found at the link below.

[http://www.hscic.gov.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HESpatient-ID/pdf/HES\\_ID\\_Replacement\\_Nov09.pdf](http://www.hscic.gov.uk/media/1370/HES-Hospital-Episode-Statistics-Replacement-of-the-HESpatient-ID/pdf/HES_ID_Replacement_Nov09.pdf)

**Table 2: Number of individuals in HES data sets, as identified by HES ID during the period April 2016-May 2016 who can be identified as present in the DID for the same period**

| <b>HES Data set (provisional)</b> | <b>Number of individuals<sup>3</sup> in the data set</b> | <b>Number of individuals in the data set who are also present in DID</b> |
|-----------------------------------|--|--|
| HES Admitted Patient Care         | 2,231,740  | 1,053,379  |
| HES Accident and Emergency        | 2,872,428  | 1,281,525  |
| HES Outpatient                    | 9,194,931  | 2,498,218  |

Not all individuals who access hospital services are expected to have also had at least one imaging procedure within a given period; therefore conclusions cannot easily be drawn from the data presented in Table 2. These figures are provided as a reference point for users of the linkage. A patient may be present in any combination of the HES data sets therefore these figures cannot be added together to give a total count of individuals in HES.

<sup>3</sup> Patients who have only ever attended hospital as day or night cases or who have had only one episode that is recorded as unfinished, during the period, are included in these counts.

# Understanding the data

## HES (Hospital Episode Statistics)

### What is Hospital Episode Statistics?

HES is a data warehouse containing details of all admissions (episodes of care), outpatient appointments and A&E attendances at NHS hospitals in England, where these have been recorded by the provider.

This data is collected during a patient's time at hospital and is submitted to allow hospitals to be paid for the care they deliver. HES data is designed to enable secondary use, that is use for non-clinical purposes, of this administrative data.

It is a records-based system that covers all NHS trusts in England, including acute hospitals, primary care trusts and mental health trusts. Strict statistical disclosure control in accordance with the HES Analysis Guide is applied to all published HES data. This suppresses small numbers to stop people identifying themselves and others, thereby ensuring patient confidentiality.

Further information relating to the latest published annual HES data sets are available within the links below. Users are also advised to read the accompanying HES data quality reports.

#### **HES Inpatients:**

<http://www.digital.nhs.gov.uk/pubs/hes1415>

#### **HES Outpatients:**

<http://www.digital.nhs.uk/pubs/outpatientactivity1415>

#### **HES Accident and Emergency:**

<http://www.digital.nhs.uk/pubs/aandeattendance1415>

Information about the processing of HES data and HES data quality can be found here <http://www.digital.nhs.uk/article/1825/The-processing-cycle-and-HES-data-quality>

## **DID (Diagnostic Imaging Dataset)**

### **What is the Diagnostic Imaging Dataset?**

DID is a data set which captures details about imaging tests carried out on NHS patients in England. It was introduced to monitor progress on Improving Outcomes: A Strategy for Cancer (IOSC). This strategy, published 12 January 2011, set out how the Government, NHS and public can help prevent cancer, improve the quality and efficiency of cancer services and move towards achieving outcomes that rival the best. To achieve that ambition, it will be essential to prevent more cancers developing in the first place and to ensure they are diagnosed while the cancer is at an earlier stage to increase the scope for successful treatment. Within that, GPs need easy access to the right diagnostic tests to help them to diagnose or exclude cancer earlier. Therefore the DID reports on imaging activity, referral source and timeliness.

This administrative data is extracted from each imaging provider's Radiological Information System (RIS) and contains details about the patient, the referral for imaging, the provider of the imaging and the imaging procedure itself, including dates of referral, test and report. The DID system uses the data extracted from providers' RIS to derive additional information which makes the data set more suitable for secondary use. Data is available from April 2012 and the first year of data exhibits some data quality issues, ranging from coverage in certain months to validity of some data fields.

Submitters of DID data have a three month period in which they can revise previously submitted data or add additional records. This three month period is based on date of test; for example, the records for tests carried out in April can be submitted and/or revised in May, June and July.

NHS England produces a monthly publication on the DID data set which looks at certain types of imaging (selected modalities) and imaging procedures used for early diagnosis of cancer. The data included each month is that which is no longer subject to revision under the three month submission window. For example a monthly publication released in August includes all tests carried out up until the end of April.

A more detailed publication on the 2014/15 final dataset for DID was published by NHS England in October 2015.

The DID data available in an extract for analysis and linkage to HES is aligned with data used for the NHS England monthly publication.

NHS England publications about DID are available from the following link:  
<http://www.england.nhs.uk/statistics/statistical-work-areas/diagnostic-imaging-dataset/>

Users are advised to read the commentary and tables in conjunction with the technical report which highlights data quality issues which may cause limitations.

## **HES-DID Linkage**

### **What is the Hospital Episode Statistics and Diagnostic Imaging Dataset linkage?**

HES has a person index which is used to identify individuals who access hospital services while keeping their personal details, such as date of birth, residential postcode, NHS number etc. confidential. For example, when new HES data for an event (such as an inpatient episode of care) is received, the individual's personal details are checked against the existing index to find their HES ID, if no match is found a new ID is created for that individual.

DID does not currently have a patient index. Personal information for each record in DID is held securely at the HSCIC. This includes NHS number, date of birth, gender and postcode.

The linkage matches the HES patient index to the personal information in DID using the individual's NHS number, date of birth, gender, and postcode, which are also stored in the HES Index. It enables the linkage of all patients' imaging records from April 2012 with the same patients' recorded interactions with secondary care services covering Inpatients, Outpatients and Accident & Emergency. The output of this process is a table which contains, for each DID record identifier, the pseudonymised HES ID that the associated personal identifying fields were linked to and the score to indicate the strength of the match between these two sets of identifiers. This is referred to as the HES-DID bridging file.



## Linkage methodology

The patient identifiers are matched using a specific linkage algorithm in which each unique combination of identifiers from DID are passed through the matching process in nine different passes, each pass having different (less stringent) matching criteria defined in Tables 3 and 4. Each set of DID identifiers which is matched to a HES ID is given a 'match rank' that indicates which pass of the algorithm it matched at. This match rank gives an indication of the confidence of the match. The HES ID is then added to the non-identifiable DID record through the use of a linkage key between DID personal identifiable data and the anonymised imaging record.

All records in DID are passed through the algorithm, this includes records for children from birth onwards. Therefore no age filters are required during any part of the matching or subsequent analysis.

## Quality of HES-DID Linkage

The HES-DID bridging file is created through trying to link all known HES IDs with all unique combinations of personal information in DID, and is not restricted by time period allowing linkage of records in any period of time. At the time of this linkage being produced, the linkage file contained 33,296,734 HES IDs that matched to at least one unique combination of personal information from DID. The number of HES IDs that matched by match rank can be seen in table 3 below (the full definition of each match rank can be found in table 4). Please note that a single HES ID may match to multiple combinations of personal information in DID, at different match ranks and so the total HES IDs which matched to at least one record will not equal the sum of HES IDs by match rank.

**Table 3: Matching criteria for each of the nine match ranks**

| Match Rank           | Matched HES IDs   | NHS Number | Date of Birth | Gender | Postcode |
|----------------------|-------------------|------------|---------------|--------|----------|
| 1a                   | 30,097,097        | Exact      | Exact         | Exact  | Exact    |
| 1b                   | 2,264,381         | Exact      | Exact         |        | Exact    |
| 2                    | 2,710,193         | Exact      | Exact         | Exact  |          |
| 3                    | 25,304            | Exact      | Partial       | Exact  | Exact    |
| 4                    | 6,861             | Exact      | Partial       | Exact  |          |
| 5                    | 333,423           | Exact      |               |        | Exact    |
| 6                    | 766,560           |            | Exact         | Exact  | Exact    |
| 7                    | 46,457            |            | Exact         | Exact  | Exact    |
| 8                    | 831,805           | Exact      |               |        |          |
| <b>Total Matched</b> | <b>33,296,734</b> |            |               |        |          |

**Table 4: Match Rank Definitions**

| Match Rank | Definition   |
|------------|--|
| 1a         | Where all the fields NHS Number, Date of Birth, Gender and Postcode match exactly in both the HES Patient index and the DID data.  |
| 1b         | Fields in both HES index and DID data match exactly for NHS Number, Date of birth and Postcode.  |
| 2          | Fields in both HES index and DID data match exactly for NHS Number, Date of birth and Gender.  |
| 3          | Fields in both HES index and DID data match exactly for NHS Number, Gender and Postcode, there is also a partial match between the Date of birth records.  |
| 4          | Fields in both HES index and DID data match exactly for NHS Number and Gender , there is also a partial match between the Date of birth records.   |
| 5          | Fields in both HES index and DID data match exactly for NHS Number and Postcode.   |
| 6          | Fields in both HES index and DID data match exactly for Date of birth, Gender and Postcode. Where the NHS number is 'Null' in either ID or equal in both and Date of birth is not 1st January and the Postcode is not in the communal established list. <sup>4</sup> |
| 7          | Fields in both HES index and DID data match exactly for Date of birth, Gender and Postcode. Where the NHS number is 'Null' in either ID or equal in both and Date of birth is not 1st January.   |
| 8          | Fields in both HES index and DID data match exactly for NHS Number only.   |

There are currently 43,089,125 unique combinations of personal information in DID. Of these, 1,726,958 have not yet been matched to any HES ID. This is a match rate of 96.0%.

## Duplicate matches

Data quality issues exist in both the DID data and in HES ID. In DID there is no patient index and none of the fields used in the matching process are mandatory, therefore the same person could be represented by different combinations of the identifiable data. In some cases there are errors or problems with the identifiable data in DID due to the collection being relatively new and submitters working to refine their data extraction and preparation methods.

In the HES Index there can be instances where the same person has been given two or more different HES IDs and also where the same HES ID has been matched to different people.

These issues can lead to duplicates in the HES-DID matching process, where the combinations of identifiable data which represent the same person in DID can be matched to different HES IDs or where the same HES ID can be matched to identifiable data that may represent different people. In general such duplicates and poor quality matches have been found to be present at Match Rank 8, which matches only on NHS number. A user of the data may wish to exclude matches at Match Rank 8 to reduce the risk of including duplicate matches, however the high level analysis described in this report used all matched data.

<sup>4</sup> <http://systems.digital.nhs.uk/data/ods/datadownloads/onsdata>

## Excluded data

No filters have been applied to DID or HES data during this analysis; the linkage uses all data available, including any HES records where the appointment was not attended. Both data sets cover the entire population receiving NHS funded secondary care services from a provider in England, including children. As this report looks primarily at the number of individuals recorded in HES, regular day and night attenders and unfinished episodes were included in the analysis. Filters are usually applied to HES data to exclude these cases when the data is being analysed at episode level.

The anonymised DID extract used to provide the counts in Table 1 excludes records where the procedure has been identified as a sensitive procedure. Most imaging procedures that can be used to diagnose a condition or disease that is classed as sensitive are also used to diagnose other non-sensitive conditions. Furthermore DID data does not include any information about the reason a patient was referred for imaging. However there are a small number of procedures that are only ever used in diagnosis or intervention of fertility related conditions. These have been deemed sensitive records in DID following a clinical review and are therefore not available in the anonymised DID extract. The imaging procedure codes that have been classed as sensitive are as follows:

National Interim Clinical Imaging Procedure (NICIP)

FHYSG

UFOTA

UFOTV

UHYSTC

Systematized Nomenclature of Medicine Clinical Terms (SNOMED – CT) Codes

168884001

241489004

252939009

432787004

Records present in the Admitted Patient Care data set which have no admission or discharge dates recorded have not been included in the analysis for this report. This affects 2,048 records.

# Information and technology for better health and care

[www.digital.nhs.uk](http://www.digital.nhs.uk)

0300 303 5678

[enquiries@nhsdigital.nhs.uk](mailto:enquiries@nhsdigital.nhs.uk)

 [@nhsdigital](https://twitter.com/nhsdigital)

ISBN: 978-1-78386-830-8

This publication may be requested  
in large print or other formats.

Published by NHS Digital, part of the  
Government Statistical Service

NHS Digital is the trading name of the  
Health and Social Care Information Centre.

Copyright © 2016

## OGI

You may re-use this document/publication (not including logos)  
free of charge in any format or medium, under the terms of the Open  
Government Licence v3.0.

To view this licence visit

[www.nationalarchives.gov.uk/doc/open-government-licence](http://www.nationalarchives.gov.uk/doc/open-government-licence)

or write to the Information Policy Team, The National Archives,  
Kew, Richmond, Surrey, TW9 4DU;

or email: [psi@nationalarchives.gsi.gov.uk](mailto:psi@nationalarchives.gsi.gov.uk)