General Details	
Dataset Title:	Hospital Admissions: Summary Statistics, 2007/08
Time Period of Dataset:	1 April 2007 – 31 March 2008 (Financial Year)
Geographic Coverage:	England
Lowest Area Output:	Middle Super Output Area (MSOA)
Supplier:	NHS Information Centre for Health and Social Care, formerly Health and Social Care Information Centre (HSCIC)
Department:	HES
National Statistics Data?	Other Official Statistic
Revisions:	None

Data Quality

This document provides a range of information that describes the quality of the data and details any points that should be noted when using the data.

ONS has developed Guidelines For Measuring Statistical Quality ; these are based upon the six European Statistical Service (ESS) Dimensions of Quality developed by Eurostat. The dimensions are:

- Relevance
- Accuracy
- Timeliness and Punctuality
- Accessibility and Clarity
- Comparability
- Coherence

About the dataset **(including the quality dimensions: Relevance and Timeliness and Punctuality)**

Hospital Episodes Statistics (HES) is a data warehouse containing records for episodes of admitted patient care to NHS hospitals in England. It includes information on the patient, when and where they were treated, diagnoses and operations, type of admission and various supplementary and derived data. The information has many uses and supports a wide variety of work within the Department of Health, other government departments, the NHS, Public Health Observatories, medical research and the general public. HES provides the only consistent, national database of all admitted patient care in NHS Trusts in England. Activity not covered includes Accident and Emergency attendances, admission of English patients to hospitals outside of England (unless NHS commissioned and recorded) and primary care.

The HES data provided here is for Finished Admission Episodes (FAE) with counts being produced for England at Government Office Region (GOR), Local Authority (LA) and Middle Super Output Area level for the 2007/08 financial year.

The information covers the following specified diagnosis, cause and operative procedures which generally have a higher prevalence amongst older members of the population:

1. Diagnosis - Coronary Heart Disease (CHD)
2. Diagnosis - Cerebrovascular Disease (including Stroke)
3. Diagnosis - Cancer (excluding non-melanoma skin cancer)
4. External Cause - Falls (basic accidental falls)
5. Operation - Coronary Artery Bypass Graft (CABG) and Percutaneous Transluminal Coronary Angioplasty (PTCA) (Heart)
6. Operation - Hip Replacement (Total and hybrid hip replacements)
7. Operation - Knee Replacement (Total knee replacement)
8. Operation - Cataract

Cancer, coronary heart disease (CHD) and cerebrovascular disease are major causes of death (accounting for around two-thirds of all deaths) and ill-health. There is a strong social gradient in these diseases, with incidence and mortality higher among more deprived groups.

HES annual data are generally published around December for the previous financial year with the data being made available on the Neighbourhood Statistics website as soon as possible after publication. The current

time series starts with the 2002/2003 dataset.

Other Hospital Episode Statistics are available on the Neighbourhood Statistics website including; age and sex breakdown information 'Hospital Admissions by Age and Sex.

How the data are collected

Hospital Episode Statistics (HES) provides information about the admitted patient treatment delivered by NHS hospitals in England and is collected as a part of the hospital admission process. The data includes private patients treated in NHS hospitals, patients who were resident outside of England and care delivered by the independent sector funded by the NHS, however is should be noted that this dataset only includes data for patients resident in England. This includes day case surgery. Data for out-patient appointments are also collected in HES but not included in this data. NHS Trusts submit their data to the Secondary Uses Service (SUS) for onward transmission to commissioners and HES is built from a national, annual extract of these records. The data are then cleaned, validated and new fields derived. (For further information, please see the 'Validation and Quality Assurance' section).

Figures have not been adjusted for shortfalls in data (i.e. the data are ungrossed).

Concepts and Definitions

Finished Admission Episode (FAE): A finished admission episode is the first period of in-patient care under one consultant within one healthcare provider. A FAE is counted irrespective of diagnosis or treatment, but mothers giving birth in hospital, babies delivered in hospital and patients formally detained under the provisions of mental health legislation or long-term (over one year) psychiatric patients have been excluded.

It is not uncommon for a given person to be admitted to hospital more than once during a year. It is for this reason, that the number of FAE's shown in the dataset may exceed the number of individual persons treated within the year.

In order to assist in the estimation of disease incidence resulting in hospitalisation, and also the analysis of clinical practice, the average number of admission episodes attributable to individual persons has been approximated by analysing patient records submitted during the 2007/08 financial year.

Coronary Heart Disease (CHD) – 1.63 in-patient admissions per person annually (e.g. every 16 admission episodes with a mention of CHD as a diagnosis will, on average, be generated as a result of 10 individuals admitted with CHD).

Cerebrovascular Disease (including stroke) – 1.29 in-patient admissions per person annually (e.g. every 13 admission episodes with a mention of Cerebrovascular Disease (including Stroke) as a diagnosis will, on average, be generated as a result of 10 individuals admitted with a stroke).

Cancer (excluding non-melanoma skin cancer) – 3.38 in-patient admissions per person annually (e.g. every 34 admission episodes with a mention of Cancer as a diagnosis will, on average, be generated as a result of 10 individuals admitted with cancer).

Falls - 1.10 in-patient admissions per person annually (e.g. approximately every 11 admission episodes due to a fall will, on average, be generated as a result of 10 individuals being admitted for a fall).

Note: the ratios give an estimation of the number of finished admission episodes likely to be generated by each admitted patient during the year. This includes data for patients where their age or/and sex was unknown. Episode/person ratios have not been calculated for the operation groups, as they would be less meaningful. Note, however, that in the case of Cataract and Hip operations, it is not uncommon for the same person to undergo two such treatments during a year.

A patient could have a FAE for diagnoses or operations different to those listed. For example, they may be in hospital for a diagnosis other than a stroke, coronary heart attack or cancer, or may have had an operation for something other than a hip replacement, cataract treatment or an operation involving the heart. This is why the variable totals in this dataset will not sum up to the total FAE count.

All Diagnoses – count of episodes: These figures represent a count of all FAE's where the diagnosis was mentioned in any of the 20 diagnosis fields in a HES record.

Primary Diagnosis: The primary diagnosis is the first of these 20 diagnosis fields in the HES data set and provides the main reason why the patient was in hospital.

Secondary Diagnosis: The remaining 19 of the diagnosis fields are the secondary diagnosis fields in HES that show other diagnoses relevant to the episode of care.

All Operations – Count of episodes These figures represent a count of all FAE's where the procedure was mentioned in any of the 24 operation fields in a HES record.

Main Operation: The main operation is the first recorded operative procedure in the HES data set and is usually the most resource intensive procedure performed during the episode.

Patient counts: Patient counts are based on the unique patient identifier HESID. This identifier is derived based on patient's date of birth, postcode, sex, local patient identifier and NHS number, using an algorithm.

Provider: This is the organisation that provided the admitted patient care, typically an NHS, or Primary Care, Trust.

Grossing: Figures published by HES may be adjusted to take account of shortfalls (or duplication) in the records received, and also for missing diagnoses and operations codes. This generally results in a slight uplifting of the national counts. However, grossing is generally not appropriate at a low geographic level and data for Neighbourhood Statistics are ungrossed.

Datayear: Each HES database table contains records relating to a particular reporting year (the period 1st April to the 31st March). This is often referred to as the HES datayear.

Data Classifications	
Standard Classifications used (if any):	All Diagnoses coded according to the **International Classification of Diseases and Related Health Problems – Tenth Revision (ICD-10).** More information on ICD-10 is available. The following codes are included in this dataset: Codes in the range of I20 to I25 (Coronary Heart Disease). Codes in the range of I60 to I69 (Cerebrovascular Disease including Stroke). Codes in the range of C00 to C97 (excluding C44) (Cancer excluding non-melanoma skin cancer). Codes in the range of W00 to W19 (Basic accidental falls) All Operations coded according to the **Classification of Surgical Operations and Procedures, Fourth Revision (OPCS-4.4).** More information about OPCS-4 is available. The following codes are included in this dataset: Codes in the range of K40 to K44 (Coronary Artery Bypass Graft). Codes from K49, K50.1, and K75 (Percutaneous Transluminal Coronary Angioplasty. Codes in the range of W37 to W39 and W93 to W95 (Total and hybrid hip replacements). Codes in the range of W40 to W42 (Total knee replacement) Codes in the range of C71 to C75 (Cataract).
Validation and Quality Assurance **(including the quality dimensions: Accuracy, Comparability and Coherence)**	

HES go through several stages of data quality checks and measures to make the data as useable as possible. A number of checks are employed throughout the year to monitor the data and a dialogue exists between trusts and the HES data quality team to help resolve problems before the end of each financial year. HES cleaning processes are then applied to pick up anomalies in the data set.

- <u>Manual cleaning</u>
  This is a set of cleaning rules individually specified (usually at trust level), applied to the data to correct any anomalies. These are mainly to remove duplicates or reassign data appearing under the wrong trust code. Some cleans are specified by the trust, for example to correct data where a system problem has made it impossible for the trust to correct the data.

- <u>Auto cleaning and derivation</u>
  A set of rules are applied to the data to 'standardise' it into a format that is easy for analysts to use, eg. by overwriting inappropriate values that are seen to be inconsistent and replacing invalid values with a default code for 'not known'. At the same time, new fields are derived (e.g. age of patient, duration of elective wait, length of stay).

Nevertheless, HES is reliant on trusts to provide complete, accurate and reliably coded data and some inaccuracies will remain where it is not possible to compensate for shortcomings. Further information on HES data quality can be found on the HES website. There are a number of data quality indicators for HES data that are relevant to these tables, although they measure the validity rather than the accuracy of the data. These are listed below with their score in 2007/08:

- Record Linkage Component (examines the validity of date of birth, gender and the postcode entries in records): 99%
- Diagnosis Component (assesses the proportion of missing or invalid primary diagnosis codes): 99%
- Operation Codes Component (assesses the proportion of invalid operative procedure codes): 100%

Above results indicate that the data quality of the fields used to produce 2007/08 table were reliable. Records with 'not known' or 'not specified' sex values were excluded from the diagnosis and operations grouped data (this accounted for 0.008% of the Finished Admission Episodes).
Records with 'not known' or 'missing' age values were excluded from the diagnosis and operations grouped data (this accounted for 0.002% of the Finished Admission Episodes).

Note that 'All Finished Admission' episodes does include 'not known' or 'not specified' sex values as well as 'not known' or 'missing' age values.

Changes to Coding Classifications - OPCS4
This dataset was referenced to the same versions of ICD-10 for diagnosis groups and OPCS4.3 for procedure groups (as the other HES datasets).

Operative procedure codes have been revised in 2007/08 to OPCS4.4. All codes that were in OPCS4.2 and OPCS4.3 remain in OPCS4.4, although the OPCS4.4 codes enable the recording of interventions and procedures which were not possible in OPCS4.2 and OPCS4.3. There have been no updates of the clinical codes for the procedure groups used in this analysis.

Additionally some activity may have been coded under different codes in OPCS4.2, OPCS4.3 and OPCS4.4. These changes need to be borne in mind when analysing time series and may explain any growth over time. More information about OPCS4 changes on the Connecting for Health website (www.connectingforhealth.nhs.uk).

Geographic Referencing

From 2002/03, all datasets have been referenced to the boundaries in place at the time of collection. Since there have been no boundary changes at Local Authority (LA) level in England that affect comparability and the time series.

The geographical area to which the FAE belonged was derived from the postcode of the patient, where given and where English FAEs were listed in this dataset only if the patient had an English postcode. This occurred for around 99% of relevant cases). Records with missing postcodes, or those not matched to the ONS geographic reference data postcode file for Neighbourhood Statistics, were excluded from the data.

In a small number of cases HES references the Middle Layer Super Output Area to a different Local Authority than the reference postcode directory. This affects less than 0.05% of total England cases, and, in the worse affected LA, less than 0.5% of cases. The data provided has been corrected in order to preserve comparability to the reference postcode directory. This does mean that some of the figures provided for a selected number of Local Authorities will not be comparable to other Hospital Episode Statistics at Local Authority level.

Disclosure Control

When publishing statistics that are produced by the Government Statistical Service reasonable steps are taken to protect confidentiality. In this datasets some/all of the variables have been given protection through the use of Statistical Disclosure Control techniques.

Every effort has been made by the NHS Information Centre to ensure the data are not disclosive. A statistical disclosure control software tool called Tau Argus has been used to protect potentially disclosive cells in this dataset. Data have been suppressed in this dataset to protect both the confidentiality of individual information and the potential statistical instability caused by low counts.

ONS has carried out a number of checks to ensure that confidentiality has been maintained within the dataset. In accordance with standard procedures these data have been reviewed and approved for release.

Sources for Further Information or Advice
(including the quality dimension: Accessibility and Clarity)

Tabulations covering a wide range of NHS in-patient activity, plus explanations of the content and further information on HES, may be downloaded from the Hospital Episode Statistics web pages.

Specific information about this dataset can be obtained from:
SUS/HES Analysis Team
The NHS Information Centre for Health and Social Care
Trevelyan Square
Boar Lane
Leeds
LS1 6AE

Telephone: 0845 300 6016
e-mail: hes.questions@ic.nhs.uk