Department of Health, Hospital Episode Statistics

HES

The Book

October 2000 edition

Document number:	HES-002
Issue:	4.3
Date:	October 2000
HES version	6.0

This document is distributed by:

Department of Health
Hospital Episode Statistics (SD2HES)
Skipton House
Room 430B
London
SE1 6LH

To obtain a copy, telephone 020 79725683, fax 020 79725662 or email sd2hes@doh.gov.uk, and ask to be put on the distribution list.

Contents

Contents 3
Preface 5
Related Documents 5
Conventions Used 5

1. What is HES? 6

2. Why HES? 7

2.1 Supporting Government 7
2.2 Supporting the NHS 7
2.3 Supporting Medical Research 7
2.4 Supporting the Public 7

3. HES Records 8

3.1 Record Format 8
3.2 Types of Record 9
3.3 HES Fields 9

4. Episodes, Spells and Patients 10

4.1 Transfers 10
4.2 Calculating Totals 10

5. Brief HEStory of Time 11

5.1 Duration of Episodes and Spells 11
5.2 Year Boundaries 11
5.3 Hospital Stays That Straddle HES Years 12
5.4 Example 12

6. Method of Admission and Waiting Time 15

6.1 Emergency Admissions 15
6.2 Elective Admissions 15
6.3 Waiting Time 15
6.4 Typical Waiting Time Enquiry 16

7. Patient Classification 17

8. Clinical Coding 18

8.1 Diagnosis 18
8.2 Operative Procedures 18

9. Maternity Data 19

9.1 Delivery and Birth Records 19
9.2 Well Babies 20
9.3 Home Births and Births in Non-NHS Hospitals 20

10. Psychiatric Data 21

10.1 Psychiatric Census 21

11. Data Quality 22

11.1 Monitoring Quality 22
11.2 Verification 22

11.3 Autocleaning 22
11.4 Validation 23
11.5 Grossing 23
11.6 Reporting to Data Providers 23
12. Accessing HES Data 24
12.1 Custom Analyses 24
13. Analysing HES Data 25
13.1 Analysis Standards 25
13.2 Counting Diagnoses 26
13.3 Counting Operative Procedures 26
14. Further Information 27

Preface

This document is a comprehensive introduction to the Hospital Episode Statistics (HES) system. It is intended for anyone who needs to understand the principles of HES and how its data can be used.

You should read the whole document, because each new term is described only on the first occasion it appears.

Readers are assumed to have access to the related documents listed below.

Related Documents

[1] HES Data Dictionary, HES-008, available at: http://tap.ccta.gov.uk/doh/hes_dd.nsf

[2] NHS Data Manual*

[3] NHS Data Dictionary*

[4] NHS CDS Manual*

[5] Specifying Extracts, HES-006

[6] Specifying Tabulations, HES-007

[7] ICD-9, Manual of the International Statistical Classification of Diseases, Injuries, and Causes of Death, based on the recommendations of the Ninth Revision Conference (ISBN 92 4 154 005 2)

[8] ICD-10, International Classification of Diseases and Related Health Problems, Tenth Revision (ISBN 92 4 154 419 8, 420 1, 421 X)

[9] OPCS4, Tabular List of the Classification of Surgical Operations and Procedures, Fourth Revision (ISBN 0 11 691 295 2)

*Documents [2], [3] and [4] are available via the NHS net at: http://nww.standards.nhsia.nhs.uk/ds/index.htm

Conventions Used

HES data items (fields) are shown in bold, for example **classpat**.

1. What is HES?

The Hospital Episodes Statistics (HES) system is a powerful database containing personal, medical and administrative details of all patients admitted to, and treated in, NHS hospitals in England. This information has many uses including:

- Policy development
- Illustrating variations in health status and health delivery through time and across geographic areas
- Producing comparative statistics to assist in performance management
- Medical research
- Helping to determine how much of the taxpayers money should be spent on health care, and how it should be distributed

To increase access to HES, tables summarising the data are published annually by the Department. Some of these may be downloaded free of charge from the Internet at www.doh.gov.uk/hes.

The records for the database are collected from all hospital NHS Trust providers of in-patient care in England. (Northern Ireland, Scotland and Wales have their own systems.) This data is currently relayed via the NHS-Wide Clearing Service (NWCS). Extracts are taken quarterly from the NWCS database and sent to the contractor appointed by DoH to run the HES system.

HES data is split into years, which run from the 1 April to the 31 March in the following year. After the end of a HES year, the data is subjected to checks before being made available to HES users. During a year, some users have access to provisional, quarterly data.

Over 11 million HES records are now generated each year, of which about 1.2 million are for maternity records for mothers and babies (including home and private deliveries). In addition, there are around 200,000 records for patients who are still in hospital at the end of the HES year; of these, some 27,000 are for psychiatric patients. There is at least one record for each patient's stay in hospital.

2. Why HES?

HES supports a wide range of work in the Department and for NHS management. The following sections cite major examples. In addition, there is a great deal of interest in HES data from external researchers, professional bodies and commercial organisations.

2.1 Supporting Government

The Public Expenditure Survey (PES) model is used for negotiating the funds for the Health Service with the Treasury (c.35billion). HES data is used in these negotiations with the Treasury to procure funding and inform the Departmental process of allocating the awarded funds within the NHS.

HES is used to answer Parliamentary Questions.

HES supports statutory requirements for the submission of mental illness or mental health data.

HES supports work on health monitoring: the impact of diseases and environmental conditions.

2.2 Supporting the NHS

HES is used to provide measures of performance, for example:

- High Level Performance Indicators (HLPI), including analysis of quarterly data giving provisional, in year results.
- Direct outcome measures or clinical indicators (CIs) of hospital care, for example, readmission rates.
- National Data Quality Indicator (NDQI).
- Public Health Outcome Indicators (PHOI).

The HLPIs, CIs and NDQI form part of the Performance Assessment Framework.

HES provides clinical and administrative information, for example:

- HES allows analyses of specific conditions and operations performed in the Health Service, by age and sex if required.
- HES provides maternity services information.
- HES provides detailed information about psychiatric patients and their length of stay.
- HES is used to estimate occupancy rates by comparing occupied bed days in HES with available bed days.

2.3 Supporting Medical Research

HES provides analyses over time for epidemiological and other research work.

2.4 Supporting the Public

HES is used to support both Community Care and Children's Issues.

3. HES Records

Over 11 million HES records are now collected annually from the NHS. They are constructed from data collected locally by the NHS via Patient Administration Systems (PAS) or Hospital Information Systems (HISS).

3.1 Record Format

Each general record contains 82 items of information (fields). The information includes some of the patient's personal details (age, sex and usual place of residence), information about the patient's admission to hospital (for example, an emergency case, or from a waiting list), and clinical data such as the diagnoses and details of any operations. The following diagram illustrates the type of data in HES fields:

epistat	procode	purcode	sex	marstat	dob	diag_1	oper_1	epidur
3	RCC00	9AF21	1	8	07-DEC-1930	H269	C751	0
3	RCJAT	QDP00	2	8	10-OCT-1964	R69X6	-	4
3	RLB00	QDE00	2	8	11-MAY-1970	R69X8	-	3
3	RLN00	QDN00	2	8	02-JAN-1969	O471	-	0
3	RR403	9EJ68	2	8	04-JUN-1998	Z380	-	5
3	RCC00	9AR34	2	8	15-FEB-1926	K449	G451	0
3	RCF00	9AP38	2	8	04-APR-1951	N811	Q089	10

Alphanumeric codes, for example, purchasers

Numeric codes

Dates, for example, date of birth

Clinical information, for example, diagnoses from ICD-10 and operations from OPCS4

Periods in days, for example, episode duration

Figure 1 Record Contents

HES is at present a subset of the Admitted Patient Care Contract Data Set (APC-CDS) as exchanged between admitted patient care providers and commissioners of that care. The CDS is defined in NHS Data Manual (see reference [2]).

The 100 or so items in the APC-CDS are required to support the procedure of health care commissioning.

For information on past data years, see the data dictionary (reference [1]). Variations in the way HES records have been collected, and changes in the structure of NHS organisations mean that comparisons between data years can be difficult.

3.2 Types of Record

In addition to the general record, augmented care data, maternity data and psychiatric data are stored in additional fields, termed *tails*.

- For augmented care since October 1997, the fields are only present for episodes containing an augmented care period. Augmented care comprises intensive care and high dependency care.
- For maternity data, the additional fields are only present for delivery and birth records (see Section 9.1).
- For psychiatric data, the additional fields are only present for certain psychiatric records (see Section 10).

3.3 HES Fields

Some of the HES fields are derived from the contents of other fields. An example is duration of episode (**epidur**) which is derived by subtracting the start date of the episode (**epistart**) from the end date (**epiend**). Some derived fields, for example **cause**, are generated by the HES system itself, and so do not exist within the hospital systems.

HES fields are broadly categorised as either numeric, alphanumeric or date.

HES allows for the use of certain special characters within fields that would otherwise only contain numeric values. These include & (ampersand) followed by spaces, meaning 'not known'. If a value is missing from a purely numeric or date field, a special code is inserted during processing.

The NHS Data Manual contains descriptions of CMDS fields, including those that are intended to be derived by NHS systems.

Note HES fields do not necessarily have exactly the same names as the CMDS fields.

The HES fields are fully documented in the HES data dictionary (reference [1]), which also contains summaries of changes from year to year and other information.

4. Episodes, Spells and Patients

Virtually every hospital in-patient is assigned to a consultant, who is responsible for the patient's treatment. A patient's period of care under a consultant is termed a *consultant episode*. In the majority of cases, a patient is treated by just one consultant during the stay in hospital, known as a *spell*. For these patients, there is only one episode, and therefore only one HES record, which contains details of the spell.

If primary responsibility for a patient is transferred from one consultant to another during a spell, which happens in about 5% of cases, a new HES record must be completed. The patient's spell in hospital, now comprises more than one episode, with the **epiend** of the first being the same as the **epistart** of the second. To aid analysis of spells, episodes are given an order number (**epiorder**); the first is 01, the second is 02, and so on. This means that the total number of episodes is equal to or greater than the number of spells.

4.1 Transfers

When a patient is transferred from one hospital to another, he is discharged from the first and admitted to the second. This shift in responsibility dictates that a new spell must be started following the transfer. In such cases, discharge date (**disdate**) on the last episode of the spell in the first hospital is usually the same as admission date (**admidate**) on the first episode of the new spell in the second hospital.

Note The term *hospital* means hospital provider; this may be an NHS Trust that comprises a number of separate hospital sites and buildings. Nationally agreed definitions for the terms *Hospital Provider* and *Health Care Provider* are held within the NHS Data Dictionary.

If a patient is moved from one site to another within the same hospital provider, the spell continues. However, if there is a change in the patient's consultant, a new episode begins within the same spell.

4.2 Calculating Totals

Aggregated totals of episodes and spells need careful interpretation. For example, if a patient leaves hospital after a two episode spell in July, and is then re-admitted a few days later, a new hospital spell commences. The first episode of this spell is numbered 01, not 03, even though it is the patient's third episode in the same HES year. Counting episodes and spells is not the same as counting patients. A person who is seriously ill may have a number of separate stays in hospital in a single year.

Counting episodes or spells does not produce the number of hospital in-patients during a HES year, because individual patients get counted more than once. From the 1997/1998 HES year, HES can link spells for the same patient, and a patient count is possible. For earlier years, an analysis requiring a patient count must attempt to link spells for the same patient, for example, by considering date of birth, sex and postcode.

5. Brief HEStory of Time

Time is a very important measure of efficiency and effectiveness within the health service. The duration of episodes and spells is often required in analyses.

5.1 Duration of Episodes and Spells

The duration of an episode is calculated by subtracting **epistart** from **epiend**. The result - a figure representing a whole number of days - is placed in the **epidur** field.

To calculate the duration of a spell, HES subtracts the **admidate** from the **disdate** to give **speldur**.

For the spells that consist of one episode, **speldur** is the same as **epidur**. Where there is more than one episode in a spell, **admidate** is copied forward from the first, to all subsequent episodes. The final episode therefore contains both **admidate** and **disdate**, allowing the spell duration to be calculated.

Because **epidur** is simply the difference between two dates, the figures obtained often require careful interpretation. For example, if someone is admitted to hospital in the early hours of the morning and is discharged later the same day, **epistart** and **epiend** contain the same date. The duration of that episode is therefore zero days, even though the patient may have occupied a bed for several hours. An **epidur** of zero days also applies to day cases or where someone is admitted to hospital for an operation, but the operation is cancelled, and the patient is sent home later the same day.

5.2 Year Boundaries

The file for each HES year contains records for every episode where **epistart** or **epiend** fall within that period. In most cases, patients are both admitted and discharged from hospital during the same year; the median duration of a spell in the acute sector is currently two days. This overall figure conceals the fact that for certain consultant specialties, for example geriatric medicine, the median is considerably higher.

Around 98% of episodes end during the HES year in which they started, and are recorded as finished consultant episodes (FCEs). Where someone is not discharged before the end of the year, the hospital submits a record of the unfinished episode, which does not contain any clinical data, such as diagnosis. The episode status (**epistat**) field indicates whether an episode was finished at the end of the HES year, except for patients in the psychiatric census (**rectype** = 41).

5.3 Hospital Stays That Straddle HES Years

Some people enter hospital in one HES year, but do not leave until the next. These patients generate an unfinished record at the end of the HES year in which they were admitted. These unfinished records are partial; they do not contain any diagnosis or operative procedure, and should not therefore be used in clinical analysis. However, the consultant specialty (**mainspef**) and treatment specialty (**tretspef**) are included.

When the patient is discharged, a separate, finished HES record is constructed for the second HES year. This has the same **epistart** as the first record, showing that the patient was admitted in the previous HES year.

In many HES analyses, the unfinished records are ignored. This avoids the possibility of double counting patients when data for two or more consecutive years is being formulated. However, there are four main cases where unfinished records are included:

- To find out how many people were admitted to hospital in a particular year
- To calculate waiting time statistics for those admitted during the year
- To calculate the total number of days that beds were occupied by patients in a given year (bed days). This figure (**bedyear**), which is a useful measure of hospital activity for administrators and policy divisions, must include the duration of unfinished records.
- To give a snapshot of the number of patients under psychiatric specialties at the end of the year (for the Disabled Persons Act).

5.4 Example

The following diagram illustrates ten spells of different composition.

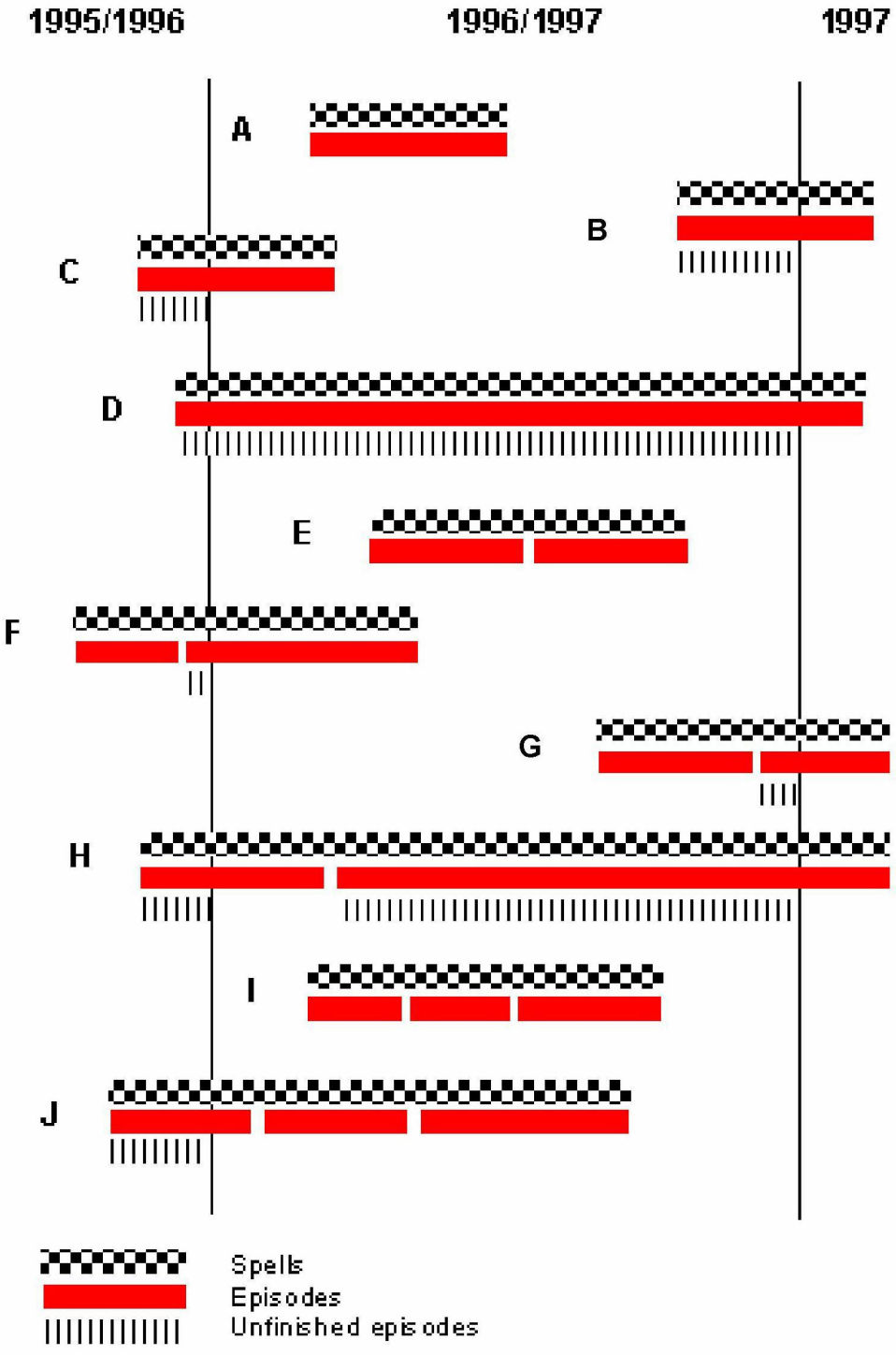

Figure 2 Episodes and Spells

Some comprise a single episode, others multiple episodes. The time frame spans three HES years. The implications for 1996/1997 are:

- Spell A is for a patient who was both admitted and discharged during the year. Furthermore, the patient remained under the care of the same consultant for the whole of the stay. There is only one HES record for A, the whole spell being a single FCE.
- E and I are similar, in that these spells both began and ended within the year. However, as primary responsibility for the patient in E was

transferred to another consultant during the spell, there are two records, one for each separate FCE. For I, there are three.

- Spell B straddles 1996/1997 and 1997/1998. As the patient is still in hospital on 31/3/97, an unfinished record is constructed for 1996/1997, which contains no clinical information. As spell B ends during 1997/1998, a full record for the whole of this FCE, including events such as the admission, which actually took place during the previous year, is submitted for the 1997/1998 HES file.
- Spell C is a single episode spell which began in 1995/1996 and leads to a full record of the FCE being submitted in 1996/1997.
- Spell D, a stay of around 13 months, is for a patient who spent the whole of 1996/1997 in hospital, but only an unfinished record was submitted for that year. A full record, giving all clinical and administrative information for the spell is in the 1997/1998 file. Note that there is also an unfinished record in the 1995/1996 file. This is an example of when you would want to include unfinished episodes, as this has a significant impact on **bedyear**.
- Spell H is more complicated. Superficially, it is similar, but whereas the patient in D remained under the care of the same consultant, H is a two episode spell. As the first episode of H finished in 1996/1997, there is a full record for this in the 1996/1997 file. For the second episode, there is an unfinished record in 1996/1997 and a full record in 1997/1998. Adding the unfinished record for 1995/1996 brings the total for this one spell to 4.
- Spells F, G and J illustrate further variations for spells of two or three episodes.

Note Spell A represents 94% of hospital stays. The rest represent only 6%.

6. Method of Admission and Waiting Time

The method of admission field (**admimeth**) contains one of 15 possible codes. These fall into four groups:

- Emergency
- Elective
- Others
- Maternity

6.1 Emergency Admissions

When somebody must be admitted to hospital immediately (other than where a woman is admitted only because she is about to give birth) it is classified as an emergency. Although this term conjures visions of accidents, it is also possible for a patient to be admitted immediately from an outpatient or other clinic. The consultant in charge of the clinic may decide that the case is so serious that there is no alternative to immediate in-patient treatment.

6.2 Elective Admissions

An admission is elective if the patient has been waiting for treatment. Most patients admitted in this way have been waiting for a hospital resource, such as a bed or a particular consultant, to become available. They would therefore have been included on a hospital waiting list.

A smaller group, who are not included on the waiting list, are those who are waiting for clinical reasons; these admissions are known as *planned*. For example, a patient, whose broken leg was mended by means of a steel pin, would return to hospital a few weeks later to have the pin removed. In this case it is necessary to allow time for the bone to heal, but the patient does not have to occupy a hospital bed while this is happening. The patient is simply waiting for nature to take its course, and the consultant would have stated in advance the best time for the patient to return, so this is not a waiting list case.

Another example of a planned admission is where radiotherapy treatment is arranged for a cancer sufferer. The radiation must be delivered in a sequence of measured doses, with necessary gaps between sessions for recovery and stabilisation. Note that HES does not receive records for the admitted patient care of regular day or night attenders, which means that many planned admissions that entail a regular course of treatment are excluded.

6.3 Waiting Time

For all elective admissions, whether the patient has been waiting for a resource, or for clinical reasons, the date it was decided to admit the patient is recorded. This field is known as **elecdate**, and the difference in days between that date and the date of admission becomes the waiting time (**elecdur**). The waiting time of planned admissions is rarely analysed. It varies from case to case depending on a number of factors, and has little meaning outside complex epidemiological studies. In contrast, the waiting time of those *waiting for a resource* is important, and is the subject of numerous enquiries.

As with other admission data, **elecdate** and **elecdur** are recorded on the first episode, and then copied to any subsequent episodes in the same spell. Therefore, waiting information is available in any episode. However, when analysing waiting time, it is vital to look only at the data for the first episode of a spell (**epiorder** = 01). To use all HES records would result in data being counted more than once with respect to certain admissions, as in spells E and I in Figure 2.

An analysis of waiting time for a given HES year must include only those admissions that occurred during the year (**admidate** is during the year). To include all episodes where **epiorder** = 01, irrespective of when the patient was admitted, would result in spells C, F, H and J being erroneously included. To detect and analyse small changes in waiting times from year to year, it is important to select only those admissions that actually took place within each year.

6.4 Typical Waiting Time Enquiry

It is common for an enquirer to ask what is the average waiting time for a particular operation. This highlights three important concepts, all of which need to be considered.

- Strictly speaking, a patient on a waiting list is not waiting for a specified treatment, but merely waiting to see a doctor. Once the patient has been admitted to hospital a doctor decides on an appropriate form of treatment, or operation.

  This is a distinction that can sometimes be significant. For example, a patient with heart disease may be placed on a waiting list for surgery. However, the decision on which of the many types of heart repair is appropriate may not be made until after the patient has been admitted and undergone further tests.

  Therefore, if somebody asks what the waiting time is for a particular type of heart valve repair, the answer should be qualified by explaining that the result is very much a product of hindsight. This also applies when analysing waiting time by diagnosis. After all, it is possible for people to be admitted to hospital to find out what is wrong.

- Because unfinished records do not contain any clinical data, it is not easy to provide analyses of waiting time by diagnosis or operation with respect to patients admitted during the year.

  For example, spell B in Figure 2 shows a patient for whom there is only an unfinished record within 1996/1997. However, the consultant specialty (**mainspef**) is entered on unfinished records, so waiting time by specialty should pose no such problems.

- The average, or mean, waiting time does not necessarily give a true picture of the service provided for the majority of patients.

  This is because there is often a very small number of patients with very long waiting times. The answer is to quote the median waiting time instead or, at the very least, to qualify the average. Note that if a record is submitted with an **elecdur** of zero days, the HES system declares this invalid. The overwriting of the zero in the **elecdur** field helps ensure that average and mean waiting times calculated from **epidur** are not unduly distorted. However, if at a later date, it is necessary to identify any records originally submitted with an **elecdur** of zero days, this can be

done by comparing **elecdate** with **admidate** (to find the cases where both dates are the same).

7. Patient Classification

In HES There are three classes of patient:

- Ordinary admissions
- Day cases
- Mothers and babies using delivery facilities only

Each HES record has a patient classification field (**classpat**). The third group are extremely rare. Currently, they are recoded by the NWCS to ordinary admissions. For earlier years a few of these records exist in HES; they must be included to capture all maternity events. Almost invariably, these patients are treated as ordinary admissions, and are simply added in to the first group.

Day cases are elective in-patients who have been admitted for treatment just for the day. They are always single episode spells with a duration of zero days, whether measured by **epidur** or **speldur**. For an in-patient to be classified as a day case, it is essential that there is a clear intention for the treatment to be concluded in one day. If, unexpectedly, the patient is kept in overnight, it must be reclassified as an ordinary admission. The majority of in-patients are ordinary admissions.

There are two further patient classifications used within the NHS – *Regular day attenders* and *Regular night attenders*. However, records with these classifications are not currently included in HES.

8. Clinical Coding

Clinical data is stored using sets of alphanumeric codes. Diagnosis is coded using either ICD-9 (numeric) or, from April 1995, ICD-10 (references [7] and [8]). Operations are coded using OPCS4 (reference [9]).

8.1 Diagnosis

All HES records for finished episodes contain details of the diagnosis. Each episode has a primary diagnosis (**diag_1**) - *the main condition treated or investigated during the episode of healthcare*. There are six further diagnosis fields, the first to record any subsidiary diagnosis, leaving five for secondary diagnoses. The subsidiary and secondary fields are used to record other diseases, conditions or complications. HES only keeps the first seven diagnoses from those available within the APC-CMDS.

This means that HES is able to cope with patients who have more than one thing wrong with them, or provide further detail on the primary diagnosis. For example, where the primary diagnosis is a broken femur, osteoporosis may be recorded as a complication. Similarly, a birth event can be complicated by the baby being in an abnormal position. Here, the mother's primary diagnosis would show the method of delivery and the complication would be entered as a subsidiary diagnosis.

For full details of how **diag_#** can be coded, see the data dictionary.

8.2 Operative Procedures

HES records for finished episodes for patients who undergo operative procedures contain details of the procedures. Each episode has four operation fields (**oper_1**, **oper_2**, **oper_3** and **oper_4**). HES only keeps the first four operative procedures from those available within the APC-CMDS. Each operation field has an associated field, which records the date of the operation (**oper_dte_#**).

The main operation need not be the first, for example, major surgery is often preceded by a biopsy. The main operation should be the one that is the most resource intensive, but sometimes data is coded chronologically.

When analysing HES data for operative procedures, all four fields may need to be studied.

9. Maternity Data

Maternity data is collected for HES and stored in delivery and birth records where a birth takes place, and a non-delivery maternity record where no birth occurs.

9.1 Delivery and Birth Records

The HES record for the mother is known as a delivery record. It contains the same data as a general record but has 19 additional fields for information about the delivery. The extra fields are known colloquially as the *baby tail*. There is a separate baby tail for each birth.

At the same time, a separate birth record, which has the same format as a general record, is completed for each baby. This also has a baby tail containing exactly the same information recorded in the corresponding tail of the delivery record. Delivery and birth episodes can be either a *consultant episode* or a *midwife episode* depending on who is responsible for the patient. For HES purposes, both are treated in the same way, but can be distinguished by looking at the main specialty code (**mainspef**).

As a general rule, excepting the small group of mothers and babies who use delivery facilities only, all in-patients on admission to hospital are classified as *general* (**rectype** = 11 or 13). This rule even applies to a pregnant woman who is admitted to a maternity ward with the clear intention of giving birth. However, immediately after giving birth the mother becomes a delivery case, so the **rectype** field is suitably amended before the record is submitted as part of HES.

If a pregnant woman does *not* give birth before the episode ends, the record remains coded as general. These rules also apply when a mother gives birth during the course of a multi-episode spell, but only one of the episodes is a delivery record. This ensures that only one baby tail, or for multiple births, one set of tails, is appended to the delivery record. The following diagram shows how HES deals with multiple births.

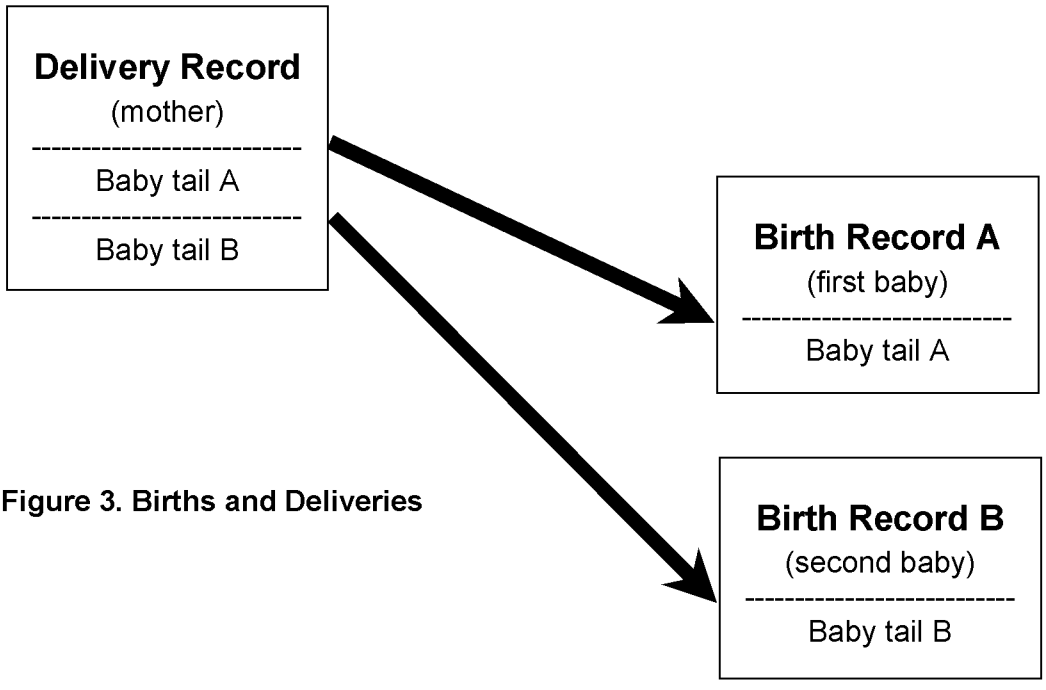

Figure 3. Births and Deliveries

Delivery and birth records have their own special methods of admission. For deliveries these are before birth (**admimeth** = 31) and following birth (**admimeth** = 32). 31 should also be used where a pregnant woman has been admitted to a maternity ward for observation, or a diagnostic procedure like an ultrasound scan.

If a baby is born en-route to hospital, **admimeth** = 32. Typically, these cases are mothers who have given birth in the ambulance or at home while waiting for it to arrive.

Nearly all delivery records are ordinary admissions (**classpat** = 1). A delivery record cannot be a day case, because a day case must have an elective method of admission. This is so, even if the mother does not stay overnight and has a recorded duration of stay of zero days.

9.2 Well Babies

Most babies born in hospital are alive and healthy, but it is not uncommon for a *neonate* (a baby aged up to, and including, 28 days) to suffer from a minor ailment like jaundice. Many HES analyses therefore require the exclusion of these *well babies*.

To assist in this, a field that indicates the level of nursing care that the baby required, is available. The field **neocare** contains neonate nursing level codes, which are recorded in the subsidiary or first available secondary diagnosis field. A well baby born in hospital is defined as one having a birth record (**epitype** = 3) which shows that a normal level of nursing care was required (**neocare** = 0), even though a minor ailment may, in some cases, be recorded in the primary diagnosis field. The HES enquiry software contains an algorithm which applies the rule automatically. For further details, including historical definitions, see the data dictionary.

A neonate who is kept in hospital, and starts a new episode, or is admitted within the first four weeks since birth is classified for HES purposes as a general in-patient. This includes foundlings (abandoned babies) and neonates admitted for social or economic reasons.

9.3 Home Births and Births in Non-NHS Hospitals

HES is intended to cover all births, so delivery records should be completed for mothers who give birth at home and in non-NHS hospitals, and birth records are completed for the babies they produce. These are often referred to as *other maternity events* and are distinguished by their record type (**rectype** = 53 and 63). These records are incomplete; they have no patient classification, date of admission, method of discharge, and so on, but they should have baby tails, completed in exactly the same way as for mothers and babies in NHS hospitals. It is normal HES practice to exclude these records from most analyses, other than those concerned with maternity issues.

10. Psychiatric Data

Psychiatric episodes are those where the main consultant specialty (**mainspef**) is mental handicap, mental illness, child and adolescent psychiatry, forensic psychiatry, psychotherapy or old age psychiatry. These are known as the psychiatric specialties. Also, general records with a psychiatric diagnosis are included in psychiatric data sets.

The field **admistat** provides information on whether a patient was previously admitted for psychiatric treatment.

A proportion of psychiatric patients remain in hospital for a considerable time, and this makes it more likely that the episode remains unfinished at the end of the HES year. Special provisions are made, as described in the next section, to avoid a lack of clinical information on these long stay psychiatric patients.

10.1 Psychiatric Census

For some *unfinished* psychiatric episodes, additional information, sometimes referred to as the psychiatric tail, is collected for patients under the care of consultants in the psychiatric specialties. This extra information is called the *Psychiatric Census*, and it enables us to obtain clinical information about long stay or formally detained psychiatric patients in psychiatric specialties which would otherwise be missing from HES.

The Psychiatric Census does not include all the unfinished psychiatric episodes within HES. It only collects information on patients in psychiatric specialties who have either:

- been formally detained under specific sections of the Mental Health Act 1983; previous mental health legislation; the Criminal Procedure (Insanity) Act 1964 or other legislation

or

- remained in hospital for a year or more continuously and are therefore considered to be long stay patients at 31 March, the last day of the HES data year; these patients must have spent the majority of their stay under the care of a consultant in the psychiatric specialties, but not necessarily all of it.

To be included in the Psychiatric Census, a patient does *not* have to be in hospital at the time of the census. As with certain other in-patients, psychiatric patients can be allowed to go home for short periods of time, as long as they agree to return to the hospital to continue treatment. This is termed *home leave*, and can be up to 28 days. The episode or spell is unfinished as the patient remains under the care of the consultant. Therefore, the patient must be counted. However, if a patient fails to return from a period of home leave, he or she is considered to have absconded. In this case the episode or spell is coded as finished.

In some cases, a patient who is formally detained may leave the hospital on *trial leave*, which can be extended indefinitely, with the hope that the patient will be well enough not to need to return to hospital. When this happens, the episode or spell is also coded as finished.

Patients who are on trial leave, or have absconded, should *not* be included in the census.

11. Data Quality

The problems of HES data are those that apply to any data process where rigorous validation is not applied to the data at source. Many of the source systems in the NHS permit individual records to be identified and updated. However these systems often allow the local use of extended or modified codes that are not nationally valid. This places an implicit responsibility on the NHS to map their locally valid values to nationally valid values in their HES submissions. There is a continuing need to cross-check the consistency of values between fields as well as validate individual values in fields.

The other dimension to data quality is that of a HES record being sent to the Department and subsequently being revised by the data provider. This can result in a discrepancy between the numbers of episodes in HES and those held on local systems.

11.1 Monitoring Quality

The quality of all data received from the NWCS is monitored by SD2HES, who liaise closely with NHS Trusts to resolve problems. This may lead to the Trust sending corrected records via the NWCS or, in exceptional cases, it may be possible for SD2HES (working in consultation with the data provider) to apply corrections to records if this is the only way of ensuring acceptable data quality.

Where the quality of data falls below an acceptable standard, the HES contractor may, after consultation with SD2HES, request a fresh submission.

To ensure that HES data is consistent and unambiguous, a number of checks are applied. Every record is interrogated: first to ascertain whether it can be accepted (verification), then to determine whether inappropriate entries should be overwritten (autocleaning) and, finally, an audit is carried out of the field contents so that a comprehensive set of quality reports can be generated (validation). The processes are described in the following sections.

11.2 Verification

To be accepted for HES, a record must contain an appropriate hospital provider code (**procode**), patient classification (**classpat**) and record type (**rectype**). Also, if the episode has finished, **epiend** must contain a date within the HES year to which the submission relates. A record which fails one or more of these checks is rejected. In some instances, error reports may enable the data provider to rectify errors and resubmit corrected data.

A verification report can be produced, which shows the number of records submitted by **rectype**, and how many of these were rejected because of a missing or inappropriate **procode**, **classpat**, **rectype** or **epiend**.

11.3 Autocleaning

Various fields within each record are analysed to see whether the entries make sense; not only in isolation, but also with reference to the contents of related fields. Where possible, some key entries that are clearly incorrect are overwritten by deriving the proper value from information contained in other parts of the record. If this cannot be done, the invalid entry is overwritten with the accepted code for ‘not known’ or ‘not available’. Some fields containing entered data, for

example, **startage** and **elecdur**, are automatically recalculated even if they are valid.

An autocleaning report can be produced, which indicates how many times the autocleaning rules were invoked and how many records were amended.

11.4 Validation

This is the process of testing the contents of individual records against a set of rules to identify errors that remain after autocleaning and that *cannot* be corrected. However, records containing such anomalies are not excluded from the HES database; they undoubtedly contain useful information. The main purpose of the validation process is to generate a report, which should prove useful in analysing data quality.

A report is generated, which summarises the quality of the submission following autocleaning, with respect to a number of individual fields, for finished and unfinished episodes. For each field, there is a total for the number of records containing an entry, plus the number of these deemed error free, and the percentage the latter represents. Finally, there is a target quality threshold, expressed as a percentage. The threshold, which is the same for all providers, is a working benchmark arrived at by assessing the importance of the field and how well it has been completed nationally in previous years.

11.5 Grossing

The extent to which HES fulfils its aim of capturing all finished consultant episodes (FCEs) is judged by comparing it with the Körner aggregate return, KP70. KP70 is completed independently of HES and provides annual totals of FCEs, thus giving a baseline against which HES can be assessed. SD2HES applies correcting factors to HES data that deviates from the baseline.

This procedure, called *grossing*, involves comparing the KP70 and HES figures by consultant specialty. A further stage in the process involves applying a second factor to compensate for missing or invalid diagnoses. Grossing is therefore applied selectively; it helps ensure consistency and comparability across the whole database.

11.6 Reporting to Data Providers

SD2HES reports frequently on the quality of data to the Chief Executives, Directors and senior managers of all NHS Trusts in England. These reports are also made available to HES users.

The quality tables contain hospital provider or Trust data within each Health Authority and Regional Office of treatment. The tables compare HES data against the Körner return, KP70.

The quality reports can be obtained from the contact address on page 2.

12. Accessing HES Data

The HES database is assembled from the submissions after verification, autocleaning, validation and any manual cleaning. The master copies of all HES records are kept securely by the HES contractor.

HES data can be accessed by:

- The on-line interrogation of the HES database. This is only available to HES users in the DoH, NHS Executive (including Regional Offices) and the HES contractor's offices. Anyone who is not a HES user can request that an enquiry be performed. The cost of enquiries depends on their complexity and other factors.
- Studying published tables, which may be viewed and downloaded from the HES website at http://www.doh.gov.uk/hes, or purchased as pre-printed volumes. These tables are comprehensive enough to answer most general queries. For example, they can be used to find out how many operations of a particular kind were carried out during the year of interest, and what was the waiting time and average length of stay in hospital for them.
- Studying extracts from the database. These can be ordered from SD2HES (see the beginning of this publication for contact details). However, all requests for extracts containing data items that might facilitate the identification of individual patients or consultants must be presented to the Security and Confidentiality Advisory Group (SCAG).

12.1 Custom Analyses

If data is not available from published tables, a custom tabulation or extract can be requested. Data tabulated at a low level, for example, by individual NHS Trust and for each specialty, may be sensitive. This could result in low numbers being suppressed, or it may be necessary to submit the request to the SCAG. Even where confidentiality is not a problem, the tabulation specification should address the following points:

- At what level (for example, England, RO, HA, provider) is the data needed?
- Is a comprehensive set of data needed at this level or is a subset adequate (for example, a certain age group)?

Enquirers should comprehend the potential size of the database and in the case of flat-file extracts, whether they have the tools for analysing the data.

Requests for custom tabulations or extracts should be made through SD2HES, as described in references [5] and [6]. The principles of HES analysis are described in Section 13.

13. Analysing HES Data

For the sake of consistency, common standards must be adopted when analysing HES data. This applies to both administrative and clinical data.

When producing figures from an analysis of HES data, the following questions are relevant:

- What is the information for and why is it required?
- What methodology should be used?

When a custom tabulation is produced, an explanation of how the result was obtained is included. In some cases the answer may not give precisely the information requested, but is the nearest approximation available given the way in which data is collected centrally.

For year on year comparisons, discontinuities in HES must be considered. HES processing is supported by reference files, which are reviewed annually to update valid codes. These coding changes, along with those in organisation codes and the geographical location of organisations, are implemented in HES processing.

Other discontinuities are the result of changes to the HES record.

Changes made during the life of the HES system are summarised in the HES data dictionary (http://tap.ccta.gov.uk/doh/hes_dd.nsf).

To overcome residual data quality problems following cleaning, enquirers can account for any invalid values within a field by requesting that all incorrect or out of range values be tabulated as a separate, 'others' row or column within the table. At the least, this provides an indication of the magnitude of any wrong coding.

13.1 Analysis Standards

For enquiries needing clinical information, only those episodes that *finished during the year* are counted. Counting FCEs means that a few patients are included who were admitted the previous year, or even earlier, but this is counterbalanced by the fact that those who remain in hospital until the following year are not counted. In addition, the standard approach is to count only the primary diagnosis.

The spells in Figure 2 that would contribute to the finished episode count are A, C, E (both episodes), F (second episode only), G (first episode only), H (first episode only), I (all three episodes) and J (all three episodes).

A consequence of the standard approach is that because the same primary diagnosis may be recorded on two or more episodes that form part of the same multi-episode spell, the count of primary diagnoses frequently exceeds the number of individual patients who have been given the primary diagnosis of interest. HES was designed to measure activity, rather than count patients.

13.2 Counting Diagnoses

Counting the instances of a particular diagnosis only where it is entered in the primary field ignores those cases where it appears in subsidiary or secondary fields. However, if the question is "How many episodes are there for *in-patient treatment of* XYZ?", this is a reasonable approach. XYZ could be entered in either the subsidiary or secondary fields of some records, but these fields do not indicate the main reason why the patient is in hospital, so they are ignored.

The subsidiary and secondary fields are used as follows:

- To elaborate on the primary diagnosis, thus providing a more complete picture of the patient's condition.
- To record other conditions or injuries that are not directly related to the primary diagnosis, but which nevertheless demand specific hospital in-patient treatment in addition to that necessary because of the primary diagnosis.
- To identify the cause in cases of accidental or deliberate injury.
- To record conditions that are nothing to do with the primary diagnosis, and do not require in-patient treatment. For example, a diabetic patient should have diabetes as a secondary diagnosis. In many cases the condition described by the secondary diagnosis influences the treatment given.

The first group shows that it is important to choose the appropriate codes, or ranges of codes when carrying out a diagnostic analysis. There is obviously little point in searching the primary field for a code that would normally be used only to *qualify* a primary diagnosis.

The second group illustrates that by following the convention of looking at the primary diagnosis only a slight under count results. For example, for a patient with a broken leg who has received multiple injuries in a car crash, the primary diagnosis might well relate to more serious head injuries.

The third group covers specific causes of a condition, which never appear in the primary diagnosis.

The fourth group explains why it is normally appropriate to look only at the primary diagnosis. If the question is, "How many records are there for patients who *have received in-patient treatment for* diabetes?" it would be important to exclude the last example from the count.

Diagnostic analysis is a complex field; the examples given above merely provide pointers. Expert advice is always available from the appropriate sections at the DoH.

13.3 Counting Operative Procedures

Operations can be counted by looking either at the primary operation field in isolation or by looking at all the operation fields. In some cases more complex approaches are demanded, for example, to count how many times a particular primary operation was accompanied by certain secondary procedures.

To count instances of comparatively simple exploratory procedures like biopsies or endoscopic examinations, it is important to look at both the primary and all the secondary operation fields.

14. Further Information

Further information can be obtained from:

NHS Data Manual (reference [2])

An overview of the guidance on the definition, collection and interpretation of the nationally agreed data sets, both HES and aggregate returns, in the NHS. It has been written for everyone who is actively involved in the collection of data and the management of information in the NHS. It is distributed as a set of Acrobat PDF files (from which it can be printed) and is published at The Data Standards Programme, NHS Information Authority, 15 Frederick Road, Birmingham B15 1JD. Telephone: 0121 625 1992 Fax: 0121 625 1999.

The NHS Data Manual can also be downloaded from: http://nww.standards.nhsia.nhs.uk/ds/index.htm

NHS Data Dictionary

This contains the nationally agreed definitions of data items and is the reference point for NHS Data Standards. It is available in Acrobat format from The Data Standards Programme, NHS Information Authority, 15 Frederick Road, Birmingham B15 1JD. Telephone: 0121 625 1992 Fax: 0121 625 1999.

The NHS Data Dictionary can also be downloaded from: http://nww.standards.nhsia.nhs.uk/ds/index.htm

The Admitted Patient Care Contract Minimum Data Set (CMDS) Edifact Message

This is defined in The NHS Commissioning Data Set (CDS) Manual, which describes the whole data set and the Edifact messages which should be exchanges between providers and purchasers. It is available from The Data Standards Programme, NHS Information Authority, 15 Frederick Road, Birmingham B15 1JD. Telephone: 0121 625 1992 Fax: 0121 625 1999.

The NHS CDS Manual can also be downloaded from: http://nww.standards.nhsia.nhs.uk/ds/index.htm

The WHO International Classification of Disease, Ninth and Tenth Revisions (references [7] and [8]). A two-volume set comprising: Volume 1, a series of codes for the classification of morbidity; Volume 2, an index. The Tenth Revision was implemented in the NHS in England from April 1995.

The OPCS Classification of Surgical Operations and Procedures, Fourth Revision (reference [9]). An HMSO publication (1990: ISBN 0 11 691295 2) containing a listing of the consolidated version of OPCS4.

Central Postcode Directory Use Guide and NHS User Postcode Directory Supplement. A description of the ONS postcode directory. For further information, contact ONS; either via the Internet (http://www.ons.gov.uk) or telephone 01329 813477 or 01329 813243.

Training materials and advice on clinical coding can be obtained from
The NHS Centre for Coding and Classification
Woodgate, Loughborough LE11 2TG. Tel: 01509 211411