ANNEX 4

HOSPITAL EPISODES STATISTICS (HES) – CONSTRUCTION OF CONTINUOUS INPATIENT (CIP) SPELLS AND ASSESSMENT OF DATA QUALITY

Contents

Overview
Section A: Construction of CIP spells
   Introduction
      What makes this spells data-set different from standard HES?
   HESID construction summary
      Overview
         Patient HESID index
      Method
   Duplicate removal
   Derived fields
   CIP spells linkage method
      Example of HES data used for CIP spells linkage
          ONS deaths linkage
   Example CIP spells
      Example 1: Cl1 – Deaths within 30 days of admission with a specified primary diagnosis
      Example 2: Cl1C – deaths within 30 days of a Coronary Artery Bypass Grafting (CABG) procedure
      Example 3: Cl4 – Emergency readmissions indicators
      Example 4: Cl5 – Discharge back to usual place of residence following emergency admission to hospital with a stroke
      Example 5: Cl7(x) – Emergency hospital admissions (primary diagnosis variants)
      Example 6: Cl79(x) – Hospital procedures (procedure or procedure/diagnosis variants)

Section B: Data quality: construction of caution marks
   Data quality requirements
      Selection of episodes for the Data Quality caution marks components
         Table 1: Selection of episodes by data completeness, coverage and death-coding
   Data quality component: data completeness
      Selection of episodes and fields evaluated
         Table 2: Summary of fields and episode selection used in data completeness for each clinical indicator
      Methodology for assessing missing/invalid data
         Table 3: Methodology for assessing missing/invalid data
   Data quality validity marks
      Table 4: Criteria for assigning data quality validity marks to organisations: completeness
      Table 5: Criteria for assigning data quality validity marks to organisations: coverage
   Data quality component: coverage
      Table 6: Criteria for assigning data quality validity marks to organisations: Death coding

Contributors
Contact details
Updated

Overview

This document describes the construction of continuous inpatient (CIP) spells that are used as the basis for all the Hospital Episode Statistics (HES) based indicators in the Compendium. There are two sections. ‘Section A’ describes the methods used in CIP spell construction, whilst ‘Section B’ describes the algorithms for creating ‘caution marks’ (measures of the data quality for each organisation and each indicator).

Section A: Construction of CIP spells

Introduction
This is a technical document describing how CIP spells are constructed from data in the Hospital Episode Statistics system. HES data are stored in episode form (an episode being a single period of care under one consultant) and in financial year groups (1st April to 31st March). This document describes the methods by which these episodes are linked into CIP spells.

The HESID (unique patient identifier) field is derived (described in the HESID construction summary section) and CIP spells are then constructed using relevant episodes (see the section on CIP spell linkage for details). Prior to this linkage, duplicate episodes are removed (see the section on duplicate removal). Clinical indicators (CIs) that include deaths outside hospital require Office for National Statistics (ONS) deaths data to be linked to the specific CIP spells for these indicators (see section on ONS linkage for details). Other useful spell information is stored within each CIP spell to make it easier to produce the indicators (see indicator specifications or the section on derived fields for details of these fields).

What makes this spells data-set different from standard HES?

The linked data-set described in this document is used to create all Compendium indicators and includes many derived fields, both at episode level, and spell level.

Finished HES episodes (mostly fields relevant to Compendium indicator analyses) are extracted from all HES years beginning with the last quarter of 1997/98 (i.e. episodes finishing from 1st January 1998) through to the most recent quarter available. Over 150 million episodes are therefore processed.

A number of non-standard fields are derived, prior to the episodes being linked into spells crossing all years. Further derived fields are then created at spell level (see the section on derived fields for details). The final data-set is tailor-made to support all Compendium indicators. The patient identifier (HESID) is mapped to correspond to the current central HESID index ensuring that episodes from different HES years can be linked into a CIP spell.

Financial year ‘views’ of the data are created by analysing all episodes having start or end dates falling within the 13-month period commencing three months prior to the period start date (e.g. from 1st January YYYY, and finishing six months after each period end (e.g. to 30th September YYYY+1). From this pool of episodes it is possible to construct not only the CIP spells that fall entirely within the year, but also the vast majority of those that straddle either the year start, or end. For example a two episode CIP that starts with an admission in March and finishes with a discharge in May, could well have a first episode end date prior to the start of the financial year. However, as the second episode of the CIP finishes within the year, both episodes must be included in the year view, within a 2-episode completed spell. This technique therefore ensures that all CIP spells, apart from a very small number relating to long stay patients, are fully constructed. The CIP spell in this example would also appear, in full, in the earlier financial year – however it would not be double-counted due to appropriate filters being applied by specific indicators (see indicator specifications and section on ‘Where CIP spells are created’ for details).

Standard HES exists in individual financial year blocks, and only episodes can currently be queried.

The provision of CIP spells alongside standard HES episodes is now being implemented by the HES team. These ‘HES spells’ are created using an algorithm similar to a previous version of the one described in this document. However the HES spells will only exist in financial year blocks, without linking between years, and will include unfinished records. This does not give the same level of cross-year comparability as the linked spells file. Unlike the spells universe accompanying standard HES, the spells used for NHS Choices Indicators are created using exactly the same methods as the spells dataset described in this document.

Other differences between this linked data-set and standard HES spells include:

- Standard HES does not support reliable quarterly analysis due to the boundaries of each financial year. The linked file overcomes this problem.
- Standard HES cannot support standard readmissions analysis where the discharge occurs in March. If a standard 28 day readmission indicator is required, the denominator based on standard HES can only include spells with discharges up to 4 March, as data for admissions after 31 March are not available. The linked file overcomes this issue.
- Standard HES does not, as a rule, have patient identifiers (HESID) updated to reflect recent additions to the central HESID index (see the section on HESID construction for details).
- Standard HES does not link episodes between years, so “broken” CIP spells will be created at the start, and end of the financial year.
- Standard HES spells include unfinished episodes, which are known to be unreliable and have therefore been excluded from the linked file (achieved by including two quarters of finished episodes from the following year).
Standard HES will not usually have residence-based fields mapped to the latest boundaries for all years (see the section on derived fields for details).

Standard HES is not expected to include deaths data from later HES episodes (see the section on ONS linkage for details).

The spells described in this document are able to support a wider range of queries than standard HES can, and spells fields can be added when required. The resulting data set is also structured in such a way that it is possible to run analyses over differing periods, i.e. analyses can be done on calendar years using the same data-set.

HESID construction summary

Overview

This section describes the means by which episodes are compared to determine which relate to the same patient. In order to identify episodes that relate to a single patient, a HESID value is assigned to every episode. If the patient identifying information recorded on two episodes is thought to belong to the same patient, the same HESID is assigned to both. Therefore, the HESID is effectively a HES patient identifier.

Patient HESID index

A patient HESID index is used to record the HESIDs that have been assigned to combinations of the data items involved in the matching steps outlined below. There is only one patient HESID index, which is independent of the data years, to facilitate linking of episodes across data years. The information from an episode is added to this index only if there are sufficient valid data items to involve the episode in one of the matching steps outlined below. Otherwise, the episode can never be matched with any other episode, and so is assigned its own unique HESID value. Note that if an episode includes enough information to attempt a match, but no match is found, the episode details are still added to the patient HESID index, because another episode may match it at some later date. For each HESID in the index, there is a corresponding LAST_ACTIVITY_DATE that contains the date where the last activity for that HESID within the HES datasets was recorded. This field is later used during Mortality Processing.

The matching algorithm is applied only to entries in the patient HESID index described above. Therefore, before matching is attempted for a data year, all relevant combinations of patient identifying items from the episodes for that data year are added to the patient HESID index, if the combinations are not already recorded. This approach ensures that both the new episodes and all previous HESID entries are involved in the match, and that previous HESIDs are reconsidered in the light of the information from the new episodes. Once the matching process is complete, the resulting HESIDs in the patient HESID index are assigned to the episodes of the data year. In outline, the matching process involves three main steps.

- The first step attempts to match episodes using the following patient identifying information:
  SEX: Sex
  DOB: Date of birth
  NEWNHSNO: NHS number.

- The second step then attempts to match episodes using:
  SEX: Sex
  DOB: Date of birth
  HOMEADD: Postcode
  PROCODDET + LOPATID: Local patient identifier within provider.

- The third step then attempts to match episodes using:
  SEX: Sex
  DOB: Date of birth
  HOMEADD: Postcode

As a general rule, matching is attempted only if all of the values required by that step are non-null and valid. The third step will only match records together if Step 1 does not determine that the records should not be matched and if one or more NEWNHSNO (NHS Number) in the records to be matched is null.

Due to the requirement for HESID compatibility across multiple data-years, a new field “HESID3C” is created prior to linking episodes, and is then used as the main patient identifier. This simply makes sure that all years of data have patient identifiers updated to match the patient identifiers which are current in the central “HESID index”.

Duplicate removal

It has become evident that episodes are often duplicated (or appear multiple times) in the HES system. These episodes would compromise the quality of the CIP spells’ construction, and are therefore removed.

The method for removing duplicates is defined as:-
Order all episodes in the financial years (FY) episodes data by the fields listed below (in the order shown).
If one or more consecutive episodes are identical on the fields listed below then use the one with the most recent SUBDATE, discarding all others.

Fields used during identical matching:

- HESID3C (HESID mapped consistently across all years using the latest central patient HESID index)
- FYEAR
- EPISTAT DESCENDING (note that only finished episodes exist in the current dataset)
- EPISTART
- EPIORDER
- EPIEND
- ADMIDATE
- ADMIMETH
- ADMISORC
- CLASSPAT
- DISDATE
- DISDEST
- DISMETH
- EPITYPE
- HATREAT
- MAINSPEF
- PROCODET
- PROCODE3
- RESRO
- ROTREAT
- TRETSPEF

We have removed NEWNHSNO, RESHA and STARTAGE from previous specifications as they are derived from, or are fields which are used in HESID construction.

Derived fields

Some generally useful information is stored from key points in a CIP spell. Extra fields are also created for each indicator that requires non-standard fields. See the individual specifications for details of these fields.

General spell information:

- RESPCTC PCT of residence (postcode derived).
- RESPCTO PCT of residence – old boundaries (postcode derived).
- RESLADSTC Local authority district of residence (postcode derived).
- RESSTHAC Strategic Health Authority of residence (postcode derived).

Both of these fields are derived in all years by looking up the postcode using the postcode reference data for the most recent FY. If the postcode is not found, then they are set to NULL.
- **SPELLRESLADSTC**: A copy of the RESLADSTC field. For each CIP spell, this field is populated by RESLADSTC from the earliest episode in a CIP spell where a valid code was found, else set to NULL.

- **SPELLRESPCTC**: A copy of the RESPCTC field. For each CIP spell, this field is populated by RESPCTC from the earliest episode in a CIP spell where a valid code was found, else set to NULL.

- **PROCODETC**: Populated by PROCODET for 2003/04 or later, and PROCODE3 earlier. To ensure the same list of organisations exists in each year a number of rules exist: General rule: if a provider code ends '-X', 'TC' or 'T1' then only he first three characters are used to determine PROCODETC. Special case: if PROCODET = RXF09 then PROCODETC = RXF.

```plaintext
IF FYEAR >= 2003/04
  IF FYEAR = 2005/06
    IF the first three characters of PROCODET = RXK
      THEN PROCODETC = RXK
    ELSE the first three characters of PROCODET = RG3
      THEN PROCODETC = RG3
  END

  IF FYEAR = 2006/07
    IF the first three characters of PROCODET = RCZ
      THEN PROCODETC = RCZ
    ELSE the first three characters of PROCODET = RW9
      THEN PROCODETC = RW9
    ELSE the first three characters of PROCODET = NT7
      THEN PROCODETC = Matching SITECODE
  END

  IF the first three characters of PROCODET = 899
    THEN PROCODETC = 899
  ELSE the first three characters of PROCODET = RJD
    THEN PROCODETC = 'RJD'
  ELSE the first three characters of PROCODET = RN5
    THEN PROCODETC = 'RN5'

  OTHERWISE
    PROCODETC = PROCODET
END

IF FYEAR < 2003/04
  IF FYEAR < 2003/04 AND PROCODE3 in (8A1,8CV)
    THEN PROCODETC = PROCODE
  ELSE FYEAR = 2002/03 AND PROCODE3 in (8A4,8CR,8CW)
    THEN PROCODETC = PROCODE
  ELSE FYEAR = 2002/03 AND PROCODE3 in (8C4)
    THEN PROCODETC = 8C437

  OTHERWISE
    PROCODETC = PROCODE3
END

Note: This is not a mapping exercise, it is necessary largely because of the move between PROCODE3 and PROCODET.

- **SPELL_DISDATE**: Discharge date from last episode in a CIP spell. If this is NULL then the EPIEND from the last episode.

- **SPELADMIDATE**: Admission date from first episode in a CIP spell. If this is NULL then the EPISTART from the first episode.

- **SPELL_DISMETH**: Discharge method from last episode in a CIP spell.
• **READMIN**
  Number of days to the first following CIP spell that starts with an EPIORDER of ‘1’ (in first episode) for the same person, set to blank if a following spell for the same person is not found.

• **READMETH**
  Admission method of first episode in the first following CIP spell that starts with an EPIORDER of ‘1’ (in first episode), set to blank if a following spell for the same person is not found.

• **READMAIN**
  Main specialty of the first episode in the first following CIP spell that starts with an EPIORDER of ‘1’ (in first episode), set to blank if a following spell for the same person is not found.

• **READDIAGP**
  Primary diagnosis from the first episode in a readmission spell.

• **READPROCODETC**
  Provider code from the first episode in a readmission spell.

• **READTYPE**
  Episode type of the first episode in the first following CIP spell that starts with an EPIORDER of ‘1’ (in first episode), set to blank if a following spell for the same person is not found.

• **CFLAG**
  Equals 1 if a diagnosis of cancer or a record of chemotherapy is found anywhere within an FCE for that patient (HESID) with an admission date within 365 days prior to the admission date of the CIP spell. Cancer is defined as International Classification of Diseases version 10 codes C00-C97, D37-D48, Z511.

• **PBCV2**
  Programme Budgetary Code Version 2. Derived mainly from primary diagnosis. Episodes are assigned to one of 69 groups (programme budget categories, subcategories including 1 group indicating invalid diagnosis coding.

• **SDGROUP**
  Index of multiple deprivation score (IMD 2007) grouped into 7 bands (values 1 through 7). This is derived from the patient’s postcode and assigns a deprivation score for the geographical area in which the patient lives, to the patient. If no IMD lookup can be found, a value of 8 is assigned. For patients who were not resident in England, a value of 9 is assigned (grouping not applicable). If a group cannot be determined because the postcode was not recorded, a value of 10 is assigned.

• **SDGROUP5**
  Index of multiple deprivation score (IMD 2007) grouped into 5 bands (values 1 through 5). See SDGROUP, above, for other values.

• **HESID3C**
  See the section ‘HESID Construction Summary’ for details. HESID3C is the current HESID based on the patient HESID central index at the time of building the database.

**CIP spells linkage method**

A CIP spell is defined as a continuous period of inpatient care within NHS or Independent providers. It may include a transfer from one provider to another. Episodes with a matching HESID are assumed to relate to the same patient.

The following description refers to the example of HES data given below and illustrates how CIP spells are linked:

Episodes that are put forward for linkage (after duplicate removal) into CIP spells will be sorted by HESID, FYEAR, EPISTAT DESCENDING, EPISTART, EPIORDER, and EPIEND.

Episodes which do not start or end in the yearly data-set are excluded from consideration for inclusion in CIP spells, the data-set date range is from 1 January YYYY – 30 September YYYY+1.

To determine when the episode after a CIP spell has occurred (i.e. the end of the CIP spell) the following procedure is used. Starting with the second episode, progress down the list asking the following questions:
If the episode order is not "1" and the previous episode discharge method is not between "1" and "5" (inclusive) and the HESID is the same as the previous episode’s HESID then it is not the end of the CIP spell, continue to the next episode;

Else
If the previous episode’s discharge destination is between ‘51’ and ‘53’ (inclusive) OR the admission source of current episode is between ‘51’ and ‘53’ OR admission method of current episode is ‘81’ and HESID is the same as the previous episode’s HESID and episode order is ‘1’ and admission date is valid (i.e. a positive number of days since 31/3/1990) and the previous episode’s discharge date is valid (i.e. a positive number of days since 31/3/1990) and admission date minus the previous episode’s discharge date is <= 2 (includes negatives) then it is a transfer, continue to the next episode;

Else it is the end of the CIP spell.

 Spells which span financial year boundaries are included, in full, in each financial year they span. A spell is considered to be in a financial year if its admission date (SPELADMIDATE) and discharge date (SPELL_DISDATE) indicate that the spell should exist in the year.

Note: The algorithms are reviewed annually in the light of experience. For example, data files containing the nine-year data set (FY 1998/99 – 2006/07) were based on a previous version of the above algorithm and have been updated for FY1998/99 - FY2007/08 analysis.
Example of HES data used for CIP spells linkage

Data Sorted by these fields in this order:

<table>
<thead>
<tr>
<th>HESID</th>
<th>FYEAR</th>
<th>EPISTAT (DESC)</th>
<th>EPISTART</th>
<th>EPIORDER</th>
<th>EPIEND</th>
<th>ADMIMETH</th>
<th>ADMISORC</th>
<th>ADMIDATE</th>
<th>DISDATE</th>
<th>DISMETH</th>
<th>DISDEST</th>
<th>STARTAGE</th>
<th>RESPECTC*</th>
<th>RESLADSTC*</th>
</tr>
</thead>
<tbody>
<tr>
<td>345</td>
<td>2002</td>
<td>3</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>19</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2002</td>
<td>3</td>
<td>12</td>
<td>2(*)</td>
<td>23</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2003</td>
<td>3</td>
<td>22</td>
<td>3(*)</td>
<td>24</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
<tr>
<td>345</td>
<td>2004</td>
<td>1</td>
<td>10</td>
<td>1(*)</td>
<td>14</td>
<td>21</td>
<td>98</td>
<td>12</td>
<td>12</td>
<td>8</td>
<td>98</td>
<td>65</td>
<td>5E1</td>
<td>00AL</td>
</tr>
</tbody>
</table>

Other columns with fields relating to operations, diagnoses etc, as appropriate for the indicator:

HESID must be the same for these episodes to be in the same CIP spell. Otherwise it is the end of the spell.

* EPIORDER = 1 signifies first episode in a provider spell.

** Transfer occurs as:
(Previous Disdest = 51,52,53 or next admimeth = 81 or next admisorc = 51,52,53) AND the next episode has:
same HESID AND
EPIORDER = “1” AND
ADMIDATE (next episode) minus DISDATE (current episode) <= 2.

ONS deaths linkage

All deaths indicators include deaths that occur after discharge. As the HES Database does not contain information about deaths that occur after discharge, deaths data from the Office for National Statistics (ONS) is obtained and linked to the HES data.

Two tables are created, one containing information about deaths recorded within HES (DISMETH = 4) referred to as HESID_PSEUDO_MORTALITY and one containing the information received from ONS know as ONS_MORTALITY

A) HESID_PSEUDO_MORTALITY

HESID_PSEUDO_MORTALITY contains the following fields from HES:

HESID
DOD: recorded as the DISDATE (discharge date) of each episode in HES with a DISMETH (discharge method) = 4 or EPIEND (Episode End Date) of the episode where the DISDATE is invalid.

If duplicate HESIDs exist within this table, the record with the latest SUBDATE (submission date) is used and the other records for that HESID are deleted. If records have the same HESID and subdate,
the record with the latest DISDATE is used and the other records are deleted. If duplicates still occur, the record with the highest EPIKEY is used.

B) ONS_MORTALITY

The death data received from ONS is linked to the HESID index and stored to the ONS_MORTALITY table using the following method:

The matching process involves two main steps, which are the same as Steps 1 and 3 used for the matching of entries within the Patient HESID Index and rely on information recorded during the Patient HESID matching process. All of the notes that apply to the matching are as described for HESID above.

The first step (which corresponds to the first step of the Patient HESID matching process) is “driven by” NHS Number, and attempts to perform a match using the following further patient identifying information used to “check” the matches:

- **SEX**: Sex (Exact match)
- **DOB**: Date of Birth (Partial match)
- **NEWNHSNO**: NHS Number (Exact match)
- **HOMEADD**: Postcode

The second step is driven by Date of Birth, and uses the following for checks:

- **SEX**: Sex (Exact match)
- **DOB**: Date of Birth (Exact match)
- **NEWNHSNO**: NHS Number
- **HOMEADD**: Postcode (Partial match)

However, as a general rule, matching is attempted only if all of the values required by that step are non-null and valid.

A match will only occur when the LAST_ACTIVITY_DATE recorded against the matched HESID is not later than the 3 days after the date of death.

Unlike the HESID matching (which matches as many records as possible), the intention is to link at most one death to each HESID. Therefore, it becomes possible to rank matches, and use only the best one for each HESID. The ranking of matches (within the context of the above two steps) is as follows, where each match is “exact” unless specified otherwise:

1. NHS Number, Sex, Date of Birth, and Postcode;
2. NHS Number, Sex, and Date of Birth;
3. NHS Number, Sex, Date of Birth (partial match), and Postcode;
4. NHS Number, Sex, and Date of Birth (partial match);
5. NHS Number and Postcode;
6. Date of Birth, Sex, and Postcode, where the NHS Number does not contradict the match (i.e. where either the HESID or the ONS Mortality record (or both) has no valid NHS Number), and the Date of Birth is not 1st January (01/01/yyyy);
7. Date of Birth, Sex, and Postcode, where the Date of Birth is not 1st January (01/01/yyyy).

Whenever an attempt is made to assign more than one date of death to a single HESID, the best match (the one with the lowest rank) above is used. In the event of two matches with equal rank, the details associated with the later date of death are used. To support this, a Mortality Match Rank is recorded as part of the HESID Mortality data.

Depending on the steps used during the HESID matching, an ONS Mortality record may match more than one HESID. In the absence of any definitive way of deciding which one is correct, the inconsistency is simply ignored, and the single death is applied to every matching HESID. Match Rank 5 is present so that the algorithm is not totally dependent on Date of Birth and Sex. However, note that it still does not allow matching of mortality date to any activity record that does not have a valid Sex and/or Date of Birth, because such a record is not represented by Patient Match Keys in the Patient HESID Index.
C) Combine HESID_PSEUDO_MORTALITY and ONS_MORTALITY to create the MORTALITY_MASTER table

These two tables containing Date of Death data linked to HESID are combined to create a master list of deaths information. This is done by appending the HESID_PSEUDO_MORTALITY information to the ONS_MORTALITY table.

The next stage is to remove all duplicate HES IDs from the appended file.

Where pseudo and ONS mortality records share the same HESID and do not contradict each other, the pseudo-record is removed. Where there is a contradiction between the two records, the pseudo-death record takes preference except when the date of death in the ONS record is between 0 and 3 days prior to the date of death in the pseudo-death record. In this case the ONS record will be used and the other removed. This is because patients may be officially discharged on a later date from death (i.e. after a weekend).

D) Apply this information to the CIP spells for mortality based indicators:

For each FY, HES/ONS deaths will be linked to the CIP spells data using the following method:

Let CiNX represent any mortality based indicator reference, i.e. 3a, 1c etc.

For CiNX, loop through all the CIP spells that would be counted towards the denominator for the indicator being analysed.

All CIP spells not selected by the above will have the ONS linked fields set to empty for CiNX.

Loop through the selected CIP spells
If the spell is not the last spell for a patient in the selection, set the ONS linked fields as follows:
- CiNX_OSPELL_DISMETH = SPELL_DISMETH
- CiNX_OSPELL_DISDATE = SPELL_DISDATE
- CiNX_ONSFLAG = 0 (NOT UPDATED)

If the CIP spell is the last for a person within the CiNX selection then:
  If the spell ends in a death (SPELL_DISMETH=4) then
    CiNX_OSPELL_DISMETH = SPELL_DISMETH
    CiNX_OSPELL_DISDATE = SPELL_DISDATE
    CiNX_ONSFLAG = 0 (NOT UPDATED)
  Else (no match is found, there is no identified death)
    CiNX_OSPELL_DISMETH = 4
    CiNX_OSPELL_DISDATE = DOD from the mortality file
    CiNX_ONSFLAG = 1 (ONS UPDATED)

END

Note: The algorithms are reviewed annually in the light of experience. For example, data files containing the nine-year data set (FY 1998/99 – 2006/07) were based on a previous version of the above algorithm and have been updated for FY1998/99 - FY 2007/08 analysis.
**Where CIP spells are created**

In the examples below, "|-----|

| | | | | represents a single episode.

Episodes are linked across years, each spell which spans a FY boundary will appear, in full, in each year which it spans. Filters on each indicator are used to ensure that the CI reports do not count a CIP spell in one FY and the same CIP spell in another. This is achieved through the use of the following fields:

- **DISDATE of the last episode in the spell** - Readmission CI’s have a filter on the DISDATE field, ensuring all discharge CIP spells end in the year of analysis.

- **EPISTART of the first episode in the spell** - Admissions indicators require that that spells start within the year.

- **OPDATE_NN (Procedure date)** – Procedure based indicators require that the procedure occurs within the year of analysis.

<table>
<thead>
<tr>
<th>Eg 1</th>
<th>HES Year 1</th>
<th>HES Year 2</th>
<th>HES Year 3</th>
<th>CIP spell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>included in</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eg 8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Eg 1

|------|

CIP spell included in Year 1.

Eg 2

|------|

CIP spell included in Year 1 and Year 2 (constructed from the finished HES record from year 2).

Eg 3

|------|

Three episode CIP spell included in Year 1 and Year 2.

Eg 4

|------|

Three episode CIP spell included in Year 1 and Year 2.

Eg 5

|------|

Three episode CIP spell included in Year 1 and Year 2.

Eg 6

|------|

CIP spell included in Year 2.

Eg 7

|------------------|

CIP spell included in Years 1, 2 & 3 (constructed from the finished HES record from year 3).

Eg 8

|------------------|

Eleven episode CIP spell included in Year 1, 2 & 3.
Example CIP spells

Note: Spells are counted to the first valid organisation (LA, PCO) found in the spell (see definition of SPELLRESLADSTC and SPELLRESPCTC).

For trust based counts, the rules are different (i.e. for readmissions indicators we count the spells to the trust of discharge). See individual indicator specifications for details.

The following examples show sample spells that may be included in the indicators. See each indicator specification for a complete list of exclusions.

**These examples use the following:**
- **s** = valid primary diagnosis as defined in the indicator specification
- **z** = primary diagnosis other than those defined in the indicator specification
- **y** = discharge method of 1-3 (live discharge)
- **d** = discharge method of 4 (death)
- **e** = episode in a CIP spell that starts with an emergency admission
- **o** = first valid organization found in the spell
Example 1: CI1 - Deaths within 30 days of admission with a specified primary diagnosis

A set of three episodes with a primary diagnosis as defined in the indicator specification in the first episode of the CIP spell, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place and the patient was discharged alive.

Transfer

\[
\begin{array}{c|c|c|c}
\text{e01} & \text{e02} & \text{e03} \\
\hline
s & o & z \\
\end{array}
\]

Trust 1 | Trust 2

This CIP spell appears in the denominator for the indicator, counted to the organisation 'o', which was looked up via the postcode coded in episode 1, trust 1. If the postcode coded in episode 1, trust 1 did not result in a valid organisation (for instance if it was invalid), then the first valid organisation may have been found in a later episode in the spell (see following examples).

If the diagnosis from trust 1, episode 1 ('s') and trust 1, episode 2 ('z') occurred in the other order, then this spell would not be included, as the valid diagnosis ('s') would not have occurred in the first episode in the spell.

If episode 3, trust 2 ended in death AND the date of death was 0-29 days from the date of admission then the spell will also appear in the numerator. The date and fact of death may be taken from ONS data (see section on ONS linkage).

After episodes are linked into CIP spells, they are then filtered for inclusion in this indicator:

<table>
<thead>
<tr>
<th>HESID</th>
<th>Episode</th>
<th>Epistart</th>
<th>Epistor</th>
<th>Episend</th>
<th>Classpat</th>
<th>Epitype</th>
<th>Admimeth</th>
<th>Admidate</th>
<th>Diadest</th>
<th>Disdate</th>
<th>Startage</th>
<th>Resladstc</th>
<th>Diag_1</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>01/07/2002</td>
<td>1</td>
<td>16/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>Sxx</td>
<td>First valid local authority of residence. Valid diagnosis in list for this indicator</td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>2</td>
<td>17/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>Zxx</td>
<td>-</td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>3</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>1</td>
<td>51</td>
<td>23/07/2002</td>
<td>67</td>
<td>00AF</td>
<td>-</td>
</tr>
<tr>
<td>49</td>
<td>23/07/2002</td>
<td>1</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>51</td>
<td>23/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>-</td>
<td>Transfer</td>
</tr>
<tr>
<td>49</td>
<td>24/07/2002</td>
<td>2</td>
<td>24/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>49</td>
<td>25/07/2002</td>
<td>3</td>
<td>26/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>1</td>
<td>19</td>
<td>26/07/2002</td>
<td>67</td>
<td>00AF</td>
<td>-</td>
</tr>
</tbody>
</table>

Fields needed for calculating indicator after spell creation:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Startage</th>
<th>Epistart</th>
<th>Epistor</th>
<th>Classpat</th>
<th>Epitype</th>
<th>Admimeth</th>
<th>Admidate</th>
<th>Diadest</th>
<th>Disdate</th>
<th>Spellresladstc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>01/07/2002</td>
<td>3</td>
<td>1/07/2002</td>
<td>Sxxx</td>
<td>26/7/2002</td>
</tr>
</tbody>
</table>
Example 2: CI1C - Deaths within 30 days of a Coronary Artery Bypass Grafting (CABG) procedure
A set of three episodes, with a CABG procedure in episode 1, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place, with a PTCA procedure in episode 1, and the patient was discharged alive.

c=CABG procedure
p=PTCA procedure
h=Heart valve procedure

\[
\begin{align*}
\text{Transfer} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Trust 1} & \quad \text{c} & \quad \text{p} & \quad \text{o} & \quad \text{y} \\
\text{Trust 2} &
\end{align*}
\]

This CIP spell appears in the denominator for the indicator, counted to the organisation ‘o’, which was looked up via the postcode coded in episode 2, trust 2.

If the CABG from trust 1, episode 1 and the PTCA from trust 2, episode 1 occurred in the other order, then this spell would not be included, as the CABG must occur first.

If episode 3, trust 2 ended in death AND the date of death was 0-29 days from the procedure date of the CABG, then the spell will also appear in the numerator. The date and fact of death may be taken from ONS data (see section on ONS linkage).

If a heart valve procedure ‘h’ occurred anywhere in the spell, then the spell would be excluded from the indicator.

Example 3: CI 4 - Emergency readmissions indicators
A set of three episodes, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place and the patient was discharged alive. The patient was readmitted as an emergency to trust 2 where a further 3 episode spell took place.

\[
\begin{align*}
\text{Spell 1} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Transfer} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Readmission} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Spell 2} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Trust 1} & \quad \text{y} & \quad \text{y} \quad \text{y} \\
\text{Trust 2} &
\end{align*}
\]

Spell 1 appears in the denominator for the indicator, counted to organisation ‘o’, which was looked up in episode 1, trust 1.

The number of days between discharge from spell 1 and readmission to spell 2 is calculated to decide whether the emergency readmission should be counted.

Example 4: CI5 - Discharge back to usual place of residence following emergency admission to hospital with a stroke
A set of three episodes with a primary diagnosis of stroke in the first episode of the CIP spell, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place and the patient died more than 2 days after the original admission.

\[
\begin{align*}
\text{Transfer} & \quad \text{e01} \quad \text{e02} \quad \text{e03} \\
\text{Trust 1} & \quad \text{s} & \quad \text{z} \quad \text{y} \\
\text{Trust 2} &
\end{align*}
\]

This CIP spell is not counted in the numerator because the patient died. It appears in the denominator and is counted to organisation ‘o’, which was looked up via the postcode coded in episode 1, trust 2.
If death ('d') occurred within 0-2 days of admission, this spell would be excluded from the denominator as a proxy for those unlikely to be discharged home due to the severity of the condition.

After episodes are linked into CIP spells, they are then filtered for inclusion in this indicator:

<table>
<thead>
<tr>
<th>HESID</th>
<th>Start</th>
<th>Episod</th>
<th>End</th>
<th>Classat</th>
<th>EnlTyp</th>
<th>Admimeth</th>
<th>Admisorc</th>
<th>Admidate</th>
<th>Amsdate</th>
<th>Dismeth</th>
<th>Disdate</th>
<th>Stagtag</th>
<th>Reaslast</th>
<th>Diag_1</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>01/07/2002</td>
<td>1</td>
<td>16/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>Sxx x</td>
<td>Valid diagnosis in list for this indicator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>2</td>
<td>17/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>Zxx x</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>23/07/2002</td>
<td>3</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>51</td>
<td>23/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>First valid local authority of residence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>24/07/2002</td>
<td>2</td>
<td>24/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>8</td>
<td>98</td>
<td>67</td>
<td>00AF</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>25/07/2002</td>
<td>3</td>
<td>26/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>1</td>
<td>19</td>
<td>26/07/2002</td>
<td>67</td>
<td>00AF</td>
<td>Discharged to usual place of residence within 56 days of admission</td>
<td></td>
</tr>
</tbody>
</table>

Fields needed for calculating indicator after spell creation:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Start</th>
<th>Episod</th>
<th>Classat</th>
<th>EnlTyp</th>
<th>Admimeth</th>
<th>Admisorc</th>
<th>Admidate</th>
<th>Amsdate</th>
<th>Epsstart</th>
<th>Episend</th>
<th>Admisdate</th>
<th>Dismeth</th>
<th>Disdate</th>
<th>Stagtag</th>
<th>Reaslast</th>
<th>Diag_1</th>
<th>Spellselas</th>
<th>Spellelas2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>67</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>3</td>
<td>01/07/2002</td>
<td>Sxxx</td>
<td>19</td>
<td>26/07/2002</td>
<td>1</td>
<td>00AF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Example 5: CI7(x) – Emergency hospital admissions (primary diagnosis variants)**

A set of three episodes with a primary diagnosis as defined in the indicator specification in the first episode of the CIP spell, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place and the patient was discharged alive.

This CIP spell appears in the numerator for the indicator, counted to the organisation ‘o’, which was looked up via the postcode coded in episode 3, trust 1.
If the diagnosis from trust 1, episode 1 ('s') and trust 1, episode 2 ('z') occurred in the other order, then this spell would not be included as the valid diagnosis ('s') would not have occurred in the first episode in the spell.

After episodes are linked into CIP spells, they are then filtered for inclusion in this indicator:

<table>
<thead>
<tr>
<th>HESID</th>
<th>Epistart</th>
<th>Epiorder</th>
<th>Epiep</th>
<th>Classpat</th>
<th>Epietype</th>
<th>Admimeth</th>
<th>Admisorc</th>
<th>Admidate</th>
<th>Distest</th>
<th>Distale</th>
<th>Startage</th>
<th>Resladst</th>
<th>Diag_1</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>01/07/2002</td>
<td>1</td>
<td>16/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>98</td>
<td>7</td>
<td>Sxx</td>
<td>Valid diagnosis in list for this indicator</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>2</td>
<td>17/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>98</td>
<td>7</td>
<td>Zxx</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>3</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>51</td>
<td>7</td>
<td>23/07/2002</td>
<td>-</td>
<td>First valid local authority of residence</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>23/07/2002</td>
<td>1</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>51</td>
<td>23/07/2002</td>
<td>98</td>
<td>7</td>
<td>00 AF</td>
<td>-</td>
<td>Transfer</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>24/07/2002</td>
<td>2</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>98</td>
<td>7</td>
<td>00 AF</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>25/07/2002</td>
<td>3</td>
<td>26/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>19</td>
<td>7</td>
<td>26/07/2002</td>
<td>00 AF</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fields needed for calculating indicator after spell creation:

<table>
<thead>
<tr>
<th>Sex</th>
<th>Startage</th>
<th>Epiorder</th>
<th>Classpat</th>
<th>Epietype</th>
<th>Admimeth</th>
<th>Admisorc</th>
<th>Epistart</th>
<th>Diag_1</th>
<th>Spellresladstc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>1/07/2002</td>
<td>Sxxx</td>
<td>00AF</td>
<td></td>
</tr>
</tbody>
</table>

The denominator for CI7 and its variants is taken from population estimates (see indicator specifications).
Example 6: CI7(x) - Hospital procedures (procedure or procedure/diagnosis variants)

This example relates specifically to CI7o (Hospital procedures: Lower limb amputations in diabetic patients), which is a procedure & diagnosis variant.

A set of three episodes with a diagnosis of diabetes in the second episode of the CIP spell, ending in a transfer (as defined in the section on CIP spell construction) to another Trust where a further three episodes took place, during which a lower limb amputation was coded and the patient was discharged alive. The diagnosis and procedure can be in any order in any episode in the spell.

\[ a= \text{lower limb amputation} \]

\[ \text{Transfer} \]

\[ \begin{array}{cccc}
|&|&|&|
\end{array} \]

This CIP spell appears in the numerator for the indicator, counted to the organisation ‘o’, which was looked up via the postcode coded in episode 1, trust 1.

After episodes are linked into CIP spells, they are then filtered for inclusion in this indicator:

<table>
<thead>
<tr>
<th>HESID</th>
<th>Epistart</th>
<th>Epipor</th>
<th>Epipend</th>
<th>Classpat</th>
<th>Eplype</th>
<th>Admitmeth</th>
<th>Adminorc</th>
<th>Admidade</th>
<th>Disid</th>
<th>Diagid</th>
<th>Startage</th>
<th>Resladst</th>
<th>Diag_NN</th>
<th>Oper_NN</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>01/07/2002</td>
<td>1</td>
<td>16/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>98</td>
<td>7</td>
<td>00AF</td>
<td>-</td>
<td>-</td>
<td>First valid local authority of residence</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>2</td>
<td>17/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>98</td>
<td>7</td>
<td>00AF</td>
<td>Sx</td>
<td>xx</td>
<td>Valid diagnosis in list for this indicator</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>17/07/2002</td>
<td>3</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>19</td>
<td>01/07/2002</td>
<td>51</td>
<td>23/07/2002</td>
<td>7</td>
<td>00AF</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>23/07/2002</td>
<td>1</td>
<td>23/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>51</td>
<td>23/07/2002</td>
<td>98</td>
<td>7</td>
<td>00AF</td>
<td>-</td>
<td>Ax</td>
<td>Transfer. Valid procedure in list for this indicator</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>24/07/2002</td>
<td>2</td>
<td>24/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>98</td>
<td>7</td>
<td>00AF</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>25/07/2002</td>
<td>3</td>
<td>26/07/2002</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>19</td>
<td>23/07/2002</td>
<td>19</td>
<td>26/07/2002</td>
<td>7</td>
<td>00AF</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fields needed for calculating indicator after spell creation:

<table>
<thead>
<tr>
<th>Fields from the first episode in CIP spell</th>
<th>First valid RESLADST, Diagnosis and Procedure*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Startage</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
</tr>
<tr>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>
The procedure and diagnosis could have occurred in any procedure/diagnosis position in any episode in the spell. The procedure, diagnosis and RESLADSTC do not have to occur in the same episode as each other.

For ci7(x) procedure variants, the algorithm is as in Example 6 but without the requirement for any specified diagnosis to be coded.

The denominator for CI7 and its variants is taken from population estimates (see indicator specifications).
Section B: Data quality: construction of caution marks

Data quality requirements

All the indicator data files contain caution marks for each organisation for each year. These flag whether issues concerning data quality (DQ) should be taken into account when interpreting indicator values. There are three types of criteria applied to the data in creating these caution marks. The first is completeness, a measure of whether records have valid codes in the data fields used for the calculation of the indicators. The second is coverage, which is a measure of whether the HES data for the organisation are in line with the previous or following years’ data. Low coverage can indicate that data are missing from HES, while high coverage can indicate undetected partial duplication of records (e.g. where updated and incomplete versions of the same record are retained). Exact duplicates are removed from the HES data before data quality is assessed. The third criterion, death coding, only applies to indicators involving deaths. The death coding measures whether episodes ending in death have a valid diagnosis. This is also a proxy for completeness of the operative procedures fields. As there is no other direct way of testing this, it is assumed that organisations with poor completeness of diagnoses in such records will have poor completeness of operative procedure coding. The three components will be explained further in the following sections.

In addition to the three measures of data quality, further marks are supplied to assist with comparisons over time. These include small numbers marks ‘N’ where the denominator is considered to be low (<200 for generic indicators, <50 for condition-specific indicators), and ‘D’ where either the numerator or the denominator is considered to be below the data disclosure threshold (0-4). A mappings file has been produced specifically to underpin the provider (trust) data (see the section on Provider Mappings for details).

Selection of episodes for the DQ caution marks components

Different episode selections apply to the data completeness and death-coding component of different indicators, i.e. different selections of data, as shown in table 1. Coverage applies equally to all the clinical indicators.

Table 1: Selection of episodes by data completeness, coverage and death-coding

<table>
<thead>
<tr>
<th>Conditions (HES code in brackets)</th>
<th>Data completeness for each indicator (see table 2 for codes)</th>
<th>Coverage</th>
<th>Death Coding 1a/1b/1c/1d</th>
<th>Death Coding 2a/3a/3b</th>
<th>Death Coding VFM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Episode type</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General (1)</td>
<td>Delivery (2)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Delivery (2)</td>
<td>Birth (3)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Birth (3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Episode status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Finished (3)</td>
<td>Unfinished (1)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Patient classification</td>
<td>Ordinary admissions (1)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Ordinary admissions (1)</td>
<td>Day cases (2)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Day cases (2)</td>
<td>Mothers and babies (5)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Mothers and babies (5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discharge method</td>
<td>Deceased (4)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>
Data quality component: data completeness

Testing for data completeness involves the measurement of the proportion of episodes with invalid/missing values in the fields used in the clinical indicator (CI) constructions. Each CI uses a different selection of episodes and fields, hence the data completeness component has a different specification for each indicator (see table 2). If any of the fields in that episode contain missing or invalid data relevant to HESID construction, spell linkage or the CI’s construction, that whole episode is marked as having invalid/missing data, but only for that CI.

Total numbers and percentages of episodes with missing/invalid values are produced by organisation (SHA, PCO, LA, TRUST) for each clinical indicator and year.

Selection of episodes and fields evaluated

The fields scanned in the data completeness component can be grouped in three sets:

- Construction of HESID.
- Construction of spell linkage.
- Additional fields in the CI’s construction.

Note 1: Some fields that are involved in the HESID construction or in the CI’s construction are not assessed in the data completeness component. This is due to the following reasons:

- Operation date – opdte_nn: although used in the specification of some of the CIs, it can be replaced by episode start date or admission date if it is missing.
- Subsidiary diagnosis – diag_nn: not essential anywhere, but if present and valid it will be used to identify cancer cases.
- Treatment specialty – tretspef: is only used to backfill mainspef and is not used for data completeness as mainspef covers it.
- Local patient Id - Lopatid: Although it is used in the HESID construction there is no invalid value.

Note 2: Newnhsno\(^2\) & Resladstc/Postdist\(^3\) is a combined check, which marks an episode as invalid when both Newnhsno and Resladstc/Postdist are missing/invalid. The construction of HESID requires that either the NHS number or the postcode is valid. If both are invalid the algorithm will not be able to match the patient over time. HESID is used to identify readmissions and construct CIP spells for all CIs. Also NHS number and postcode are needed to link the ONS death data with the hospital episodes for death indicators.

\(^2\) NHS number
\(^3\) Resladstc is local authority district, Postdist is postcode district.
## Table 2: Summary of fields and episode selection used in data completeness for each clinical indicator

| CIs included | Clinical indicator short name | Selection (only FINISHED (epistat = 3) episodes are included. Applies to all indicators) | Startage | Sex | Newrhnsno & Residst | Admimag | Admisdte | Disdate | Dismeth | Disdest | Epistart | Epiend | Episrch | Episrhns | Epiorder | Oper 1-24 | Diag 1 | Diag 7-10 | Mlinnfal |
|--------------|-------------------------------|------------------------------------------------------------------------------------------------|---------|-----|---------------------|---------|----------|---------|---------|---------|----------|--------|--------|---------|---------|----------|--------|---------|---------|---------|
| CI1a/b       | Death within 30 days of surgery (non-elective admissions) | classpat = 1, epitype = 1,2 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | 
| CI1c         | Death within 30 days of a heart bypass operation | classpat = 1,5, epitype = 1,2 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI4a         | Emergency readmission to hospital following discharge (adults) | classpat = 1, epitype ≥1 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | 
| CI4f         | Emergency readmission to hospital following treatment for a fractured hip | classpat = 1, epitype ≥1 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI9ef and other CI9(x) indicators | Readmission to hospital following a psychiatric discharge (adults/older people) | classpat = 1,2, epitype = 1 | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | √ | | 
| Value For Money Indicator | Deaths within 30 days of admission | classpat = 1,2,5, epitype = 1,2 | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI7(x) diagnosis variants | Admissions for selected ICD codes | classpat = 1, epitype = 1 | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI7(x) procedure variants (no daycase) | Admissions for selected OPCS4 codes | classpat = 1, epitype = 1 | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI7(x) procedure variants (inc daycase) | Admissions for selected OPCS4 codes | classpat = 1,2, epitype = 1 | √ | √ | √ | √ | √ | √ | √ | √ | | 
| CI7(x) procedure and diagnosis variants | Admissions for selected OPCS4 and ICD10 codes | classpat = 1 epitype = 1 | √ | √ | √ | √ | √ | √ | √ | √ | |
Methodology for assessing missing/invalid data

The method for assessing whether a particular field contains missing or invalid data is summarised in table 3 below.

**Table 3: Methodology for assessing missing/invalid data by field.**

<table>
<thead>
<tr>
<th>Field</th>
<th>Missing/invalid data</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>= 0 (Not known)</td>
<td></td>
</tr>
<tr>
<td>Startage - Age at start of episode</td>
<td>Null</td>
<td></td>
</tr>
<tr>
<td>Newnhsno - NHS number &amp; Resladst - Local authority district Together with Postdist – Postcode district</td>
<td>Missing/invalid NHS number if = Y (Not known)</td>
<td>Missing/invalid but not ‘ZZ99’ (mainly non-UK resident, see note below)</td>
</tr>
<tr>
<td>Admimeth - Method of admission</td>
<td>= 99 (Not known)</td>
<td></td>
</tr>
<tr>
<td>Admidate - Admission date</td>
<td>Not between 1/1/1930 and end of stated financial year</td>
<td></td>
</tr>
<tr>
<td>Admisorc - Source of admission</td>
<td>= 99 (Not known)</td>
<td></td>
</tr>
<tr>
<td>Disdate - Discharge date</td>
<td>Not within stated financial year</td>
<td>Dismeth between 1 and 5 and epistat = 3 (finished)</td>
</tr>
<tr>
<td>Dismeth - Method of discharge</td>
<td>= 9 (Not known)</td>
<td>Epistat = 3</td>
</tr>
<tr>
<td>Disdest - Discharge destination</td>
<td>= 99 (Not known)</td>
<td>Dismeth between 1 and 5 and epistat = 3</td>
</tr>
<tr>
<td>Epistart - Episode start date</td>
<td>Not between 1/1/1930 and end of stated financial year</td>
<td></td>
</tr>
<tr>
<td>Epiend - Episode end date</td>
<td>Not within stated financial year or epiend-epistart</td>
<td>Epistat = 3</td>
</tr>
<tr>
<td>Epiorder - Episode order</td>
<td>= 99 (Not known)</td>
<td>Epistat = 3</td>
</tr>
<tr>
<td>Oper1–24 - Operative procedures #1–24</td>
<td>= &amp; (Not known)</td>
<td>Epistat = 3</td>
</tr>
<tr>
<td>Diag_1 – Primary diagnosis</td>
<td>= R69X, R95-R99 or 6th character = ‘A’</td>
<td>Epistat = 3</td>
</tr>
<tr>
<td>Mainspef - Main specialty</td>
<td>= &amp; (Not known)</td>
<td></td>
</tr>
</tbody>
</table>

Note: The postcode field is a sensitive data item and its access is very restricted. To check its validity derived fields are used: Resladst (Local authority district) and Postdist (Postcode district). Resladst equal to ‘Y’ identifies all missing/invalid postcodes. Records from patients who are non-UK resident do not have a UK postcode and are not considered to be failures. These records need to be excluded from the selection of invalid/missing records. This is done selecting Postdist different from ‘ZZ99’. Birth records (EPI TYPE=3, patients aged less than 3 months at the start of the episode (STARTAGE), episodes with a diagnosis in the range Z311, Z312, Z313, Z318 and episodes with a procedure code of Q383 or one starting with Q13 are excluded from this validity check.

The primary diagnosis check looks for unknown causes of morbidity (R69X) and mortality (R95-R99). It also looks for asterisks. Some diagnosis codes (dagger) need to be completed with another diagnosis as the secondary diagnosis (asterisks). Asterisks codes are identified by an ‘A’ at the 6th character, and they should not be in the primary diagnosis field.

**Data quality validity marks**

As data completeness is assessed specifically for the requirements of each indicator, organisations that do not have adequate data quality for one indicator may pass on another indicator. See table 4 below for data validity.

---

4 An episode is marked as invalid if both Newnhsno and Resladst are missing/invalid.

5 Resladst (local authority district) and Postdist (postcode district) are derived form the postcode. If the postcode is invalid/missing Resladst is set to ‘Y’ (not known), Postdist is used to exclude patients who are not domiciled on UK.

6 The modulus 11 algorithm is used to check the validity of the NHS number. In addition, the following known ‘dummy’ codes are considered to be invalid: XXXXXXXXXXX or X00000000X (where X is 0-9), 2333455667.

7 Unknown and non-specified causes of morbidity.

8 Unknown and non-specified causes of mortality.

9 A diagnostic code with an ‘A’ in the sixth character position identifies an asterisk code, which should not occur in the primary diagnosis.
criteria. The thresholds are applied to the % of episodes missing/invalid data, rounded to the nearest whole number (and similarly for other data quality components).

**Table 4: Criteria for assigning data quality validity marks to organisations: completeness**

<table>
<thead>
<tr>
<th>Quality mark</th>
<th>Description</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Adequate data quality validity</td>
<td>20% or less of FCEs with missing/invalid codes for acute readmissions, admissions, and deaths indicators.</td>
</tr>
<tr>
<td>Q</td>
<td>Low data quality – caution because of low record quality</td>
<td>More than 20% of FCEs with missing/invalid codes for acute readmissions, admissions, and deaths indicators.</td>
</tr>
</tbody>
</table>

**Data quality component: coverage**

Coverage tests for a discrepancy in the number of HES records received compared with a previous or following year’s HES records. A measure of coverage is calculated for each organisation and each year. In addition to shortfalls in data submitted to HES in any one year, organisation mappings between years could cause the coverage component for an organisation to be flagged.

**Table 5: Criteria for assigning data quality validity marks to organisations: coverage**

<table>
<thead>
<tr>
<th>Quality mark</th>
<th>Description</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Adequate data quality coverage</td>
<td>At least 80% coverage for acute and specialist trust readmission, admission and death indicators.</td>
</tr>
<tr>
<td>Q</td>
<td>Low data quality – caution because of low record coverage</td>
<td>Less than 80% coverage for acute and specialist trust readmission, admission, and death indicators.</td>
</tr>
</tbody>
</table>

**Data quality component: death coding**

There is a third DQ caution marks component - death-coding - that checks the quality of the diagnostic coding of death episodes. The data completeness component checks the validity of the main diagnosis for all episodes, including death episodes. However, missing diagnoses in the death episodes could be masked by the other episodes.

There is no direct test for whether the coding of an operation has been missed in a death episode. It is assumed that if the diagnosis coding of a death is missing, any associated operations are missed too. Therefore, the percentage of missing diagnoses on deaths is a proxy for the percentage of missing operations on deaths. If a trust only codes CABG operations for patients that survived, leaving the procedure field for patients who died uncoded, they will appear to have a zero 'in-hospital death rate' for CABG irrespective of the actual death rate.

The death-coding component only applies to those indicators where death is involved, deaths following a heart bypass operation (CI1C) and deaths following non-elective surgical procedures (CI1A) for example. The percentage of episodes ending in death in hospital without a diagnosis must be lower than 20% for those organisations to be considered having adequate data quality for such indicators. See table 6 below for the quality marks criteria.
**Table 6: Criteria for assigning data quality validity marks to organisations: Death coding**

<table>
<thead>
<tr>
<th>Quality mark</th>
<th>Description</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>Adequate clinical coding of deaths</td>
<td>20% or less of episodes ending in death with primary diagnoses missing/invalid</td>
</tr>
<tr>
<td>Q</td>
<td>Low data quality – caution because of poor clinical coding of deaths</td>
<td>More than 20% of episodes ending in death with primary diagnoses missing/invalid</td>
</tr>
</tbody>
</table>

**Provider Mappings**

A reference file ('Compendium Annex 4 HES Mappings Dec 08 v1.1' (www.nchod.nhs.uk, Methods, Annex 4)) has been created, and maintained over time, to show organisational change that affects each provider that appears in the HES database. A ‘mapping’ may occur when either two providers merge, or a provider splits into two or more parts i.e. the mental health part of a trust may be moved to a different provider. The indicators are analysed based on providers as they exist within each financial year. No attempt is made to reallocate activity if there has been a change between years. The mappings reference file provides an alert plus details of the organizational change between years that should be considered when interpreting trend data.

For each provider code that appears in the PROCODETC field (see indicator specifications or the section on derived fields for details of these fields), the following information is presented for financial years 1998/99 through to 2007/08:

- The Strategic Health Authority (if known) of the provider. This is not year-specific, the SHA does not vary between years.
- The cluster (based on clusters used in the Healthcare Commission star ratings) of the provider. Note that the cluster can vary between years, due to re-classifications i.e. because of a mapping. Where a cluster is not known for a provider within a financial year, an attempt to assign one is made by looking at previous/following years labels, mappings, information from external sources and/or case-mix.
- Three flags: does the trust appear in HES in the previous financial year – Y or N; does the trust appear in HES in the current financial year – Y or N; does the trust appear in HES in the following financial year – Y or N.
- Up to six columns for mappings affecting the provider between the financial year and the next financial year.
- Up to six columns for notes relating to the mappings columns.

The following information is taken from this mappings file and used to construct provider level aggregate data:

- The trust cluster label (varies between years).

The following information is taken from this mappings file and presented in indicator output files:

- A flag is added in each year if the trust cluster changes at any time within the 10 year series.
- A flag is added to a year if there is a mapping involving the provider code between the previous and current year, or between the current and following year.

**Contributors**

Craig Spence, Northgate Information Solutions
Azim Lakhani, National Centre for Health Outcomes Development
James Coles, CASPE Research
Kate Wooster, Northgate Information Solutions

**Contact details**

Northgate Information Solutions, People Building 2, Maylands Avenue, Hemel Hempstead HP2 4NW. Email kate.wooster@northgate-is.com

**Updated:** October 2009