hscic Health & Social Care
Information Centre

A Guide to Linked Mortality Data from Hospital Episode Statistics and the Office for National Statistics

June 2015

Version History

Version Number	Changes	Date Issued
1		14/Dec/2010
1.1	Modified Appendix C to include algorithm history details	01/Mar/2011
2	Updated text on linkage methodology, match ranks and death record used Added illustration of linkage methodology Updated match rank and death record used charts with latest data Added diagram to explain death record used Updated HES-ONS date of death comparison table with latest data Modified the frequency and timelines table	05/Aug/2011
2.1	Modified text on 'Linkage Methodology' and 'Matching ONS data to HES' sections. Minor formatting/syntax corrections	23/Aug/2011
2.2	Reviewed for accuracy and to transfer to new HSCIC branded template	13/May/2013
2.3	Reviewed and modified Figure 1 Reviewed and modified Table 1 Removed Appendix C	03/Jun/2015

Contents

Introduction

Hospital Episode Statistics (HES) data contains clinical information on patients' hospital activity and treatment such as diagnoses mentions and procedures. It also includes information on the circumstances under which a patient was admitted (such as elective, emergency etc) or discharged (discharged with clinical consent, died etc). Information on patients who die in hospital can be analysed on the basis of primary diagnosis but HES data alone cannot be used to identify the cause of death. In addition, HES data cannot be used to obtain information on patients who died after discharge from hospital.

The Office of National Statistics' (ONS) mortality data is a richer source of information on deaths than HES, with data on the place of death and the original underlying cause of death which takes into account information provided by medical practitioners and/or coroners. Linking ONS mortality data to HES data permits the analysis of deaths in and outside hospital for all patients with a record in HES. It is also a rich source of data for analysis on a wide range of subjects including outcomes of hospital care such as postoperative mortality (deaths within 'n' days of surgery) and tracking performance of health care providers and outcomes of treatment over time.

Linkage methodology

The ONS mortality data is linked to HES by matching person identifiable data in the ONS mortality dataset with patient identifiers in HES (the HESID index). The linkage process results in assigning a unique HES patient identifier (HESID) to the ONS death record. The HESID is present in all HES data sets (Accident & Emergency, Admitted Patient Care and Outpatients), enabling patients to be tracked in a confidential way. Read more about the HESID and its methodology on the processing cycle and HES data quality page of the HSCIC website which provides detailed information on creation of HESIDs for patients.
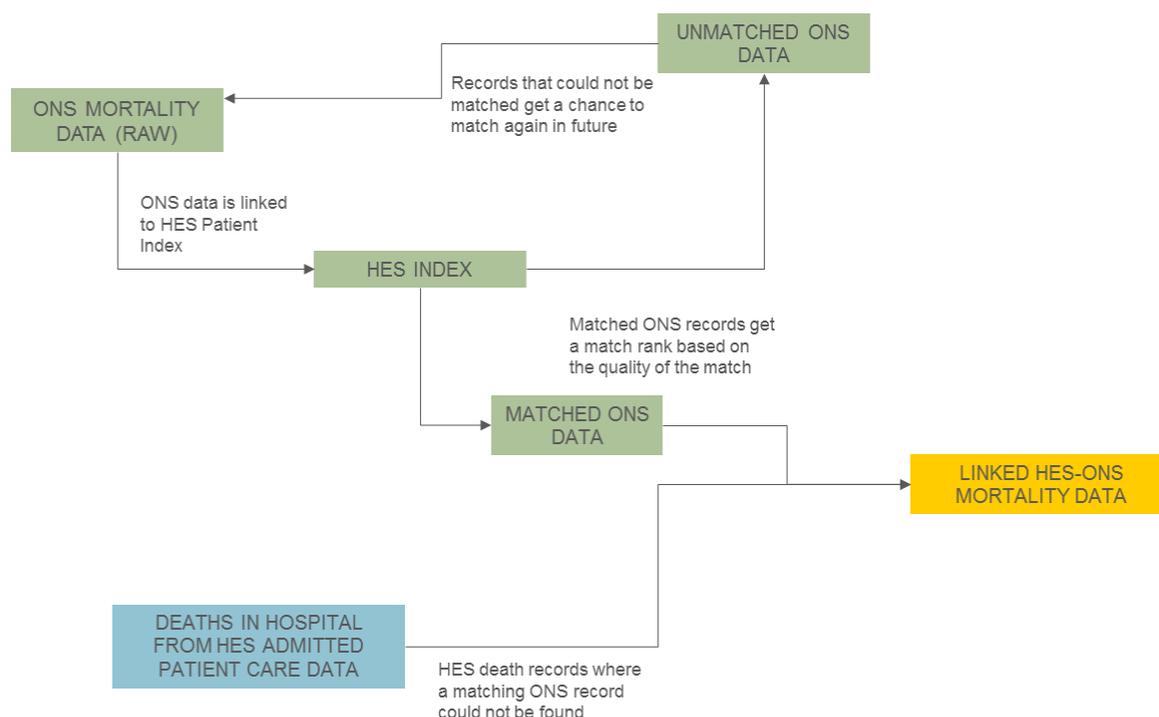
The latest processed HES data is always used for linkage. It should be noted that only ONS records that can successfully match to a patient in HES are included in the linked dataset. ONS records that cannot be matched are rejected and given opportunity to match every subsequent month when the latest death registrations from ONS are available. An ONS mortality record can match to a HESID based on 8 different criteria (refer to the section 'Matching ONS data to HES' in this document on page 8). If an ONS record matches to more than one HESID the best quality match is selected. The process results in the assignment of a unique HES patient identifier (HESID) to the ONS death records.

HES captures all activity in English hospitals, but the ONS mortality file contains all deaths in England and Wales. So when the linkage happens it is possible to capture Welsh residents treated in England.

The ONS records that have been linked to HES are later merged (refer to the section 'Merging ONS and HES mortality data' in this document on page 10) with death records from the HES admitted patient care (APC) dataset. These records are identified by the hospital discharge method - 'dead'. This is done to ensure that death records of all patients who were discharged dead from hospital and not available in the ONS dataset or that could not be linked to the ONS dataset are made available in the linked mortality dataset.

HES captures all activity in English hospitals, but the ONS mortality data contains all deaths registered in England and Wales. So when the data is linked it is possible to capture Welsh residents treated in England. The deaths of Welsh residents who haven't been treated by an English provider will remain unmatched. Poor quality person identifiers in either HES or ONS can also cause linkage to fail. These two reasons account for account for the majority of unmatched ONS records.

Figure 1 below is a high level illustration of the linkage process.

The latest data from both HES and ONS is always used for linkage. HES data is always cumulative for the current financial year. This can cause some activity records to change or disappear when providers of healthcare modify or remove submitted data from a previous month of the same financial year. Further, an annual refresh of the provisional HES data is produced to provide a finalised version of the provisional monthly data.

The linkage methodology uses ONS mortality data based on registered deaths. Registered deaths are deaths that were registered in a period as opposed to death occurrences which occurred in that period. Registered deaths are available on a monthly basis and therefore are timelier compared to death occurrences. Sometimes the year in which a death is registered may not be the same as the year in which a death occurred such as when a coroner's investigation is required e.g. if a death is suspicious. Until a coroner's investigation is completed, a death cannot be registered.

ONS publishes mortality data based on calendar year of death registration. Just as in HES, ONS provisional data has not been subject to full quality assurance and may not contain all deaths which were registered or which occurred during the period. Deaths which occurred in a given year may be registered in a subsequent year. The linkage process remains open to accept these registrations so figures are subject to change. Figures may also change due to occasional upgrades to the linkage algorithm to enhance the way it operates. For more information on the data periods that get linked, refer to the section 'Frequency and timelines for receiving and publishing data' in this document on page 14.

Further, an annual refresh of ONS data is received each year covering the most recent calendar year, to provide a finalised position on the monthly data. On reconciling the monthly and annual refreshed ONS mortality data for the calendar year 2009, it was observed that there was less than 1% difference in data, implying the coverage and quality of the submitted monthly data is high.

Figure 2: Counts of deaths received from ONS for i) monthly data ii) annual refresh data for the calendar year 2009. The count is broken down by month of registration of death.

Comparison of 2009 ONS mortality annual refresh against provisional data
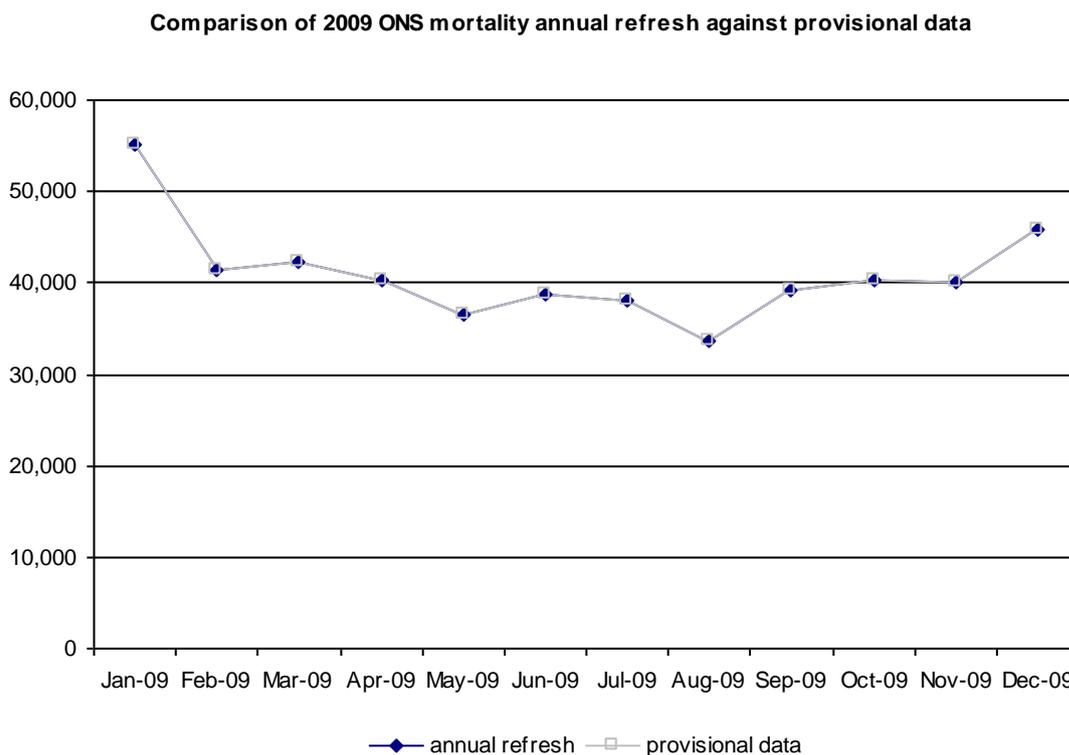
Rules used to obtain unique records for deaths in hospital

Each month a query is run against the latest cumulative HES data within the current data year to obtain deaths in hospitals. In the unlikely event that the same HESID is included in more than one death record, all but one of the records with the same HESID is removed. The following rules are applied sequentially to determine which records are deleted:

1. Delete all records sharing the same HESID except the one with the latest discharge date (and with a valid episode end date [EPIEND]).
2. If duplicates still remain, delete all records sharing the same HESID except the one with the latest submission date.
3. If duplicates still remain, delete all records sharing the same HESID except the one with the highest unique episode identifier [EPIKEY].

Matching ONS data to HES

The mortality data received from ONS each month is matched with patient data in HES. As a result most ONS death records are assigned a unique HES patient identifier (HESID), while the others are rejected. It is to be noted that ONS data for a particular month is available before HES data for the same month (refer to the section 'Frequency and timelines for receiving and publishing data' in this document on page 14). Every subsequent month the rejected ONS records get a chance to match again with the HESID index. This ensures that the unmatched ONS records are not permanently rejected. It also allows any HES records that have changed in subsequent monthly submissions another chance to link to ONS records.

The match rank obtained from the HES-ONS linkage process is an indication of the level of confidence that an ONS death record has been correctly matched to a patient in HES. Matching is performed by comparing patient identifiable fields, such as date of birth, sex, NHS number and/or postcode, which are present in both HES and ONS. The lowest rank (1) is considered the best quality match and the higher rank (8) the lowest quality match.

The match ranks used are as follows:
- Match rank 1: Exact match of DOB, SEX, NHSNO and POSTCODE; if no match is found then
- Match rank 2: Exact match of DOB, SEX, NHSNO; if no match is found then
- Match rank 3: Partial match of DOB (refer to Appendix A) and exact match of SEX, NHSNO and
- POSTCODE; if no match found then
- Match rank 4: Partial match of DOB, and exact match of SEX, NHSNO; if no match found then
- Match rank 5: Exact match of POSTCODE and NHSNO; if no match found then
- Match rank 6: Exact match of DOB, SEX and POSTCODE where NHSNO does not contradict the match and DOB is not 1 January and the POSTCODE is not in the 'ignore' list (communal establishments such as hospitals, prisons, army barracks, etc).
- Match rank 7: Exact match of DOB, SEX and POSTCODE where NHSNO does not contradict the match and DOB is not 1 January.
- Match rank 8: Exact match of DOB, SEX and POSTCODE where DOB is not 1 January.

Figure 3 below is an illustration of the fields used in for each matching criteria.

RANKS

NHSNO	DOB	SEX	HOMEADD	1
NHSNO	DOB	SEX	HOMEADD	2
NHSNO	DOB	SEX	HOMEADD	3
NHSNO	DOB	SEX	HOMEADD	4
NHSNO	DOB	SEX	HOMEADD	5
X NHSNO	DOB	SEX	HOMEADD	6
X NHSNO	DOB	SEX	HOMEADD	7
NHSNO	DOB	SEX	HOMEADD	8

Depending on the steps used during the HESID matching, it is possible for an ONS mortality record to match more than one HESID. The record with the best match rank is always used in the linked dataset. It is highly unlikely that all matches have the same match rank. In the absence of a definitive way of deciding which match is best, the inconsistency is tolerated and the single HESID is applied to each death record. Match rank 0 indicates that the death record is present in HES only, because an ONS record could not be matched to HES, or the death record is not available in the ONS dataset.

Figure 4, below, shows the percentage of records corresponding to each match rank in the linked dataset in May 2011. This is obtained by linking the mortality data in ONS from January 1998 to February 2011 and the HESID Index updated with HES data until January 2011.
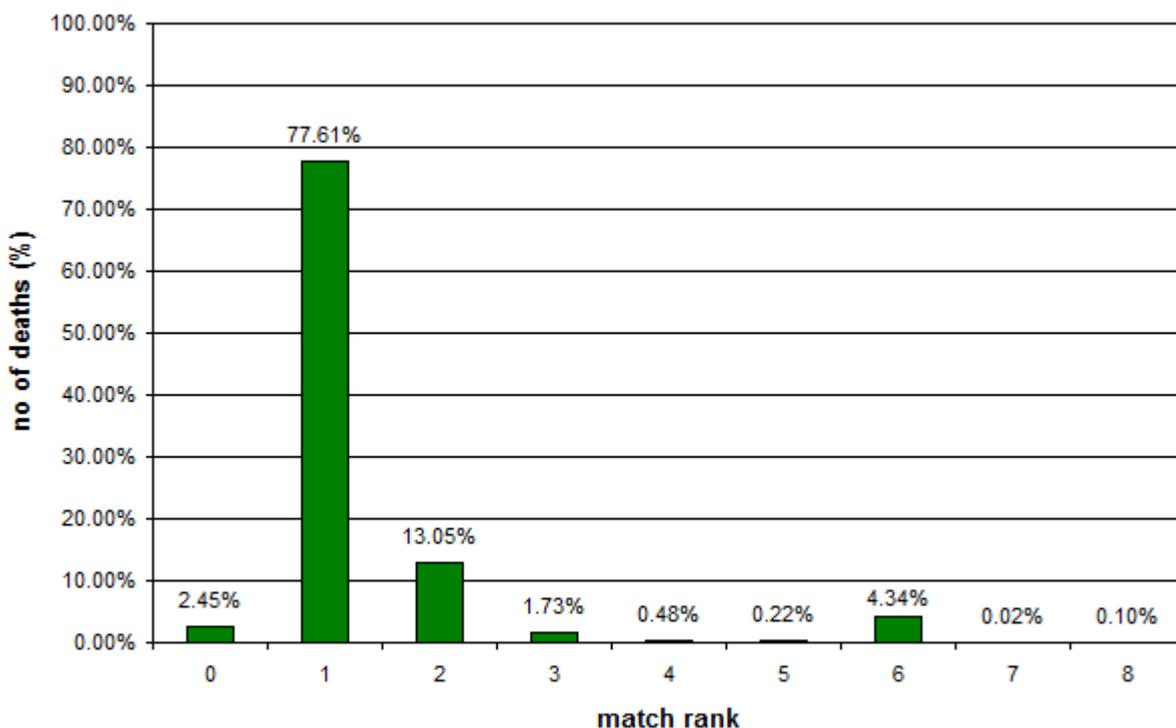
Merging ONS and HES mortality data

The deaths recorded in HES and the ONS mortality data are merged to create the linked dataset so that there is only one record per patient. The death record used (DRU), which is a derived field in the linked dataset, indicates the source of each death record in the linked dataset. There are five distinct values (HES1, HES2, ONS1, ONS2, and MIX1) for DRU. Table 1 outlines the meaning of these values and indicates which record should take preference under a range of scenarios. The record that does not take preference is removed from the final data.

Table 1: Explanation of death record used in the linked dataset				
Source	**ONS mortality date**	**DRU**	**Match rank**	**Death record from**
ONS only	n/a	ONS2	1 through 8	ONS - implies that the death was recorded only in ONS
HES only	n/a	HES1	0	HES – contradictory death info, hence information is present HES takes precedence
Both	**> 3 days after HES-**mortality date	HES2	1 through 8	HES - where contradictory date of death death information is present HES takes precedence
Both	1–3 days inclusive after HES-mortality date	MIX1	1 through 8	ONS, but date of death from HES - proximity of dates suggests slight recording error
Both	0–3 days inclusive before HES-mortality date	ONS1	1 through 8	ONS - implies delay in patient leaving hospital when dead
Both	**> 3 days before HES-**mortality date	HES2	0	HES – contradictory death info, hence information is present HES takes precedence

If DRU is ONS2, it means that the death record was present only in ONS at the time of linkage. If DRU is HES1, it indicates that the death record was present only in HES at the time of linkage. Many deaths initially recorded with only a HES-generated mortality record (DRU = HES1) often get updated with an ONS record at a later date. This might be due to the delays in registering the death as in the case of inquests by a coroner. When there is a matching ONS record the DRU changes in the linked dataset from HES1 to ONS1/ MIX1/ HES2, depending on the date of death in ONS. The death record that is low in preference is removed from the linked dataset.

ONS mortality records contain a richer set of information about the death than HES records. Where HES and ONS mortality records share the same HESID and date of death, the ONS record takes preference (DRU = ONS1). In cases where the ONS mortality data is used in the linked dataset (DRU = ONS1, ONS2 or MIX1), the match rank will always be between 1 and 8. In all cases where HES mortality data is used (DRU = HES1 or HES2) in the linked dataset, the match rank is recorded as 0 as no matching data is present in ONS (DRU = HES1) or the data in ONS is of too poor a quality to create a match (DRU = HES2).

Figure 5 below illustrates how DRU is defined using the ONS date of death as the reference.
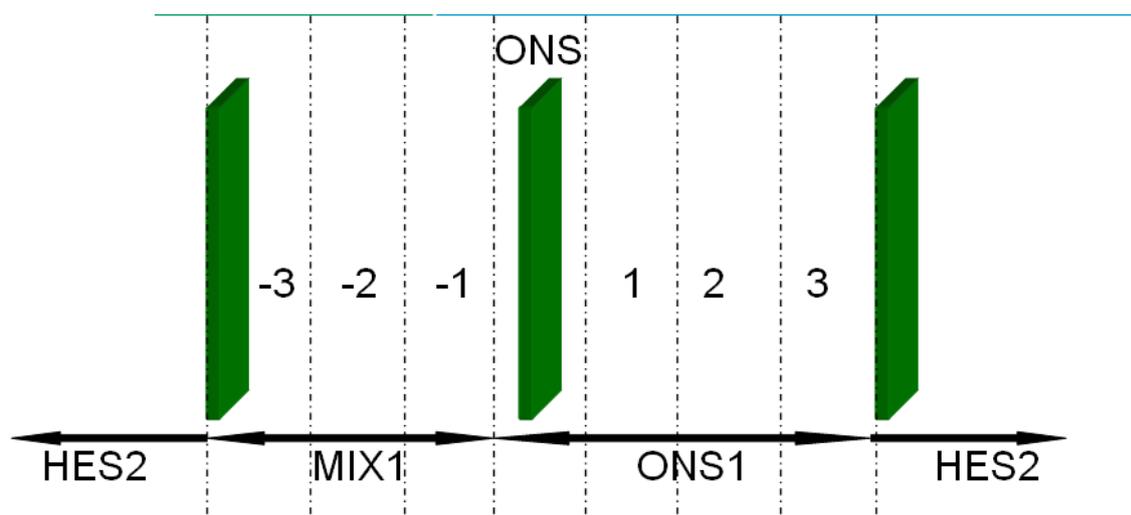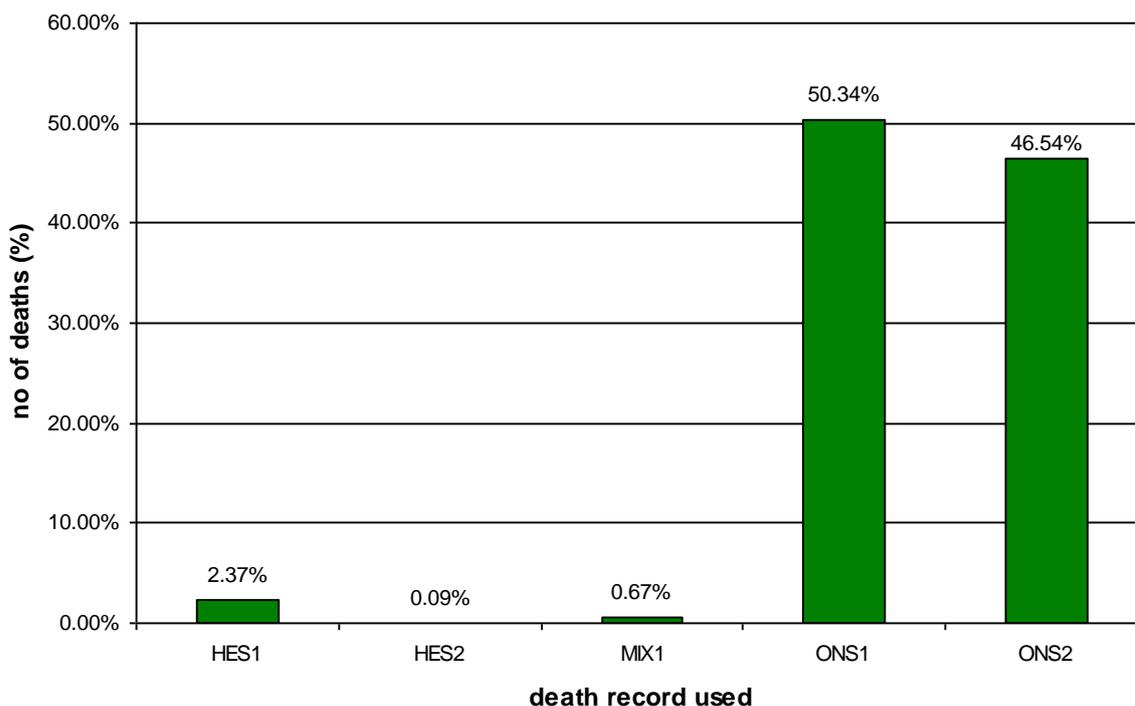

Figure 6, below, illustrates the percentage of records corresponding to each death record used in the mortality in May 2011. This is obtained by linking the mortality data in ONS from January 1998 to February 2011 and the HESID Index updated with HES data until January 2011.

From Figure 6 we can infer that 46.54% of deaths in the ONS dataset have successfully linked to a person in HES but has no death record in HES. 51% of deaths are recorded both in HES and ONS. Additionally, 2.37% of deaths are recorded only in HES or could not be successfully linked to a death record from ONS, although it is mandatory that every death should be registered within five days of occurrence. Some of the reasons for deaths not appearing in ONS may be:

- Deaths referred to coroners for inquests
- Provisional nature of monthly ONS data, which is not yet finalised
- The HES and ONS record not linking successfully due to missing or inaccurately recorded patient identifiers

When a patient dies in hospital, the discharge date is taken as the date of death. Table 2 shows the difference between the ONS date of death and the HES date of death. This table is based on the linked mortality data as of May 2011.

Table 2: Comparison of dates of death in HES and ONS based on linked mortality data as on May 2011

Difference in days	Number	%	Description
< 3	15,138	0.48%	
-3	709	0.02%	HES date of death before ONS
-2	1,834	0.06%	date of death
-1	39,327	1.25%	
0	2,963,205	94.40%	same date of death in both
1	107,837	3.44%	
2	5,697	0.18%	HES date of death after ONS
3	2,071	0.07%	date of death
> 3	3,124	0.10%	
	3,138,942		

Table 2 shows that while the vast majority of dates of death within the two datasets agree there is a significant minority where the dates are different. There may be a number of reasons for this:
- The patient may have died late at night and the hospital were unable to record the discharge until the next day
- The patient may have died on a previous day but was not released until tests were performed on a subsequent day
- There was a data input error, meaning that the discharge date was incorrect.

Flagging patients in the linked dataset with subsequent activity in HES

On rare occasions patients may appear to have activity in HES after the mortality record indicates that they have died. This is a data quality issue, either in the patient identifiers (causing an incorrect data linkage between HES and ONS), or due to a patient being incorrectly recorded in HES. The linked dataset flags these records as having activity in HES after the date of death.

These flagged records are available to customers. They are flagged and not removed because it is possible that the activity was incorrectly recorded in HES – for example, where a patient had an outpatient appointment, but died before the appointment, resulting in the data being incorrectly sent by the patient administration system (PAS). Often, such records appear in the monthly HES publications, but disappear after the HES annual refresh, as providers correct their submissions. In these cases the flag will be removed once the submission is corrected.

Mortality records of patients having more than one HESID

Hospital activity data submitted by providers of healthcare is, at times, incomplete or incorrect. This can cause the system to create new HESIDs for patients who already have a HESID, resulting in multiple HESIDs for some patients. The HESID creation process underwent changes in the past which resulted in the creation of a unique HESID for some patients with multiple HESIDs.

In the mortality linkage process we deal with such HESIDs by maintaining separate death records in the linked dataset for both new and old HESIDs, but with the same mortality data. Therefore duplicates are intentionally maintained in the linked dataset. The reason for retaining duplicate records is that users of the mortality data can link it to HES data extracts taken at different points of time, even if the patient's HESID has changed over time.

Accessing the data

Customers can request extracts of linked HES-ONS mortality data from the Health and Social Care Information Centre's Data Access Request Service.

The release of information has separate terms and conditions which are outlined in this section of the HSCIC website. For advice and guidance please contact the HSCIC contact centre on 0300 303 5678 or via email: enquiries@hscic.gov.uk

Frequency and timelines for receiving and publishing data

HES and ONS mortality data is refreshed every month, with a subsequent annual refresh of the full year's data. The annual refresh period is a financial year for HES data and calendar year for ONS data. On a monthly basis the HSCIC receives cumulative activity data for the current financial year from HES, whereas it receives just one month's worth of data from ONS.

The provisional monthly ONS mortality data predominantly contains deaths which were registered in the past two months and a few from earlier periods. This ensures that up-to-date information on deaths registered with ONS is available for linkage with deaths recorded in HES. The annual refresh of ONS mortality data contains all deaths registered in the calendar year and has been subject to quality assurance. It is usually available in the second half of the subsequent calendar year for linkage with HES.

NHS trusts and other healthcare organisations submit activity data every month, which is then run through a series of data quality checks and cleaning procedures before it is published. Due to the time taken for submission and processing there is a lag of three to four months before monthly HES data is published. The annual refresh of the HES admitted patient care dataset is usually published around 6 months after end of the financial year.

Table 3: the latest available linked HES-ONS mortality data as of May 2013

Month	HES data for Jan-13 Hospital activity	ONS mortality data available for linkage with HES
Jan-13	Activity in hospital	
Feb-13	First submission deadline (Reconciliation) for Jan-13 hospital activity data is by the end of Feb-13. The completeness level of Jan-13 data is low at this point.	
Mar-13	Second submission deadline (post-reconciliation) for Jan-13 hospital activity data is by the end of Mar-13. The completeness level of Jan-13 data at this point in time is considered acceptable for linkage.	
Apr-13	HES data processing for Jan-13 activity data HES-ONS mortality data linkage for Jan-13 data	Death Registrations: Jan 1998 to 95% Feb-13 and 50% Mar-13 Death Occurrences: Jan 1998 to 90% Feb-13 and 40% Mar-13
May-13	Jan-13 hospital activity data is published by the HSCIC	

The table above shows that hospital activity (HES) data for January 2013 is submitted by end of March 2013 and processed in April 2013. This is the latest available HES data that will be used for linking to ONS mortality data in April 2013. The latest available ONS mortality data at this point in time contains death registrations from Jan 1998 to approximately 50% of March 2013. Based on this, at the beginning of July 2013 the linked HES-ONS mortality data is expected to contain death registrations from January 1998 to May 2013 linked to March 2013 hospital activity data.

Appendix A – Notes on valid data values

The algorithm compares a range of fields to match records in ONS with that in HES. To aid data quality, the data is subjected to a range of validation rules to ensure that records are valid and in the correct format.

A date of birth is valid if:
- it is not null
- it is a valid date
- it is no earlier than 1895/01/01
- it is not later than the end of the current data year.

An NHS number is valid if:
- it is not null
- it consists of exactly ten digits
- the ten digits are not all the same
- it is not of the format 'n00000000n' (where the first and last digits are the same)
- it is not the dummy/default value '2333455667'
- the check digit is correct.

A postcode is valid if:
- it is not null
- it is exactly eight characters long
- it is of the format AXXX_9AA, AXX__9AA, or AX___9AA, where A is any upper-case alphabetic character (A–Z), X is any upper-case alphanumeric character (A–Z, 0–9), 9 is any digit (0_9), and _ is a space
- it does not start with 'ZZ'.

In addition to the validation checks detailed above, there are a number of further criteria that are applied to the data.

Local patient ID

For matching purposes, all zeros and spaces are removed from local patient identifiers, which cover local PAS or case note numbers.

Date of birth partial matching

Criteria applied to the data are:
- Neither DOB is 1901/01/01
- Neither DOB is 1899/12/31
- The two DOB values are no more than 14 years apart
- The two DOB values are the same
- *or* two components (ie YYYY, MM, DD) of the two DOB values match
- *or* two components of the two DOB values match when the MM and DD parts of one of them are swapped.

Appendix B – Fields in the linked dataset

The list of fields from the linked dataset which customers could request using the extract service is given below.

Appendix B: Fields in the linked dataset	
Field name	**Description**
pseudo_hesid	This field uniquely identifies a patient across all data years. It is generated by matching records for the same patient using a combination of NHS number, local patient identifier, provider code, postcode, sex and date of birth. Customers who request an extract will receive a unique version of the PSEUDO_HESID called the EXTRACT_HESID.
dod	Date of death
resstha	Strategic health authority of usual residence of deceased
respct	Primary care trust of usual residence of deceased
sex	Sex
communal_establishment	Place of death code 00001–99999 / H = home / E = elsewhere
nhs_indicator	Indicates whether the communal establishment code refers to an NHS establishment, referring to the physical building rather than the service. 1 = NHS establishment 2 = Non-NHS establishment
cause_of_death	Underlying cause of death
cause_of_death _non_neonatal_1 to cause_of_death_non_neonatal _15	Cause of death mentions for non-neonatal deaths (deaths occurring after 28 days of life)
match_rank	Indicates the strength of the match between the ONS and HES records.
death_record_used	Indicates whether the death record in the linked dataset was taken from ONS or HES
subsequent activity	If activity is recorded in HES after the date of death the date of this activity will be displayed in this field

Glossary

A&E	Accident and emergency
Death record used	This field indicates whether the death record in the HES-ONS linked dataset is from HES or ONS
EPIEND	Date when an episode ended
EPIKEY	Unique identifier for each episode recorded in HES
HES	Hospital Episode Statistics – the central repository of NHS activity information including admitted patient care (APC) or inpatient (IP), outpatient
HESID	Anonymised unique identifier for a patient in HES.
HESID index	The central reference table that maps individual patient keys to HESID or PSEUDO HESID.
IP	Inpatients: patients who were admitted to hospitals, also known as the admitted patient care (APC) dataset.
LOPATID	Local patient identifier, unique only for patients within the trust.
Match rank	A number that indicates the quality of match between a death record in ONS and a patient in HES, 1 being the highest quality and 8 the lowest.
HSCIC	Health and Social Care Information Centre
NHSNO	Unique identifier for a patient within the NHS.
ONS	The Office for National Statistics.
PAS	Patient administration system. Healthcare providers enter data through these systems which ends up in HES.
Patient key	A pseudonymised version of the HESID. Customers who request HES-ONS linked extracts get only the PSEUDO_HESIDs.
Postoperative mortality	Deaths after surgery (could be in or outside hospital).
PSEUDO_HESID	A pseudonymised version of the HESID. Customers who request HES-ONS linked extracts get only the PSEUDO_HESIDs.
SUS	Secondary Uses Service: the single source of comprehensive data to enable a range of reporting and analysis needs. The data from the providers of healthcare come directly to SUS before they are extracted into HES.