

Hospital Episode Statistics (HES) Analysis Guide

Published December 2019

Information and technology
for better health and care



Digital

Document Management

Revision History

Version	Date	Summary of Changes
1.0	30/12/2014	Updated following comments from group and uploaded to web.
2.0	March 2015	Table 1 updated – small number release summary
3.0	September 2018	Organisation name updated throughout. Revised disclosure control methodology incorporated into all guidelines relating to small numbers
4.0	June 2019	Updated guidance on ‘publishing’ HES data and amended formatting for new NHS Digital style guide
5.0	December 2019	Updated guidance on table 1, section 6 and updates to outdated references

Contents

Document Management	2
Revision History	2
1 Introduction and background	5
1.1 Who is the HES Analysis Guide for?	5
1.2 The purpose of the HES Analysis Guide	5
2 Analysing HES data	6
2.1 Clinical classifications	6
2.2 Episodes and spells	6
2.3 Type of admission	7
2.4 Unfinished records	7
2.5 Outpatient Appointments	7
2.6 A&E Attendances	8
3 Integrity issues	8
3.1 Year-on-year analysis	8
3.2 Data quality	10
4 Sensitivities of HES data and protecting the individual	10
4.1 Patient information	10
4.2 Clinician-related data	10
4.3 Are individuals identified?	11
4.4 Are the data securely kept?	12
4.5 CAG, IGARD and other legal issues	12
5 Guidelines on releasing HES information	13
5.1 Fundamental principles of releasing HES data	13
5.2 Recognition of an individual and breaching confidentiality	13
5.3 General release guidelines	14
5.4 Release of tabulations with small numbers or record level extracts	14
Disclosure Control Summary of release guidelines (Table 1)	15
5.5 Provisional monthly data	16

6 Further explanations	18
6.1 Disclosure control	18
6.2 Deaths	21
6.3 Re-admissions	22
6.4 Mental health	23
6.5 Abortions	23
6.6 IVF and AIDS	24
6.7 Postcode	24
6.8 Census Output Areas	25
6.9 What if the request fails the guidelines but is still required?	25
Appendix A	26
A.1 Small numbers	26
Small Numbers Summary of release guidelines (Appendix A - Table 1) (prior to September 2018)	28

1 Introduction and background

1.1 Who is the HES Analysis Guide for?

1.1.1 The Hospital Episode Statistics (HES) Analysis Guide is intended to be a useful reference guide for everyone with access to HES data. This includes everyone:

- responsible for processing HES data, e.g. staff at the Health and Social Care Information Centre (NHS DIGITAL)
- querying HES data via the Data Access Environment (DAE) e.g. analysts at NHS Digital, the Department of Health and Social Care, NHS England and NHS Improvement and other users of the HES data warehouse
- requesting extracts and tabulations of HES data
- using HES data for research.
- reading our official statistics publications

1.1.2 All users with access to record level HES data will have signed an agreement enabling their access; this Guide is to be used in conjunction with that agreement.

1.2 The purpose of the HES Analysis Guide

1.2.1 NHS Digital seeks to make information available to serve the public interest, while protecting the confidentiality of those described by it, such as patients, GPs, consultants and other clinicians. It is important that the people who are described by the data are confident that their data is being:

- collected and held legally
- handled confidentially, with due care and respect for privacy
- managed in accordance with relevant legislation, e.g. Data Protection Act 2018.

1.2.2 Given the scope of HES data and the variety of uses to which it could be put, it is essential that the security procedures within this document are understood and applied consistently. Therefore, this Guide:

- presents an overview of the data available in HES
- describes the sensitivities of some of this data
- defines the limits up to which data can be released without risk of encountering problems with those sensitivities
- gives rules for releasing HES data
- outlines some of the data quality and integrity issues.

2 Analysing HES data

2.1 Clinical classifications

2.1.1 HES uses the World Health Organization's ICD-10 (International Classification of Diseases and Related Health Problems) classification to record diagnosis information. The OPCS-4 (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, 4th Revision) classification is used to record details of any procedures or interventions performed, e.g. hip replacements.

2.2 Episodes and spells

2.1.2 It is important to remember that each admitted patient record is for a continuous period of care (episode) administered within a particular consultant specialty at a single hospital provider. Therefore, if a patient is transferred to another consultant or to a different provider during a spell of treatment (the total time a patient is in hospital, from admission to discharge) a new record is generated. This means that not all stays in hospital will be represented by a single HES record; this is why there are more finished consultant episodes than finished admission episodes.

Order of episodes in a spell

The Episode Order (EPIORDER) field shows you the order that episodes occur in a spell. The HES record detailing the first episode within a spell of treatment (the 'admission episode') will have an EPIORDER value of '1'.

2.3 Type of admission

The Patient Classification (CLASSPAT) field shows the type of admission.

Ordinary admissions have a classpat value of '1' and day cases have a CLASSPAT value of '2'.

Regular day or night attenders (CLASSPAT=3 or 4) are available in later HES years (from 2002-03) but are generally excluded from analysis as they are not always submitted by providers.

Less frequently used within the HES dataset is CLASSPAT=5 (mothers using delivery facilities only).

2.4 Unfinished records

The Episode Status (EPISTAT) field shows whether the patient was still undergoing treatment on the last day of the HES data year (31/03/yyyy). Such cases, of which there are about 250,000 per data year, give rise to what is termed an unfinished record. These records often don't contain any useful clinical data as it's normally added when the episode is finished (an important consideration when requesting extracts or tabulations or analysing data).

Unfinished records have an EPISTAT value of '1'. Most other records, i.e. those detailing episodes which ended within the data year, have an EPISTAT value of '3'. We only use finished episodes (EPISTAT=3) in our analyses.

2.5 Outpatient Appointments

Each outpatient record represents a single appointment at a consultant clinic in a single hospital provider. Patients may see the same consultant, or members of the consultant's team, on more than one occasion during the course of a single year. They may have appointments at more than one consultant clinic in the same or different specialties, or in different providers, for the same or different conditions.

Non-attendances

If you are interested in non-attendance rates you may wish to only include records from those trusts recording non-attendances (these records have not been returned by all trusts).

For DNA (did not attend) data you should only look at records with attendtype values of 4, 5, 6, 24 or 25 (which looks at first or subsequent or not known appointments and tele-consultations).

Cancellations

Requests for patient cancellations should be restricted to attendtype in the range 7, 8, 9, 27 or 28 (which looks at first or subsequent or not known appointments and tele-consultations).

Requests for hospital cancellations should be restricted to attendtype in the range 10, 11, 12, 30 or 31 (which looks at first or subsequent or not known appointments and tele-consultations).

2.6 A&E Attendances

Each A&E record represents an attendance at an A&E department.

3 Integrity issues

3.1 Year-on-year analysis

In general, the quality of HES data has been improving over time. This means that an analysis produced using 1989-90 admitted patient data (the oldest available) is not strictly comparable with a similar analysis run against data for later years. You must bear this in mind when producing a time series analysis, as small year-on-year changes may well be a product of shortfalls in the earlier years and should not automatically be interpreted as trends in treatment practice or activity.

There have been several other changes that also need to be taken into consideration. For example:

- There have been several changes in the organisation of the NHS. Some changes are structural, e.g. transfer of Directly Managed Units (DMUs) into hospital trusts, abolition of Regional Health Authorities (RHAs) and subsequent Regional Offices (ROs), and the introduction of Clinical Commissioning Groups (CCGs). Other changes relate to the geographic boundaries defining these structures.
- The clinical classification used to record diagnoses changed in April 1995-96 when the World Health Organization's ICD-10 classification of diseases and related health problems started to be used for coding diagnoses in place of the

ICD-9 classification (used from 1989-90 to 1994-95). An updated 4th edition of ICD-10 came into use in 2012.

- A statistical classification of surgical operations was introduced in the UK in 1944. Over subsequent years it was periodically revised, culminating in the introduction of the OPCS-4.2 classification, which was fully implemented across the NHS in 1990.
- The OPCS-4 classification is reviewed annually by the NHS Classifications Service of NHS Digital to ensure it continues to reflect current clinical practice in the NHS. Every review improves the accuracy of NHS coded clinical detail and the quality of NHS coded data and supports the needs of all NHS users to improve planning and delivery of patient care. A list of which OPCS 4 version applies to each data year can be found at:

https://hscic.kahootz.com/connect.ti/t_c_home/view?objectId=14270896#14270896

- Data items in the HES data warehouse have changed over the years (e.g. 'ethnic category' has been added and 'category of patient' is no longer used).
- The values within data items can change over time due to changes made to the commissioning datasets that are used to submit HES data.
- The introduction of Payment by Results, increased private sector involvement in the delivery of secondary care and the increase in the number of single specialty A&E departments, walk-in centres and minor injury units will have all affected trends. Payment by Results is a system whereby hospitals are paid for the number of patient treatments, known as activity, they perform and the complexity of these treatments. It was introduced in a phased way from the middle of the last decade onwards. In order to get paid correctly, hospitals need to record the activity they perform and the clinical codes that outline the patients' conditions and treatment. This has provided a major financial incentive for hospitals to ensure all of the activity they perform and the clinical coding is fully recorded. This improved recording of information captured by HES could be one of the factors leading to the reported activity increases.
- The NHS has seen increases in real terms expenditure throughout the period for which HES data is available. In the earlier years of the period, the year on year increase in this expenditure was higher than in the most recent years. The period has also seen a rise in the number of emergency admissions. One factor contributing to this is likely to be the increased demand on health services from an ageing population. Alongside this there has been the introduction of observation or medical assessment units at many hospitals to which patients arriving in Accident and Emergency departments are admitted, often for around a day, to enable observation and tests to be performed on them.
- Extra care should be taken when looking at clinical data, as changes in NHS practices (such as the introduction of new procedures and interventions) can have an effect on changes through time.

These points should be shared with anyone using the data and appropriate warnings included in any analysis.

3.2 Data quality

Data quality notes are published alongside our HES publications highlighting any known issues with that HES data.

4 Sensitivities of HES data and protecting the individual

4.1 Patient information

4.1.1 The data held within HES relate to the care of individual patients while in hospital. Some clinical information, such as certain diagnoses (e.g. AIDS) or procedures (e.g. abortions), could potentially be very sensitive if related to an individual patient. While all health information is sensitive, some data items can be more sensitive than others or can more easily enable identification of individuals. For example:

- the set of data items including postcode, sex and date of birth (collectively referred to as 'the potential identifiers') could, in combination and with outside knowledge, help to identify an individual patient
- the NHS number or the healthcare organisation's local patient identifier can uniquely identify a patient if the relationship between the data item and the patient is known. This relationship is not kept within HES.

4.1.2 In HES, the patient's ethnic group is not a restricted data item. The patient has the option to select their ethnic group or not to state the ethnic group to which they consider they belong.

4.2 Clinician-related data

4.2.1 In addition to patient-specific data items, there is sensitivity around practitioner-specific data. This includes the General Medical Council (GMC) code of the

consultant and the registered GP. These data are routinely available but only in an anonymised format.

4.2.2 Even anonymised data may, however, be recognised as relating to known practitioners and should therefore be used with care. For example:

- The consultant code in HES may refer to a consultant-led team of doctors or an individual consultant
- Data aggregated by specialty may, in certain circumstances, report on an individual consultant
- In a single-handed GP practice the GP Practice Code will identify an individual GP.

4.3 Are individuals identified?

4.3.1 The privacy of the individual is one of the basic principles behind the HES Analysis Guide. Measures are in place to make sure that the identity of a patient is never discovered using HES, even though potential identifiers (see 4.1) are held. These data items are held for grouping and linkage purposes in order to identify and group conditions and analyse treatment patterns and patient pathways.

4.3.2 Patients' NHS numbers were added to the HES dataset in 1997-98 to reduce the dependence for access to potential identifiers. However, the NHS number is now treated as a potential identifier and is therefore handled as a sensitive data item in HES. Where direct access to this and other sensitive data items is allowed, they are normally anonymised. This allows differentiation of patients for the purposes of comparative analysis, while still concealing the potential identity of individual patients.

4.3.3 PSEUDO_HESID is a data field in HES. It was established to give each patient in the HES data warehouse an identifier that is unique to them and HES. This was implemented to prevent much of the need for the disclosure of potential identifiers purely for matching purposes, although they will still be needed for matching with data outside of HES, e.g. Office for National Statistics (ONS) mortality data, which are routinely linked to HES.

4.3.4 The PSEUDO_HESID can be used to link together records for a single patient from 1997-98 onwards, and has the potential to link admitted patient to

outpatient records from 2003-04 onwards and to A&E records from 2007-08 onwards.

- 4.3.5 The PSEUDO_HESID succeeds the HESID and has been developed to bring it into line with general encryption standards under the guidance of the former Ethics and Confidentiality Committee (ECC) which is now the Confidentiality Advisory Group (CAG). The algorithm, which derives the PSEUDO_HESID, has been enhanced to match as many patients as possible, even in cases in which data quality is poor.
- 4.3.6 To meet CAG's standards on data sharing, customers who request PSEUDO_HESID as part of an extract will receive a field called EXTRACT_HESID which is unique to the customer.
- 4.3.7 Most users of the HES data warehouse will never need to access the potential identifiers. Any user who wishes to gain access to identifiable or sensitive items must have a legal basis to do so (Consent, Statutory Powers or Section 251 approval via CAG and Independent Group Advising on Releases of Data (IGARD) approval for any sensitive data items) and will need to sign up to additional conditions to gain access.
- 4.3.8 Where users are granted access to sensitive data items they must not query, extract, store or link these data items with their equivalent anonymised data items. On no account should an attempt be made to decipher the process of creating anonymised data items. Such an act would be treated as a serious breach of security.

4.4 Are the data securely kept?

- 4.4.1 HES data is stored in a secure data centre where there is restricted access to the environment. The data is held securely under a combination of software and hardware controls.

4.5 CAG, IGARD and other legal issues

- 4.5.1 HES data is not obtained directly from the patients but from hospital provider systems via the Secondary Uses Service (SUS). It is not feasible to obtain explicit consent from patients for the multitude of uses to which HES data can

be put. NHS Digital has statutory powers through the Health and Social Care Act 2012 to receive store and process Patient Identifiable data. Access to this level of data by another organisation would require a legal basis either via

Statutory Powers, Consent or Section 251 approval administered by the Confidentiality Advisory Group (CAG)

4.5.2 The Independent Group Advising on Releases of Data (IGARD) was set up to review applications for Hospital Episode Statistics (HES) and Medical Research Information Services (MRIS) where sensitive data items are being requested. IGARD is an independent group hosted by NHS Digital.

4.5.3 HES data are deemed 'personal' under the terms of the Data Protection Act 2018 and are therefore subject to its fair processing requirements. The HES team has taken legal advice to make sure that the ever-developing Data Access Environment and service complies with the provisions of the Act.

5 Guidelines on releasing HES information

5.1 Fundamental principles of releasing HES data

5.1.1 There are three basic principles to follow when sharing, releasing or publishing HES data:

1. Protecting the privacy and confidentiality of individuals.
2. Making the data available, both routinely and on request, as widely as possible (subject to NHS Digital policy, the HES Analysis Guidance Document and current legislation).
3. Seeking advice (from a colleague, a senior officer or a member of the HES team) whenever there is any doubt about whether data can be released.

5.2 Recognition of an individual and breaching confidentiality

5.2.1 All users of HES data must consider the risk of identifying individuals in their analyses prior to sharing, release or publication. You should be on the alert for any rare and unintentional breach of confidence, such as responding to a query relating to a news item that may add more information to that already in the public domain.

5.2.2 Data should always be released at a high enough level of aggregation to prevent others being able to 'recognise' a particular individual. If you recognise an individual while carrying out an analysis you must exercise professionalism and respect their confidentiality. If you believe this

identification could easily be made by others you should alert a member of the Information Governance team at NHS Digital for further advice [via enquiries@nhsdigital.nhs.uk]. Similarly, if you recognise an individual from existing analyses you should alert the author of the analysis as well as NHS Digital Information Governance team.

5.2.3 While appropriate handling of an accidental recognition is acceptable, the consequences of deliberately breaching confidentiality could be severe. Action would be taken by NHS Digital against both your organisation and you as an individual.

5.3 General release guidelines

5.3.1 Table 1 summarises the general rules for protecting the identity of individuals - data classified as 'OK' in the table may be shared, released or published without restriction. Note that where more than one rule is relevant, apply the strictest. It also highlights where further consideration is necessary before sharing, releasing or publishing data. The rules are explained further in the paragraphs following the table.

5.3.2 All analyses must state that the data source (Hospital Episode Statistics, NHS Digital).

5.4 Release of tabulations with small numbers or record level extracts

5.4.1 Where the release of small numbers is necessary, authority and advice must be sought from the Information Governance team at NHS Digital. Where approved, a data sharing agreement (DSA) will need to be completed. Guidance on the release of small numbers is listed below

Disclosure Control Summary of release guidelines (Table 1)

Description	Detail	Level of Aggregation	
		National	Sub-National (Regional, Area Team (AT), Commissioning Region (CR), LSOA, MSOA, CCG, GP Practice, Trust, Provider etc.)
Small Numbers:	1 - 7 acceptable in any individual cell	OK (agreed by IGARD)	No, follow disclosure control rules and suppress
Any other Numbers:	0 and any number greater than 7 in any individual cell	OK (agreed by IGARD)	No, follow disclosure control rules and round to the nearest 5, zeroes can be displayed Consult NHS Digital if a suitable alternative analysis cannot be found within the guidelines or where you are unsure of the geographical area and the request is still required (this may need approval from the IGARD)
<p>In addition to small number in a cell, certain diagnoses and procedures are especially sensitive. The clinical codes for these are currently being reviewed so please use your judgement if you are aware of codes not on the list that are perceived as being sensitive</p>			
Deaths:		OK, but ensure footnotes are given with output	Must consult with NHS Digital before issuing

Surgical Procedures:	Abortions: Q09.1, Q10.1, Q10.2, Q11.1, Q11.2, Q11.3, Q11.5, Q11.6, Q14.1-Q14.6, Q14.8, Q14.9, R03.1, R03.2, R03.8, R03.9	There are restrictions on using and releasing abortion statistics and official statistics for abortions are published by the Department of Health	
	Neurosurgery for mental disorders	OK	OK – see section 6.5
	Electroconvulsive Therapy (ECT): A83		OK – see section 6.5
	Other	OK	OK

Description	Detail	Level of Aggregation	
		National	Sub-National (Regional, Area Team (AT), Commissioning Region (CR), LSOA, MSOA, CCG, GP Practice, Trust, Provider etc.)
Diagnoses:	HIV: B20 - B24, Z20.6, Z21 and Z71.7,	OK	No
	Abortions: O04 to O08	There are restrictions on using and releasing abortion statistics; official statistics for abortions are published by the Department of Health. <i>(See section 6.5 for further details)</i>	

Sexually Transmitted Infections: A50
to A64 inclusive

OK

No

IVF:
Z31.2

OK

No

Other:

OK

OK

Consultant, Registered GP, Referrer: Data are available in anonymised form, although the use of this still needs care as consultants can be recognisable from the anonymised data if there is sufficient other information available. Use of original (raw) codes required IGARD approval.

NB – The above list of codes is currently under review. The last revision (as listed above) is from 1999 but with current ICD10 codes.

5.5 Provisional monthly data

5.5.1 Each month NHS Digital publishes provisional headline information from monthly HES data and makes the underlying data available to HES users.

These data are provisional and they are subject to changes and revisions each month. They should be treated as estimates until the final National Statistics Annual Publications.

Each month we publish:

- high level summary of key figures
- headline figures for special topics
- monthly time series
- data quality notes
- data completeness notes
- background information about the monthly data.

5.5.2 The data completeness note clearly identifies the scale of revisions to figures from earlier months. The monthly data quality note describes limitations on appropriate use of the data including a brief description of any particular issues with the latest month's data. It is planned that most identified quality issues with monthly data will not prevent publication, but that we will make users aware of them, investigate them and update users about them through the data quality note published with the subsequent monthly publication. Users should be aware that while we will highlight known data quality issues, but due to tight timescales and the volume of data within HES, it will not be possible to identify all data quality issues. We will however welcome feedback from users if you identify any data quality issues for investigation. Once the headline publication is released, more detailed monthly HES data is made available to analysis and extract service customers.

5.5.3 Provisional monthly data is only available for a given year until the final annual figures have been published for that year. (A monthly breakdown of the final annual figures will still be available.) A chart containing 25 months of data is published each month containing finalised annual HES data for previous years as well as monthly provisional data for the latest year. Special topics are also covered in each monthly publication and we welcome suggestions as to which topics users would like to see covered in future months. The topic which is going to be included each month will be preannounced, alongside the finalised publication date, on NHS Digital publication calendar.

5.5.4 Potential users of monthly data should be aware of the following features of the data:

- They are not usually manually cleaned
- The latest month will always be incomplete as the monthly data is taken before all providers have to submit their data to SUS (i.e. at the PbR reconciliation date) □
Monthly data are taken cumulatively so earlier months' data are improved with each successive month
- Month 13 data (which will also contain the earlier eleven months) cannot be used to draw firm conclusions about annual activity, as it will be updated by the HES Annual Refresh.

Considering the points above, all monthly data must be clearly labelled 'provisional' and due consideration must be given to fitness for purpose.

5.5.5 The headline information that is made publicly available is for the period up to 6 weeks before the publication month. For example, in October 2019 NHS Digital published provisional monthly information for the period up to August 2019.

6 Further explanations

6.1 Disclosure control

6.1.1 'Small numbers' in HES are the numbers 1 to 7. Sub-national analyses are more likely to contain small numbers, which might facilitate identification of individual patients, especially at a local level. They might also allow identification of a hospital consultant, where local knowledge identifies a single consultant treating patients in a particular specialty. In these instances, disclosure control rules will need to be followed. Please note that disclosure control only needs to be applied to values relating to individuals. No disclosure control is required for values not relating to individuals such as a count of providers.

6.1.2 Small numbers are not necessarily a problem when they cover a broad geographical area, because the patient would not normally be identifiable (Table 1 above shows the acceptable levels). Therefore, in general practise relating to national level data, no disclosure control will be required either for small numerators or denominators.

However, data that are likely to be more sensitive, e.g. deaths (see 6.2.1), should still be treated with care if they are likely to identify individuals. Small numbers within local authorities (LAs), wards, postcode districts, CCGs providers and trusts may allow identification of patients and should not be published, released or shared.

6.1.3 When publishing, releasing or sharing HES data, you must make sure that;

National level data

No disclosure control required for small numbers.

(Restrictions on certain diagnoses and procedures as set out in table 1 still apply.)

Breakdowns below national level

The following steps will be applied to reduce the risk of identifying individuals from small numbers.

- 1) If the national total is between 1 and 7 (inclusive), no sub-national breakdown will be displayed.
- 2) If the national total is greater than or equal to 8
 - a. Sub-national counts between 1 and 7 (inclusive) will be displayed as '**'.
 - b. Zeroes will be unchanged.

Digital

c. All other counts will be rounded to the nearest 5.

For example, sub-national numbers will appear after disclosure control as:

Before Disclosure Control	0	1	2	3	4	5	6	7	8	9	10	11	12	13
After Disclosure Control	0	*	*	*	*	*	*	*	10	10	10	10	10	15

Please note that row/column totals will be calculated and then rounded, which means that the total of rounded values may differ from the rounded total, as in the following example:

	M	F	U	Total
A	5	10	4	19
B	12	17	11	40
C	8	8	16	32

After disclosure control has taken place this table will become:

	M	F	U	Total
A	*	10	*	20
B	10	15	10	40
C	10	10	15	30

The most common calculations using counts of HES records are percentages or rates. Unless the denominator for a percentage is relatively large, it is often possible to work out the numerator and denominator because only one possible pair of values would give this percentage (especially if the percentage is displayed to a higher degree of precision such as 2 decimal places).

Disclosure control will therefore also need to be applied to percentages. If it is necessary to include a percentage breakdown at sub-national level the following should be applied:

- Where the numerator or denominator is between 1 and 7 (inclusive), no percentage or rate is calculated, and a '*' will be displayed.
- Where the numerator is zero, the percentage will be 0%.

Digital

- Where the unrounded numerator and denominator are greater than or equal to 8, a percentage or rate is calculated using the rounded numerator or denominator. It is important to ensure the rounding takes place **before** the percentage is calculated.

6.1.4 Presentation of suppressed percentages should adhere to the following rules.

Unrounded counts				Rounded counts				Percentages		
	M	F	Total		M	F	Total		M	F
A	0	16	16	A	0	15	15	A	0%	100%
B	5	2	7	B	*	*	*	B	*	*
C	5	12	17	C	*	10	15	C	*	67%
D	9	12	21	D	10	10	20	D	50%	50%
E	8	14	22	E	10	15	20	E	50%	75%

Please note that because row/column totals will be calculated and then rounded, the total of rounded values may differ from the rounded total, and row/column percentages may sum to more than 100%, as in row E above.

Users should be aware that percentage calculated using the rounded numerator and denominator will differ from the percentage that could be calculated using unrounded numbers, by quite a large degree for smaller denominators.

For additional examples on the presentation of percentages subject to disclosure control please refer to the full Disclosure Control Methodology document.

For additional analysis and calculations such as mean, median or mode the following rules apply:

- At national level, no disclosure control is required
- At sub-national level, where the number of values used is between 1 and 7 inclusive, no calculation is made and a '*' is displayed.
- At sub-national level, where the number of values used is greater than or equal to 8, the calculated value is displayed.
- For complex calculations such as standardised rates, confidence intervals and regression models, no disclosure control is required.

6.1.5 The rules on disclosure control of low cell counts should be considered wherever small numbers are encountered, irrespective of whether the count is directly a count of patients. The rules cover several types of analysis (e.g. episodes, admissions and deaths) and measures based on small numbers, such as bed days. While a bed day

measure may not appear to be disclosive, a small number of bed days may imply a small number of cases so similar disclosure control is needed.

6.1.6 Certain other measures, such as average times waited or length of stay, appear not to give any disclosive information on the number of cases, but at times they may do so, e.g. a mean of 5 days with up to 5 cases implies no case exceeded 25 days. In such cases, the averages might not be disclosive, but judgement still needs to be taken as to whether they imply something more about individual cases.

6.1.7 An alternative to suppressing values from 1 and 7 is to consider a higher level of aggregation for one or more items, e.g. move from trust level to Area

Team/Commissioning region of treatment, or from diagnosis at the 4-character level to the 3-character level, or group using wider age bands. A higher level of aggregation is the preferred option if several cells are affected by the disclosure control rule.

6.1.8 Another option is to provide the data at the requested low level (if necessary, for purpose), but anonymising the level of aggregation, i.e. replace identifying codes or labels with arbitrary reference numbers.

6.1.9 In addition to this, as detailed in the small number table, there are a number of diagnosis and procedure codes which are covered by the small numbers' guidance. This list is currently under review and may be subject to change once ratified. Advice should be sought from NHS Digital if there are any doubts around any potentially sensitive ICD10 or OPCS codes.

6.2 Deaths

6.2.1 Deaths analysis can be sourced from the HES dataset itself or from ONS death registration records supplied by ONS for linkage with HES. The DISMETH (Discharge method) and DISDATE (Discharge date) fields in HES are limited in that they do not capture deaths occurring post discharge from hospital. This can cause incompleteness or bias in some comparative analyses. Linking HES with the ONS registered deaths data addresses this issue.

6.2.2 ONS mortality data is present in public records and could therefore lead back to a patient being identified in a HES analysis thus making the HES contents more vulnerable to disclosure of additional facts about the individual otherwise not in the public domain. DMSG reviewed this situation and concluded that adherence to the HES small numbers policy is sufficient to address any small additional risk posed by

the linkage. However, users of ONS mortality data in HES should take extra care to ensure that any output tables run no risk of disclosure.

6.2.3 Queries concerning deaths should always be handled with caution. Requests for counts of deaths by operative procedure and/or diagnosis for individual NHS trusts are especially sensitive.

6.2.4 HES deaths from whichever source data should not be used to infer the quality of care provided by NHS trusts; many factors not included within the HES dataset will have an effect on the care of the patient. If HES deaths data are used as part of such comparative analysis users should ensure that the data is suitably footnoted.

6.2.5 If you have access to mortality data derived from ONS then you must ensure that you are adhering to the terms of the ONS data access agreement you signed.

6.3 Re-admissions

6.3.1 Queries on re-admissions (to the same or another hospital) can be carried out using the HESID data item. However, analysis of re-admissions figures on their own cannot inform on quality of care, since other factors not described in the HES data can affect the care of the patient.

6.3.2 It is important to note that re-admissions can result from a variety of reasons and may not be clinically related to an earlier admission. However, most indicators assume that all providers are equally affected by this and other confounding factors to make the analysis meaningful comparatively if not in absolute terms.

6.4 Consultant team / clinician data

6.4.1 Since 1997-98, the consultant code has been included with each episode record. It gives the general medical council (GMC) code of the lead consultant responsible for the care of the patient. Analysis carried out with this data item is referred to as being 'by consultant team' as it incorporates (although does not currently distinguish between) the work of the whole team including junior doctors, anaesthetists, nurse practitioners and other members of staff.

6.4.2 Consultant code is currently handled as a sensitive data item (See Appendix B) and is routinely available only in an anonymised format. However, even the anonymised consultant data should be used with care if there is sufficient information to recognise

individual consultants, for example if there are only a few people within a specialty in a Trust. Should the original codes (i.e. non-anonymised) be required this will need approval from IGARD.

6.5 Mental health

6.5.1 All analyses involving either neurosurgery for mental disorders or electroconvulsive therapy (ECT) should be treated with caution and caveated as following: Data relating to ECT procedures/neurosurgery should be treated with caution as variation in the setting of the procedure impacts the recording of it, i.e. some trusts will perform as a day case, some as an out-patient procedure etc. ECT is often given as a course of treatment so any counts of procedures or episodes should not be described as patient counts

6.6 Abortions

6.6.1 There are restrictions on using and releasing abortion statistics and official statistics for abortions are published by the Department of Health and Social Care. The numbers of abortions recorded in HES data are different to published official abortion statistics and hence, any release of HES abortion data should be done with extreme caution, following the disclosure guidance of abortion statistics and released with appropriate caveats. Guidance on the release of abortion statistics can be obtained by contacting the Department of Health and Social Care Abortion Statistics team on 0207 972 5537 or emailing abortion.statistics@dhsc.gov.uk. Annual statistics for abortion can be found at www.gov.uk, search for 'abortion statistics'. Data for earlier years can be requested by email.

6.6.2 Some areas of interest as described in the guidelines:

- Small numbers and cells that may breach confidentiality:
 - o From consideration of the risk scenarios the unsafe cells for abortion statistics are those for which:
 - The count is zero unless no other value is logically possible
 - The count is below 5 for a Government Office Region in England, the country of Wales or a larger area
 - It is less than 10 and the area concerned is smaller than Government Office Region in England or the country of Wales
 - It is less than 10 and the variables are considered highly sensitive
 - The count is associated with at most 2 practitioners

- It is associated with at most 2 hospitals.
- Other highly sensitive variables:
 - Young ages (<15)
 - Late gestation (over 24 weeks)
 - Procedure by gestation
 - Medical conditions.

Please note that in the event of any requirement to access small numbers and/or sensitive information relating to abortions you must contact the Abortion Statistics department at the Department of Health: abortion.statistics@dhsc.gov.uk

6.7 IVF and AIDS

6.7.1 IVF data are included in the HES dataset, but all identifiers should be removed. The source of official statistics on IVF and fertility treatment is the Human Fertilisation and Embryology Authority (HFEA) [<http://www.hfea.gov.uk>].

6.7.2 AIDS data have special sensitivities. The source of official statistics, epidemiology and surveillance of infectious diseases (including HIV/AIDS) is the Centre for Infections (CfI) at the Health Protection Agency (HPA) [<http://www.hpa.org.uk/Topics/InfectiousDiseases/InfectionsAZ/HIV/>].

6.8 Postcode

6.8.1 The full postcode of a patient's home address is an identifiable data item and requires Consent or CAG approval for its use. The postcode district (the first part of the postcode only, e.g. W1, N14 and GU25) is generally available to users of the Data Access Environment.

6.8.2 Many geographic fields have been derived from postcode and are routinely available.

See the HES Data Dictionary:

[<http://www.hscic.gov.uk/hesdatadictionary>]

6.9 Census Output Areas

- 6.9.1 Output Areas (OAs) were devised by ONS to identify small geographical areas that cover similar population sizes, according to the 2001 Census. They are defined by adjacent postcodes as at census day. There are 165,665 OAs in England. Too small for meaningful statistics in their own right, their main purpose is as building blocks for a range of other measures.
- 6.9.2 The full ten-character census output area code (OACODE) is a sensitive data item in HES. Access to this field requires IGARD approval. The lowest level of census output code routinely available in HES is OACODE6 (six-character census output area code).
- 6.9.3 The increased number of small-area codes routinely available in HES, now including oacode6 as well as CURRWARD (current ward) and WARD81 (frozen census ward 1981) or WARD91 (frozen census ward 1991), raises concerns that yet smaller areas might be identified by differencing. This could enable identification of selected small slithers of land (in HES, these might represent single postcodes).
- 6.9.4 As part of their conditions of access, users of the HES data warehouse must not try to produce smaller areas by differencing.

6.10 What if the request fails the guidelines but is still required?

- 6.10.1 There may be exceptional circumstances where access to sensitive data items or small numbers is needed for an analysis. There are procedures in place for when such circumstances contravene the guidelines presented here. This may include seeking approval from IGARD.
- 6.10.2 In the case of extracts and tabulations, data requests that contravene the guidelines in this Guide will need to be fully justified, scrutinised by the related data custodian and submitted to NHS Digital for consideration before being released.

Appendix A

Below are the rules on small numbers that were used and were valid for HES data up to September 2018. These are displayed to explain how previous published analyses had suppression applied. The below rules will also still be applied to those tabulations where a data sharing agreement was signed prior to the release of the updated documentation.

A.1 Small numbers

A.1.1 'Small numbers' in HES are the numbers 1 to 5. Low-level analyses are more likely to contain small numbers, which might facilitate identification of individual patients, especially at a local level. They might also allow identification of a hospital consultant, where local knowledge identifies a single consultant treating patients in a particular specialty.

A.1.2 Small numbers are not necessarily a problem when they cover a broad geographical area, because the patient would not normally be identifiable (Table 1 above shows the acceptable levels). However, data that are likely to be more sensitive, e.g. deaths (see 6.2.1), should still be treated with care if they are likely to identify individuals. Small numbers within local authorities (LAs), wards, postcode districts, CCGs providers and trusts may allow identification of patients and should not be published/released.

A.1.3 When publishing/releasing HES data, you must make sure that cell values from 1 to 5 are suppressed at a local level to prevent possible identification of individuals from small counts within the table. Zeros (0) do not need to be suppressed. If only one cell requires cell suppression, you must suppress at least one other component cell (the next smallest) to avoid calculation of suppressed values from the totals. You should replace these values with '*' and add a note:

*' in this table means a figure between 1 and 5

A.1.4 The rules on suppression of low cell counts should be considered wherever small numbers are encountered, irrespective of whether the count is directly a count of patients. The rules cover several types of analysis (e.g. episodes, admissions and deaths) and measures based on small numbers, such as bed days. While a bed day measure may not appear to be disclosive, a small number of bed days may imply a small number of cases so similar suppression is needed.

A.1.5 Certain other measures, such as average times waited or length of stay, appear not to give any disclosive information on the number of cases, but at times they may do so, e.g. a mean of 5 days with up to 5 cases implies no case exceeded 25 days. In such

cases, the averages might not be disclosive, but judgement still needs to be taken as to whether they imply something more about individual cases.

A.1.6 An alternative to suppressing values from 1 and 5 is to consider a higher level of aggregation for one or more items, e.g. move from trust level to Area

Team/Commissioning region of treatment, or from diagnosis at the 4-character level to the 3-character level, or group using wider age bands. A higher level of aggregation is the preferred option if several cells are affected by the suppression rule.

A.1.7 Another option is to provide the data at the requested low level (if necessary for purpose), but anonymising the level of aggregation, i.e. replace identifying codes or labels with arbitrary reference numbers.

A.1.8 In addition to this, as detailed in the small number table, there are a number of diagnosis and procedure codes which are covered by the small numbers guidance. This list is currently under review and may be subject to change once ratified. Advice should be sought from the HSCIC if there are any doubts around any potentially sensitive ICD10 or OPCS codes.

Small Numbers Summary of release guidelines (Appendix A - Table 1) (prior to September 2018)

Description	Detail	Level of Aggregation	
		National, Regional, Area Team (AT), Commissioning Region (CR), Strategic Clinical Networks (SCN)	Postcode District, Local Area, Electoral Ward, LSOA, MSOA, CCG, GP Practice, Trust, Provider
Small Numbers:	1 - 5 acceptable in any individual cell	OK (agreed by IGARD)	Generally no, suppress Consult NHS Digital if a suitable alternative analysis cannot be found within the guidelines or where you are unsure of the geographical area and the request is still required (this may need approval from the IGARD)
In addition to small number in a cell, certain diagnoses and procedures are especially sensitive. The clinical codes for these are currently being reviewed so please use your judgement if you are aware of codes not on the list that are perceived as being sensitive			
Deaths: issuing with output		OK, but ensure footnotes are given	Must consult with NHS Digital before
Surgical Procedures:	Abortions: Q09.1, Q10.1, Q10.2, Q11.1, Q11.2, Q11.3, Q14.1-Q14.6, Q14.8, Q14.9, R03.1, R03.2, R03.8, R03.9 Neurosurgery for mental disorders Electroconvulsive Therapy (ECT): A83	There are restrictions on using and releasing abortion statistics and official statistics for abortions are published by the Department of Health Q11.5, Q11.6, OK – see section 6.5 OK – see section 6.5	

Other

OK

OK

Description	Detail	Level of Aggregation	
		National, Regional, Area Team (AT), Commissioning Region (CR), Strategic Clinical Networks (SCN)	Postcode District, Local Area, Electoral Ward, LSOA, MSOA, CCG, GP Practice, Trust, Provider
Diagnoses:	HIV: B20 - B24, Z20.6, Z21 and Z71.7,	OK	No
	Abortions: O04 to O08	There are restrictions on using and releasing abortion statistics; official statistics for abortions are published by the Department of Health. <i>(See section 6.6 for further details)</i>	
	Sexually Transmitted Infections: A50 to A64 inclusive	OK	No
	IVF: Z31.2	No	No
	Other:	OK	OK
Registered GP, Referrer:	Data are available in anonymised form, although the use of this still needs care as consultants can be recognisable from the anonymised data if there is sufficient other information available. Use of original (raw) codes required IGARD approval.		

NB – The above list of codes is currently under review. The last revision (as listed above) is from 1999 but with current ICD10 codes.