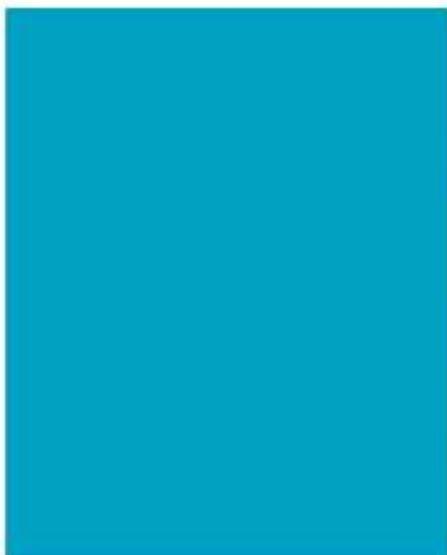


Care Episode Statistics: Technical Specification of the GP Extract



NHS England INFORMATION READER BOX

Directorate

Medical	Operations	Patients and Information
Nursing	Policy	Commissioning Development
Finance	Human Resources	

Publications Gateway Reference:

126

Document Purpose	Guidance
Document Name	Care Episode Statistics: Technical Specification of the GP Extract
Author	NHS England
Publication Date	29 May 2013
Target Audience	GPs
Additional Circulation List	
Description	This document provides the technical specification for release 1 of the general practice data extraction for the care.data service.
Cross Reference	N/A
Superseded Docs (if applicable)	
Action Required	
Timing/Deadlines (if applicable)	
Contact Details for further information	Geraint Lewis Patients and Information Quarry House Leeds

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the intranet is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the intranet.

NHS England

Care Episode Statistics: Technical Specification of the GP Extract

First published: May 2013

Contents

Foreword from the Chief Data Officer	6
1. Introduction	8
2. Benefits and Risks	10
3. Information Requirements	14
4. Data Management.....	16
Appendix A: List of fields and selection criteria from “Everyone Counts”	22
Appendix B: List of Sensitive Codes	26
Appendix C: Glossary of Terms	27

“The effective collection, analysis and dissemination of relevant information is essential for swift identification and prevention of substandard services; facilitating accountability; provision of accessible and relevant information to the public; and supporting patient choice of treatment.”

Robert Francis QC
Report of the Mid Staffordshire NHS Foundation Trust Public Inquiry

Foreword from the Chief Data Officer



Some of the information systems we use in the English NHS are ranked among the very best in the world. Our GP practices have employed electronic health records for decades; the *GP2GP* system now allows patients to transfer their entire clinical record seamlessly when they change GP practices; and soon, the *Electronic Prescription Service* will be transmitting prescriptions between any GP practice and any pharmacy in the country.

Similarly, the data systems we use behind the scenes are world class in many respects. For example, the *Hospital Episodes Statistics* (HES) system held by the Health and Social Care Information Centre (HSCIC) collates information about every NHS-funded hospital admission and outpatient attendance in England. HES data underpin the decisions made by commissioners and managers across the health service, and HES has been used in tens of thousands of audits and peer-reviewed research studies, which have led to direct improvements in the quality and efficiency of health care.

However, to reduce unexplained variations in care, and for patients to participate more fully in their care, considerable changes are needed to NHS information systems. In particular, the use of information must become far more transparent, and clinical data need to flow more safely and more effectively across the health and care system. At the moment, for example, the IT systems used in one part of the NHS seldom link to those in another. And, as its name suggests, the HES system is currently limited to hospital data: there is no national equivalent to HES for primary, community, or social care data.

A modern data system for the health and care system

Over the course of the next few years, we will establish a new, modern data system for the NHS in England. Known as *care.data*, its purpose will be to provide timely, accurate information to citizens, clinicians and commissioners about the treatments and care provided by the NHS.

The aims of the *care.data* programme are six-fold: first, to support patients' choice; second, to advance customer services; third, to promote greater transparency; fourth, to improve outcomes; fifth, to increase accountability; and finally, to drive economic growth by making England the default location for world-class health services research. The underlying data within *care.data* will all be collected routinely as part of the care process. For the first time, these data will be linked for patients nationwide along their entire continuum of health and social care.

Because most NHS consultations occur in primary care, it is vital that data from GP systems be incorporated into *care.data* as soon as possible. One of the first goals of the *care.data* programme, therefore, will be to expand HES. Hospital Episode Statistics (HES) will become *Care Episode Statistics* (CES) by incorporating and linking individual-level data across all care settings. This expansion will ensure that commissioners obtain a more complete and balanced picture of the care being delivered to patients, since information will now be linked across care settings at a person-level. Since the analysis of linked data underpins the commissioning of integrated care, CES will play vital role in supporting commissioners to reduce unnecessary fragmentation in care.

NHS England will review the care.data primary care extract with a range of stakeholders after 12 months.

Care Episode Statistics

The first new building block of the care.data programme will be the inclusion within CES of primary care data drawn from all GP practices. At a local level, analogous linked data systems have been established over the past few years in several parts of the country. However, the time has now come for us to standardise these systems. This change is essential for providing commissioners, such as NHS England and clinical commissioning groups, with the data that they require for analysing relationships between the needs of communities they serve, the services they commission for patients, and the health outcomes that result.

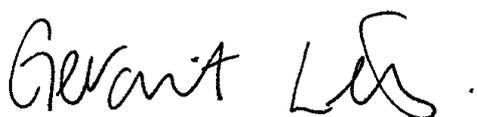
Protecting Confidentiality

The entire care.data system, including CES, will conform to the same strict standards of data security and confidentiality that have governed the use of HES for many years. For example, it will be fully compliant with human rights legislation, data protection legislation, and with the common law duties of confidentiality wherever they apply. It will also comply with the recommendations of the second Caldicott Information Governance Review and with the NHS Constitution.

As the technical guidance set out in this document explains, we will not be requiring any patient names or addresses to be extracted; nor will we be requesting any free text; nor any retrospective data prior to 1 April 2013. However, this document does signal how we will require GP practices to provide coded, patient-level information to the Health and Social Care Information Centre, so that primary care data can be incorporated into CES. This information will include diagnostic codes, investigation results, and information about NHS prescriptions.

As with HES data, *aggregated* CES data will be published openly; however, *record-level* CES data will only be made available pseudonymously to accredited organisations, such as a clinical commissioning group, under the terms of a legally binding agreement. Each user will be required to sign a data sharing and re-use agreement, which specifies what data are to be made available to the user, and the purpose for using the data extract.

As a public health doctor who has been involved in the use and analysis of NHS and social care data for many years, I am extremely excited about the potential of the care.data programme to increase transparency within the health and social care system to encourage greater participation by patients and citizens. I am confident that that CES will rapidly become seen as an invaluable resource for clinicians, commissioners, and researchers as they work to improve the quality, efficiency, and equity of the health service for us all.



Geraint Lewis FRCP FFPH

Chief Data Officer

1. Introduction

Purpose

This document provides the technical details for *Release 1* of the general practice data extraction for the care.data service. This extract was announced in the NHS planning guidance, *Everyone Counts*,¹ published in December 2012. The aims of the extract include increasing transparency and improving patient outcomes. The extract will contribute to the *Care Episode Statistics* (CES) database, which will form a core part of the new care.data service – a modern data service for the health and social care system.

The specification described in this document was recommended to proceed in March 2013 by the Independent Advisory Group (IAG) of the General Practice Extract Service (GPES). This recommendation was based on the satisfaction of three conditions the IAG set out at their meeting on 14 February 2013, one of which was a review of the codes by the Joint General Practice Information Technology Committee (JGPITC) of the British Medical Association (BMA) and the Royal College of General Practitioners (RCGP).

The current document begins by setting out the background to the primary care extract for the care.data service. In chapter 2, we describe the anticipated benefits, as well as the risks associated with the extract and how those risks will be mitigated. In chapter 3, we provide details of the data extract itself, followed in chapter 4 with a description of how the data will be managed by the Health and Social Care Information Centre. Finally, a series of appendices provide additional details of which codes will and will not be extracted, together with a glossary of terms used in this document.

Background

NHS England is responsible for supporting clinical commissioning groups (CCGs) in commissioning NHS services, and also for directly commissioning primary care services and specialised services. The Government's *Mandate to the Commissioning Board* (2012) describes how:

9.2 NHS England will be directly commissioning NHS services provided by GPs, dentists, community pharmacists and community opticians; specialised care; health services for people in custody; and military health. This offers a great opportunity to improve standards and national consistency, for example in services for people with rare conditions. The Board has an important responsibility to drive improvements in the quality of primary care, reflecting the vital role that stronger primary care will play in supporting delivery of objectives across this mandate.

9.3 The Department will hold the Board to account for the quality of its direct commissioning, and how well it is working with clinical commissioners, health and wellbeing boards, and local healthcare professionals. An objective is to ensure that, whether NHS care is commissioned nationally by the Board or locally by clinical commissioning groups, the results – the quality and value of the services – should be measured and published in a similar way, including against the relevant areas of the NHS Outcomes Framework. Success will be measured not only by the average level of improvement but also by progress in reducing health inequalities and unjustified variation.

¹ <http://www.commissioningboard.nhs.uk/everyonecounts/>

To ensure that it can meet these responsibilities, NHS England is working with partners such as Public Health England and the Care Quality Commission to establish the care.data service, in order to:

- **improve outcomes**, by monitoring against the Outcomes Frameworks
- **increase accountability** in the health service by making data more widely available
- **promote greater transparency**, for instance in support of local service planning
- **support patients' choice** of service provider and treatment, by making comparative data publicly available
- **advance customer services**, so that services are planned around the patient
- **drive economic growth** through the effective use of linked data, for example by making England the default location for world-class health services research.

An accompanying document to the planning guidance, entitled “*Better data, informed commissioning, driving improved outcomes: clinical data sets*”, defined an outline data set that would be used in the care.data service (see Appendix A). Developed with input from clinical commissioners, this set out the types of data that GP practices would be required to submit to the Health and Social Care Information Centre (HSCIC) for inclusion in CES.²

GP practices may choose to provide these data using the *General Practice Extraction Service* (GPES); alternatively, they may wish to use another secure method for transmitting the data to the HSCIC. If a GP practice wishes to supply the required information to the HSCIC using another secure method, then the HSCIC must first agree this alternative method with the practice.

Public Awareness and support for GP practices

As reiterated in the report of the Caldicott2 review of information governance⁵, patients must be made aware of how their data are being used. As the data controllers of primary care data, GP practices are in the best position to raise awareness with patients, and practices should make information available to patients about data disclosures to the HSCIC. NHS England will support GPs by providing materials to promote public awareness of this data extract. We have worked closely with the BMA and the RCGP to co-develop materials for patients, as well as supporting materials for GP practices, including a helpline. These resources will be made available to practices well in advance of their first data extract.

Patient objections

As set out in the NHS Constitution,^{2,3} individuals have a right to object to the disclosure of their personal confidential data, and they have a right to have any reasonable objections respected.⁴ In order for an individual to object, they must be given as much information about how their data will be used outside the context of direct care. Such information should be made

² Under Section 259 of the H & SC Act 2012 the disclosure is mandated, however, it has been agreed that individual patient objections will generally be respected.

³ <http://www.nhs.uk/choiceintheNHS/Rightsandpledges/NHSConstitution/Pages/Overview.aspx>

⁴ See NHS Constitution Handbook page 58 <http://www.gov.uk/government/publications/the-nhs-constitution-for-england>

⁵ Caldicott2, <http://caldicott2.dh.gov.uk/>

available not only through websites but also through other publicity materials made available to individuals, such as posters and leaflets.

More recently, the Secretary of State for Health has given a commitment that in relation to data held in GP records, individuals' objections to disclosure to the HSCIC will be respected other than in exceptional circumstances.⁶ The guidance we are providing to GP practices sets out in detail how this commitment will be implemented. Briefly, there will be two types of objections:

1. Objection 1 - the individual does not want any personal confidential data flowing to the HSCIC from the GP record. In this case, the GP practice must not disclose the individual's personal confidential data to the HSCIC. This objection will be implemented through the use of specified Read codes.
2. Objection 2 - (in development) the individual is content for their personal confidential data to flow to the HSCIC but does not want their personal confidential data to be disclosed to other organisations by the HSCIC. In this case, the GP practice must ensure that the objection is coded in the GP record using specified Read codes so that an objection "flag" is transmitted to the HSCIC in any subsequent collections. This objection "flag" can be extracted by the HSCIC when determining whether an individual's personal confidential data will be disclosed by the HSCIC depending on the legal basis for the disclosure. Again, as indicated above, the objection will be respected in all but exceptional circumstances; however the data will be used and made available in aggregate and/or pseudonymised form.

2. Benefits and Risks

In the autumn of 2012, a group of GPs and local commissioning representatives considered the anticipated benefits and the potential risks associated with the proposed GP extract submitted into CES.

Benefits

The consultation participants expected the care.data GP extract to deliver the following benefits to communities, patients and primary care clinicians.

For communities:

1. Communities have a right to expect their health and social care services to be designed according to a comprehensive assessment of the needs of the local population. Such "needs assessments" should incorporate information about the health priorities of entire communities, not just 'sickness priorities'. Therefore, information from GP practice systems is vital in providing a whole picture of health needs and of health care provision. Indeed, the majority of patient consultations occur within primary care and general practice systems hold rich data about health needs of a population.
2. For communities to be more active in planning, redesigning and prioritising local care services, they need information that reflects the end-to-end process of care (i.e., from the time of first presentation with a symptom through to the end of treatment).
3. Linked data are a prerequisite of integrated care. As more services are provided in the community, so the need for primary care data will become increasingly important.

⁶ Launch of Caldicott2 Review on 26 April 2013

For patients:

4. Patients have a right to know that any unexplained variations in how care is provided will be identified and addressed swiftly.
5. Patients should be confident that services are planned in a coordinated way, and that planning should be centred on patient pathways rather than on an organisation-by-organisation basis.
6. Patients will expect clinical outcomes to be measured in transparent ways, based on an appropriate set of data.

For GPs and other primary care clinicians:

7. The consultation respondents felt that there would be major benefits to GPs arising from the linkage of primary and secondary care data. In particular, they felt that analyses of the patterns of care received by patients with long-term conditions would lead to significant improvements in care.

The major domains of benefits to GPs are set out in Table 1, together with associated examples.

Table 1: Anticipated benefits arising from the GP extract

Anticipated Benefit		Examples
1)	Improved monitoring of outcomes through linkage between primary and secondary care	<ul style="list-style-type: none"> • What is the route to diagnosis? (e.g., routine, urgent, appointments, A&E, screening) • What is the route from diagnosis? (e.g., bed utilisation, complications, rehabilitation, etc.)
2)	Improved monitoring of performance through linkage between primary and secondary care	<ul style="list-style-type: none"> • Assess the types of patient who are attending A&E either frequently or with complaints that could be seen in the practice
3)	Earlier diagnosis of illness	<ul style="list-style-type: none"> • Identify the number of patients recorded on QOF disease registers compared with expected prevalence
4)	Improving the contribution of primary care to wider CCG outcomes	<ul style="list-style-type: none"> • What are the outcomes at an individual practice level for measures in the CCG Outcomes Indicator Set?
5)	Improved data quality	<ul style="list-style-type: none"> • Ethnic category, referrals in GP records
6)	Monitoring and understanding trends	<ul style="list-style-type: none"> • Trends in A&E attendances and unplanned admissions and readmissions
7)	Predictive modelling	<ul style="list-style-type: none"> • Models such as PARR, PARR-30
8)	Evaluation of preventive services and interventions	<ul style="list-style-type: none"> • Propensity-matched and prognostic-matched controls
9)	Exploring patient pathways	<ul style="list-style-type: none"> • Health and social care timelines • End-of-life pathways
10)	Detecting unwarranted variation	<ul style="list-style-type: none"> • Analyses of variation

All of the consultation participants acknowledged that the information requirements for CCGs were still emerging, and therefore that more time was needed to understand more fully the potential benefits of using linked data. Accordingly, the GP extract specification set out in this document represents the first iteration of the requirement specified by NHS England.

It is important to note that this first release constrains the uses to which the data may be put while commissioners and providers identify their needs more clearly, and while work progresses on improving data quality. In particular, under this first release, the outputs of CES will be provided in aggregated or pseudonymous forms only. In future, patients might be able to consent to having their linked, person-identifiable data viewed by themselves, their carers, their clinicians or researchers; however, in this initial release, no person-identifiable data will be released by the HSCIC.

Risks and Risk Mitigation

The HSCIC will develop and maintain the data extracted from GP systems, which will be refreshed monthly. The HSCIC has highly secure systems in place for storing and processing such data, systems that are fully compliant with latest UK and European standards. The security of these systems is the most important priority of the HSCIC, as is reflected in its accountabilities and statutory responsibilities.

Decisions on access to CES data are the responsibility of the HSCIC. As with HES data, *aggregated* CES data will be published openly; however, *record-level* CES data will only be made available pseudonymously to nominated users under the terms of a legally binding agreement. Each user will be required to sign a data sharing and re-use agreement, which specifies what data are to be made available to the user, and the purpose for using the data extract.

By providing such accredited users with extracts of the pseudonymous CES file, there is a risk of malicious re-identification of patients from inference (a so-called “jigsaw attack”). This risk will need to be taken into account when the HSCIC determines what views of the data should be made available to different accredited users. A thorough assessment of re-identification risk will be conducted in response to all requests to access CES data. These assessments will ensure that any data are released in accordance with the principles set out in *Anonymisation: managing data protection risk code of practice* (Information Commissioners’ Office, 2012) and the *Anonymisation Standard for Publishing Health and Social Care Data* (approved by the Information Standards Board) where applicable.

3. Information Requirements

This chapter specifies the data that GP practices will be required to submit to the HSCIC for inclusion in the care.data service.

Patient Cohort

The cohort consists of all patients registered at a general practice in England, excluding those individuals who have objected to their personal confidential data being extracted from GP systems.

The major business objectives for NHS England are to support CCGs in improving outcomes and to support improvements in outcomes for directly commissioned services, including primary care and specialised services. In order to achieve these objectives, data are needed for the entire patient population (although recognising that 100 per cent data capture is not feasible for a number of reasons, including instances where a patient has objected). The selection criteria described in the following section have been adopted to ensure that only those data items needed to support these commissioning purposes are extracted.

Data Content

The data to be extracted are detailed in Appendix A1. They are comprised of four blocks of data, namely:

- Demographics (including identifiers)
- Clinical Data
- Referrals
- NHS Prescriptions.

NHS England does not propose extracting any free text: only coded information, such as Read codes, will be extracted.

Selection Criteria

The criteria specified below are based on current specifications, such as the Quality and Outcomes Framework (QOF) codes and long-term conditions (LTC) codes. For Release 1, the following three-way classification of event types has been applied:

- 1) Entries that are needed to support primary and/or secondary care commissioning and which are likely to be of sufficient quality for immediate use
- 2) Entries that are likely to be useful for commissioning but are considered to be of variable quality and completeness. The aim is to use these data items to drive improvements in recording, which in turn should promote their use in commissioning activities. These data can currently be used for:
 - Analysing the GP extract itself (e.g., determining the prevalence of data, the consistency of coding, etc.)
 - Analysing linked data, which would allow for an assessment of quality across patient pathways

A data quality review will be established by the HSCIC to facilitate the analysis of data, considering aspects such as completeness, validity, and consistency. Joint work with the RCGP will enable this analysis to contribute to the development of record-keeping standards and training support.

- 3) Those entries that are recognised as not being needed for the purpose of commissioning, (e.g. administrative codes for recording items such as “visit to solicitor”). Such entries are not specified in this document and therefore will not be extracted.

Only those events recorded post-April 2013 will be collected. Collection will be on a monthly basis, selecting new events that meet the criteria.

In addition, supporting meta-data will need to be extracted, together with a record of attribution and the provenance of the data. Clinical data entries containing so-called “sensitive” codes will not be extracted (see Appendix B). However, we are mindful that the existence of such lists risks perpetuating the stigma associated with the conditions that are listed. There have been calls by certain patient groups to end this type of special treatment. NHS England therefore plans to work with the RCGP, the BMA and other groups to address this issue in due course.

The list of relevant Read / CTV3 codes was developed by the clinical informatics team at the HSCIC. The JGPITC of the RCGP and the BMA reviewed the codes, and the details of their review were communicated to the Independent Advisory Group for the General Practice Extraction Service (GPES), to support their consideration of whether to recommend the extraction to proceed.

Use of the data

A data reporting workstream will be established with GP and CCG input to consider which data outputs will be most useful for commissioners. In addition, NHS England plans to engage with patients and the public on the types and formats of information that will be provided. For this first release, outputs will not be identifiable. Aggregated outputs will be published through standard aggregate reports, and record-level extracts will be made available to commissioners and providers in pseudonymous form (pseudonymisation is the process of distinguishing individuals in a dataset by using a unique identifier that does not reveal their ‘real world’ identity).

4. Data Management

This chapter describes the processing arrangements and the proposed outputs from Release 1. In the diagram below,

- **Green** flows of data will be published in aggregated form, with small number suppression, in line with guidance published by the Information Commissioner's Office (ICO) [Anonymisation: managing data protection risk code of practice, ICO, 2012];
- **Amber** data will only be disclosed by the HSCIC to specific users such as NHS commissioners and providers. Although amber data do not include identifiers, there is a potential risk of malicious re-identification through so-called 'jigsaw attack'. In view of this risk, such data will only be disclosed in a controlled environment where robust safeguards are in place, as set out in the ICO guidance and the Information Governance Review (Caldicott2). All such disclosures will be on the basis of legally-binding agreements, and will be either in the form of aggregate reports and/or pseudonymous, record-level data;
- There will be no identifiable (**Red**) disclosures by the HSCIC in relation to care.data in Release 1

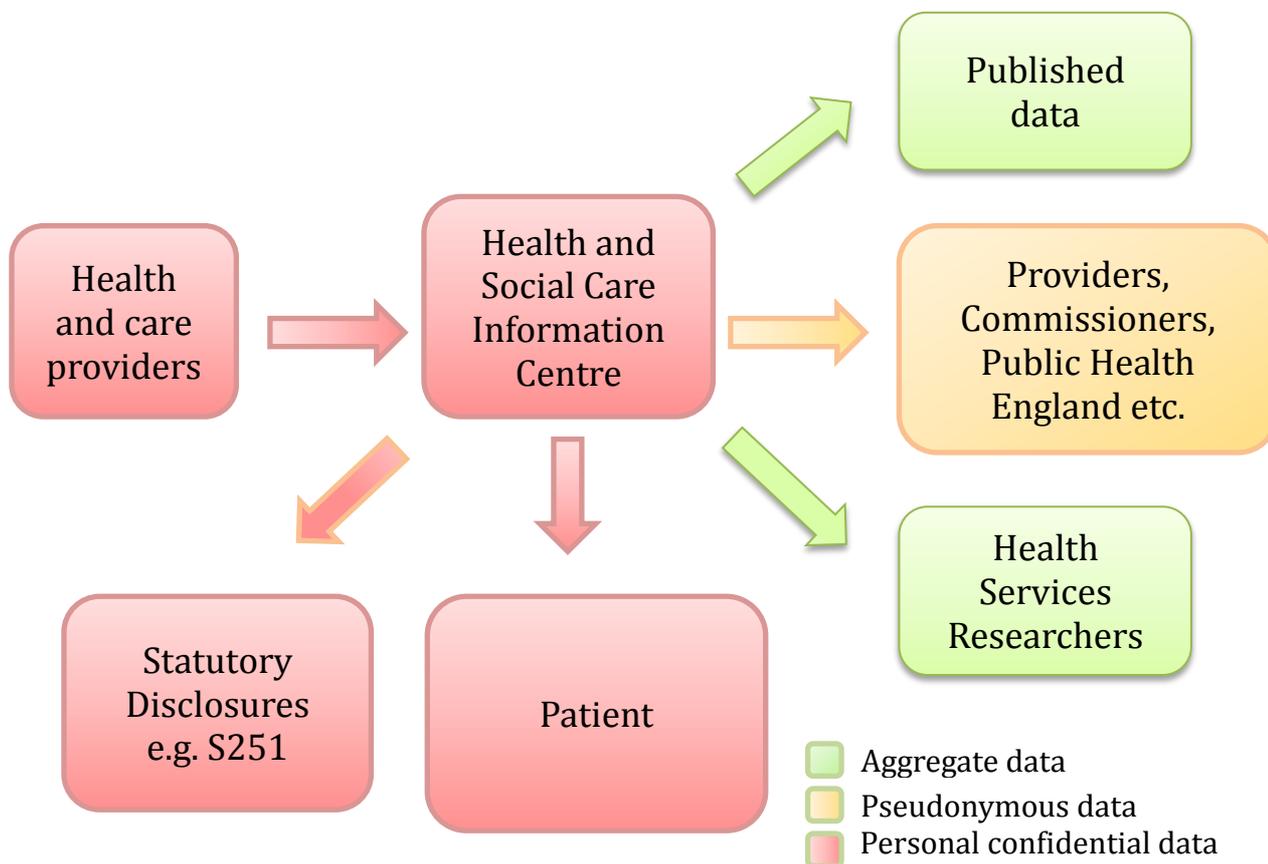


Figure showing care.data input and output feeds for Release 1.

NHS England anticipates that some proposals for the disclosure of identifiable data will be included in future releases (indicated in red in the diagram); however, such proposals would be:

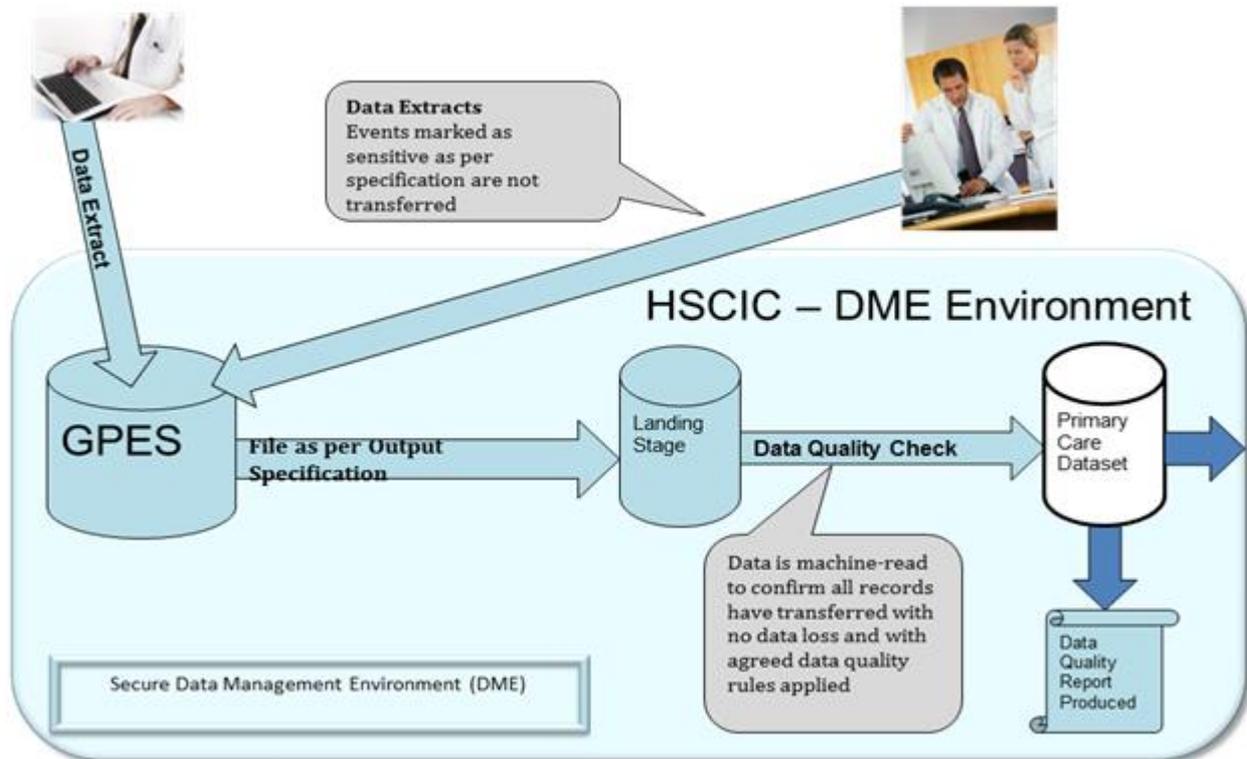
- (a) only where there is a legal basis (e.g., on the basis of patient consent or under the section 251 regulations,³⁷ and approved by the Secretary of State or by the Health Research Authority on the recommendation of the Confidentiality Advisory Group⁸); and
- (b) in line with recommendations from the Information Governance Review (Caldicott2); and
- (c) in accordance with the NHS Constitution; and
- (d) as discussed and agreed with relevant stakeholders.

⁷ The Health Service (Control of Patient Information) regulations 2002 SI1438

⁸ <http://www.hra.nhs.uk/hra-confidentiality-advisory-group/>

Processing of data

The data extracted from GP systems will be stored in the *Data Management Environment* (DME) within the HSCIC. The GP extract will be gathered in the central GPES server (GPET-Q) and forwarded to a transient HSCIC Data store. Data will be held in the *Transient Data Store* (shown below as the “primary care dataset” in the diagram below) to enable linkage, processing and data quality analysis; they will then be deleted and over-written by the following month’s data.



Data Linkage

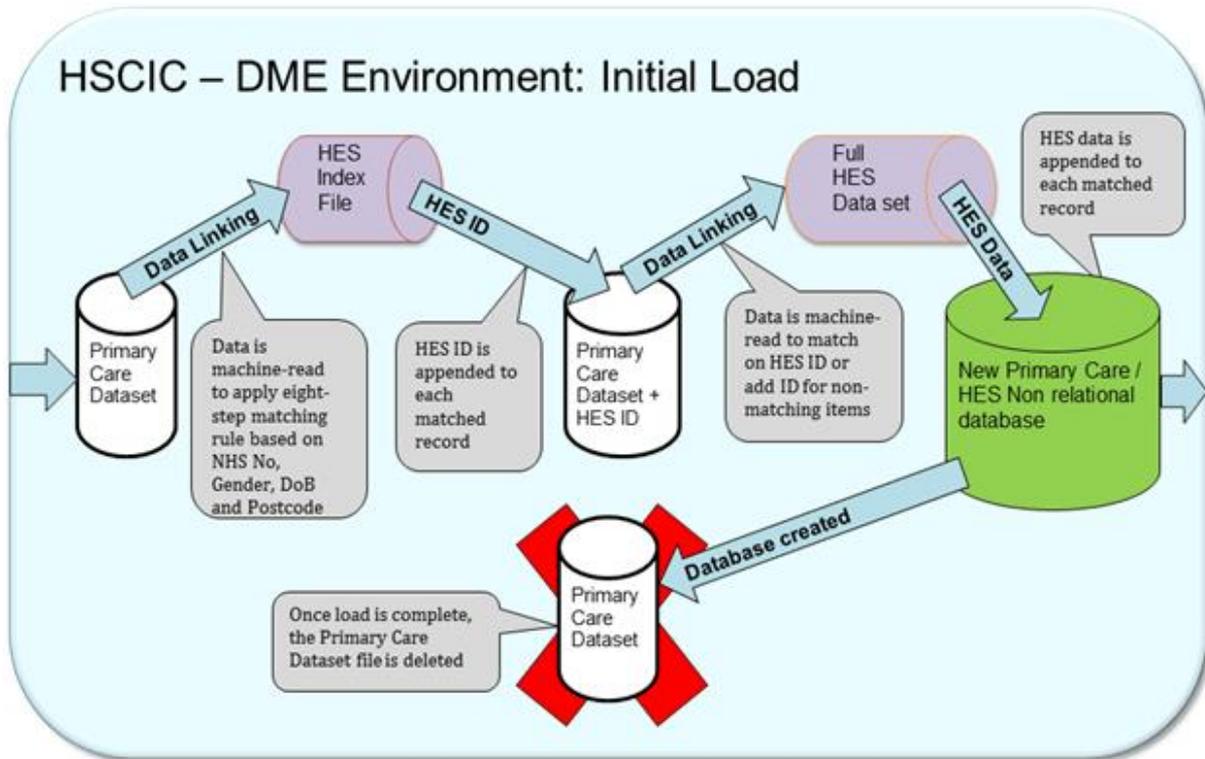
The HSCIC’s *Data Linkage Service* will establish data linkages with secondary care data. Linkage will take place at a patient level using the following four patient identifiers: NHS number, date of birth, gender and postcode. These four data-matching items will be sufficient to provide a very good matching rate. Based on current experience, a rate of above 90 per cent is expected initially, with improvements in the matching rate expected to follow.

An automated process will then take place to confirm that all records have transferred as expected, with no data corruption. Note that this entire process is performed by a machine (i.e., no humans will view the person-identifiable data). The primary care data set will then be machine-read to apply an 8-step matching rule, based on patient identifiers, against the HES index file.

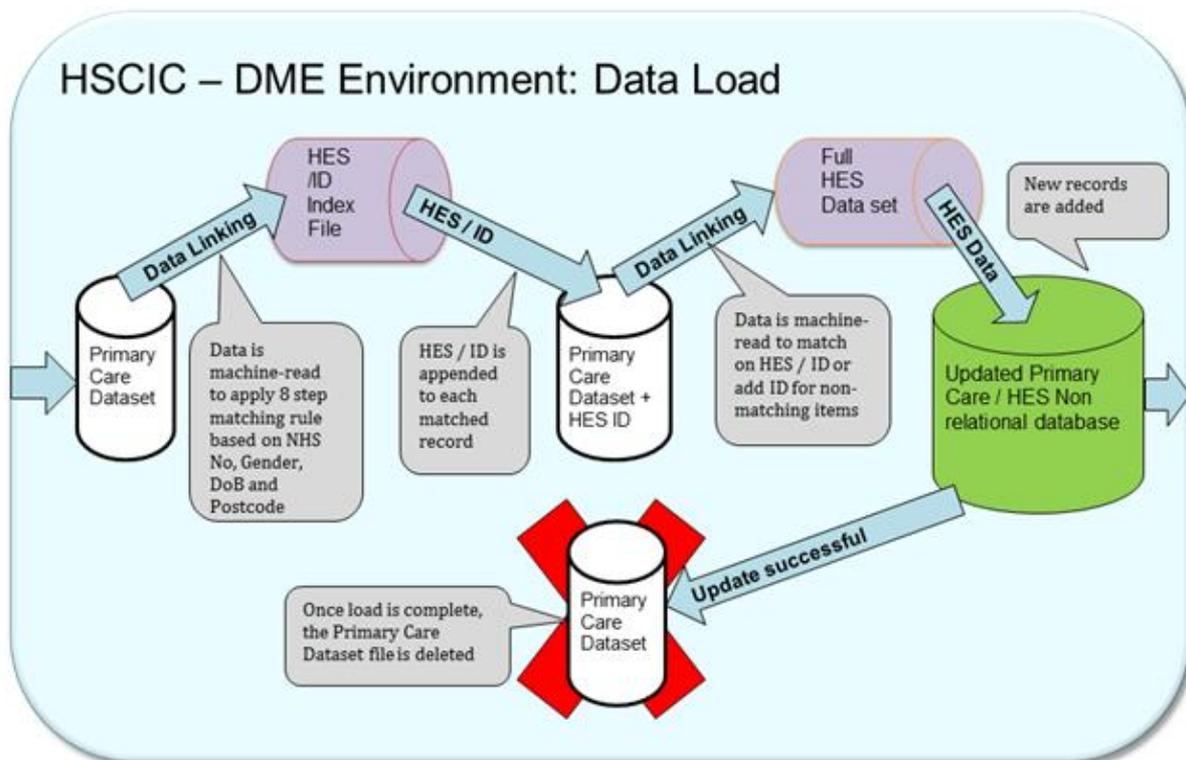
For those individuals with a match, the HES ID will be appended to the primary care dataset, and the updated primary care dataset will then be machine read to confirm the match rate and to determine that the matching has been successful. For each record, a machine process will then match the HES ID to the HES dataset to form the linked primary and secondary care database (CES). For those records without a match, a new ID or pseudonym will be created and the data will be retained for subsequent matching. This is because, for commissioning purposes, it is important to understand activity undertaken (or indeed not undertaken) in all

care settings. The “delta load” approach (by which only new events are uploaded) requires such data to be retained, to enable subsequent linkage.

The initial load process is shown in the figure below.



There will be a monthly “delta load” of new entries, for which the subsequent load process will then be as follows:



As described above, once the linking process has been completed, the data in the “primary care dataset” will be deleted.

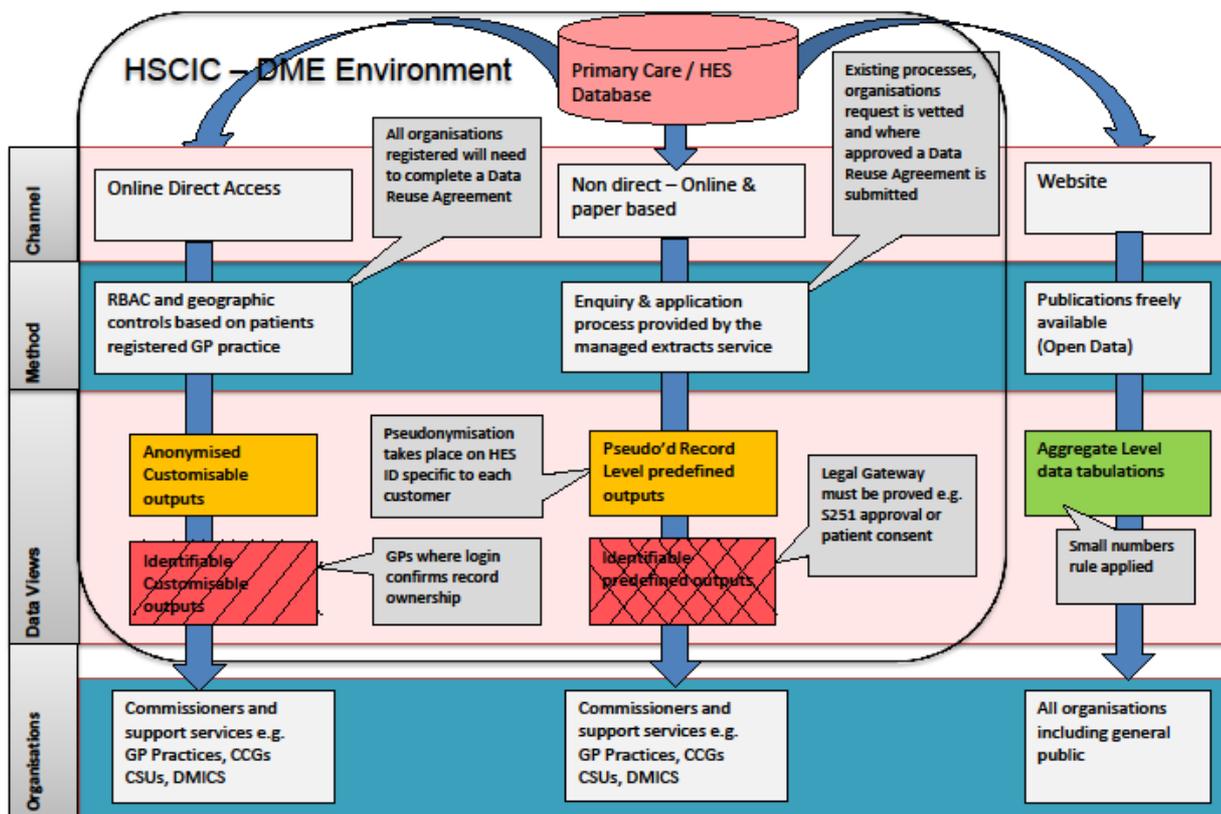
Outputs from DME

A set of standard data extract views will be created. All requests for extracts will be processed through the *Managed Extracts Service* as part of the *HSCIC Data Linkage Team*, with the exception of a potential online access route available only to commissioners and the organisations supporting the commissioners, which will have its access governed by role-based rules to ensure that there is no risk of re-identification.

Extracts will be provided in accordance with agreed information governance processes. Where a request for an extract is authorised, the nominated user will receive a copy of the extract, with the CES IDs being specific to the accredited user making that request.

The outputs are illustrated below. Commissioners will be able to receive outputs within a controlled environment in aggregate or pseudonymised form. Other users (e.g. researchers) will only have access to aggregated information under Release 1. Each nominated user will sign a data sharing agreement, which sets out the rules and conditions under which the data can be used. Although there exists the potential for identifiable outputs from the HSCIC in future releases, this will only ever occur where a legal basis exists (such as explicit patient consent or under section 251 of the NHS Act 2006 and its supporting regulations), and in line with the Caldicott2 review of information governance and the NHS Constitution.

Release 2 will further consider the outputs to be provided from care.data to each of the receiving types of organisation.



Appendix A: List of fields and selection criteria from “Everyone Counts”

The list in Appendix A1 sets out the data items to be extracted in Release 1. This list was based on local specifications, such as those used in Liverpool and North East London. The JGPITC of the RCGP and BMA reviewed the specific codes. On 27 March 2013, the Independent Advisory Group (IAG) of the GP Extraction Service (GPES) recommended that Release 1 should proceed to extraction.

The list of codes can be downloaded from the GPES website: <http://www.hscic.gov.uk/gpes>

Appendix A1: List of fields

Patient data

- *NHS Number*
- *Date of Birth*
- *Gender*
- *PostCode*
- *EthnicityCode*
- *Registration Status*
- *RegistrationDate*
- *DeRegistrationDate*
- *Date of Death*

Event data

- *Date of event*
- *Read code*
- *Rubric*
- *Value associated with the read code*
- *Further value*
- *Clinician (HCP – GP or other Healthcare professional responsible)*

Referral data (the fields shown in red will not be available through GPES in Release 1)

- *ReasonForThisReferral*
- *StatusOfThisReferral*
- *DateOfThisReferral*
- *DateOfResponseForThisReferral*
- *ReferralDischargeDate*
- *ReferrerName*
- *SourceForThisReferral*
- *ReferralUrgency*
- *Clinician*

Prescriptions

- *Patient_Id*
- *Medication_Readv2_Code*
- *Medication_Readv3_Code*
- *Medication_Name*
- *Medication_Dosage*

- *Medication_Quantity*
- *Medication_Start*
- *Medication_Expiry*
- *Repeat_Prescription*
- *Prescriber_Id*
- *Organisation_Id*

Appendix A2: Selection criteria identifying the types of condition / diagnosis:

Diagnosis Group

- *Diabetes (most recent & date)*
- *Retinopathy (most recent & date)*
- *CKD (most recent & date)*
- *Hypertension (most recent & date)*
- *COPD (latest of COPD codes mild, mod, severe & Date)*
- *Asthma (most recent & date)*
- *CHD (most recent & date)*
- *Stroke/TIA (most recent & date)*
- *PVD (most recent & date)*
- *MI (most recent & date)*
- *Angina (most recent & date)*
- *HF (most recent & date)*
- *AF (most recent & date)*
- *Chronic Liver disease (most recent & date)*
- *Bladder Cancer (most recent & date)*
- *Breast Cancer (most recent & date)*
- *Cervical Cancer (most recent & date)*
- *Bowel Cancer (most recent & date)*
- *Prostate Cancer (most recent & date)*
- *Skin Cancer (most recent & date)*
- *Other cancer (most recent & date)*
- *Schizophrenia (most recent & date)*
- *Bipolar Affective disorder (most recent & date)*
- *Other psychoses (most recent & date)*
- *Dementia (most recent & date)*
- *Depression (most recent & date)*
- *Anxiety (most recent & date)*
- *Palliative care register (most recent & date)*
- *Supportive care register (most recent & date)*
- *CHD (flu 1)*
- *COPD (flu 2)*
- *CKD (flu 3)*
- *Chronic Liver disease (flu 4)*
- *Diabetes (flu 5)*
- *Immunosuppression (flu 6)*
- *Stroke/TIA (flu 7)*
- *Patients with Chronic Degenerative Neurological Disease, Cerebral Palsy or MS (flu 8)*
- *Patient pregnant (flu 9)*

- Pulmonary embolism

Health group

- Cholesterol (value) (most recent & date)
- HDL Cholesterol (value) (most recent & date)
- LDL cholesterol (value) (most recent & date)
- BP diastolic (value) (most recent & date)
- BP systolic (value) (most recent & date)
- BMI (value) (most recent & date)
- Hba1c (value) (most recent & date)
- eGFR (value) (most recent & date)
- FEV1/FVC (value) (most recent & date)
- Microalbuminuria (code) (most recent & date)
- Microalbuminuria (value) (most recent & date)
- ACR (value) (most recent & date)
- Proteinuria (code) (most recent & date)
- Smoking Status (most recent & date)
- Alcohol consumption (most recent & date)
- Spirometry (or declined contraindicated) (code) (most recent & date)
- Spirometry (value) (most recent & date)
- BS/Framingham/Q-Risk (code) (most recent & date)
- BS/Framingham/Q-Risk (value) (most recent & date)
- Pulse rate O/E(most recent & date)
- Family history (most recent & date)

Review/interventions group

- Flu vaccination (most recent & date)
- 1st Dtap/Hib vaccinations (most recent & date)
- 2nd Dtap/Hib vaccinations (most recent & date)
- 3rd Dtap/Hib vaccinations (most recent & date)
- MMR 1st dose (most recent & date)
- MMR Booster (most recent & date)
- HPV Vaccinations (most recent & date)
- Mental Health Review (most recent & date)
- Mental Health Care plan (most recent & date)
- Dementia Review (most recent & date)
- Depression Review (most recent & date)
- Depression screening (most recent & date)
- Bowel Cancer Screening (latest of bowel cancer screening declined) (most recent & date)
- Diabetes Review (most recent & date)
- Diabetes care setting (most recent & date)
- Referred to dietician (most recent & date)
- Exercise advice (most recent & date)
- Dietary advice (most recent & date)
- Diabetic foot review (most recent & date)
- Diabetic neuropathy testing (most recent & date)

- *Smoking cessation advice (most recent & date)*
- *Alcohol consumption advice (most recent & date)*
- *CHD review (most recent & date)*
- *Medication review (most recent & date)*
- *Referred to pulmonary Rehabilitation (most recent & date)*
- *Commenced pulmonary rehab (most recent & date)*
- *Completed pulmonary rehab (most recent & date)*
- *Self-management plan (most recent & date)*
- *NHS Health checks (most recent & date)*

Exceptions group

- *Flu vac contraindicated/declined (most recent & date)*
- *DNA bowel Screening (most recent & date)*
- *CHD Exception reporting (most recent & date)*
- *Stroke/TIA Exception reporting (most recent & date)*
- *Hypertension Exception reporting (most recent & date)*
- *Diabetes Exception reporting (most recent & date)*
- *COPD Exception reporting (most recent & date)*
- *Mental Health Exception reporting (most recent & date)*
- *CKD Exception reporting (most recent & date)*
- *Asthma Exception reporting (most recent & date)*
- *CVD Risk assessment declined (most recent & date)*
- *Cholesterol exceptions (most recent & date)*
- *BP exceptions (most recent & date)*
- *HbA1c exceptions (most recent & date)*
- *DNA cardiac clinic (most recent & date)*
- *DNA NHS Health Checks (most recent & date)*
- *MMR contraindicated/declined (most recent & date)*
- *HPV vaccination declined (most recent & date)*

Appendix B: List of Sensitive Codes

Clinical data entries containing the “sensitive” codes listed below will be de-selected. However, NHS England is mindful that the existence of such lists risks perpetuating the stigma associated with the conditions that are listed. There have been calls by certain patient groups to end this type of special treatment. NHS England therefore plans to work with the RCGP, the BMA and other groups to consider this complex issue in due course.

HIV & AIDS:
13N5. or 43C% or 43WK. or 43d5. or 43h2. or 43W7. or 43W8. or 4J34. or 62b.. or 65P8. or 65QA. or 65VE. Or 67I2. or 6827 or 8CAE. or A788% or A789% or AyuC4 or Eu024 or R109. or ZV018 or ZV019 or ZV01A or ZV19B or ZV6D4 or ZV737
Sexually transmitted infections:
1415 or 43U% or A9% or A780. or A78A. or A78A3 or A78AW or A78AX or 65P7. or 65Q9. or 6832 or A7812 or L172% or ZV016 or ZV028 or ZV745 or EGTON34
Termination of pregnancy:
1543% or 6776 or 7E066 or 7E070 or 7E071 or 7E084 or 7E085 or 7E086 or 8M6.. or 956% or 9Ea% or 8H7W. or L05% or L06%
IVF treatment:
ZV26% or 8C8% or 7E0A% or 7E1F2
Marital Status: 133%
Complaints: 9U%
Convictions and imprisonment:
13H9. or 13HQ. or 13I71 or 6992 or T776. or ZV4J4 or ZV4J5 or ZV625
Abuse (physical, psychological or sexual, by others):
14X. or 1J3. or SN55. or SN571 or TL7. or TLx4. or ZV19C or ZV19D or ZV19E or ZV19F or ZV19G or ZV19H or ZV19J or ZV19K or ZV4F9 or ZV4G4 or ZV4G5 or ZV612

Appendix C: Glossary of Terms

Term	Definition
Aggregated	Statistical data about several individuals that have been combined to show general trends or values without identifying individuals within the data.
Anonymised	Data in a form that does not identify individuals and where there is no reasonable chance that the recipient could infer identities from that data.
Controlled environment	Where access to data is restricted and the data are not released into the public domain
Data controller	An individual or an organisation that – either alone or jointly – determines the purposes for which and the manner in which any personal data are, or are to be, processed. Data controllers must ensure that any processing of personal data for which they are responsible complies with the Data Protection Act (DPA)
Identifiable	See Personal confidential data. The term ‘identifiable data’ is used as a specific, grammatically constrained, statement to describe data about identified or identifiable individuals. ‘Personal’ includes the DPA definition of personal data but it is expanded to include deceased as well as living people, and ‘confidential’ includes both “given in confidence” and “that which is owed a duty of confidence” and is adapted to include “sensitive” as defined in the Data Protection Act
Identifier	An item of data, which by itself or in combination with other identifiers enables an individual to be identified.
Personal data	Data that relate to a living individual who can be identified from those data, or from those data and other information that is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.
Processing	Processing in relation to information or data, means obtaining, recording or holding the information or data or carrying out any operation or set of operations on the information or data, including: (a) organisation, adaptation or alteration of the information or data, (b) retrieval, consultation or use of the information or data, (c) disclosure of the information or data by transmission, dissemination or otherwise making available, or (d) alignment, combination, blocking, erasure or destruction of the information or data.
Pseudonymised	The process of distinguishing individuals in a dataset by using a unique identifier that does not reveal their ‘real world’ identity.
Re-identification	The process of analysing data or combining them with other data with the result that individuals become identifiable. Sometimes termed ‘de-anonymisation’ as a synonym.

© Crown copyright 2013
First published May 2013

.