nuffieldtrust

DPIA – Processing HES / CSDS Data for Quantitative Research

26th April 2022

Introduction

Completed By	Sarah Scobie / Tony Harbon	**Date**	**7/03/2022**
Department	Research / DPO		

Project Title:	Processing HES Data For Research

Step 1: Explain the project in broad terms

1.1. What are the aims and objectives of the data processing? - Why was the need for a DPIA identified?

The Nuffield Trust undertakes research and analysis using Hospital Episode Statistics (HES) and Community Services (CSDS) data. There are a number of common ways in which the data is used:

• Health care activity data such as emergency department attendances, admissions and re-admissions, are important (though imperfect) proxies for health outcomes. Tracking changes in these events over time enables analysis of the impact of changes in health services.

• Detailed analysis of particular health events can be used to identify particular issues with quality of care, for example as part of the harm project.

• Variation between hospitals or areas in use of services can identify populations where there are gaps in care, and also areas delivering high quality care from which the NHS can learn more widely.

• Evaluation of healthcare innovations can be made more robust by using matched case-control analysis – comparing outcomes or trends in a service being evaluated with similar patients elsewhere.

Processing includes:

• Descriptive analysis of patterns and trends in hospital activity

• Development of outcome measures or risk factors from hospital data

• Comparison of outcomes and risks over time and between cohorts of patients

A DPIA is required because the HES and Community Services data being processed is classed as Sensitive under the GDPR and UK DPA.

Step 2: Describe the processing

2.1. Describe the information flow: how will you collect, use, store and delete data? What is the source of the data? Will you be sharing data with anyone? You might find it useful to refer to a flow diagram or another way of describing data flows. What types of processing identified as likely high risk are involved?
NHS Digital disseminates monthly data extracts of pseudonymised HES Emergency Care, OP, and APC data, as well as quarterly CSDS data to the Nuffield Trust by Secure Electronic File Transfer. The data will only be processed by individuals who: i) are approved researchers working under a substantive contract at the Nuffield Trust or by processors. All research staff are subject to confidentiality requirements to access data to support business objectives and required to complete mandatory data security training annually or. ii) are employed by Nuffield Trust as specialist third party consultants. Where these consultants help to determine how the data should be processed an application will be made to have them added as joint controllers. Whilst the nature of detailed analysis in relation to each project varies, the broad context of processing is consistent. In summary :- • The data is downloaded from NHS Digital to the Trust's Research Server. The server is held on-site, and access is restricted to named individuals according to The Nuffield Trust's security policy using Microsoft Role Based Access Control (RBAC). • The data is held within separate folders within the server. • Remote access to the database is permitted, but only through Citrix via Multi-Factor Authentication (MFA) (so processing is still carried out on site), and with local printing and downloading disabled. • Only staff who have signed a confidentiality agreement and have received IG training are permitted access. • All access to individual files is recorded, and a sample audited to investigate the existence of any adverse incidents, and ensure that appropriate access has been maintained. • The researcher will view the data and select a specific cohort for each individual study. Commonly a process will initially take place to define the particular cohort of interest in terms of e.g. individual diagnostic codes or procedure codes. The researchers will use routinely available filter definitions where possible, but may amend these based on the nature of each study's group of interest. Depending on the research a similar control group may be established. • The individual researcher then analyses the data, before applying the relevant disclosure controls to any output. Software used will be SAS, R and Stata; typically this will involve analysis on several outcome measures, risk adjustment and the construction of control groups. • No record level data would be linked to this dataset, but it may be combined with publicly available demographic or geographic data, for example in relation to local Trust performance • Outputs consist of aggregate data (or indicator/statistical data) only. Data used for research is securely erased using government accredited software when it is no longer required.

2.2. Describe the scope of the processing: what is the nature of the data, and does it include special category or criminal offence data? How much data will you be collecting and using? How often? How long will you keep it? How many individuals are affected? What geographical area does it cover?

HES data includes the following types of information

- Demographic characteristics of patients
- Service characteristics, eg hospital of treatment
- Treatment received (eg procedures) and medical conditions (eg diagnoses)
- Information about the patient's pathway eg dates of admission, discharge and appointment, mode or arrival and discharge

As such, HES data includes sensitive information.

There is no criminal offence data.

As above, information is received monthly from NHS digital, and retained in accordance with the DSA with NHS Digital.

The HES data set covers England – volumes of data are approximately as follows:

- Outpatient attendances/year – 120 million
- A&E attendances/year – 22 million
- Inpatient episodes/year – 17 million

The Community Services Dataset (CSDS) is a person-level data set which covers publicly funded community services delivered to all ages in England. Services could include health visiting, end of life care, intermediate care, arts therapy, hearing, podiatry speech and language therapy, and much more. These services could take place in settings such as health centres, Sure Start centres, day care facilities, schools or community centres, mobile facilities, or a patient's own home.

The data set includes information on: demographics, social and personal circumstances, breastfeeding and nutrition, care events and screening activity, diagnoses, including long-term conditions and disabilities, and results of any scored assessments carried out.
As such, CSDS includes sensitive information. Information is received quarterly from NHS Digital and retained in accordance with the DSA with NHS Digital.

The volumes of data covered by the CSDS are approximately as follows:
- Referrals/month: 1.3 million
- Care contacts/month: 8 million

Community Services data is disseminated quarterly under a data sharing agreement with NHS Digital.

2.3. Describe the context of the processing: what is the nature of your relationship with the individuals? How much control will they have? Would they expect you to use their data in this way? Do they include children or other vulnerable groups? Are there prior concerns over this type of processing or security flaws? Is it novel in any way? What is the current state of technology in this area? Are there any current issues of public concern that you should factor in? Are you signed up to any approved code of conduct or certification scheme (once any have been approved)?

As secondary users of NHS data, patients will generally not be aware of the specific analysis undertaken by the Nuffield Trust. Individuals have the ability to opt-out of data sharing. The ability to share data in this way is set out in the Health and Social Care Act.

The data set covers all age groups including children.

There are no specific prior concerns over the processing or security flaws.

The analysis will use established analysis methods for large data sets (eg regression).

Current state of technology is not applicable to this work.

There are no specific issues of public concern.

The Nuffield Trust operates an ISMS which conforms to the ISO 270001 standard for processing this information.

Clear requirements for all organisations engaging in Healthcare research are set out by the Healthcare Research Authority.

2.4. Describe the purpose of the processing: what do you want to achieve? What is the intended effect on individuals? What are the benefits of the processing for you, and more broadly?

The Nuffield Trust aims to improve the quality of health care in the UK by providing evidence-based research and policy analysis and informing and generating debate. The analysis of HES and CSDS data enables this to be achieved in a number of ways:

• Tracking changes in outcomes and activity over time provides evidence about the impact of changes in health services, which can be used to assess the effectiveness of health policies, and identify improvements in health policies or how they are implemented. These improvements have the potential to improve the quality of care for patients, and ensure best use is made of public money.

• Detailed analysis of particular health events can be used to identify particular issues with quality of care, for example harms to patients. This provides information to support NHS organisations with improving care.

• Variation between hospitals or areas in use of services can identify populations where there are gaps in care, and also areas delivering high quality care from which the NHS can learn more widely.

• Evaluation of healthcare innovations can be made more robust by using matched case-control analysis – comparing outcomes or trends in a service being evaluated with similar patients elsewhere. This enables us to understand which interventions are effective in improving health services, and which should be adopted more widely (or stopped).

In all such work, The Nuffield Trust analyses patterns of hospital activity by area, by year, by condition or by provider, developing comparative analyses and standardising for a range of episode level, or patient level variables – such as age, the presence of a long terms condition, prior patterns of use. The analyses commonly follow the health and care of a well-defined cohort of individuals over a lengthy period of time. Such analyses require complex processing for fair comparisons and to capture activity for whole populations – something that only nationally collated data can provide.

Step 3: Consultation process

3.1. Consider how to consult with relevant stakeholders: describe when and how you will seek individuals' views – or justify why it's not appropriate to do so. Who else do you need to involve within your organisation? Do you plan to consult IT (eg to set up Secure File Transfer), or IG (to draw up data sharing agreements)?
The Nuffield Trust draws on patient and public stakeholders at an individual project level, ensuring that our research questions, the evidence we apply, and the recommendations we make take proper account of the experience and needs of the people who use services. At a strategic level we are establishing a partnership with two or more organisations which represent the voice of patients and the public, to provided challenge and support as we set our medium to long term work programme, ensuring that we are considering issues which are truly relevant to those who use services.

Step 4: Assess necessity and proportionality

4.1. Describe compliance and proportionality measures, in particular: what is your lawful basis for processing? Does the processing actually achieve your purpose? Is there another way to achieve the same outcome? How will you prevent function creep? How will you ensure data quality and data minimisation? What information will you give individuals? How will you help to support their rights? What measures do you take to ensure processors comply? How do you safeguard any international transfers?
We are satisfied that the interests of the data subjects do not override our legitimate interests; that they would reasonably expect the processing and it would not cause unjustified harm. The data subjects interests and fundamental rights are protected through appropriate minimisation of fields and patient records being processed; pseudonymisation to minimise any risk of identifying individuals; protection of the data in a secure environment, and guaranteeing secure destruction at any stage at the request of NHS Digital or after a defined period on completion of the project. Nuffield Trust's lawful basis of processing under the GDPR is Article 9(2)(j) processing for scientific research purposes - as its condition for processing the special category data of the participants. It ensures it has addressed the safeguards set out in Article 89(1) of the UK GDPR and in section 19 of the DPA 2018. Due to the nature of the work we do, our analysis requires data covering the whole population, in order to provide evidence at a national level, and to enable complex analysis requiring comparisons of sub-groups within the population. An analysis plan is prepared for each project, setting out the data requirements and methods. In each case, the use of the HES / CSDS data is necessary and proportionate to the purpose of the project and that the minimum amount of data necessary is used - this will include consideration of the necessity for use of each individual HES dataset; the number of years of data; the sizes of any cohorts or control cohorts derived from the data, and the inclusion and exclusion criteria (such as presence of specific diagnostic or procedure codes). Appropriate safeguards are in place to protect confidentiality; minimise risks of re-identification and use of excessive data beyond necessity. This is described more fully in our DSA with NHS Digital. The Nuffield Trust website hosts project pages outlining the types of analysis work we are undertaking, as well as our privacy notice which describes the types of data we use. The Nuffield Trust does not share or transfer data internationally.

Step 5: Identify and assess risks

Describe source of risk and nature of potential impact on individuals. Include associated compliance and corporate risks as necessary.	**Likelihood of harm**	**Severity of harm**	**Overall Risk**
1. Users or administrators may deliberately steal or release data from the processing environment.	**Possible**	**Significant**	**Medium**
2. The IT service Third party providers/contractors accidentally or deliberately removes data from the processing environment.	**Possible**	**Significant**	**Medium**
3. A physical intruder could steal equipment related to the processing environment.	**Possible**	**Significant**	**Medium**
4. External (Internet) hackers breach external defences and gain access to the processing environment, resulting in theft or release of data.	**Possible**	**Significant**	**Medium**
5. A hacker within wireless range could breach the wireless networks made available by the Trust, and gain access to data.	**Remote**	**Significant**	**Low**
6. Sensitive Research Data is accidentally transferred on to the Company Drive	**Possible**	**Minimal**	**Low**
7. Unauthorised Access to the Research environment/Data	**Possible**	**Significant**	**Medium**
8. Insufficient security measures applied to the Trust Systems	**Possible**	**Severe**	**High**
9. Firewall allows internet access to research server	**Possible**	**Severe**	**High**
10. There is limited or no management control or oversight	**Possible**	**Severe**	**High**

Step 6: Identify measures to reduce risk

Identify additional measures you could take to reduce or eliminate risks identified as medium or high risk in step 5.				
Risk	**Options to reduce or eliminate risk**	**Effect On Risk**	**Residual Risk**	**Measure Approved**
1	1. Ensure users training is adequate 2. Ensure procedures make accidental loss unlikely 3. Technical export controls (Remote Desktop Protocol (RDP) restrictions) 4. Remote Working Guidelines	**Reduced**	**Low**	**Yes**
2	1. Ensure technical export controls are in place 2. Monitor compliance with Policy and procedures 3. Confidentiality agreement sets out user obligations	**Reduced**	**Low**	**Yes**
3	1. Ensure physical security measures are adequate to prevent intrusion, or detect break-ins quickly. 2. ACL Policy applied 3. Encryption (AES 256 bit) applied to server and backups	**Reduced**	**Low**	**Yes**
4	1. Confirm that boundary controls are effective 2. Limit exposed services 3. Regular penetration testing	**Reduced**	**Low**	**Yes**
5	1. Ensure the wireless configuration is robust 2. Seek independent validation (as part of penetration testing)	**Reduced**	**Low**	**Yes**
6	1. IG Training 2. System and file auditing 3. File filtering through firewall	**Reduced**	**Low**	**Yes**
7	1. Access Control Policy 2. ISMS documentation 3. Physical Security Policy 4. Data Encryption	**Reduced**	**Low**	**Yes**
8	1. ISMS documentation 2. Implementation of ISMS Policies	**Reduced**	**Low**	**Yes**
9	1. Acceptable Use of Internet Policy 2. Patch Management Policy 3. Firewall Configuration Rules 4. Penetration Testing	**Reduced**	**Low**	**Yes**

10	1. ISMS documentation 2. Information Governance Committee (IGC) 3. Leadership Team (LT) 4. IG Training 5. Annual Board Report 6. Regular Key Issues Report 7. IGC Reporting to Leadership Team	**Reduced**	**Low**	**Yes**

Step 7: Sign Off And Record Outcomes

Item	Name / date	Notes
Approved by:	Tony Harbon 2/3/22	Integrate actions back into project plan, with date and responsibility for completion
Residual Risks approved by:	No high risks	If accepting any residual high risk, consult the ICO before going ahead
DPO Advice Provided: Yes	Tony Harbon 2/3/22	DPO should advise on compliance, step 6 measures and whether processing can proceed
Summary of DPO Advice: Processing can proceed subject to compliance with ISMS provisions.		
DPO advice accepted / overruled by:		Reasons must be given if DPO advice is overruled
Comments		