
Our response to ‘Protecting Health and Care Information: A consultation on proposals to introduce new regulations’

Introduction

We, the Health Foundation, the King’s Fund and the Nuffield Trust, are voluntary sector organisations working to provide high-quality evidence in support of health and care services across the UK. Our work has made a direct contribution to thinking on national and local health and care policies. We have an established record in working with national government agencies, research funders and local commissioners and providers of care to broaden the evidence base on health services, which in turn results in benefits for patients and service users.

In recent years the ability to use information to analyse and understand healthcare has increased exponentially. The safe and appropriate use of large datasets containing person-level information is increasingly important to running health systems across the world. Such information is also central to research and analysis to understand how providers can improve services and patient outcomes.

“The information contained in medical records has the power to improve care as much as any breakthrough drug”

Ara Darzi (*Health Service Journal*, 2014)

The NHS has a key asset in terms of its information in that it goes beyond single hospital or care providers. The ability to span whole populations and providers of care over time makes its information amongst the best in the world. This has already enabled large amounts of world-leading research (Wellcome Trust, 2014). In the past, the use of hospital episodes statistics (HES) data has proved invaluable in helping to assess the quality of care in the NHS, including identifying lapses in the quality of care; it played an important role in uncovering issues in hospital care at Bristol and Mid Staffordshire NHS trusts. Without the use of these data, both situations would have taken far longer to identify.

Our organisations have undertaken significant amounts of research using de-identified (pseudonymised) person-level data, including work at the Nuffield Trust on developing methods of resource allocation; risk stratification tools for case finding; linkage of health and social care data to explore service use at the end of life; and evaluations of interventions aimed at reducing the need for hospital admissions. Most of these studies have developed new ways of looking at information in applied research settings. Many have received

funding support from the Department of Health or the NHS. Often they used de-identified person-level datasets not available through the Health & Social Care Information Centre (HSCIC), such as local authority social care records, GP records and the Patient Episode Database for Wales (PEDW).

To give just one example, one quarter of cases of bowel cancer are not diagnosed until a patient reaches A&E with symptoms. By linking de-identified GP, hospital and Cancer Registry records the Nuffield Trust was able to examine the service use of patients with bowel cancer before diagnosis. This allowed for the exploration of the potential of detecting more cases of bowel cancer in primary care, thereby increasing the chances of survival (Sheringham and others, 2014). The Nuffield Trust has included other examples of how it has used person-level data in the briefing accompanying this consultation response (Nuffield Trust, 2014).

In undertaking these studies the Health Foundation, the King's Fund and the Nuffield Trust have worked closely with the government and national agencies to ensure that we have access to the information needed – and that we use it appropriately within both legislative requirements and expected good practice of that time. We welcome moves to strengthen and clarify these mechanisms, and support increased transparency in the system. We particularly recognise that our ability to conduct research in this way, and the associated benefits to patients, depend on the systems governing the use of these data being able to maintain public confidence. To this end, it is important to bring clarity to the protocol for sharing data, for both those inside and outside the service, and to provide a clear narrative to the public about how their health data are used to improve the health of the population and the health service. Security features must not be presented in isolation, but alongside the tangible benefits that data linkage offers to patients, service users and the whole population.

Our response is encapsulated in the following four points, each of which is discussed in detail below with answers to the specific consultation questions:

- Appropriate strategies for accreditation are the key to managing the risk and benefits.
- A clear and consistent framework for use of all de-identified person-level data, which protects confidentiality, facilitates new research, and can be built upon by other sectors is essential.
- Active and proportionate monitoring, not high volumes of regulations, are the key to a secure and effective system.
- Enabling novel research through de-identified data linked at the patient level has massive potential to improve health for all and reduce health inequalities.

In summary, we must use this opportunity to build a robust, transparent system of data exchange that both protects confidentiality of individuals and makes use of the huge potential that de-identified person-level data have to improve care and health for all.

Appropriate strategies for accreditation are the key to managing the risk and benefits

Q4. Should there be any restrictions as to the type of body which might become (in whole or in part) an ASH [Accredited Safe Haven], for example, a social enterprise, a private sector body or a commercial provider (working under a data processor contract)? Please let us know what you think.

Rather than focusing on the type of body which might become an Accredited Safe Haven (ASH), we believe that access to ASH status should be determined by an organisation's ability to meet the required information governance standards and that its purpose as an ASH will be consistent with the Care Act 2014 stipulation that NHS information will only be used for the benefit of healthcare.

We are concerned that voluntary sector organisations such as ourselves, with a strong track record, might be denied access to ASH status on the arbitrary basis that we do not fit a particular organisational 'type'. If this were to be the case, it could put at risk many vital research projects. It could also cause the NHS to miss out on new opportunities, as the reality is that the most innovative approaches to using data do not flow from statutory agencies (who are often tied down to routine agendas).

It is critical that the approval mechanisms put in place have the following characteristics:

- The guidance is clear, unambiguous and applies to all applicants.
- A programme of tools is available to help people, for example consent guidance, secure data apps and information tools for data requests.
- Form-filling is streamlined and coordinated.

In addition, we suggest that policy-makers consider how applicants' past records of secure information handling can be taken into account in the application process to be granted ASH status.

Q5. Is there a maximum number of accredited safe havens that you would consider to be acceptable? Please give your reasons.

The consultation document already acknowledges doubts about HSCIC's current capacity to deliver the volume of data linkage requests that are anticipated (paragraph 31). It is clear that a number of ASHs will be required for some time to come.

The total number of ASHs should be governed by two factors: the number of organisations that succeed in meeting and maintaining the stringent information governance requirements to become an ASH; and the number able to sustain themselves by meeting the demand for their services. If in time the largest ASHs are able to accommodate total demand, then the number of ASHs will shrink naturally. This is a much more practical and equitable arrangement than selecting an arbitrary number.

It is also worth acknowledging that in some cases, where only local datasets are being used, it will be more convenient for all parties to work with the local ASH, than it will be to work with a national body such as HSCIC.

A clear and consistent framework for use of all de-identified person-level data, which protects confidentiality, facilitates new research, and can be built upon by other sectors is essential

Q1. Are these purposes the right ones? Are there any other purposes that it is acceptable for an ASH to use data for? Please set out what you think the purposes should be.

We believe that the list of purposes for which information may flow **to** an ASH (paragraph 26) should explicitly include evaluation of service interventions. While this may be considered to fall under the ‘audit, monitoring and analysing’ heading, we believe that the increasing importance of robust evaluation of new approaches for delivering health services – and the additional challenges that it often presents, such as linking novel participant datasets from third sector providers – merit explicit recognition of activity endorsed elsewhere in the consultation document (paragraph 54).

We note that linked data have the capability to describe a patient’s journey through care. In a world where integration between service delivery models is so important, and where new ways of organising care are appearing, for example capitation-based payment to providers, ASHs may need to link data from a wide range of sources. This may involve person-level data with unique access requirements (Office for National Statistics mortality data), from local authorities (social care or housing data), small third sector providers (interventions delivered by charities) or data from outside the care sector (such as benefits data from the Department of Work and Pensions). This diversity emphasises two points: first that the purpose of an ASH should be to securely link **any** person-level data to which access has been properly negotiated; and second that the framework given to ASHs must be capable of functioning as a unified framework for data sharing across sectors.

Similarly, while we understand that other legal data-sharing measures will still be valid (paragraph 25), we believe that policy-makers must push towards a single, all-inclusive framework for the exchange of person-level data. Where other legal approaches to data sharing exist, they must be acknowledged in the framework so that organisations wishing to exchange data are aware of their options and can make an informed choice when selecting an approach. Piecemeal guidance creates ambiguity, weakens the robustness of the system, hinders valid research and diminishes public confidence.

Q9. What are your views of the controls set out above?

We believe that the strong controls governing the release of data from an ASH are appropriate, especially the inclusion of strong controls on the handling and use of the data on those receiving them.

A key challenge for the new regulations is to find an appropriate balance between the responsibilities and penalties for data suppliers and data recipients. We believe that the proposals have broadly the right balance by requiring that HSCIC and ASHs must ‘satisfy themselves’ that data recipients are not able to re-identify data, while data recipients are subject to a meaningful set of fines and reputational risk. However, the balance should be reassessed as the regulations progress to ensure that data providers are not unduly disincentivised to share data for reasons of patient benefit at negligible levels of risk. We strongly support the recognition that appropriately pseudonymised data used for research, service evaluation and other approved purposes are not ‘personal data’ within the meaning of the Data Protection Act.

The consultation document asserts that “it is almost impossible to be 100% certain that any data could not potentially be used to identify individuals” and, quite rightly, specifies that controls should be put in place to minimise this risk. While it is crucial that these controls manage the risk of individuals being re-identified, it is just as crucial that the data are not needlessly obfuscated to a level at which they are no longer useful for research. Examples might include dates of service use being truncated to months, making it much harder for researchers to establish patterns of care or to evaluate the impact of an intervention relative to the participant’s start date. It is essential that the nature of these controls is consulted on widely, and that they are sufficiently flexible to enable high-quality research on a variety of topics.

The supporting guidance to the Anonymisation Standard for Publishing Health and Social Care Data (The Information Centre, 2013) eloquently outlines the criteria that might be used by ASHs to decide when datasets are appropriately pseudonymised. While this guidance demonstrates the context-specific nature of this decision, we are mindful that exercising the controls must not place too great a burden on HSCIC and the ASHs. Consideration should be given to establishing rules-of-thumb about what constitutes acceptably pseudonymised information within a given context, with examples of datasets that are satisfactory.

One example may be large, population-level datasets that meet the following requirements:

- Patient name, address and identification number (such as NHS number) do not appear.
- Date of birth only exists in month and year format.
- Post code sector or lower super output area may be present, but full post code is not.
- Identifiable fields (such as NHS number) required for linkage of pseudonymised data sets by the data recipient have been encrypted using an approved one-way encryption algorithm.
- The proposed recipient (or organisations with which they would be permitted to share the data) is unlikely to receive another dataset that would enable identifiable fields to be linked.

The requirement about the data recipient being ‘unlikely’ to receive another dataset that would render the original data identifiable is important, and is consistent with the Data Protection Act. It should, however, be as clear as possible and interpreted in line with the proposed penalties on data recipients.

Smaller datasets (less than population level) may not lend themselves to as simple a definition of 'acceptably pseudonymised' as that given above. For example, a pseudonymised dataset that is limited to a small number of patients with a rare condition may be at higher risk of re-identification than one that contains the same fields for the entire population of a defined geographical area. In these situations, it is likely that some judgement will be required of HSCIC or the ASHs. Scientific approaches to assessing re-identification risk are being developed and could be considered for the longer term.

It is also essential that the administration of requests functions smoothly and efficiently – something that has not always been the case. The administrative burden on the Confidential Advisory Group (CAG) needs to be recognised and it needs to be supported to deliver swift responses, which could be achieved by devolved decision-making in straightforward cases. We also ask policy-makers to consider how organisations hoping to receive new information from an ASH might be able to refer back to their previous successful applications and avoid repeating the entire process – contingent on their past record of safe information handling.

We note that other aspects of working with ASHs – particularly the process by which CAG are involved for approving information requests – are being developed separately. We would like to stress again the value of putting in place a single, comprehensive set of guidance for all parties making use of person-level data.

Q10. What are your views on the level of the civil penalty that we should set for any breach of these controls?

Penalties must be proportionate to the seriousness of the breach, large enough to give the public confidence in the system and meaningful to the organisations in breach (that is, flexible to apply the same level of relative punishment to small research organisations as large corporations). It is also worth noting that in many cases withdrawal of the right to use data, immediately and in the future, will have more significant impact than financial penalties.

Active and proportionate monitoring, not high volumes of regulations, are the key to a secure and effective system

Q2. Are there any other regulatory controls that you think should be imposed? [On the operation of an ASH]

Building on a clear operational framework and a robust accreditation process, the body regulating functioning ASHs must actively monitor compliance and take swift action if standards lapse. We believe that this can be best executed by a limited, tightly-defined set of regulations that are fully understood by both the regulator and the ASHs. Excessive volumes of requirements create confusion for ASHs and extra burden for the regulator.

Q3. What are your views on the maximum amount of the civil penalty that we should set for breach of the controls proposed above in relation to ASHs?

See answer to Q10.

Q11. Are there any other controls that you think should be imposed? If so, please set out what you think these should be. [On receiving information from an ASH]

See answers to Q9 and Q10. It is also worth considering how the controls can be applied proportionally for organisations and individuals with strong track records in receiving data from ASHs and handling it correctly. This may take the form of being able to skip certain sections of the application process, or project-level approval being waived when organisations have successfully completed similar projects in the past.

In addition to placing controls on those using data from ASHs, policy-makers should also consider what mechanisms they will put in place to ensure that the new systems are operating smoothly. This should cover both effective and timely administration (Parliamentary review or the National Audit Office could be possibilities) and what right those receiving data (or the ASHs themselves) have to apply if they feel that the controls have been applied incorrectly.

Enabling novel research through de-identified data linked at the patient level has massive potential to improve health for all and reduce health inequalities

Q13. Do you have any views on the proposals in relation to the Secretary of State for Health's duty in relation to reducing health inequalities? If so, please tell us about them.

Routine person-level administrative datasets capture many interactions that a population has with the health and care system. This means that they represent the service use of marginalised groups much more accurately than the opt-in research cohort, which even when selected at random will be exposed to bias from the type of person that will choose to opt-in or can even be contacted in the first place.

For this reason, analysis based on these datasets represents a huge opportunity to address health inequalities. These efforts will be hampered unless enabling systems are put in place to allow secure access to de-identified health records linked across multiple data sources.

Likewise, as health and care services are increasingly delivered by networks of integrated providers, it has never been more important that commissioners, analysts and researchers have access to person-centred linked data to understand the quality and effectiveness of those services.

References

Wellcome Trust (2014) *Sharing Your Patient Record Can Help Researchers Save and Improve Lives*
www.patientrecords.org.uk

Health Service Journal (2014) Darzi: Electronic records are the next 'breakthrough drug'.
www.hsj.co.uk/comment/darzi-electronic-records-are-the-next-breakthrough-drug/5071976.article#.U9i29PldXzx

Nuffield Trust (2014) *Use of Patient Level Data by the Nuffield Trust*.

Sheringham JR, Georghiou T, Chitnis XA and Bardsley M (2014) 'Comparing primary and secondary health-care use between diagnostic routes before a colorectal cancer diagnosis: Cohort study using linked data', *British Journal of Cancer*, doi: 10.1038/bjc.2014.424.

The Information Centre (2013) *Anonymisation Standard for Publishing Health and Social Care Data. Supporting Guidance: Drawing the line between identifying and non-identifying data.* www.isb.nhs.uk/documents/isb-1523/amd-20-2010/1523202010guid.pdf

Supporting information

Roberts A and Charlesworth A (2014) *A Decade of Austerity in Wales? The funding pressures facing the NHS in Wales to 2025/26*. Nuffield Trust. www.nuffieldtrust.org.uk/publications/decade-austerity-wales

Bardsley M, Steventon A, Smith J and Dixon J (2013) *Evaluating Integrated and Community-based Care: How do we know what works?* Nuffield Trust. www.nuffieldtrust.org.uk/publications/evaluating-integrated-and-community-based-care-how-do-we-know-what-works