

Advice on Risk Prediction and Stratification Activities

June 2012

Introduction

The National Information Governance Board has considered and provided advice to a number of NHS organisations on a range of Risk Prediction and Stratification Activities in relation to the information governance implications. This includes both the preliminary analysis of data and the subsequent contact with the individuals in higher risk population groups with a view to offering them additional services.

The purpose of this document is to provide guidance for NHS, social care and partner agencies on how such activities can be undertaken with a secure legal basis i.e. compliant with the common law duty of confidentiality, Data Protection and Human Rights Acts, and sets out the key information governance principles that should be followed. All of the activities considered have been in relation to adult services; however the same principles would also be applicable to children's services, albeit that it would often be parental consent that would be sought.

This guidance reflects the current situation at the time of publication. It may be subject to change following the recommendations of the Information Governance Review and/or the development of Statutory Instrument legislation issued in support of the Health and Social Care Act 2012. This guidance will therefore be subject to frequent review in line with these developments and re-issued as and when changes are made.

Background

The NHS, with a number of partner agencies such as the Nuffield Trust and more recently Social Care organisations, have been engaged in risk prediction and stratification activities, such as PARR + (Prediction And Reducing Re-admission Plus tool) and the Combined Model Predictive Risk Model (see Annex 1). This has been with a view to offering additional support services to those identified as being at high risk. The nature of the risk in question would vary according to the different kinds of analyses undertaken. These projects have sought to link primary and secondary care health data and more recently, to link these data with social care data.

These activities essentially have two stages:

- 1) Gathering, linking and analysing data about a population in order to identify the target population.
- 2) Contacting the target population to offer them additional support. This is with the intention of preventing and reducing the likelihood of hospitalisation, adverse events and other poor health outcomes.

National Information Governance Board for Health and Social Care

Gathering, linking and analysing the data

The first stage involves gathering, linking and analysing personal data about a large group of people, not all of whom will ultimately be targeted for additional services and thus benefit. In order to process personal data lawfully for these purposes therefore, this first stage of gathering, linking and analysing data should only occur:

- 1) With the explicit consent of all the individuals whose personal data are to be processed; or
- 2) By using pseudonymised data¹.

Which route, consent or pseudonymisation, is appropriate will depend on the following:

- a) If the analysis is not being used to select individuals who would benefit from additional interventions but only to help with local service planning at a population level then pseudonymised or anonymised data should be used.
- b) If the linkage only relates to health data i.e. linkage of primary and secondary care data then again it should be feasible for this initial stage of analysis to use pseudonymised data.
- c) If the intention is to link health and social care data then as this involves disclosure of personal data across health and social care sector boundaries, generally this should only be undertaken with consent.
- d) An exception to this would be where a third party (such as the Nuffield Trust) was receiving and linking pseudonymised health and social care data together to generate a “risk score”. In this situation, the third party would not have access to the identifiable data and could only disclose the derived risk score to the relevant health and social care organisations provided care was taken to ensure that additional information could not be inferred from the score.

Other aspects to consider:

- Where a new service is offered to the general population then consent may well be feasible.

¹ Pseudonymisation is the technical process of replacing person identifiers within a dataset with other values (pseudonyms) available to the data user, from which the identities of individuals cannot be intrinsically inferred, for example, replacing a NHS number with another random number, replacing a name with a code or replacing an address with a location code. Pseudonyms themselves should not contain any information that could identify the individual to which they relate (e.g. should not be made up of characters from the date of birth etc.). Where the key is held to unlock the pseudonyms and re-identify the data set, pseudonymised data remains personal data under the terms of the DPA. See <http://www.connectingforhealth.nhs.uk/systemsandservices/pseudo>

- If the intention, however, is to target services to those at highest risk then generally pseudonymised data should be used to undertake the initial data analyses. This will ensure that patients and service users are not given false expectations about services, which may not be made available to them. This will also ensure that any additional services are applied to best effect for those at the highest risk. Most people will be understanding about the need for services to be targeted at those with greatest need.
- Where the intention is to link both health and social care data then obtaining consent prior to linkage may be the most appropriate route because of the subsequent disclosure and use of the linked data by both health and social care organisations (other than just the risk score). This is a key principle agreed by the NIGB in its consideration of the Common Assessment Framework (CAF) demonstrator programme and other data-sharing across health and social care to ensure there is a secure legal basis for the disclosures involved.
- It must be recognised that pseudonymised data is still personal data where the data controller holds the pseudonymisation key or holds the data in an identifiable form or could otherwise identify the individuals.
- Individuals' dissent must be respected where data is processed in order to anonymise or pseudonymise it, and to pseudonymised data which is still personal data.

The NIGB is aware that analyses may be undertaken by a wide range of organisations including: NHS and Social Care organisations, GPs and independent sector bodies, Acute and Mental Health services. The same principle, that consent should be obtained first or pseudonymised data used, applies to all bodies engaged in this activity.

Exceptionally, where neither route of consent or pseudonymisation are practicable then seeking support under the Health Service (Control of Patient Information) Regulations 2002² for the use of patient data may be appropriate. However, because using pseudonymised data has been demonstrated as practicable³ then robust justification would need to be provided as to why it was not feasible to use it for the particular analyses being proposed. Factors that may be relevant are whether there are language barriers, sensory or capacity impairment which would mean that relevant individuals could be significantly disadvantaged if they were omitted from the analysis and it is not feasible to include them otherwise. It should also be remembered, however, that such populations are often in contact with services and therefore it should be feasible, both to inform them and seek their consent. Where

² The regulations are given effect under Section 251 of the NHS Act 2006.

³ The King's Fund Risk Prediction algorithm project (PARR) which has been developed further by the Nuffield Trust.

individuals lack capacity their data may be used provided the relevant professional deems it to be in their best interests.

Where the intention is to identify a specific target population to offer them additional services, once the relevant group have been selected the pseudonymisation would need to be reversed (but only for the selected individuals). This re-identification process should be undertaken by a member of the relevant care team so there is no breach of confidence.

Contacting the target population

It is essential that the first point of contact with patients is made in the most appropriate way both to ensure good take up of the additional support offered and to maintain public trust in how their personal and confidential information is used by the NHS and social care organisations. The initial contact should be made by someone from a care team known to the patient, such as their GP practice or social care team depending on the primary focus of the analysis and risks being predicted. Contact may be in person, in writing or by telephone.

Consideration also needs to be given to how best to make contact where the population group in question may have impaired capacity or communication difficulties.

These and the following principles apply both to where additional services are to be offered and to where consent needs to be sought to link and analyse data prior to potentially offering additional services.

Data Controller and Data Processor arrangements

A Data Controller, as defined in the Data Protection Act 1998⁴ must ensure that any processing of personal data that they are responsible for complies with the 8 principles of the Act. Failure to do so carries a risk of enforcement action by the Information Commissioner.

A Data Controller may commission the services of a Data Processor⁵ to process personal data on their behalf. Such arrangements may range from a local agreement between the organisations involved, where one will take the lead for processing data on behalf of the others; or where 3rd party systems are procured and the supplier is the Data Processor. In all cases certain conditions set out in the 7th Principle apply to Data Controller and Data Processor arrangements including the requirement for a contract that clearly sets out what the Data Processor is allowed to do with the personal data and all necessary organisational and technical security measures they are required to take.

⁴ A person who (either alone or jointly or in common with other Data Controllers) determines the purposes for which and the manner in which any personal data are, or are to be, processed.

⁵ A person (other than an employee of the Data Controller) who processes personal data on behalf of the Data Controller

The Data Controller is accountable and retains the liability for all personal data processing activities carried out under their control, even when that processing is carried out by a Data Processor under contract. It is therefore important to identify and keep a clear focus on roles and responsibilities when processing personal data for Risk Prediction and Stratification purposes.

Use of Secondary Use Services (SUS) data

The use of SUS data for Risk Prediction and Stratification purposes relies on section 251 approval provided by the Ethics and Confidentiality Committee (ECC) of the National Information Governance Board⁶. The SUS s251 approval includes the use of the Commissioning Data Set (CDS) data for certain healthcare management purposes, but does not make explicit mention of use of the data for risk prediction and stratification purposes⁷. The terms of how SUS data can be used for this purpose under this approval is specified within the application as stated below and, in order to rely on this approval as the lawful basis for processing, it is important to keep within its specific boundaries. It is also important to note that all of the principles of the Data Protection Act still apply, in particular the Fair and Lawful requirements set out in the 1st principle.

Extract of the stated purposes of CDS:

The CDS data stored in SUS underpins the NHS commissioning process (including Practice Based Commissioning, Payment by Results and the 18 week referral to initiative (RTT) NHS Constitution obligation), and planning of services as well as performance management, which in turn support the provision, auditing and monitoring of patient care and treatment. The data will continue to be extracted in patient-identifiable form from SUS by local NHS organisations with access restricted by rigorous controls. Attention is drawn to the use of CDS data for RTT purposes, as SUS is the only mechanism for tracking patients in their pathways across organisational and geographic boundaries.

Section 251 (S251) approval is requested to enable the disclosure of patient identifiable data from care providers to the Department of Health (DH) (in the form of SUS) and disclosure from the DH to the NHS commissioning bodies, through allowing the extraction of CDS based data from SUS to proceed.

The key principles listed in the following section still apply to the SUS data sets in as much as the data should be pseudonymised for the purpose of analysis and subsequent re-identification should be restricted to a member of the relevant care team so there is no breach of confidence. The further use of that personal data by the relevant care team for a direct care purpose moves it out of the scope of secondary use.

⁶ National Health Service Act 2006 <http://www.legislation.gov.uk/ukpga/2006/41/section/251>

⁷ S251 application reference No. 250 details on the s251 register <http://www.nigb.nhs.uk/s251/registerapp>

Key Principles

The NIGB has identified the following key principles⁸.

Key Principles

- There needs to be transparency so that patients and social care service users know who is providing the service, and who and under what conditions organisations and staff, not already involved in their care can access their personal information.
- The information provided to those who may use the service (open access) should be clear and sufficiently detailed to inform them how the service will operate, including:
 - a) Who is providing the service;
 - b) The type of information which will be disclosed to the provider;
 - c) How consent will be obtained;
 - d) That the minimum information necessary will be held on the provider's systems, securely and confidentially and only for as long as they provide the service;
 - e) That the information can only be used for the purpose of providing this service and cannot be used for any other without agreement from the data controller⁹;
 - f) Where they can obtain more information and with whom they can discuss concerns.

[See NIGB's Guide to producing patient and service user information materials about what may be useful to include <http://www.nigb.nhs.uk/advice/infomaterials>].

⁸ These principles were initially identified in its consideration of West Kent Care Call, a service delivered for the PCT by BUPA (see Annex for an outline of this service).

⁹ The data controller is the organisation which determines the manner and purposes of the processing. They may either do this by processing the data directly or through a contract with one or more data processors which will perform the processing on their behalf and according to their instructions. All of the obligations in the Data Protection Act are on the data controller with the intention that the relevant safeguards will be included in the contractual arrangements with data processors.

http://www.ico.gov.uk/for_organisations/data_protection/the_guide/key_definitions.aspx

- The initial approach to patients and social care service users should be from an
- organisation with which the patient has a clinical / care relationship.
- The contract with any provider must include appropriate information governance and security standards¹⁰ to ensure compliance with data protection requirements and should include compliance with the NHS Care Record Guarantee.
- There needs to be valid consent from the patient or service user to process the information. This should be explicit, or in some instances implicit, provided there is a sufficient basis to support the validity of the consent¹¹. Sharing across health and social care boundaries should be with explicit consent.
- Where consent has been implied people must be informed and given a reasonable time period in which to opt out prior to disclosure of confidential information and a mechanism for dissent to be recorded and implemented.
- Where a service is provided for people with long-term conditions or belonging to particular risk groups, any population based data processing should be done using pseudonymised data and the pseudonymisation keys used for linkage and re-identification must be held within the most relevant NHS or social care organisation responsible for their care.
- If eligible patients do not opt to use the additional services then their data should be removed from IT systems after a reasonable period (typically four months); unless there are other legitimate purposes for which the data are being held.
- Care should be taken to ensure that individuals are not excluded from additional services based purely on the results of such risk scores as they are not of sufficient reliability to provide automated decision-making or decision support.

¹⁰ In line with the requirements of the 7th Data Protection Principle: Appropriate technical and organisational measures shall be taken against unauthorised or unlawful processing of personal data and against accidental loss or destruction of, or damage to, personal data. This is further defined in paragraph 11 of Schedule 1 and in ICO guidance.
http://www.ico.gov.uk/for_organisations/data_protection/the_guide/the_principles.aspx

¹¹ In general, the same tests of validity apply to both implied and explicit consent, namely that consent is informed, specific to the circumstances, that the individual has capacity and that it is freely given. For explicit consent the individual gives a positive indication of their agreement. For implied consent this is by not opting out having been informed and given the opportunity to do so.

Annex 1

The King's Fund Risk Prediction approach

The solution that was agreed for the King's Fund algorithm project was that the analysis software would include an encryption programme and that the staff member running the programme would do so without accessing the identifiable data held within the data file. The output would be a series of files (messages), one for each GP practice that would contain the data relating to patients registered with the practice in encrypted form. These output files would be sent to the practice or other clinicians already providing care and treatment to the patients concerned. The decryption keys would be held by the PCT and sent separately to the relevant health professionals once they had confirmed receipt of the encrypted files. Clinicians would then be able to view the information about their own patients and offer additional support or services to the highest risk patients.

Combined Model

The Combined Model was developed by the King's Fund, Health Dialog, and New York University to predict risk of emergency admission to hospital across all patients in a given health economy. It exists in the form of technical guidance that the NHS could use to implement risk prediction systems for extracting GP and hospital data, and which would identify patients both at risk of re-hospitalisation and more importantly, patients at risk of admission but who have never had a hospital admission. The model allows risk stratification for patients with long term and other specific conditions.

- The main difference to PARR is that the Combined Model uses GP and other secondary care data and therefore is deemed to be predictive for a whole registered population and could identify patients who have not had a hospital admission, so that interventions can be put in place at an earlier stage.
- The Combined Model is not a downloadable tool and can only be implemented in PCTs that are already (or are working towards) integrating storage of primary and secondary care data, such as through data warehousing and safe havens.

West Kent Care Call¹²

This service was developed by BUPA for West Kent PCT and involved BUPA receiving separate streams of demographic data and pseudonymised clinical data. All patients in the area were written to and provided with a four week opportunity to opt out of the programme, prior to the commencement of any clinical data flowing to BUPA. No clinical data was provided for those who opted out and their demographic data was also removed with a flag being held on their GP record of their opt out. BUPA then undertook a range of analyses using pseudonymised data on behalf of the PCT and for those patients who had not opted out and who were identified as at risk, BUPA nursing staff then provided telephone support to them with explicit consent obtained for the service at this first contact. Whilst the core aim was to target those at risk it was a service which was open to the whole adult population of the area.

The Nuffield Trust - Predicting people who will start intensive social care

The Nuffield Trust has conducted a study to test the feasibility of building a model to predict individuals who will start receiving intensive forms of social care (such as admission to a care home or the start of intensive home care). Their models were based on linked, pseudonymised data from health and social care systems. The researchers concluded that a next important step would be to pilot models in order to test the usefulness of risk scores as a method for targeting additional resources aimed at preventing loss of independence and the need for intensive social care use. Where the intention is to target services on the basis of linked health and social care data, consideration needs to be given to the long term use and sharing of the linked data. It is therefore particularly important to provide individuals with information regarding the use of their data and to obtain their explicit consent prior to this data being linked, if the intention is to share the linked data rather than just the derived risk score.

¹² This service has since been discontinued.

One approach for the implementation of such a model may be:

- a) Provide individuals with information about the use of their data. This should include an explanation of how the scores are being arrived at through analysis of linked anonymised information, and contact details if they wish to discuss any concerns, either about the score or about the use of their information.
- b) Pseudonymise the data sets from health and social care
- c) Link pseudonymised data sets and generate risk scores
- d) Re-identify high-risk patients /service users. Re-identification should only relate to the risk score and not the data on which it is based.
- e) If the focus of the analysis is to identify social care risk, the initial contact should be made by the social care team if already known to the patient; otherwise initial contact should be made by another care team already known to the patient, such as the GP practice.
- f) If there is a desire to share more detailed information to support better integrated care of the individual then explicit consent should be obtained to do so at this point.

Anticipatory care planning and integration – a primary care pilot study aimed at reducing unplanned hospitalisation

Br J Gen Pract 2012;DOI: 10.3399/bjgp12X625175

A cohort study of a service intervention in a general practice and primary care team in Scotland. The research paper was published in the British Journal of General Practice June 2011



Risk prediction
study.pdf