

please send comments to sam@medconfidential.org and contact prior to use for current text
this version: January 2015

Delivering your promise to patients: opting out of secondary uses doesn't affect your care

[Introduction](#)

[Invoice Reconciliation](#)

[How can the Secure Data Facility support Risk Stratification?](#)

[Generating Risk stratification models](#)

[Using Risk stratification models](#)

Introduction

Patients who are given the choice of secondary use of their data, primarily under the banner of research, are routinely told that their care will not be affected if they opt out.

While this may be substantively true, research based tools are beginning to use individual level data to offer capabilities to commissioners and GPs that mean that the promise can be broken. The promise to patients that their care will not be affected must be met.

Similarly, as a result of adequate tools and data products for other secondary uses, there has been a tendency to reuse detailed individual level records where a properly designed aggregated dataset would be

MedConfidential has previously introduced papers¹ on the concept and implementation of a safe setting², to meet the needs of bona fide research to access individual level patient records in a safe, consensual and transparent manner. We now update the risk stratification and invoice reconciliation sub-annexes in light of many informative conversations about details.

Those conversations have shown that such an approach is feasible, with no lack of will or imagination on the part of those who would be required to implement parts. However, we have seen no significant leadership from the institutional drivers, beyond the repeated concerns highlighted by IIGOP³.

There is an absolute need for GP system suppliers to be involved, facilitating the interface between the models created by research and the application of a particular model to the

¹ <https://medconfidential.org/2014/the-department-of-health-protecting-personal-health-and-care-data/>

² formerly described as the Health Research Remote Data Laboratory (HRRDL) now the HSCIC Secure Data Facility.

³ <https://www.gov.uk/government/publications/iigop-annual-report-2014>

please send comments to sam@medconfidential.org and contact prior to use for current text
this version: January 2015

entire registration list of a particular practice (or, in the case of commissioners, the aggregated counts of people in an area).

We see no reason that data used for commissioning purposes, for invoice reconciliation, or for risk stratification, is exempt from consideration under the principles of balancing safe, consensual and transparent.

please send comments to sam@medconfidential.org and contact prior to use for current text
this version: January 2015

Invoice Reconciliation

The privacy of those receiving any treatment should not, and need not, be compromised for the generic payment of invoices. Carefully designed data products should allow this to be done without reference to individuals. Such data products will be sensitive due to the necessary presence of small counts, but they need not be individually identifying.

As such, data products produced for the purposes of invoice reconciliation should be about treatments, not individuals. Several individuals may be included in a single treatment count, and several treatments may be attributed to an individual. Such counts operate at an organisation level, for a stated period of time. As aggregated counts on treatments, the data is not an individual level record, and will have no direct consent implications.

Invoices, then specify which treatments, and which time period(s), are being billed for.

Reconciliation looks to see that invoices do not exceed those prescribed, for each time period.

Where treatments being invoiced diverge from those being prescribed, for any specified time, they can then be more closely examined. This should limit the scope for fraud within the invoicing system, as it is treatments that are charged for and controlled, not individuals receiving treatment.

Additionally, all non-machine-readable documents (ie, invoices capable of being printed), must include a machine readable barcode⁴, to minimise errors from repeated retranscription of code.

Apart from the bank and specific contact details of the organisations involved, the details of any invoice, which didn't include small numbers, should be entirely publishable.

Such a model should also allow effective generation of open data around spending on items and counts in line with the NHS England and Government commitments to transparency.

It is possible to define a system for invoice reconciliation. That NHS England has put no effort or resources into it does not affect that process.

⁴ or QR code, which allows more information to be encoded.

please send comments to sam@medconfidential.org and contact prior to use for current text
this version: January 2015

How can the Secure Data Facility support Risk Stratification?

Risk stratification is roughly separable into two highly distinct activities. Firstly, design of strata - the creation of models, interactions, treatment effects; and secondly the application of those models to a specific area and population, which is far more prevalent than their creation.

Neither of these are direct care, and attempting to use identifiable data with a direct care justification is questionable for the reasons NHS England explain in the past attempts to create Accredited Safe Havens⁵.

Deconflating the different aspects of risk stratification, to separate out the various data needs for the different users at different steps in the process, the differing purposes, and hence the required different data products, allows for much clearer data flows, audit and consent.

This can be safe, consensual and transparent; the choice is whether or not it will be.

Generating Risk stratification models

The generation of models, the testing of hypotheses to form them, the development of publications justifying them, is research. The process for a model developed by MRC is almost indistinguishable from a model for the Mayor of London's office. The models should be subject to publication, peer-review, etc. As such, that development should take place within a data laboratory, on the rich, detailed patient data required to distinguish between (or at least, potentially investigate) correlation and causality. As development works on models of a population, it can use fully consented data

As part of creating a risk stratification model within a safe setting, part of the requirement can be to design the data specification required for the model to be run. Those outputs should be aggregated, and area (CCG) specific, and can then be made approved by HSCIC before being copied to a different facility for access by users of a risk stratification model.

Where a model needs to include data from other sources, and many of them will, the HSCIC's safe setting, when operational, should be welcomed into the club of equivalent government safe settings (ONS, MoJ, HMRC, DWP) with cross-departmental acceptance of standards, such that data of lower sensitivity could be available in the HSCIC safe setting for that particular project. This is established process, and while it can not be confirmed before such a laboratory is operational, it would be a cause for public concern if HSCIC's design for the Facility did not exceed the standards of the other facilities available in Government.

⁵ such as in the 2014 DH ASH consultation.

please send comments to sam@medconfidential.org and contact prior to use for current text
this version: January 2015

Using Risk stratification models

Most organisations will not generate their own models from scratch, but apply models developed elsewhere to their own areas. Even organisations which build their own models will likely draw these distinctions internally, with separate research and commissioning functions which interact but are operationally distinct.

Models developed for stratification can include the specification of the data they need for any particular area of interest. Such specifications will be aggregated statistical outputs, with up to some number of dimensions of data tabulated. While those tables will include small numbers, and so can not be used in uncontrolled environments, they are aggregated figures about an area, and should not be individual level data.

As a result, there will still need to be a controlled environment for an organisation to look at the risk stratification data that they are interested in, consider different iterations and boundaries, based on the specification of the model using constraints designed by the model's designer. No model allows an unlimited amount of variation, which is why different models are required for different purposes.

However, models can be used within a safe setting appropriate to the data input to the model (ie the custom data product designed by the model creator), which will be of much lower sensitivity than individual level patient data. The primary output of the application of a model to an area is not a list of patients, but a set of types and thresholds of treatments and counts, to enter into other documents - it will likely be an aggregated dataset in several dimensions without small number suppression having been applied.

Where appropriate, after agreement, those types and thresholds can then be run by GPs against their full patient dataset, which will include individuals who have opted out of secondary uses for data sharing. They are not included in the generation of the model, but by careful and correct data design.

Such a design has been discussed as part of the emerging “big data” approach to health data. Whereas the question of general research requires access to individual-level data⁶, specific research questions can be answered using particular data products designed for the purpose.

⁶ DataSHIELD: resolving a conflict in contemporary bioscience—performing a pooled analysis of individual-level data without sharing the data. <http://ije.oxfordjournals.org/content/39/5/1372.full.pdf+html>