Blurring health data widens sharing options

Being able to share health records is not just about passing patient details between clinicians who deal with the same patient; the data is also valuable for research and analysis. Rob Navarro of Sapior explained at the BCSHIF meeting in April how his company is looking at blurring health records so that data can be used by a wider community. Helen Boddy reports.

To predict the most likely reasons for emergency service call-outs, reliable data sets of health problems need to be analysed. To do so, it is not necessary to know the details of individual patients' illnesses, as long as they can be grouped together.

There are various other scenarios where government, research institutes, commercial developers of healthcare products and other bodies could use health data to improve healthcare services through research and analysis.

'There is quite a lot of value in health data,' said Rob. 'There are a lot of new applications to analyse data patterns. Pharmaceutical companies want to see if their drugs make a difference. There is also an emerging trend to develop preventative medical reminders, for instance for families disposed to certain hereditary health conditions.'

Sapior has therefore developed a way of blurring health records data that would make individual patient record data unidentifiable so that it can be released to a wide set of government agencies and commercial bodies. The important thing is that it is not possible to deduce or infer from this data anything about an individual patient, while keeping it meaningful in terms of being able to distinguish, for example, geographical patterns.

Rob made it very clear that what Sapior is looking at is not aimed at clinical use of data where NHS staff need to be able to see details about the patient sitting in front of them. In that case blurring of data would not be appropriate. For staff looking at actual patient data, he suggested instead that details would be kept secure by role- and time-based control of who could access the system, encryption and audit logs, and surveillance of which data is accessed and for what it is used.

When data is to be made available to a wider community, there is no need for it to include patient-identifiable data. Rob explained, however, that simply pseudonymising data was not secure enough as 'inference' could be used to work out patient details by linking some information to other known information, such as details publicly available on the electoral roll. Inference attacking was publicly demonstrated in August 2006 when pseudonymised search records were released by AOL only to then be identified by journalists using these methods.

'Inference attacks are the biggest weakness,' explained Rob. 'For example, if you know when someone has visited a doctor, you can work out which record must belong to that person if no-one else has visited the same variety of healthcare providers on exactly the same sequence of days. It is our thesis that if we deal with inference attacks we can open up the data to a wider group.'

Sapior has therefore developed a solution in which similar values in all fields are grouped and then blurred to make each record non-unique. The technology radically extends 'k-anonymity'. The idea is that at least 'k' records are made the same with just one identifying field being different.

K is derived from the risk of illicit re-identification associated with the user of the data – the greater the risk, the bigger the k, and the more blurred the data is that they receive. Therefore the riskier the recipient, the more blurred the data would be that they receive.

'If records are ambiguous we can stop inference attacks from happening,' claimed Rob.

Sapior is running a trial of its system this year with a commercial company and is looking to run a trial with an NHS body (the Information Centre for Health & Social Care, for example). Anyone interested should contact Robert.Navarro@sapior.com.