Out of Hours

Healthcare datasets:

ethical concerns

A number of tools are about to be launched that will permit researchers, for the first time, to search the computerised medical records held by the NHS. Such a vast repository of data should enable researchers to identify whether a drug or other intervention really does work, or in what circumstances it works best. It will also enable researchers to discover how many patients are being treated for particular conditions; or perhaps to identify those who are not being properly treated. It may even identify hitherto unknown medical conditions (perhaps hypotension will soon be recognised in the UK?). Thus there is potential benefit to be gained by such research, but there are ethical problems too — yet these tend to be ignored. I see worrying parallels with previous ethics scandals that I had hoped we had learned from. In particular, it seems that the lessons of Alder Hey¹ have been forgotten.

CLINICAL SYSTEMS

Many clinical computing systems in the UK are poised to offer their own (as they see it) computerised medical record datasets up to researchers. For instance, TPP's SystmOne's version is called ResearchOne and Vision practices have Clinical Practice Research Datalink (CPRD, formerly General Practice Research Database [GPRD]) (www.cprd.com). In addition there is the generic General Practice Extraction Service (GPES; www.hscic.gov.uk/gpes) which has been developed by the Health and Social Care Information Centre and is intended to work on all the clinical systems. Most practices will come across this very soon as it will be used to replace the Quality Management and Analysis System from the 2013/2014 financial year and facilitate data transfers to the Calculating Quality Reporting Service in order to permit payment to practices for their Quality and Outcome Framework achievements.

Practices will be asked to opt-in to research on a study-by-study basis (CPRD will additionally pay practices for the administrative work entailed). Computer-savvy patients will be able to say which studies may use their data. But this is where the ethical issues really begin.

Patients will apparently be able to look at certain websites to find out what research is planned and what data items are going to be used in the particular extractions. However

"I see worrying parallels with previous ethics scandals that I had hoped we had learned from. In particular, it seems that the lessons of Alder Hey have been forgotten."

it is unclear how a patient may opt-in or out of individual studies: computer codes to exclude patients (such as XaZ89) are very broad in their scope, excluding patients not just from the instant research but from all such research. This 'broad consent' model is very different from the 'informed consent' approach that research ethics has traditionally relied on to ensure adequate consent, and it fails to recognise the special nature of computerised health data.

ANONYMOUS?

Patients may be told that if they opt-in to research it will only be for 'fully anonymised' data. This may sound very reassuring, but it actually means only that certain grossly explicit identifiable data — the patient's name, address, postcode, date of birth, and NHS number — will not be passed on to the researchers. If this seems reasonable it is to fail to see the richness of health data. One must realise that the NHS number (if not the other clearly identifiable 'data bits') will be needed to link datasets, such as hospital and GP records, so this data has to be collected before it is removed in some way. But even then what is left may very well remain always potentially identifiable, as the House of Lords has recognised.² Ouellet confirms:

*'... even after stripping personal health data of direct personal identifiers, the resultant information is typically so rich that the risk of indirectly revealing the individual's identity is sufficiently high to require that the de-identified data be treated as identifiable'*³

An example will illustrate the issue. Suppose I know my new neighbour's

daughter is visiting on Saturday coming; the daughter will be 32 on that day. I can now easily work out her date of birth and from other information my neighbour tells me I learn that this neighbour gave birth to the daughter at age 28. I further discover, in conversation, that the daughter is now diabetic like her mother, and that the mother had her influenza jab yesterday at her doctor's surgery. She even rhetorically asks me 'Isn't Dr Holiday a nice man?' and so indicates her particular surgery. Now let us imagine that I have access to the 'de-identified' health records of all patients with diabetes over 50 years old in the region. I can now search for females aged 59–61 years who have diabetes and who had a 'flu vaccination on such-and-such a date and who gave birth to a female child between 1978 and 1982. In fact, if one had the whole country's 'anonymised' health records at one's disposal one could filter-out all records matching the criteria. I would probably find my neighbour's record using only some of the information she so innocently disclosed.

Furthermore, although a patient's postcode is deemed to be potentially identifiable data and thus not to be disclosed in an 'anonymous' record share (it would narrow down a record to about a dozen or so properties), the first half of the postcode alone is not regarded as a personal identifier. However even a broad postal area code can help narrow down a search involving massive (millions of records) datasets. Thus 'lives in DGxx [first half of a postcode], is female, a diabetic, had a vaccine of type xxx on dd/mm/yyyy, collected a prescription of metformin at XYZ Pharmacy on dd/mm/yyyy ...' or any combination of these or

"... health data cannot ever be guaranteed to be truly anonymised."

“... patients give information to clinicians in the expectation that by telling the whole story they can get the most appropriate help. If they had to edit their story before discussing it ... they could not know if they had given a sufficiently full picture for the clinician to offer the most appropriate help ...”

similar facts may be used to find a target. The point is that health data cannot ever be guaranteed to be truly anonymised because it all depends on what you know or can find out in the future about your target. Armstrong for example, cites how:

[w]ith remarkable ease, new generation software can take even anonymised data, match it with other lists, and effectively re-identify many individuals.”⁴

If a patient is not prepared to permit their data to be used for those studies they consider to be not worthwhile but then changes their mind, and then changes it back again: how may such changed preferences about the use of data be implemented? And how can that patient check that his/her data were not included in a particular search/extraction? It is after all ‘anonymous’. Who is responsible for administering the hundreds of data searches that could potentially be made over a year? Once one has given one’s dataset how can one stop it being used for other studies that were not anticipated when the data were first provided to a group of researchers? There is the basis for an ethical minefield here.

SCANDAL

Such imminent new use of patient data has potential parallels with the Alder Hey scandal — if we are not careful — in that patients may not know exactly what their consent (should they give it) will really mean. Just as in Alder Hey, and elsewhere, by agreeing to donate ‘tissue samples’ to the pathology department, patients or family members of the deceased did not realise that by ‘tissue sample’ any part, or all, of a human cadaver was caught by the term. So patients are unlikely to understand that by ‘de-identified’ or ‘anonymised’ this may not really mean that that patient cannot subsequently be identified and his or her records read in their entirety.

There is also the potential for identifiable information to be deliberately collected

about the patient: even without their permission. Cancer registries already have access to identifiable patient records without the patient realising. Such practices will continue and burgeon. The common law duty of confidentiality that governs identifiable records (we have seen that in reality fully-anonymised data are still always potentially identifiable and so something of a misnomer) can, where researchers can convince a body (from April 2013, the Ethics and Confidentiality Committee [ECC] of the Health Research Authority)⁵ that the research is so worthwhile, be set aside in the interests of society as a whole, and where seeking consent would be impractical or too burdensome for the researchers.

We must not forget either that patients give information to clinicians in the expectation that by telling the whole story they can get the most appropriate help. If they had to edit their story before discussing it with the doctor or nurse they could not know if they had given a sufficiently full picture for the clinician to offer the most appropriate help and the loss of trust implied would surely have adverse health consequences. This state of affairs must not be allowed to happen.

DEBATE

The problem with all of this complexity though is, fortunately, simple enough to tackle and can be addressed by public discussion. Given the chance, most patients would probably be happy to consent to their records being used for research (even taking into account that they may not really be truly anonymous). But not to allow them to know about these matters is wrong. Discussion must begin, and it must start sooner rather than later, before the research trawls begin, not afterwards, and in public, not in the consulting room please (10 minutes isn’t enough). It is easy to create a scare story here — and I don’t want to — but I do want debate. If too many people drop out — and it is likely that these will

ADDRESS FOR CORRESPONDENCE

Stephen Humphreys

The Garden City Practice, 11 Guessens Road, Welwyn Garden City, AL8 6QW, UK.

E-mail: stephenhumphreys@nhs.net

be clusters of people from population sub-groups comprising of the less-educated and lower-income groups (the same who tend to opt-out of the influenza vaccination)⁶ — then the data used in the research could be compromised: a fact which can only be compounded by there being different datasets available to researchers. I am all for the research, but all for a proper public discussion about it too.

Stephen Humphreys,

Practice Manager, The Garden City Practice, Welwyn Garden City.

DOI: 10.3399/bjgp13X668230

REFERENCES

1. BBC News. *Organ scandal background*. <http://news.bbc.co.uk/1/hi/1136723.stm> [accessed 29 Apr 2013].
2. House of Lords. *Common Services Agency v. Scottish Information Commissioner*. UKHL 47. House of Lords, 2008.
3. Ouellet R. Privacy issues and the Canadian Medical Association Health Information Privacy Code. In: Flood CM (ed.). *Data data everywhere: access and accountability?* Montreal: McGill-Queen’s University Press, 2011: 93–110.
4. Armstrong W. Getting lost in doing good: a societal reality check. In: Flood CM (ed.). *Data data everywhere: access and accountability?* Montreal: McGill-Queen’s University Press, 2011: 119–134.
5. The Health Service (Control of Patient Information) Regulations 2002, S.I. 2002/1438.
6. Tamblin R. Balancing safety, quality, security, and privacy: the case of prescription drugs. In: Flood CM (ed.). *Data data everywhere: access and accountability?* Montreal: McGill-Queen’s University Press, 2011: 53–70.