International Population Data Linkage Conference

Linking Data – Improving Lives

This was Farr sponsored Conference that HeRC PPI members attended held at one of the Swansea University Campuses – this one was on the edge of the sea and the weather was good.

Prior to the Conference we attended a work shop on public engagement for researchers, who had to produce a summary of their research firstly to produce an interesting title and then a 300-word description, aimed at a newspaper article. The four researchers had to do this in discussion with PPI participants and their resulting text was then scrutinized by a journalist who wrote on scientific research (she was not a scientist) and she then generated headline stories from what she heard for the different types of newspapers that are published in the UK. This process resulted in completely total revisions from the researchers, who were surprised how difficult it was to express their studies accurately in understandable language. Both researchers and the PPI members were equally surprised by the misleading and distorted headlines a journalist could then produce from the agreed text, which resulted in major revisions. It was a useful learning exercise for all and the researchers all reported on the benefit of the learning from this exercise.

The Conference theme was '***Linking Data-improving lives***' and as an International conference had papers from around the world. These papers demonstrated what could be done by linking data which was held in different locations to produce important and useful research outcomes. A few examples from the conference were:

- How good are statins at preventing heart disease?
- Improving understanding of multiple sclerosis by gathering and analyzing data.
- Did the smoking ban work?
- Using data to improve the uptake of life-saving heart procedures in the UK.
- Using data to study and improve the treatment of teenage anxiety.
- Maternal weight gain between pregnancies and childhood body mass index - using sibling analysis to address confounding by shared lifestyle.
- Improving primary care prescribing safety using routine data for intervention and evaluation: four trials in 500+ practices.
- Comparison of child mortality by characteristics at birth in England and in Sweden using linked administrative data.

I will describe a number of themes of possible interest to HeRC PPI members, in view of comments made about data security in our H@PPI meetings, and which researchers within HeRC need to think about:

1. One theme was that many members of the public felt that the data was already linked together, whereas researchers showed how difficult it was to get access to data, let alone to link data sets together. It was not generally understood that in the health area that patients GP and hospital records were not usually linked (Salford is one exception). This of course results in risks to

patients and one of our members described the result of having to transfer his own results between different specialist doctors to ensure good informed treatment.

2. As it was an International Conference I will limit my comments on available datasets and who might have access to them to UK examples and primarily England, since it is different to some extent in Scotland, Wales and N Ireland. For health data in England and Wales the major source is NHS Digital (rename for what was HSCIC) and for administrative data this is co-ordinated through the Administrative Data Research Network. This Administrative Data Network has branches in England, Wales and Scotland and its key theme is 'Better knowledge…better society'.

3. NHS Digital gave a presentation on how they worked to provide data to researchers. The key theme for me was the security of their data, since they effectively run all the main NHS systems and this contains all our names plus data that they extract from GP records according to the legal requirements and the wishes of the NHS. This is all identifiable data.

   How they hold their data safely is a large subject and the different issues are covered in the National Data Guardians review in detail – see 4 below.

   This presentation described how NHS Digital can combine datasets internally and then make these datasets anonymous before making them available to authorized researchers. These researchers are required to keep these datasets secure and we have seen some of the security within HeRC.

   So the key points are how and where these datasets are combined and kept and who has access to them. Such authorized research normally has to be agreed and authorized by HRA Ethics Committees and where the data contains identifiable patient data that has to go through an additional ethics approval by the 'Confidentiality Advisory Group'. All these approvals are published and available for all to see what is happening. However fully anonymised datasets can be analysed without the need for HRA Ethics approval.

   So my view, as an ethics committee member, is that the risks of loss of personal information is very low and when compared to the benefits and there is clearly a major benefit to this research using patient data. Use of a fully anonymised set of data allowed me to identify a group in Cheshire that succeeds in training type 1 diabetes patients compared to the lack of success in all other CCG's.

4. Not discussed at this International Conference was the report by the National Data Guardian Dame Fiona Caldicott entitled 'Review of Data Security, Consents and Opt-Outs', which we briefly discussed at our last H@PPI meeting. At the same time there is also the EU Regulation on 'On the protection of natural persons with regard to the processing of personal data and on the free movement of such data', which enters into force on 25th May 2018 and replaces the existing Directive.

The Fiona Caldicott consultation paper is well worth reading to understand the complexity and she in her opening comments states:

> 'A key aspect of this work ***must be a dialogue with the public***. We owe it to citizens to enable them to understand data usage as fully as they wish, and ensure that information about how data is accessed, by whom, and for what purposes, is available. This work is part of a wider dialogue that should be conducted on data use across different sectors. Health and social care data, although unique, cannot be isolated from that discussion.'

The report has been turned into a 17-page consultation document, where submission of comments closes on 7th September 2016. I raised it at the conference and was informed that HeRC would be making a submission but not whether any PPI members would be involved in that process. The consultation document is well worth reading.

5. The main suggestions in the Caldicott report is to allow opt-outs from identifiable data being used except in cases where there is a legal requirement:

   - the Care Quality Commission, which has powers of inspection and entry to require documents, information and records;
   - the HSCIC, the statutory safe haven, which has powers to collect information when directed by the Secretary of State or NHS England;
   - the NHS Counter Fraud Service, which has powers to prevent, detect and prosecute fraud in the NHS;
   - investigations by regulators of professionals;
   - coroners' investigations into the circumstances of a death, i.e. if the death occurred in a violent manner or in custody;
   - health professionals must report notifiable diseases, including food poisoning;
   - the Chief Medical Officer must be notified of termination of pregnancy;
   - employers must report deaths, major injuries and accidents to the Health and Safety Executive;
   - information must be provided to the police when requested to help identify a driver alleged to have committed a traffic offence; or to help prevent an act of terrorism or prosecuting a terrorist;
   - information must be shared for child or vulnerable adult safeguarding purposes; and health

   or alternatively for issues of major public importance such as:

   > 'preventing and responding to natural disaster; monitoring and control of important diseases in humans such as TB and diseases of epidemic potential such as Ebola; infections that pass between animals and humans such as the zika virus; and for chemical, biological, radiological and nuclear events. It would also include personal confidential data for monitoring and control of communicable diseases and other risks to public health.'

   However, for anonymised data she considers that there should be no opt-out and outlines her reasons as follows:

> ‘The Information Commissioner’s Office has a Code of Practice that establishes how data may be sufficiently anonymised that it may be used in controlled circumstances without breaching anyone’s privacy. The ICO independently monitors the Code.
>
> The Health and Social Care Information Centre, as the statutory safe haven for the health and social care system, will anonymise personal confidential information it holds and share it with those that are authorised to use it.
>
> By using anonymised data, NHS managers and researchers will have less need to use people’s personal confidential information and less justification for doing so.’

6. There were numerous papers at the conference on how to fully anonymise data so that there was never any chance of re-identification being made. Equally as this was a wide ranging conference there were also papers describing how datasets could be linked even when there was no common identifying field. So this then raises questions for a non-expert as to whether so called anonymized data could be re-identified. So what is possible, since one hears that this is all possible and we hear about how much insurance companies already know about us. As non-experts who have no access to the data or the science, we are unable to make any sensible conclusion.

   However, my simple thought is that if the anonymized data sets are not freely available then they can only be used for the purpose that they were legally allowed to be used.

   The Caldicott report also recommends and is consulting on the following suggestion:

   > ‘The Review has recommended that the Government should consider introducing stronger sanctions to protect anonymised data. These will include introducing criminal penalties for the deliberate or negligent re-identification of individuals. This is intended to give the public greater confidence that firm action will be taken as necessary to protect their personal confidential data.’

7. Mary Tully presented the results of the Citizen’s Jury held in Manchester about use of health data. She reported as we know that after 3 days of lectures and discussions that the majority of the participants was supportive of the use of their data in research (identifiable or annonymised) but she also reported that a few had hardened their attitude that their data should not be used in research. Unfortunately, there is then no detailed discussion to see why a few people’s attitudes had hardened. So that is a gap in our knowledge.

8. So my conclusion is that there is a need for a simple guide with adequate and detailed reassurance as to why patient data can safely be used with hardly any risk of loss of confidentiality of personal data. One cannot totally remove any risk of loss of data because people are people but the benefits of using the data has to be shown to be of massive benefit to patients compared to the minute risk of loss of information. The key point is to describe the detailed procedures that I find to be very encouraging, plus legal actions taken where any loss occurs.

Such a guide would assist with the #datasaveslives campaign.

Finally, I would like to thank HeRC for funding me to attend this very interesting conference.

Peter J Donnelly

September 2016