Submission to the Senate Inquiry into circumstances in which Australians' personal Medicare information has been compromised and made available for sale illegally on the "dark web."

Dr Chris Culnane, Dr Ben Rubinstein and Dr Vanessa Teague.

School of Computing and Information Systems

University of Melbourne

This submission addresses

- the security and privacy of Medicare numbers,
- some possible implications of the breach for the privacy of My Health Records and other personal information, and
- the practices and procedures for de-identification of personal Medicare information.

We would be happy to discuss any of these issues with the committee.

THE SECURITY AND PRIVACY OF MEDICARE NUMBERS

A person's Medicare number isn't an inherently sensitive piece of information. The concern is whether that information could be used by a criminal to impersonate the person and hence extract sensitive information or perform other kinds of fraud.

The HPOS system is used by a very large number of providers – it is probably impossible to guarantee that nobody ever misuses their access or leaves security holes in their system.

Attention should focus on minimizing the damage that occurs if someone learns someone else's Medicare number.

This shows that cybersecurity issues have to be thought of in the full context of all the other protocols running at once. Although no immediate harm may be caused to a person through exposure of their Medicare number, there may be significant harm from the combination of that breach with other breaches, weaknesses, or design choices.

Recommendation 1: It is probably not possible to keep Medicare numbers secret while also making them available to everyone who needs to use them for billing. Rather than trying to prevent any information leaking from the HPOS system, it would be better to focus on ensuring that a Medicare number cannot be used as a proof of identity to commit fraud or gain access to more sensitive information.

DE-IDENTIFICATION OF MEDICARE INFORMATION

We were the team that showed it was possible to decrypt supplier IDs in the MBS-PBS 10% sample open data. The MBS-PBS dataset release was motivated by convincing arguments about the utility of that data for medical research that saves lives. We strongly support this sort of research, and the general aim of informing public policy and inspiring innovation with scientific analysis of data. The question is how to engineer that without destroying privacy.

Our work on re-identifying supplier numbers in the MBS-PBS de-identified data release is only one of a long history of successful re-identifications in the literature. Secure de-identification of rich data is probably not possible without substantially degrading the data. Re-identification will only become easier as more information is released. It is very unlikely that even the most well-informed and well-intentioned set of guidelines on de-identification can guarantee privacy protections appropriate for sensitive data such as the MBS-PBS sample while retaining the usefulness of the data. The department of Prime Minister and Cabinet's "Process for publishing sensitive unit record level public data as open data"[1] is notably not a specific technical process. We do not believe any such secure process exists.

Recommendation 2: Sensitive unit-record level data, particularly when that data contains detailed information about each individual, cannot be securely de-identified without substantially degrading the data. It should not be released publicly or transferred to non-research entities.

MY HEALTH RECORDS

My Health Records could genuinely save lives by transferring important information among a patient's healthcare providers. Like many situations on the Internet, it is important to understand both the potential benefits and the potential risks.

---

[1] https://blog.data.gov.au/sites/g/files/net626/f/process_for_publishing_open_data_dec16.pdf

We have not conducted any careful study of My Health Record privacy, but have briefly listed some of the principles that should apply. It is much harder to keep something secret if it is already shared with hundreds or thousands of people – the My Health Record permissions should be set carefully, by default, so that only those who really need access to a person's record get only those parts of it that they need to see.

The current version of the official My Health Record website, as at 28th August 2017[2], reads:

> *The department has no intention to sell de-identified data from the My Health Record system. The My Health Record legislation provides authority to the preparation and issue of de-identified reports for public health and research purposes. A framework is being developed for the secondary use of My Health Records information. … Consultation is expected to begin in the second half of 2016. The final framework will address issues, including expectations for management of data including use, storage and transfer, in order to ensure that the privacy provisions in legislation will be met. Until this framework is in place, there will be no secondary use of the data provided by The Agency.*

Note that "has no intention to sell" is an extremely weak guarantee, particularly since a record as complex as an individual's health record is extremely unlikely to be securely de-identifiable without removing most of the information from the data.

It is not appropriate to share or sell patients' health records without their consent. For consent related to genuine medical research, consideration could be given to a system of richer and more expressive consent options such as Kaye *et al's* dynamic consent model.[3] An interactive online health record like My Health Record would be a perfect vehicle for giving patients more information and control over who reads their record.

The shift from opt-in to opt-out is concerning because it means that patients do not choose the visibility settings for their health records when they are established, and that patients who do not wish to trust the system with their details may not be able to prevent the uploading of their data in advance. It is too late to tighten privacy protections after the data has been read. This is a complex challenge, because of course there are risks associated with *not* sharing a patient's critically important health information with the relevant healthcare providers.

A stronger and clearer guarantee of patient privacy would probably improve patient trust, increase uptake and hence extend the benefits of My Health Records to more people. Of course this guarantee would need to be backed up by genuine strong technical protections against unauthorized or inappropriate access.

Recommendations:

3. My Health Records should not be made available to anyone not directly connected to the patient's care, except (with the patient's consent) genuine scientific researchers in a secure research environment under ethical oversight.

4. "De-identified" versions of sensitive unit-record level data should be understood to be probably easily re-identifiable, and should be secured appropriately.

---

[2] https://myhealthrecord.gov.au/internet/mhr/publishing.nsf/Content/privacy?OpenDocument&cat=Privacy%20and%20Security

[3] http://www.nature.com/ejhg/journal/v23/n2/full/ejhg201471a.html?foxtrotcallback=true

5. It might be possible to publish safely some aggregate statistics, such as the frequencies of particular illnesses or prescriptions. The government should explore techniques for Differential Privacy in the release of aggregate statistics. The government should be entirely open about what data is passed to whom under what circumstances, particularly for non-public disclosures to commercial entities.