Submission to the Senate Inquiry into the My Health Record System

Dr Chris Culnane, A/Prof Benjamin Rubinstein and Dr Vanessa Teague
The University of Melbourne

September 14, 2018

Our expertise spans technological and mathematical aspects of data security and privacy. This submission highlights privacy and security concerns, particularly the arrangements for third-party access. This submission contains our opinions as researchers with relevant expertise, not the official position of The University of Melbourne.

1 Introduction

It is undeniable that there are benefits to an individual of maintaining an electronic health record. There are also significant societal benefits from genuine medical research on large medical datasets. However, a patient may wish to derive the individual benefits of their own health record without accepting the risks associated with secondary uses. The conflation of the two aims confuses the discussion, because some uses of an individual's record are for that individual's benefit, some are for wider societal benefit with no intended bad consequences for the individual, and some are for wider societal benefit with potential negative repurcussions for the individual. It is difficult to see any evidence of the system having been designed with security and privacy as primary objectives, as it appears that access has been prioritised over everything else. If we think of My Health Record as a public health research platform, then that design goal might have seemed reasonable, but if individuals are to believe that their participation is to their benefit, much greater emphasis needs to be placed on security and privacy. We risk losing the possible individual and societal benefits of an electronic health record by eroding patients' trust that their data will be used for their benefit, rather than against them.

The consent model being applied to My Health Record is insufficient, and the more so because of the switch to an opt-out consent model for the system as a whole. It cannot be claimed that "Informed Consent" has been obtained, when in reality all that has been provided is a right to opt-out within a narrow timeframe.

The decision to allow a separate opt-out for secondary uses is a very positive one, though of course the same argument for opt-in vs opt-out applies to that decision as it does to participation in the system as a whole. According to the plain english companion piece on secondary uses of MHR data, "If a person does not opt-out, their consent is implied." However, the option to opt out doesn't adequately resolve the issue of consent because patients may not have carefully informed themselves of the implications, or may not even be aware of the option to opt out. There are existing techniques for improved consent, in particular Dynamic Consent. Such approaches could offer significant benefits.

We welcome the mention in the *Framework to Guide the Secondary Uses of My Health Records* of the possibility of introducing a Dynamic Consent model, and also the promise not to publish My Health Record data as open data. Both of these are very positive, but they are not enough to guarantee protection of the data.

2 The risks of re-identification

The *Framework to Guide the Secondary Uses of My Health Records*[1] assumes it is possible to compute a "risk of re-identification," using the techniques of O'Keefe et al[2], which in turn rely on k-anonymity. Unfortunately, de-identification of detailed unit-record level data does not work, at least not without substantially reducing the scientific value of that data. This is explained in detail in our report for the Office of the Victorian Information Commissioner[3]

2.1 Public re-identification of MBS-PBS data & implications for My Health

Our research team identified both suppliers and patients in the Department of Health's de-identified MBS and PBS dataset, which was published openly online in 2016.[4] The dataset included very little demographic information about patients, only their year of birth, state and gender. As such, a naïve calculation of the "risk of re-identification" must have suggested that the risk was very low. Unfortunately, like numerous other studies in data re-identification, we could show that individuals are identifiable based on the data available: a few points of information about dates of childbirth or (other) surgeries are sufficient to identify many patients. Such demonstrations are a simple matter of knowing very few facts about the person (for example, retrieved from online news stories) and running straightforward database queries to find how many patients in the sample match.

We do not understand the OAIC's conclusion[5] that patients were not "reasonably identifiable" by law because of the technical difficulty of re-identification, the absence of complete confidence in all cases, and the fact that only patients with "unique or rare attributes" can be identified. The technical difficulty of finding patients is within the reach of a competent high school student. Re-identification can be made with high confidence (especially for patients with multiple data points or rare conditions) in many cases. Almost all individuals have unique linkable attributes if enough information about them is known.

Since our earlier paper we have identified other patients with multiple data points in the sample. Women who have had two or more children as reflected in Medicare billing are typically identifiable based on those billing dates alone.

The MBS-PBS data breach has both direct and indirect relevance to the use of de-identified My Health Record data. Indirectly, it indicates a continued gap in understanding and adequate technical knowledge of data privacy within the Australian government.

More directly, the presence of the identifiable MBS-PBS data for 10% of the population is now a resource that an attacker could leverage in My Health Record re-identification. For example, a patient's MBS-PBS data could be re-identified based on childbirth dates from before their My Health Record Data commenced. From the MBS-PBS data other information about the person could be inferred, such as chronic conditions, medications, and recent medical events. This inferred information could then be used for successful re-identification of a de-identified My Health Record.

2.2 The risk of non-public re-identification

We welcome the undertaking in the Framerwork for secondary uses of my health data not to publish My Health Record data as open data—this means that there will be less open re-identification. However, it does not make re-identification any harder for those who receive it through other channels. Public information (as described above) is the minimum that might be used for re-identification. Data that is effectively de-identified against a university researcher may nevertheless be very easily re-identifiable by a large corporation that holds detailed data about the patient's web searches, payments, photos and email messages. Even if neither

[1] `https://www.health.gov.au/internet/main/publishing.nsf/Content/F98C37D22E65A79BCA2582820006F1CF/$File/MHR_2nd_Use_Framework_2018_ACC_AW3.pdf`

[2] `https://publications.csiro.au/rpr/download?pid=csiro:EP173122&dsid=DS2`

[3] `https://ovic.vic.gov.au/wp-content/uploads/2018/07/Protecting-unit-record-level-personal-information.pdf`

[4] `https://arxiv.org/abs/1712.05627`

[5] `https://www.oaic.gov.au/privacy-law/commissioner-initiated-investigation-reports/publication-of-mbs-pbs-data`

dataset is public, a corporation could re-identify the My Health Record data and thus make inferences about the patient’s health.

Any information that has a significant intersection of data points with a de-identified My Health Record or MBS-PBS record could be used for re-identification. For example, Melbourne-based company NostraData collects “data about the purchases of customers of our member pharmacies and reports that information to our member pharmacies and other clients.”[6] They claim ‘NostraData captures every script from over 3,000 pharmacies nationwide”[7]. The privacy policy goes on to claim that the data is de-identified. However the effectiveness of this de-identification is unknown as there is almost no public information about it. What is known is that NostraData formed a joint venture with IMS Health (now IQVIA)[8]. IQVIA advertises a “comprehensive global data portfolio representing over 530 million non-identified patient records across 100+ markets”[9].

Re-identification becomes easier the more detailed the de-identified record and the more information available about the person. For many Australians, some companies already hold a highly detailed record about the individual’s health. There is a serious risk that a person’s inadequately de-identified My Health Record could be re-identified and used for discrimination in credit assessment, employment, insurance, or numerous other scenarios against the patient’s best interests.

3 Privacy, security and trust

It is claimed that the “The My Health Record system has the highest level of security and meets the strictest cyber security standards,”[10] and yet it is difficult to see how it implements the principle of least privilege, which is one of the most fundamental security best practices available. A broadly defined class of Health Professionals have access to the system, with no systemic limitation on their access to an individual’s record. There are only legal repercussions for unauthorised access. Given the sensitive nature of the data involved, such an approach is inadequate. The concern is not only in trusting many thousands of registered health professionals, but in the required assumption that all of their machines, networks, and credentials will remain secure in perpetuity.

It would be far better if users were empowered to set default limits on who could read their files, rather than being able to password-protect only what has already been uploaded.

This issue of having to trust a broad range of organisations that have access to the data is particularly concerning given the current state of privacy protection of sensitive medical data. How many members of the public are aware that when they get their prescriptions filled at a pharmacy, there is a good chance that a “de-identified” copy of their script is being provided to a third-party company? The sale of Medicare numbers on the darkweb[11] also demonstrates that not everyone with access to that information could be trusted to keep their systems secure and obey the rules.

3.1 Lessons from the international arena

There are many similarities between the My Health Record and the UK’s failed care.data system, including even some of its leaders. Care.data was discontinued in 2016 when an independent review found that, “broadly, the public does trust the NHS with confidential data.”[12] This followed an admission in 2014 that the “Health and Social Care Information Centre admitted giving the insurance industry the coded hospital records of millions of patients, pseudonymised, but re-identifiable by anyone with malicious intent.”[13]

[6] Nostradata Privacy Policy https://www.nostradata.com.au/Public/Home/Privacy. Last accessed 12 Sep 2018.

[7] https://pharmadispatch.com/news/a-completely-different-world

[8] http://www.fo.kit.net.au/news/ims-nostradata-deal/48458

[9] https://www.iqvia.com/solutions/real-world-value-and-outcomes/realworld-data

[10] https://www.myhealthrecord.gov.au/for-you-your-family/howtos/frequently-asked-questions

[11] https://www.theguardian.com/australia-news/2017/jul/04/the-medicare-machine-patient-details-of-any-australian-for-sale-on-d

[12] https://www.gov.uk/government/speeches/review-of-health-and-care-data-security-and-consent

[13] https://www.theguardian.com/commentisfree/2014/feb/28/care-data-is-in-chaos

Even the specific question of opt-in vs opt-out consent was debated in the UK too, with Cambridge Professor Ross Anderson commenting, “the NHS opt-outs are like Facebook’s: the defaults are wrong, the privacy mechanisms are obscure, and they get changed whenever too many people learn to use them.” [14]

One important conclusion should have carried over from the UK: without maintaining a social license through earning the trust of individual patients, potential benefits for individuals and society will go unrealised.

4 Conclusion and Recommendations

Recommendation 1 If a data source comprises unit record-level data, assume that the individuals described are identifiable even if an attempt has been made to de-identify it.

Recommendation 2 Don’t share people’s identifiable data without their consent.

[14] `https://techscience.org/a/2015081103/`