Privacy protection for health information research in New Zealand district health boards

Vithya Yogarajan, Michael Mayo, Bernhard Pfahringer

ABSTRACT

AIM: To examine the practices used by New Zealand's 20 district health boards (DHBs) to protect patient privacy when patient information is used for research, and particularly practices for de-identifying information.

METHOD: An e-mailed questionnaire survey, using New Zealand's Official Information Act to request information on the policies and practices of each DHB.

RESULTS: 19/20 DHBs (95%) responded to the survey, one of which reported that it did not provide patient information for research. 18/18 (100%) of the DHBs that reported providing patient information for research required the project to have ethics approval. 18/18 (100%) of the DHBs that offered patient data for research also required individual patient consent and/or de-identification of the information before it was used for research. 14/18 DHBs (78%) deidentified data before releasing it for research, 8/18 DHBs (48%) sought individual patient consent before releasing data for research, and 5/18 (28%) used both methods. Other measures to protect privacy included confidentiality agreements, encryption and cybersecurity procedures.

CONCLUSION: Our findings show DHBs self-report that they have sufficient measures in place to protect privacy when patient information is used for research. However, these measures need to be continuously evaluated against rapidly evolving international practices and technological developments.

New Zealand's 20 district health boards (DHBs) potentially hold a large volume of health information about the over 4.5 million New Zealanders eligible for publicly funded health services, including medical notes, prescription records, medical images and laboratory test results. These records are potentially an invaluable resource for secondary data analysis (henceforth referred to as health information research).

There are several legal and ethical codes designed to protect the safety and privacy of patients involved in health information research. The Health and Disability Commissioner's *Code of Health and Disability Services Consumers' Rights* Regulations 1996 guarantees the rights of anyone receiving health and disability services in New Zealand. These rights include the right to have privacy respected, and the code specifically states that it also applies to those involved in research and teaching.[1]

The Health Information Privacy Code 1994 (HPIC) governs how any agency that uses health information—such as a DHB—collects, stores and uses that information, among other things.[2] The Health Research Council's *Health research and privacy: Guidance notes for health researchers and ethics committees* gives detailed guidance on how the provisions of the HPIC apply to health research in New Zealand.[3]

The National Ethics Advisory Committee's *Ethical Guidelines for Observational Studies* provides guidance on the design and conduct of health information research projects, as well as other types of observational studies.[4] This includes guidance on when an individual patient consent should be sought, and which projects have risks

that require ethics approval from the Health and Disability Ethics Committee (HDEC). The guidelines also recommend a set of controls for projects which only use anonymous or de-identified patient information, ie, information from which individual patient identity cannot be reconstructed. There is a different—much stricter—set of controls for projects that use identified or potentially re-identifiable patient information. The guidelines note that de-identification requires the *irreversible* removal of all information that could be used to identify the patient, such as name, date of birth, address and postcode.

The United States Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule goes much further and lists 18 categories of protected health information (PHI) that must be removed for health records to be considered de-identified.[5] These requirements are arguably the worldwide gold standard in de-identification and could be regarded as equivalent to the New Zealand HPIC requirement that the individual cannot be identified.[6] A key component of the present study was determining how de-identification practices used by New Zealand DHBs compare with this gold standard.

We therefore set out to identify the methods used by DHBs to protect individual patient privacy when providing information for health information research. We particularly focused on current DHB practices in de-identifying data provided for health information research, as this is a rapidly evolving field internationally.[7]

Methods

The study design was an e-mailed questionnaire survey. Information was requested from each DHB under the Official Information Act 1982 (OIA).[8] A standard letter was emailed to the appropriate contact address at each DHB, which were identified via the Ministry of Health website and individual DHB websites. If no response was received within the timeframe of 20 working days required by the OIA, a standard reminder letter was also sent. Copies of the standard letters are available from the authors on request.

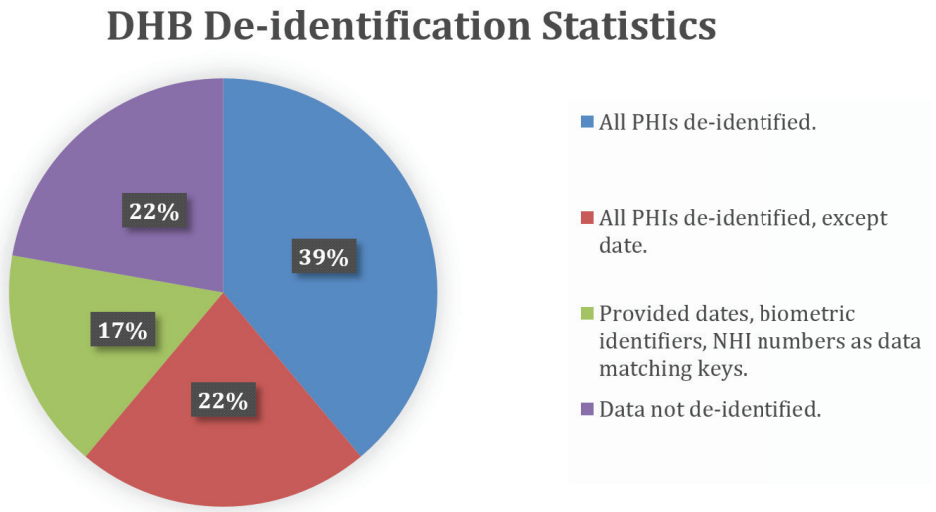
A standard set of questions were asked of each DHB, based on the 18 categories of protected health information that must be removed before health records are considered de-identified under the HIPPA Privacy Rule. Certain US-specific elements such as ZIP code and social security number were changed to their New Zealand equivalents. The full list of questions is available in the Appendix, and from the authors on request. The responses were analysed via descriptive statistics (frequencies).
Ethics approval was not required as the Official Information Act gives all New Zealand citizens the right to request information held by official bodies, including DHBs. Copies of policies, procedures and rules for decision-making by official bodies are specifically included under the remit of the OIA.[8]

Results

Out of 20 DHBs in New Zealand, 19 of them responded to the request (95% response rate). We opted not to exercise the right under the Official Information Act to appeal the one non-response to the Office of the Ombudsman, as we could not exclude the possibility that our request had not been received. To ensure that the requests and responses were as standardised as possible, we decided not to initiate verbal or written communication with the DHBs other than the two letters, although we did respond to requests for clarifications. One of the DHBs responded that it did not provide patient data for research and was excluded from further analysis.

All 18 of the remaining DHBs (100%) required research projects to go through their internal research approval processes, with referral to the Health and Disability Ethics Committee (HDEC) as needed. Interestingly, all 18 DHBs (100%) also combined ethics approval with individual patient consent and/or de-identification of the data provided, using more than one type of privacy protection.

Figure 1: Frequency distribution of the data de-identified by the DHBs.



Of the 18 DHBs, 14 (78%) provided de-identified data, either routinely, or if this was a condition specified in the research approval. The other four DHBs did not de-identify data. Among these DHBs, 7 of the 18 DHBs (39%) either de-identified all 18 categories of protected health information or did not collect them in the first place. A further four of the 18 DHBs (22%) de-identified all elements *except* dates (including dates of birth), which were supplied in the full day/month/year format. The remaining three of the 18 DHBs (17%) provided dates, biometric identifiers such as height and weight, and NHI numbers as data matching keys. Figure 1 provides the breakdown of how DHBs de-identify data.

None of the DHBs that de-identified data had policies or standard processes explicitly related to the de-identification of data. Instead, they relied on combinations of their general research policies, the requirements specified in research approvals, the Health Information Privacy Code and institutional knowledge among their staff.

Out of the 18 DHBs, eight (48%) used individual patient consents before releasing data for research, either routinely or as a condition of the research approval. All three of the precautions—research approval, de-identification and individual consent—were used by 5 of the 18 DHBs (28%). Other privacy measures named by the DHBs included staff and researcher confidentiality agreements, encrypted and password protected files, and cybersecurity procedures.

Discussion

Summary of findings

Our findings show New Zealand DHBs self-report that they have sufficient processes in place to protect patient privacy in health information research. All 18 (100%) of the DHBs that confirmed they provide patient data for research use at least two of the following three precautions: research approval, de-identification of patient data and individual consent. By doing so, they facilitate potentially valuable research while complying with relevant legal and ethical codes.

Strengths and limitations of the present study

To the best of our knowledge, the present study is the first to examine health information privacy protections across New Zealand DHBs, and particularly practices related to de-identification. A key strength is the 95% response rate. The high response rate—perhaps aided by the requirement for DHBs to answer OIA requests—minimises the possibility of response bias. Using a standard set of questions allows relatively objective comparison across DHBs.

However, a potential limitation is that the findings are based on the responses given by DHBs themselves, which may be affected by legal and reputational concerns. In addition, while a standard set of questions allows objective comparison, it limits the scope for an in-depth exploration of differences between DHBs.

The present findings could be validated by future case studies that directly observe the health information research process at individual DHBs, though gaining such direct access could be difficult for security and confidentiality reasons. A potential alternative are follow-up studies that interview key informants at each DHB about how they manage privacy requirements, potentially supplementing the descriptive findings presented here with in-depth qualitative analysis.

Comparison with existing literature

The HPIC places restrictions on collecting research information from sources other than the individual concerned, such as through health records. The HPIC also restricts the use of information collected to provide healthcare for an unrelated purpose such as research.[2,3] It similarly restricts the disclosure of health information held by the DHB to other parties such as researchers from outside the DHB.[2,3] However, there are several exceptions to these restrictions.

Among these exceptions are:

i) where the individual concerned—or an authorised representative if applicable—has authorised the collection, use or disclosure of the information;

OR

ii) where the information will only be used in a form in which the individual concerned cannot be identified;

OR

iii) where the information is to be used for research purposes (for which approval by an ethics committee has been given if applicable) **and** the information will not be published in a form that could reasonably be expected to identify the individual concerned.[2,3]

These exceptions give researchers and institutions a degree of flexibility, allowing the controls placed on each project to be tailored to the risks of that project, rather than enforcing a 'one-size-fits-all' approach. The multi-pronged approach used by DHBs fits this model, with those that use individual consent in combination with research approval leaning towards the first exception, and those who use de-identification in combination with research approval towards the latter two exceptions.

The cornerstone of the approach used by the DHBs is the ethics approval process. Other authors have noted that New Zealand's ethics approval pathways need to be strengthened to meet the challenges of evaluating health information research, which has different risks to interventional research.[9] These could include stereotyping of and discrimination against individuals or communities, heightened and self-reinforcing surveillance of those perceived to be a threat, and opportunities for financial exploitation.[9–12] It is also essential to consider the emerging risks created by powerful modern algorithmic or artificial intelligence-driven data analysis techniques, so-called 'big data'. Individuals' health information could be exposed by inference, linkage with other publicly available datasets such as voter rolls and postal address data, or information that patients have shared with commercial entities to access goods and services.[9–11,13,14] Information in the modern world is also, once publicly available, essentially 'immortal', and challenging to redact.[10,15] Such information could potentially compromise the privacy not just of the patients concerned, but also their family members and descendants.[15] Many authors have argued for new models of data research oversight that take these risks into account and are soundly based on human rights principles and international law.[9,12,16] We support these approaches, which will inevitably take time to mature. In the meantime, more widespread use of individual consent for health information research and routine de-identification could support the approval process and mitigate the risks. These approaches are complementary, but each comes with its own challenges.

Individual patient consent can increase public support, as even members of the public who are not concerned with privacy are more comfortable with their data being used for research if their consent has been sought first.[17] However, individual consent can be impractical where large numbers of patients are involved, in some cases can affect the validity of the data collected, or even be harmful to patients themselves.[5] It was also important to note that the very definition of consent is affected by how data research differs from interventional or clinical observation research. For instance, can consent be given on behalf of family members or descendants whose privacy may also be affected? Are participants

comfortable with the data being reused for other purposes, even in anonymous or aggregated form, which they may not be aware of? Are they comfortable with commercial entities having access to their records, and possibly linking this with other data those entities may have collected separately?[11,15 17] Are patients even aware of the possibility of any of these things happening?

Greater use of routine de-identification can increase rates of patient consent and public support. Members of the public are more supportive of researchers having access to their health information if the information has been de-identified.[17] Routine de-identification to the standards of the HIPAA Privacy Rule also increases the possibility of collaborations with health systems, academic institutions and public agencies that follow HIPAA or equivalent standards. However, manual de-identification of large volumes of health information is extremely challenging. Specially trained personnel who are familiar with both medical data and de-identification techniques are needed. Also, the process is time-consuming and therefore expensive. Automated de-identification of medical data via machine learning (artificial intelligence) is a rapidly developing field but has not yet reached the stage where all 18 categories of information specified in the HIPAA Privacy Rule can be consistently de-identified to 95% or greater accuracy.[18,19] Given that it is still possible to re-identify individuals from 'de-identified' data, it is also important to debate whether some level of individual consent is still needed for the collection of de-identified data, or whether there is a social consensus that the risks are acceptable when weighed against the potential gains.[20,21]

Policy implications

Individual DHBs also listed other strategies for protecting research information such as confidentiality agreements, file encryption and cybersecurity measures. Future work could include evaluating how these complement the combination of research approval, individual consent and de-identification in protecting patient privacy. A possible model for a multi-layered system to protect patient privacy in health information research has been proposed previously.[22] This applies the Reason model of error prevention—widely used in patient safety initiatives—to protecting patient privacy in health information research.[23] It also adapts the 'five safes' approach used by Statistics New Zealand to protect information in the Integrated Data Infrastructure to health records.[24]

Such a nationally standardised system could benefit DHBs by reducing legal and reputational risks. It could also address public concerns that individual DHBs take varying approaches to privacy protection, potentially giving patients living in one area protections that patients living in another area may lack ('postcode privacy', if you will). The current varying approach could be considered a natural consequence of the flexibility offered by New Zealand law and the DHB model itself, which decentralises health service provision and delegates most operational decisions to locally based and (partly locally elected) DHB boards.[25] There are also currently no authoritative national guidelines on data sharing and the use of data in health information research for DHBs to draw upon.[26] Organisations such as the Data Futures Partnership are working to develop such guidelines, and such efforts should be supported.[26]

It is important that such guidelines consider the values of the New Zealand public, and thereby build social consent for the use of health records in secondary research.[26] It is also crucial that such guidelines incorporate Māori perspectives on consent, autonomy and the rights of—and obligations to—extended family (whanau).[27] Such differences may be subtle and will naturally vary between generations and individuals. However, guidelines drawn solely from the dominant Western paradigm (which places a premium on the individual as an autonomous unit) may be too restrictive for the needs of Māori, especially considering the health disparities between Māori and non-Māori.[27]

Given the international nature of healthcare, research and information flow, New Zealand's evolving health information research guidelines and research approval processes also need to be acceptable to potential international research partners and overseas regulators. However, variances in national laws and industry codes mean there is no one universally accepted set of best practices to set future standards against, or indeed for DHBs to

measure their current practices against. For example, New Zealand's HPIC, HIPAA in the US, Australia's Federal and State privacy laws and the European Union's General Data Protection Regulation each have their unique requirements.[2,5,6,28,29] While a full comparative analysis of these laws is beyond the scope of this article, all set the most stringent requirements on the protection of identifiable individuals.[2,5,6,28,29] It stands to reason that developing and implementing routine and user-friendly de-identification practices would help ensure New Zealand's health information research is internationally accepted.

Conclusion

Our findings show that DHBs self-report they have systems in place for protecting patient privacy that meet legal and ethical standards. However, these can be strengthened further to meet the challenges posed by increasingly powerful data analysis techniques. The lack of standardised policies and procedures for de-identification increases the risk that de-identification may be of variable quality. This could be addressed either by policies at the individual DHB level, or New Zealand-wide standards equivalent to the HIPAA Privacy Rule.[5]

Appendix

Survey questions

1. Does [X District Health Board] supply patient data for research?
2. Does [X District Health Board] de-identify patient data before the data are supplied for research? If yes, which of the following elements are de-identified? (Please circle all that apply).

   a) Names

   b) All geographic subdivisions smaller than DHB catchment area (eg, postal code, street address, city)

   c) All elements of date (except year)

   d) Telephone numbers

   e) Fax numbers

   f) Electronic mail (E-mail) addresses

   g) Identifiers issued by any other Government agency, such as Inland Revenue Department (IRD) numbers

   h) National Health Index (NHI) numbers or any other medical record numbers

   i) Health insurance plan beneficiary numbers

   j) Account numbers (including patient bank account or DHB client account numbers)

   k) Certificate/license numbers

   l) Vehicle identification and serial numbers, including license plate numbers

   m) Medical device identifier and serial numbers

   n) Web Universal Resource Locators (URLs)

   o) Biometric identifiers

   p) Full face photographic images and any comparable images

   q) Any other unique identifying numbers, characteristics or codes
3. Does [X District Health Board] have a written policy or policies for de-identification of patient data before the data are supplied for research? If yes, please provide one copy of each policy, or a summary of the policy or policies [maximum one page].
4. Does [X District Health Board] have a standard process (separate from that contained in a written policy or policies) that must be followed for de-identification of patient data before the data are supplied for research? If yes, please provide a description of this process [maximum one page].
5. If [X District Health Board] has neither a written policy (or policies) **or** a standard process for de-identification of patient data before the data are supplied for research, please provide a summary of the steps that are taken to protect patient confidentiality before the data are supplied for research [maximum one page].

Competing interests:
All three authors work in and research the field of automated de-identification of health records.

Author information:
Vithya Yogarajan, PhD candidate, Department of Computer Science, The University of Waikato, Hamilton, Waikato; Michael Mayo, Senior Lecturer, Department of Computer Science, The University of Waikato, Hamilton, Waikato;
Bernhard Pfahringer, Professor, Department of Computer Science, The University of Waikato, Hamilton, Waikato.
Corresponding author:
Mrs Vithya Yogarajan, Department of Computer Science, The University of Waikato
Gate 8, Hillcrest Road, Hamilton.
vyogaraj@waikato.ac.nz

URL:
http://www.nzma.org.nz/journal/read-the-journal/all-issues/2010-2019/2018/vol-131-no-1485-9-november-2018/7736

REFERENCES:

1. Office of the Health & Disability Commissioner. Code of Health and Disability Services Consumers' Rights. [Online]. Available at: http://www.hdc.org.nz/your-rights/about-the-code/code-of-health-and-disability-services-consumers-rights/ Last accessed 10 June 2018.
2. Office of the Privacy Commissioner. Health Information Privacy Code 1994. [Online]. Available at: http://privacy.org.nz/assets/Files/Codes-of-Practice-materials/Consolidated-HIPC-current-as-of-28-Sept-17.pdf Last accessed 10 June 2018.
3. Health Research Council. Health research and privacy: Guidance notes for health researchers and ethics committees. [Online]. Available at: http://www.hrc.govt.nz/sites/default/files/Research%20involving%20personal%20health%20information.pdf Last accessed 10 June 2018
4. National Ethics Advisory Committee. 2012. Ethical Guidelines for Observational Studies: Observational research, audits and related activities. Revised edition. Wellington: Ministry of Health.
5. United States Department of Health & Human Services. Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. [Online]. Available at: http://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html Last accessed 10 June 2018.
6. Office of the Privacy Commissioner. Comparison paper on health privacy laws. [Online]. Available at: http://www.privacy.org.nz/news-and-publications/books-and-articles/comparison-paper-on-health-privacy-laws-2/ Last accessed 10 June 2018.
7. Dalianis H. Ethics and Privacy of Patient Records for Clinical Text Mining Research. In Clinical Text Mining 2018 (pp. 97–108). Springer, Cham.
8. Parliamentary Counsel Office. Official Information Act 1982. [Online]. Available at: http://www.legislation.govt.nz/act/public/1982/0156/latest/DLM64785.html. Last accessed 10 June 2018.
9. Ballantyne A, Style R. Health data research in New Zealand: updating the ethical governance framework. N Z Med J. 2017 Oct; 130(1464):64–71.
10. Office of the Privacy Commissioner. Privacy and big data. [Online]. Available at: http://www.privacy.org.nz/assets/Files/Speeches-presentations/2014-09-02-Privacy-and-big-data-MSD-Speech.pdf Last accessed 16 June 2018.
11. Zook M, Barocas S, Crawford K, Keller E, Gangadharan SP, Goodman A, Hollander R, Koenig BA, Metcalf J, Narayanan A, Nelson A. Ten simple rules for responsible big data research. PLoS computational biology. 2017 Mar 30; 13(3):e1005399.
12. Vayena E, Blasimme A. Health research with big

data: time for systemic oversight. The journal of law, medicine & ethics. 2018 Mar; 46(1):119–29.

13. Roop E. Big data creates big privacy concerns. [Online] Available at: http://www.fortherecordmag.com/archives/091012p10.shtml Last accessed 16 June 2018.
14. Heffetz O, Ligett K. Privacy and data-based research. Journal of Economic Perspectives. 2014 May; 28(2):75–98.
15. Zarate OA, Brody JG, Brown P, Ramirez-Andreotta MD, Perovich L, Matz J. Balancing benefits and risks of immortal data. Hastings Center Report. 2016 Jan 1; 46(1):36–45.
16. Ploem MC. Towards an appropriate privacy regime for medical data research. European journal of health law. 2006 Apr 1; 13(1):41–63.
17. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. International journal of medical informatics. 2012 Apr 1; 81(4):279–89.
18. Stubbs A, Filannino M, Uzuner Ö. De-identification of psychiatric intake records: overview of 2016 CEGS N-GRID shared tasks track 1. J Biomed Inform. 2017 Nov 1; 75:S4–18.
19. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: Overview of 2014 i2b2/UTHealth shared task Track 1. J Biomed Inform. 2015 Dec 1; 58:S11–9.
20. Roop, E. The de-identification dilemma. [Online]. Available at: http://www.fortherecordmag.com/archives/0515p16.shtml Last accessed 17 June 2018.
21. Doll R, Asscher W, Hurley R, Langman M, Gillon R, Strachan D, Wald N, Fletcher P. Consequences for research if use of anonymised patient data breaches confidentiality. Bmj. 1999 Nov 20; 319(7221):1366.
22. Ragupathy R, Yogarajan V. Applying the Reason Model to enhance health record research in the age of 'big data'. NZ Med J 2018 July 13; 131(1478):65–67.
23. Reason J. Human error: models and management. BMJ. 2000 Mar 18; 320(7237):768–70.
24. Statistics New Zealand. How we keep IDI and LBD data safe. [Online]. [Last accessed 07 May 2018]. Available at: http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx
25. Ashton T. Recent developments in the funding and organisation of the New Zealand health system. Australia and New Zealand Health Policy. 2005 Dec; 2(1):9.
26. The Data Futures Partnership. Our work. [Online]. [Last Accessed 09 Sept 2018], Available at: http://datafutures.co.nz/our-work-2/
27. Menkes DB, Hill CJ, Horsfall M, Jaye C. Perspectives on access to personal health information in New Zealand/Aotearoa. Anthropology & medicine. 2008 Dec 1; 15(3):199–212.
28. O'Keefe CM, Connolly CJ. Privacy and the use of health data for research. Medical Journal of Australia. 2010 Nov 1; 193(9):537–41.
29. Rumbold JM, Pierscionek B. The effect of the general data protection regulation on medical research. Journal of medical Internet research. 2017 Feb; 19(2).