**Sharing Identifiable Health Information for Use in Predictive Risk Modelling Research:**

**Advice to the Minister of Health**

Executive Summary

1. The Ministry of Social Development (MSD) is leading research to establish the feasibility of a predictive risk modelling (PRM) system which, if effective, might help in identifying children at greatest risk of maltreatment.
2. You have asked NEAC to advise you on the ethical aspects of sharing identifiable health information, held by the Ministry of Health, with MSD for use in the PRM research. This document, providing NEAC’s advice, responds to your request.
3. In this document we outline the relevant ethical standards that apply to the research, and whether the research proposal appears to meet those standards. To formulate our advice, we have considered the proposed research protocol, Privacy Impact Assessment, and application to the Central Regional Health and Disability Ethics Committee (CHDEC). We were also able to meet with the lead researchers for the project.
4. NEAC has considered only the issues presented by the PRM research itself. We note that if PRM is to be operationalised, the ethical implications will be significant and far-reaching. It would be prudent to seek further ethics advice at that stage.[[1]](#footnote-1)
5. Similarly, we note that NEAC has not considered the scientific validity or quality of scientific methodology being applied in the PRM research.

**Summary of advice**

1. NEAC’s advice is as follows:
2. NEAC considers the inclusion of health data within the proposed PRM research to present the following ethical issues:
   * access to identifiable health information without participant consent
   * the use of data linkage across different agencies’ data
   * maintaining secure data storage to ensure confidentiality
   * access to the data for future research.
3. NEAC’s assessment is that the research proposal meets the ethical standards set out in our *Ethical Guidelines for Observational Studies.* We conclude that there is a broad social benefit to the research, and that the loss of privacy to participants is very small. We note however, that the minimal loss to privacy depends entirely on the rigour with which the protocols to safeguard the data are applied. We cannot emphasise strongly enough the need for MSD to ensure that the research is carried out in strict compliance with the safeguards that have been proposed.
4. Researchers undertaking observational studies are ethically obliged to publicise the study results by whatever suitable means are available. Publication of both positive and negative study results is of value. This is because research is most useful if it is shared with those who can use it – frontline professionals and policy makers in New Zealand and internationally. It is important that MSD agrees to publish the PRM research findings, favourable or otherwise, as an essential output of the project.
5. NEAC is concerned that the researchers’ proposal to delete the de-identified datasets in June 2015 will not allow for further research beyond that date. Making the dataset available for other researchers (including for replicating and confirming the study’s findings) is an ethical obligation of observational research. We note, however, that additional governance arrangements and privacy safeguards would be required to ensure appropriate storage and use of the data over a longer term.
6. NEAC notes that any proposals for future use of the de-identified linked data would require a return to the Health and Disability Ethics Committee for consideration.
7. The PRM research exemplifies the growth in opportunities for research linking identifiable health information and other administrative data. It may be timely to engage public discussion on the range of modern uses for health data, how such research might be beneficial and mechanisms for ensuring health information is used safely and appropriately into the future.

Introduction

**Context**

1. The White Paper for Vulnerable Children and associated Children’s Action Plan include a range of proposals to protect vulnerable children at risk of maltreatment. A key proposal is to develop and trial predictive risk models which, if effective, might help front-line professionals in identifying and assessing children at greatest risk of abuse or neglect.
2. The Ministry of Social Development (MSD) is leading a programme of research to establish the feasibility of population-wide predictive risk modelling (PRM) for New Zealand. It is proposed that data held by a variety of agencies, including health data held by the Ministry of Health, be added to the model over the next few months to test whether this achieves further gains in the tool’s predictive power.
3. The health data sought includes known risk factors for child maltreatment, such as maternal mental health, together with variables the researchers hypothesise will help identify child abuse and neglect in the broader (non-beneficiary) population. The research project will determine which, if any, of the health variables strengthen the existing model. A possible outcome of the research is that no health information, or only a sub-set of health factors, would be included in a final PRM.
4. You have asked NEAC to advise you on the ethical aspects of sharing identifiable health information with MSD for use in the PRM system research. We note that the Central Health and Disability Ethics Committee (CHDEC) has reviewed and approved the research proposal.

**The nature and limits of ethics and ethical standards**

1. Ethics is about how people should behave towards each other. While clear rules have their place in guiding behaviour, the exercise of judgement is unavoidable. Judgement is particularly necessary when a choice must be made between competing interests. By its nature, the judgement involved in weighing up goods and bads cannot be expressed in clear, complete and exceptionless rules.
2. Consequently, ethics is not exact. Because ethics is not exact, and because ethical principles were drawn up by fallible humans, ethical standards cannot be expected to give clear, sensible guidance in every case. Ethical standards therefore have to be sufficiently adaptable to reflect context.

Ethical considerations and key issues

1. NEAC considers a number of underlying principles to be important for the ethics of observational studies (that is, the type of research that MSD proposes to conduct). The principles most relevant to the PRM proposal are as follows.

* **Respect for people**: respecting people, and their rights, requires respecting their capacity for self-determination (including the right to give informed consent to participate in research or to refuse) and the right to control or limit access to personal information. Respect for people also requires that people who are dependent or vulnerable be afforded security against harm.
* **Justice**: within a population there should be a fair distribution of burdens and benefits. Stigmatising individuals or groups, including through the use or disclosure of research findings, should be avoided.
* **Beneficence and non-maleficence**: the risks that a study poses should be reasonable in light of the anticipated benefits. However, not all ethical considerations weigh equally. A study can be ethically acceptable even if some expectations are not met, such as confidentiality, provided the potential benefits outweigh the risks and the risks can be minimised.

1. In our view, the areas of ethical concern presented by the PRM proposal are:

* access to identifiable health information without participant consent
* the use of data linkage across different government agencies’ data
* maintaining secure data storage to ensure confidentiality
* access to the data for future research.

1. We consider each of these items in turn below.

**Access to identifiable health information without consent**

1. Health information is recorded and stored in a variety of forms: identifiable, potentially identifiable, partially de-identified, de-identified, and anonymous (see Appendix 1 for definitions).
2. Ordinarily, individual (and sometimes collective) participant consent is required for researchers to access identifiable health information. However, international practice and NEAC’s *Ethical Guidelines for Observational Studies: Observational Research, Audits and Related Activities* (EGOS) recognise that there are circumstances when departure from this principle may be justified. Specifically, when:
   1. it is not feasible to seek consent due to the quantity or age of records involved, where the procedures required to obtain consent are likely to cause unnecessary anxiety for potential participants, or where seeking consent would prejudice the scientific validity of the study; and
   2. there would be no disadvantage to the participants, their relatives or any collectives involved; and
   3. the public interest in the research outweighs the public interest in privacy; and
   4. the researchers have justified to an ethics committee the reasons for not seeking consent, demonstrated how confidentiality will be protected, and shown that the study has the goal of protecting or advancing health.[[2]](#footnote-2)
3. We agree with the researchers’ assessment that, given the large numbers of health records proposed to be transferred to the study (multiple cohorts of 60,000), it would not be feasible to seek consent from all participants.
4. We also consider that the research itself does not pose any disadvantage to participants or their relatives, beyond loss of privacy (discussed further below). The researchers’ distinctions between the research and its potential use are helpful in clarifying that the research alone will not directly cause losses to participants. For example, the researchers are exempt from the standard MSD requirement to report any benefit fraud identified during the course of the research.[[3]](#footnote-3)

*The key test – weighing up public interest in the research against interest in privacy*

1. Less immediately certain is whether the public interest in this research outweighs the public interest in privacy. NEAC regarded the following factors as important for making a judgement:

* the nature of the health information sought by the researchers (that is, an individual’s expectation of confidentiality may be greater for some types of information than others)
* the degree to which the loss of privacy will be minimised
* the health and wider social benefits of undertaking the research, which form the ‘public interest’ in this situation.

1. The health information sought covers a wide range of variables, some of which are particularly sensitive. Variables of interest include use of maternal addiction services, maternal mood and/or anxiety disorders, and previous births. It is reasonable to assume that some individuals would prefer to keep this information private.
2. On the other hand, the research design includes strategies to limit the extent to which the data is both identifiable and accessible in identifiable form to researchers. The research protocol aims to transfer identifiable data into a de-identified form as soon as the data linkage is complete, and the process of de-identification will not be reversible. Access to the identifiable data will be limited to three research staff who, in practice, are highly unlikely to see the data in its complete form.[[4]](#footnote-4) Our assessment is that the loss of privacy for the participants is therefore very small.

*Assessing the public interest in the research*

1. At issue then, is the extent of public interest in the research proceeding. In making a judgement, we were cautious about not conflating the benefit of the research with the possible social benefit if:
   1. the research is successful; and
   2. Government chooses to operationalise the model; and
   3. the PRM tool is effective in contributing to a reduction in child abuse and neglect.
2. Most research that falls within NEAC guidelines and/or health and disability committee ethical review is aimed at improving the understanding and management of a health condition, or ensuring the safety and efficacy of health care services. Patient information is usually used in the context of ensuring the high quality delivery of health or disability services.
3. The PRM research does not, at this point, fall into this same category. The immediate goal of the research is to test whether adding health data to the dataset can improve the power of a tool meant to identify children at risk of maltreatment. A second objective is to generate new population-level knowledge about children who are the subject of abuse and neglect. This strand of research will inform the design and delivery of programmes and services aimed at preventing child maltreatment more generally. The potential public interest is therefore of a wider social nature.
4. Taking a wider social benefit perspective, NEAC concludes that there is a good degree of public interest in the research for two important reasons.
5. Firstly, as the current study will influence further decisions about the development and trialling of operational predictive risk models, it is in the public interest that the model be as robust as possible so that informed decisions can be made. NEAC agrees that developing a robust model requires careful and thorough testing of a wide range of predictors, including those which, while sensitive, have been associated in other studies with child maltreatment. Testing whether inclusion of different types of health data adds to the predictive power of the model (and conversely, whether robust predictions could be made with fewer or no health related variables) is therefore important for strengthening the quality and utility of the research.
6. Including health information at this stage of this research could also be of dual benefit – by starting with a wide range of data, the research will be able to narrow down to only those variables essential for the predictive power of the model. This would hopefully minimise the amount of sensitive data required for a fully implemented PRM, and therefore the loss of privacy that may result.
7. Secondly, the public is entitled to services that are safe and effective. When new knowledge created through research is translated into action that improves the delivery of public services, the public benefits from that research. Because the PRM project breaks new ground for use of predictive risk modelling and data linkage in understanding child maltreatment, it has the potential to create valuable new knowledge that can be used by frontline professionals and policy makers both in New Zealand and internationally. For example, the addition of health data to the model might provide a completely new window on predicting maltreatment in families that would otherwise not be picked up by other datasets or tools.
8. NEAC notes however, that these benefits will only be realised if the research findings – favourable or otherwise – are disseminated widely through publication. Indeed, the potential benefits of observational studies include providing objective results that can be used to formulate sound policy. Investigators have an ethical obligation to release information that is in the public interest, and study results should be publicised by whatever suitable means are available. The publication of both positive and negative study results is important.[[5]](#footnote-5) NEAC therefore considers that publication be an essential output of the PRM project.

*NEAC’s view: the public interest outweighs privacy considerations – with caveats*

1. As the participants will suffer no apparent disadvantage, other than disclosure of personal information into the study, we conclude that the research is of sufficient public interest to outweigh privacy considerations, **providing**:
   1. **The loss of privacy is minimised**; **and**
   2. **The findings of the research are widely disseminated through publication**.
2. NEAC’s view is that the design of the study and strategies to safeguard confidentiality of the data, if applied as proposed, are appropriate to ensure that this is the case. Provided that MSD agrees to publish the results of the research project, the ethical standard is therefore met.

**Data linkage across government agencies**

1. When researchers seek to link records containing identifiable information without consent, they should ensure that the participant’s identity is not disclosed further (apart from for linking the records) and is disposed of once the information has been linked.[[6]](#footnote-6) Identifiable information must be securely maintained and the research should have the potential to benefit to the public.
2. The research protocol proposes that the Ministry of Health provide a range of health data to MSD. As noted earlier, the health data sought, in some instances, will be highly sensitive. When linked to other data held by MSD, the range of information available about each individual would be greater than in any of the contributing datasets.
3. The research proposal contains extensive information about how the data linking will work. The MSD research team will assign each record a temporary identifier and create a match map to link these across all key agency datasets. A unique research identifier will then be assigned to the linked data set. After this point, the original identifying information will be deleted permanently.
4. Only three people – permanent staff members of MSD, bound by a Code of Conduct and confidentiality agreements – will have access to the identifying data until it is deleted. In addition, access to the data will only take place in a secure environment. NEAC understands that MSD and the Ministry of Health intend to sign a Memorandum of Understanding (MOU) outlining expectations for data security.
5. In our view, these precautions are sensible. Provided the research protocol and MOU are followed as proposed[[7]](#footnote-7), they should ensure that access to identifiable data is limited to key personnel, for as short a time as necessary. We consider that this is consistent with ethical standards.

**Secure data storage**

1. There are risks associated with collecting and storing confidential personal information. Inappropriate or inadvertent disclosure of personal information, particularly sensitive information, can cause significant distress for those affected, and result in loss of public confidence in agencies (or health providers) charged with the responsibility for the safeguarding of information.
2. The PRM proposal contains a particular risk because disclosure of the linked data could result in a greater loss of confidentiality than if the security of any individual dataset were compromised.
3. In practical terms, these risks can be minimised by deleting the identifying data following linkage, secure data warehousing and ensuring that those who seek to access the information are bound to maintain confidentiality. The research proposal builds in appropriate strategies to ensure the safe storage of data (including a separate and secure IT environment) and preservation of confidentiality.
4. However, such strategies are only as credible as the individuals implementing them. Much will depend on the actual conduct of the research. We note MSD’s plans to carry out an annual security audit, and to report the results to the Central Health and Disability Ethics Committee. However, day-to-day vigilance will also be crucial.

**Access to the datasets for future research**

1. Principle nine of the Privacy Act 1993 requires that personal information not be kept longer than necessary. The researchers propose to delete the de-identified datasets by June 2015.
2. Whilst NEAC acknowledges the good intent behind this decision, the expectation from an ethics perspective is that these datasets be retained for future use, including for quality checking at a later date. Indeed, one aspect in favour of undertaking the PRM project, in terms of public interest, is the potential for further research using the extremely rich dataset that will be created.
3. The research proposal contemplates other uses of the data, but only in the short-term. NEAC’s concern is that the June 2015 date provides too short a period for any substantive research outside of the immediate project to take place. It will also prevent any future check of the quality of the research findings.
4. We note, however, that our support for retaining the datasets does not negate our earlier emphasis on rigorous privacy safeguards. We consider the safeguards currently proposed to be satisfactory for the existing research protocol. If the datasets were to be retained for a longer period (as we recommend) we expect additional safeguards, in terms of appropriate governance and strict data storage arrangements, would need to be agreed.
5. This is particularly important as, due to the breadth of data held, individuals may still be identifiable even after the dataset is de-identified.[[8]](#footnote-8) Although this is highly unlikely while access is confined to MSD’s PRM research team, it may become more of a concern if the datasets are accessible to a larger pool of researchers.
6. If the datasets are to be used for another purpose and/or by people other than those specified in the approved research protocol, the researchers would need to submit a new (or revised) protocol to a Health and Disability Ethics Committee. We note that this standard applies for **any** other use of the dataset, pre or post June 2015.

Conclusion

1. The PRM research appears to break new ground for use of health information in predictive modelling. NEAC has been unable to identify any similar research that brings together health information with the breadth of other factors being linked in the proposed model, for a purpose that is not directly health related. Other studies of similar magnitude and characteristics in New Zealand (such as the Cancer Trends and the Census-Mortality studies) are essentially health-oriented.
2. Viewed overall, whilst the PRM research poses some risks in relation to non-consensual access to identifiable health information, stigmatisation, and data security, the researchers have proposed ways of appropriately minimising these. In relation to these concerns, the research meets NEAC’s ethical standards. However, we cannot emphasise strongly enough the need for MSD to ensure that the research is carried out in strict compliance with the protocol as submitted to the CHDEC.
3. In forming this advice, our central concern was the ethics of balancing public interest and privacy in the research. We are aware that there is a potential public interest in *not* adding health data to the project – loss of trust in the health system resulting from a substantial privacy breach (if this were to occur) would present a serious risk that New Zealanders become less inclined to provide accurate health information about themselves and their children.
4. Nonetheless, this research may contribute to better identifying and assisting children at risk of maltreatment – an aim which has broad social benefit. The knowledge gained by testing a PRM tool in this context, if shared appropriately, will also be valuable in both New Zealand and abroad. Finally, the project is exploratory in evaluating which, if any, types of health data are useful for a predictive modelling approach. This is important for informing future decisions about the model.
5. For these reasons, NEAC considers that public interest outweighs privacy concerns at this stage, provided that the data security concerns are managed and the research findings are shared.
6. We note however, that if the intention is for the PRM tool to be operationalised, the ethical issues involved would require separate and significant consideration. Without knowing precisely how such a tool might be applied, and what services would be available to address the risks identified, it would be premature for us to comment on those issues.

**The need for public discussion about the use of health data**

1. NEAC notes that this proposal, and many other longitudinal or cohort studies, involves access to identifiable health information and data linkage for research purposes. Growth in administrative databases, e-health records, blood and tissue banking, and sophisticated statistical software packages, means that the opportunities for such research are likely to increase.
2. It is unclear whether members of the public are aware of the range of uses to which such data may be put, why this might be necessary and how it might be beneficial. It may be timely to open this topic up for public discussion, with a view to ensuring public confidence in the safe and appropriate use of health information into the future.

Appendix 1: Identifiability of Health Information - Definitions

Identifiable Data allows a specific individual to be identified. Identifiers may include the individuals name, date of birth or address.

Potentially identifiable Personally identifiable data is separated from substantive data, but a potential link is maintained by assigning an arbitrary code number to each data-identifier pair before splitting them. The key allows the re-associating of the substantive data with the identifiers, under specified conditions.

Partially de-identified Data is coded with abbreviate identifiers, such as initials, date of birth, or sex. This enables the reporting clinician to re-identify the data, but it is otherwise anonymous to other recipients of the data.

De-identified Data is completely anonymised and cannot be re-identified, i.e. any link to the original identifying data has been removed.

Anonymous No identifying data has been collected from the outset.

1. NEAC understands that MSD is in the process of commissioning an ethical review of the ‘use of PRM to identify children at high risk of maltreatment’. The Terms of Reference for that review are still to be agreed. [↑](#footnote-ref-1)
2. For points a) to c) see paragraph 6.43 of EGOS. For point d), see paragraph 6.45 of the same document. [↑](#footnote-ref-2)
3. MSD Privacy Impact Assessment, p.17. [↑](#footnote-ref-3)
4. NEAC understands that the identifying personal information (names, dates of birth, addresses etc.) will be separated from the other variables during the manual stages of developing and checking the programming code used to link the data. Consequently, the researchers will see the personal information without the added variables, or the other variables without the personal information. They are likely to see few, if any, complete records. [↑](#footnote-ref-4)
5. See EGOS paragraphs 10.1 and 10.3 – 10.5. [↑](#footnote-ref-5)
6. EGOS paragraph 8.11. [↑](#footnote-ref-6)
7. Of course, the researchers should also be expected to make improvements to the data security protocols in the course of the research should practice reveal that aspects are not working or could be done better. [↑](#footnote-ref-7)
8. This is because the wide range of pieces of information about an individual might, when linked, make that person identifiable to someone with sufficient knowledge of the individual to recognise them from their data (even if their personal information has been deleted). [↑](#footnote-ref-8)