MINISTRY OF
SOCIAL DEVELOPMENT
Te Manatū Whakahiato Ora

Date: 7 February 2013 **Security Level:** IN CONFIDENCE

Privacy Impact Assessment for Population-level Child Maltreatment Research based on Linked Administrative Data

1 Introduction and overview

Introduction

1 This Privacy Impact Assessment has been prepared for the linkage of administrative data held across a range of government agencies for the purposes of research on child maltreatment.

2 The research will inform and support the implementation of the Government's White Paper on Vulnerable Children by:

- investigating the feasibility, data requirements and predictive power of population-wide Predictive Risk Models that predict the risk of a child being maltreated, focussing initially on a Predictive Risk Model that would apply to all New Zealand-born children at or around the time of birth
- generating new population-level knowledge about the prevalence of reports and substantiations of maltreatment and hospitalisations for intentional and non-intentional injuries, the degree to which the children who are the subject of these events overlap, their backgrounds and outcomes, and Māori – non-Māori differences.

3 This Privacy Impact Statement has been prepared to provide assurance that:

- the research project complies with all relevant privacy and data-use legislation and codes
- potential privacy risks and public perception risks are identified and high quality systems are put in place to mitigate those risks.

4 Data linkages will be sought between the following:

- Ministry of Social Development - benefit, care and protection, and Child Youth and Family youth justice data
- Registrar General of Births, Deaths and Marriages - birth and death information
- Ministry of Business, Innovation and Employment - migration data
- Corrections - imprisonment and community sentences data

- Ministry of Health - maternity, mental health, B4School Check, hospitalisations and injury-related mortality data
- Ministry of Education – educational participation and attainment data.

5 Data linking for any operational Predictive Risk Models raises a different and separate set of privacy and legal concerns to those that arise in the context of data linking for research. Any operational Predictive Risk Model will require its own Privacy Impact Assessment.

6 Public confidence and acceptance are concerns that need to be properly managed. In the context of data linking for research, one option for addressing these concerns and managing the risks inherent in data integration is for Statistics New Zealand to link the data and be the custodian of the resulting research dataset. This was an approach given support in a 1997 Cabinet agreement that:

> where datasets are integrated across agencies from information collected for unrelated purposes, Statistics NZ should be the custodian of these datasets in order to ensure public confidence in the protection of individual records (CAB (97) M 31/14).

7 As with the 2006-2011 Working for Families Evaluation,[1] this is not a practical solution for the current research. As a result, the data will be linked and held by MSD. All efforts will be made to emulate the data protections that Statistics New Zealand custodianship would provide.

8 At a future date, Statistics New Zealand and MSD will discuss the development and custodianship of a long-term data linkage infrastructure for the social sector, and the role that Statistics New Zealand could play in its governance.

Overview

9 This report begins with a description of the legal context. It then describes the research project and its origins, the data to be linked and the flows of information. Privacy issues associated with the collection, integration, storage, and use of the source data are discussed, along with a description of the release and access practices to be used. This discussion includes an assessment of the risks inherent in these processes, and the privacy enhancement and risk management procedures that will mitigate them. The report outlines compliance mechanisms before concluding.

[1] The Working for Families Evaluation provides a precedent where Statistics New Zealand (SNZ) agreed that the 1997 Cabinet decision should not prevent Inland Revenue undertaking the initial conceptual and prototyping linking of MSD benefit administration data and Inland Revenue tax administration data. It was agreed that the evaluation based on an integration of these data held outside of SNZ could proceed, but that consideration should be given to SNZ being future custodians of the linked dataset. In the event, SNZ did not wish to become custodians and this linked research dataset has remained with Inland Revenue.

Legal context

Privacy Act 1993

10 The Privacy Act 1993 provides protection to information about an individual and applies to every agency that deals with personal information. The 12 information privacy principles in the Privacy Act 1993 provide the foundation upon which the Privacy Act 1993 controls the collection, use, disclosure, storage and access to personal information.

11 Under the terms of the Act the Privacy Commissioner must retain the capacity to conduct an independent review in the event of a complaint. As a consequence the Privacy Commissioner is not able to approve proposals such as this in advance. However, the Privacy Commissioner is able to signal any practices that are not permitted under the Act or that might pose a problem of perceived privacy risks.

12 MSD has taken the position that any such concerns, even of perception, should be addressed in an appropriate and defensible manner. The intention is that all necessary steps are taken to comply with both the spirit and letter of the Act.

13 None of the data linkages that will be undertaken for the research constitute an information matching programme as in Part 10 of the Privacy Act as the data are not being shared for the purpose of producing or verifying information that may be used for the purpose of taking adverse action against an identifiable individual.

14 The use and disclosure of data involved in the data linkages required come within exceptions to Privacy Principles 10(f)(ii) and 11(h)(ii), in that the information will be used for research purposes and will not be published in any form that could reasonably be expected to identify the individuals concerned.

15 At the same time, other principles of the Privacy Act require:

- reasonable steps to be taken to ensure that personal information is protected from loss, unauthorised access, modification or disclosure, and other misuse (Principle 5)
- reasonable steps to be taken to check personal information is correct before it is used (Principle 8)
- personal information not to be kept longer than necessary (Principle 9).

16 The Privacy Analysis set out in Section 3 provides details of measures to ensure compliance with these requirements.

Health Information Privacy Code 1994

17 The Health Information Privacy Code recognises an expectation that health information should be treated differently from other information held about individuals. It applies specific rules to agencies within the health sector to better ensure the protection of individual privacy. With respect to health information collected, used, held and disclosed by health agencies, the code substitutes for the Privacy Principles in the Privacy Act.

18 The use and disclosure of health data involved in the data linking comes within the exception to Rule 11(2)(c)(iii) of the Health Information Privacy Code which provides for

disclosure of Health information without the authorisation of the person concerned where the agency holding the data believes on reasonable grounds that it is either not desirable or not practicable to obtain authorisation from the individual concerned and where the information is to be used for research purposes (for which approval by an ethics committee, if required, has been given) and will not be published in a form that could reasonably be expected to identify the individual concerned.

Adoption Act 1995

19 Ministry of Health maternity data is sought for the research. This data holds identifiable information on mothers and live born babies. Disclosure of linked mother-baby information to parties directly involved in a closed adoption compromises the confidentiality of the adoption and is in breach of the implicit prohibition placed on disclosing adoption information under Section 23 of the Adoption Act 1995. Such a breach could occur where:

- the research project involves directly contacting participants who have been involved in a closed adoption *and* the study releases the linked information to the participants, or
- the authorised recipient is involved in a closed adoption *and* the research project legitimately includes the recipient within the research cohort.

20 The likelihood of these scenarios occurring are extremely remote, particularly as the number of closed adoptions is understood to be low for the time period a mother-baby link is available from the maternity data (2002 onwards). However, as it is not possible to robustly identify and exclude data relating to a closed adoption, any release of identifiable linked mother-baby information poses this risk.

21 The Ministry of Health needs to balance the risk to privacy of individuals against the potential public good of the research being undertaken. As guardians of identifiable health information and of closed adoption records as defined in the Adoption Act 1955, the Ministry of Health has a responsibility to the individuals it holds information for. In the case of this research, the Ministry of Health agrees that any release of health data should include identifiable child data together with identifiable mother data, as the risk of a breach of privacy is low given the very limited access to identifiable data.

Births, Deaths, Marriages, and Relationships Registration Act 1995

22 Section 75G(1)(b)(ii)(B) of the Births, Deaths, Marriages, and Relationships Registration Act 1995 allows for disclosure of administrative information relating to births, deaths and marriages where the request for information is for the purposes of health research and the Registrar-General is satisfied that in providing the information, the public benefit outweighs the effect on individual privacy.

23 When considering requests for information made on these grounds, the Registrar-General must consult with and invite comments from the Privacy Commissioner, and take into account the number of individuals whose privacy will be affected, the degree to which each individual's privacy will be affected, whether an ethics review committee or a similar body has considered and approved the research and, if so, the persons making up the committee or body and the type of work it undertakes, how the agency, body, or person undertaking the research proposes to hold, use, and, if relevant, dispose of the information obtained.

2 Description of the Project and information flows

Description of the Project

Background

24 In 2011, as part of a research strategy focussed on the most vulnerable children, MSD developed a new capacity to link data from across its care and protection, youth justice and benefit service arms. For the first time, this allowed an examination of the numbers of unique children having contact with these different services over the course of their childhood, and estimation of the population-level period prevalence of different combinations of service usage (Centre for Social Research and Evaluation, 2012).

25 A Ministerial priority for more in-depth research based on the linked MSD data was to investigate the feasibility and the potential benefits and risks of developing and using models to prospectively identify children at high risk of substantiated maltreatment.

26 Auckland University was commissioned to carry out this research. They developed a demonstration predictive risk model for the estimated 33 percent of children who have contact with the benefit system before age two. Predictive risk models harvest administrative data to automatically generate an estimate of the risk that an individual will have an adverse outcome in the future. The model developed had fair, approaching good, power in predicting which of the young children having contact with the benefit system would be the subject of substantiated maltreatment by age five.

27 The most at-risk children identified by the model represented 37 percent of all children in the cohort studied who went on to have substantiated findings of maltreatment by age 5. This most at risk group comprised five percent of children in the cohort overall.

28 Close to two in five of the children in this most at-risk group had at least one substantiated finding of maltreatment by age five, and among those for whom a longer follow-up was possible, half had at least one substantiated finding of maltreatment after nine years (Vaithianathan *et al*., 2012).

29 The authors recommended a full ethical evaluation of predictive risk modelling and the development of an ethical framework to guide agencies in their responses to the use of automated child risk scores before any implementation. Their preliminary ethical analysis suggested that mandatory policies for high risk families need to be treated extremely cautiously. Far fewer ethical concerns were anticipated if scores are to be used to engage high risk families in voluntary services.

30 At the same time, the Auckland University findings highlighted ethical concerns about applying a Predictive Risk Model for maltreatment only to children known to the benefit system. The authors recommended broadening the research data in order to include all children, and including community-level characteristics as predictors (Vaithianathan *et al*., 2012).

31 The White Paper for Vulnerable Children includes developing and trialling predictive risk models that draw on linked social sector administrative data. These models, operationalised, would aim to assist front-line professionals in the early identification of children at highest risk of maltreatment as part of a preventive strategy. Successful

early intervention can help reduce exposure to maltreatment and its long-term health and social consequences (MacMillan *et al*., 2009; WHO, 2012).

The Project

32 The data linkage outlined in this Privacy Impact Assessment forms part of a new programme of research to support the implementation of the White Paper (the "Project"). MSD care and protection, youth justice and benefit data will be linked with administrative data held in other systems, including birth information (providing near comprehensive overage of New Zealand-born children and demographic information about the child and parents), and health, education, migration and corrections data.

33 The research will develop a population-wide research predictive risk model for identifying children at highest risk of maltreatment based on linked data, focussing initially on a model that would apply at or around the time of birth. It will assess the performance of the model and its feasibility as an operational tool, and identify which administrative data are needed in order to make the best predictions of risk. Findings will inform implementation of the White Paper, including further decisions on the development and trialling of operational predictive risk models.

34 The linked data will also be analysed to generate new population-level knowledge about children who are the subject of reported and substantiated maltreatment and hospitalisation for non-accidental injuries. Linked data can be used to explore the extent to which the same children are presenting to different services (Gilbert *et al.* 2011), and to better understand children's backgrounds and outcomes. This will include an analysis of Māori children's experiences and Māori – non-Māori differences. This strand of the research will inform the design and delivery of programmes and services aimed at preventing child maltreatment.

Information flows

Inflows of data

35 For selected cohorts of children, the research will link:

- birth and death information including identifiable information on child and parents
- migration data for child
- health data for child and mother
- benefit data for parents/caregivers
- care and protection and youth justice data for child, other children in the family, and parents/caregivers
- education data for child and parents/caregivers
- corrections data for parents/caregivers.

36 The research will entail a single inflow of data to be linked (ie. there will be no updates after the initial linkages have been formed). Figure 1 sets out the information flows.

37 In most cases, the flow of data will be uni-directional. Agencies will identify records of interest based on the date of birth of the child and pass these by way of secure transfer to MSD for linkage.

38 The exceptions to this will be Corrections and Ministry of Education data relating to parents/caregivers. In the case of these data, agencies are unable to select data to supply based on date of birth. In order to avoid seeking more information than is required for the research, the identities of the parents and caregivers of interest will be passed by way of secure transfer to these agencies together with a unique research identifier. Agencies will pass back to MSD by way of secure transfer the unique research identifier and the variables sought (identifying information will not be required).

39 There are no unique identifiers for children (eg National Health Index or National Student Number) that are used across agencies in New Zealand. Agencies' unique identifiers cannot, therefore, assist with the linking of data and will not be included in the raw agency datasets that contribute to the research. The exception to this is birth and death registration numbers which will be included in data requested from the Registrar General of Births, Deaths and Marriages and will be used to match deaths to matching births.

Secure transfer

40 Agency datasets will arrive on CD, iron key, or other secure delivery device. To manage risks as data are passed from one agency to another, all agencies will encrypt their data in accordance with the recommendation from the Office of the Privacy Commissioner. A password (of sufficient strength) to unencrypt the data will be supplied.

41 The data will be collected from the agencies by a MSD employee, or delivered by a representative of the agency. Passwords will not be written or carried with the data and no other stops will be made by the person carrying the data. Data mediums will be destroyed, or returned in the same manner. A log will be kept of external data transfers including recording the contact person at the originating agency and summarising the access and deletion arrangements for the data.

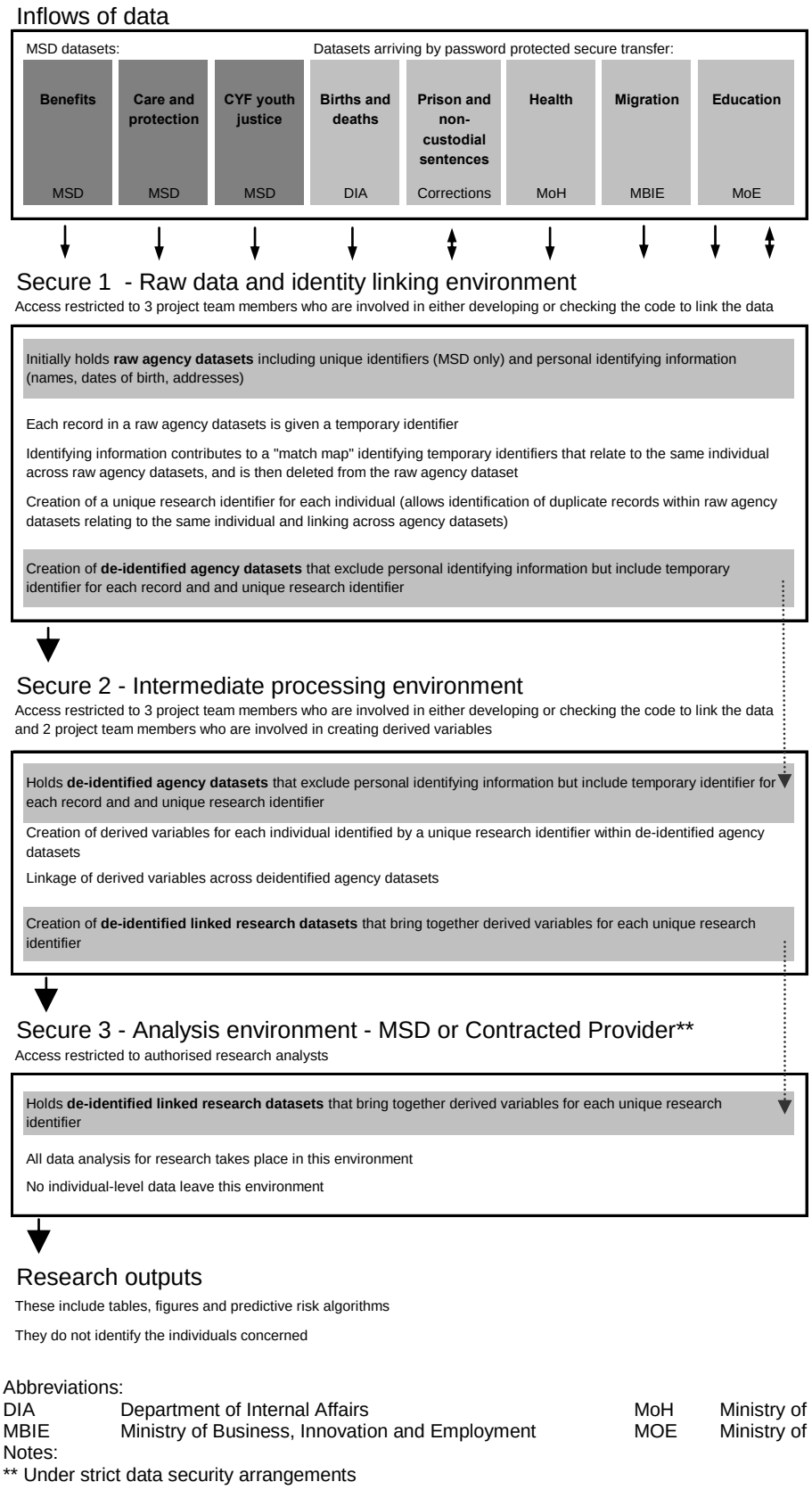
Secure holding and handling of identified data

42 MSD will create three secure data storage environments within which the data will be held, linked and analysed. Access to each of these environments will be strictly limited to members of the research team and authorised system administrators, with varying permissions depending on their roles and responsibilities. In addition to the research team, two IT systems administrators will have access to the secure environments. These staff will not have any role in linking or analysing the data.

43 Data linkage across agency datasets requires matching using name, any aliases, date of birth, and potentially other identifying variables such as address and relationship information. It is when the individual records remain identified by this information that the highest potential privacy risk exists. Any casual or intentional access to the data at this time might allow those viewing the data to access personal information.

44 To mitigate this risk, on arrival, the data will be moved to a secure "raw data and identity linking" environment with restricted access within the project team. Each record on the incoming files will be given a temporary identifier. Identifying information will contribute to the creation of a "match map" identifying temporary identifiers that relate to the same individual across all raw agency datasets, and then be deleted from the incoming raw agency data file. The match map, which holds identifying information but no other personal information, will be securely held in the "raw data and identity linking" environment until all linkages have been completed for the research and then deleted.

Figure 1 Flows of personal information



Abbreviations:

DIA	Department of Internal Affairs	MoH	Ministry of Health
MBIE	Ministry of Business, Innovation and Employment	MOE	Ministry of Education

Notes:

** Under strict data security arrangements

Manual inspection of a sample of records

45 A large task for the research will be the linkage of data. It is expected that, as a result of reporting error, administrative error or fraud, the same individual may have more than one identity within a system (eg. a child may have more than one benefit identity). This necessitates within-system linking ("de-duplication") to identify unique individuals.

46 Manual inspection of samples of records will be undertaken to access accuracy in the linking. Individual data for these samples will need to be downloaded to spreadsheets for clerical assessment of the quality of the electronic linking. These spreadsheets will be held in password encrypted zip files and stored in a folder that can be viewed only by members of the project team who are involved in this part of the project. The spreadsheets will not include any details of the variables being analysed (eg. the care and protection or birth information of the individual). They will include only details relevant to verifying the electronic linkage (ie. names, date of birth, gender, address history and relationships).

47 Any printouts to enable the checking to proceed will be securely stored, and will be destroyed immediately (by shredding) at the completion of checking.

Data analysis

48 De-identified, de-duplicated linked datasets will be created and analysed within a secure data storage environment. Research outputs will include tables and figures and predictive risk model algorithms. No individual-level data will be included in research outputs, reports or briefings.

49 Contracted providers may be engaged to analyse de-identified research datasets. This will only occur under strict security arrangements that require the provider to emulate the secure data storage environment outlined in Figure 1 and comply with all data security, data use and data destruction arrangements outlined in this document.

50 The de-identified data sets will be permanently deleted when the Research Project ends in June 2015.

51 Interagency project governance will be provided through the Vulnerable Children's Board, which is made up of social sector Chief Executives. Other research proposals that seek to draw on the de-identified data sets may be developed. In some cases, these may involve other Government agencies or contracted providers. Where proposals gain consensus approval from the Vulnerable Children's Board and ethics approval (if required) and have a Privacy Impact Assessment in place, provision will be made for de-identified linked datasets to be supplied for the purposes of the research under strict confidentiality agreements and only where those organisations are able to meet high data security standards and comply with the Privacy Act in respect of the use, disclosure and destruction of the data. Where the de-identified linked dataset to be supplied for the purposes of the research includes Ministry of Health data, approval of the Director General of Health will also be required before the data are supplied. Where it includes birth and death information, approval of the Registrar General will be required.

3 The privacy analysis

Privacy Act 1993

52 The Privacy Act 1993 aims to promote and protect individual privacy. It relates to personal information. In section 6, twelve principles relating to the collection, storage, security, access, retention, use and disclosure of personal information are outlined. Several of the principles provide for exemption on the grounds that the information is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned. Regardless of these exemptions, it is important to consider the ideals expressed by the principles.[2]

53 This section describes the privacy issues associated with using personal information in the Project, working through the Principles in turn.

Collecting and obtaining information

54 *Principle 1: Purpose of collection of personal information.* The Project draws on existing administrative data that is the necessary information collected or generated as part of the process of lawful administration of Government services. Collection of these data at source complies with Principle 1 of the Privacy Act.

55 Selection of variables for inclusion in the research has been informed by the evidence on factors associated with child maltreatment. With respect to risk and protective factors for child maltreatment, a small number of studies have examined associations between administrative data available at birth and subsequent reporting or substantiation of maltreatment (reviewed in Putnam-Hornstein and Needell, 2011).

56 There is a much larger literature on the broader risk and protective factors for child maltreatment. The most widely adopted explanatory model for child maltreatment is the ecological model which considers a number of factors, including the characteristics of the individual child and his or her family, those of the caregiver or perpetrator, the nature of the local community, and the social, economic and cultural environment (WHO, 2012; CYF, 2006).

57 At least some of these factors can be fully or partially measured, or proxied, using linked social sector administrative data. In other cases, risk and protective factors are not able to be directly measured or proxied, but may be strongly associated with other information available from administrative data. A history of other children associated with the caregiver being the subject of findings of abuse or neglect or removal from care, for example, may be associated with risk factors for perpetration that are not able to be directly measured (parenting skills and parenting stress, for example).

58 *Principle 2: Source of personal information*. Box 1 outlines the source of the data to be included in the research.

---

[2] Statistics New Zealand Data Integration Manual http://www.stats.govt.nz/about_us/policies-and-guidelines/data-integration.aspx

Box 1 Source Data

The research will draw together a range of variables, either for the purposes of linking the data, for the development of research on Predictive Risk Models, or for investigation of the backgrounds, experiences and outcomes of children notified for or experiencing different forms of substantiated child maltreatment.

Ministry of Social Development - benefit, care and protection and youth justice data

MSD benefit variables will be derived from the Benefit Dynamics Dataset (with the exception of identifying variables which will be obtained form the SWIFTT source system. MSD care and protection and youth justice variables will be derived in respect of the benefit caregiver/s in their own childhood, in respect of the child, and in respect of other children in the family. Variables of interest include:

- name, aliases, sex, date of birth and address of child and parents/caregivers (for linkage)
- sex and ethnic group of benefits caregivers, country of birth, refugee status and time in New Zealand
- partnership status of benefit caregiver/s and type of benefit received
- number of children cared for by benefit caregiver/s, birth intervals, presence of multiple birth children, age of caregiver when this child and when oldest child born
- duration of benefit receipt of caregiver/s (a proxy for the persistence of low income)
- whether caregiver/s received incapacity benefits in the past with incapacity codes that indicate substance abuse or mental health
- frequency of address changes of benefit caregiver/s
- receipt of Child Disability Allowance (indicating severe disablement) or Disability Allowance in respect of child.

For child and caregiver/s

- care and protection notifications (with Family Violence notifications separately identified), investigations, and substantiated findings of abuse, neglect or behavioural difficulties
- care and protection Family Group Conferences, Child and Family Assessments and care episodes
- Child Youth and Family youth justice referrals, Family Group Conferences, and court orders

Department of Internal Affairs - birth and death information

Birth and death information of interest includes:

- name, aliases, sex, date of birth and address of child and parents (for linkage)
- sex and ethnic group of child, citizenship status and place of birth
- still-birth, birth weight, gestation, and whether multiple birth
- whether father listed on birth certificate and parents' relationship status and number of previous children of the relationship
- parents' ages, ethnicity, birth place, and occupations
- date of death of child

Box 1 continued

Ministry of Business, Innovation and Employment - migration data

Variables of interest include for child:

- name, aliases, sex, date of birth of child (for linkage)
- dates of entry to and exit from New Zealand.

Corrections - imprisonment and sentencing data

Variables of interest include for parents/caregivers:

- dates of entry to and exit from prison
- dates of start and finish of non-custodial sentences by type

Ministry of Health - maternity, mental health, B4School Check and hospitalisations data

Variables of interest include:

- name, aliases, sex, date of birth of child and mother (for linkage)
- sex, ethnic group, residency status, place of birth, DHB and NZDep of child
- age, sex, ethnic group, residency status, DHB and NZDep of mother
- still-birth, birth weight, gestation, and whether multiple birth
- previous births of mother (self-report and administrative count)
- birth abnormality as recorded at birth
- indicator of late or no ante-natal care (noting variation in coverage over time and between DHBs)
- indicators of smoking status of mother
- maternal health service usage and indicators of mental health disorders
- maternal addiction service usage and indicators of drug and alcohol disorders
- maternal mood and/or anxiety disorders (based on pharmaceuticals and health service usage)
- child immunisation status of child and sub-status (indicating whether parental choice is a reason for non-completion)
- child B4School Check participation and selected outcomes/scores
- child hospitalisations for maltreatment-related injury, any injury, any reason, and maternal hospitalisation for assault
- injury-related child deaths

Box 1 continued

Ministry of Education – educational participation and attainment data

Variables of interest include:

- name, aliases, sex, date of birth of child (for linkage)

for child and parents/caregivers

- sex, ethnicity, years of schooling, student type
- institution name, type, school authority, school decile
- school isolation index and location variables (TA, region)
- school enrolment and attendance, qualification attained
- suspension and detention data
- participation in special education

for parents/caregivers

- highest qualification attained at start of tertiary education
- tertiary enrolment
- course pass rate (tertiary education)
- completion of tertiary education
- participation in industry training, youth guarantee programmes and their outcomes
- performance on PISA, TIMSS and other international comparative studies

59 These data will be used in the Project without informing or seeking consent from the individuals involved due to the number and age of the records.

- For most of the data required, this is allowable under exception (2)(g)(ii) to Principle 2 of the Privacy Act on the grounds that the information is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned.
- For health data, this comes within the exception to Rule 11(2)(c)(iii) of the Health Information Privacy Code, subject to approval by an ethics committee.
- For birth and death information, this is provided for under Section 75G(1)(b)(ii)(B) of the Births, Deaths, Marriages, and Relationships Registration Act 1995, subject to approval by an ethics committee, discussion with the Privacy Commissioner, and approval by the Registrar General.

60 *Principle 3: Collection of information from subject*. A Privacy Statement on benefit application forms informs benefit applicants that their information will be used by MSD for statistical and research purposes and for providing advice to Government.

61 However many of the variables required for the research relate to children who cannot be expected to have been made aware of the collection and use of data relating to them. In particular, clients of MSD's care and protection services may be the subject of a notification by a third party and may not be made aware that administrative data relating to them is being collected and held.

62 Use of these data for the Project, and data held by other agencies, is allowable under exception (4)(f)(ii) to Principle 3 of the Privacy Act on the grounds that the information is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned.

63 *Principle 4: Manner of collection of personal information.* Principle 4 of the Privacy Act requires that personal information shall not be collected by an agency (a) by unlawful means, or (b) by means that, in the circumstances of the case, (i) are unfair, or (ii) intrude to an unreasonable extent upon the personal affairs of the individual concerned.

64 A degree of intrusion is inherent and reasonable in the administration of many Government services. Collection of these data at source complies with Principle 4 of the Privacy Act.

65 A consideration that needs to be weighed in the collection together of the data for the Project is whether the level of intrusion upon the personal affairs of the individual concerned is reasonable. Our assessment is that it is, given the potential benefits of the research.

Security and access

66 *Principle 5: Storage and security of personal information.* The source and new linked individual-level data will be held in a secure research environment that sits within a secure data warehouse. Access to the data warehouse is restricted to authorised users only. Access to Project files will be further restricted to project team members and authorised IT systems administrators. Within the team, each team member will only

have access to files that are necessary for the stage of the Project they are responsible for.

67 The Project will, in the main, be carried out within MSD National Office. This is a secure office with building access restricted to authorized personnel. Computer desktops are password protected.

68 Where contracted providers are engaged to analyse de-identified research datasets for the Research Project, this will only occur under strict security arrangements that require the provider to emulate the secure data analysis environment outlined in Figure 1 and comply with all data security, data use and data destruction arrangements outlined in this document.

69 MSD team members and systems administrators are permanent employees of MSD and bound by an obligation under the MSD Code of Conduct to keep information secure and respect the privacy of personal information held by the Ministry. For the avoidance of doubt, research team members (and any contracted provider) will be required to sign a letter that acknowledges the particularly sensitive nature of the linked data and records their agreement to comply with processes put in place to protect the data from unauthorised use, misuse and disclosure. The letter will also identify the process to follow should there be any concern about the security of the data.

70 Privacy enhancing procedures have been developed in response to particular risks around the loss, unauthorized access to and misuse of data. These are described in the table below.

Principle 6 and 7: Access to personal information, Correction of personal information.

71 This proposal relates to linking administrative data that Government already holds in relation to individuals. Existing protocols for access to personal information in the source administrative systems will apply. Because identifying information will be deleted from the linked research data, individuals will not be able to access and seek correction of these data.

Accuracy

72 *Principle 8: Accuracy of personal information to be checked before use.* Steps to check the accuracy of data used in the linking process and the linking process itself include a process of checking for multiple records that relate to the same individual and a clerical assessment of the quality of the linked data. For the purposes of research, these are reasonable steps.[3] To inform the development of an operational predictive risk model that complies with Principle 8, the research will form two linkages and compare results from each.

- A *conservative* linkage. This approach to linkage is likely to be the approach taken in any operationalisation of predictive risk modelling due to the need to avoid false positive matches.

[3] Statistics New Zealand Data Integration Manual http://www.stats.govt.nz/about_us/policies-and-guidelines/data-integration.aspx

- A *less conservative* linkage. This approach supports a better linkage of data for individuals who appear under different names at different times, which can often be the case for children moving between caregivers and children whose caregivers separate or re-partner. For the purposes of research, this provides a much better collation of data. However it also increases the risk of false positive matches.

Retention

73 *Principle 9: Agency not to keep personal information for longer than necessary.* Linked individual-level de-identified research datasets will be held for as long as they are required for the purposes of research and then deleted (no later than June 2015).

74 Identifying information will be deleted at the earliest opportunity (once linkages are complete).

Use

75 *Principle 10: Limits on use of personal information.* Privacy Principle 10 provides that an agency that holds personal information that was obtained in connection with one purpose shall not use the information for any other purpose unless the agency believes, on reasonable grounds, that one of the listed exceptions to that principle apply.

76 Use of benefit information relating to adults is allowable under exception 10(e) - "that the purpose for which the information is used is directly related to the purpose in connection with which the information was obtained". This exception applies as a Privacy Statement on benefit application forms informs applicants that their information will be used by MSD for statistical and research purposes and for providing advice to Government.

77 The linking and use of administrative data relating to children from across the benefit, care and protection and youth justice systems is permissible under exception 10(f) - "that the information (ii) is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned".

78 The Project also uses personal information held by the Registrar General of Births, Deaths and Marriages, and the Ministries of Health, Education and Business, Innovation and Employment. These agencies can disclose this information under the Privacy Act due to the exception of Principle 11(h)(ii) which provides that – "An agency that holds personal information shall not disclose the information to a person or body or agency unless the agency believes, on reasonable grounds, - (h) that the personal information –
(ii) is to be used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned".

79 The use and disclosure of health data involved in the data linking comes within the exception to Rule 11(2)(c)(iii) of the Health Information Privacy Code which provides for disclosure of Health information without the authorisation of the person concerned where the agency holding the data believes on reasonable grounds that it is either not desirable or not practicable to obtain authorisation from the individual concerned and where the information is to be used for research purposes (for which approval by an ethics committee, if required, has been given) and will not be published in a form that could reasonably be expected to identify the individual concerned.

80 The linked database is a significant merging, with the potential for individuals to be concerned about their privacy and about how the information will be used. Limits on the use of the datasets and their output are detailed in this document.

81 There is also potential for a general public concern about an unreasonable level of government scrutiny of individual circumstances associated with the Project. This needs to be weighed against public interest in the study. The aim of the study is to build knowledge to inform and enhance public policy and service delivery for the most vulnerable children.

82 There is a specific concern that researchers will be able to detect apparent benefit fraud as a result of a linked dataset with unit record data, with identifiable individuals. MSD researchers have a responsibility under the MSD Code of Conduct to report any suspected fraud. Because this would involve use of the data for a purpose other than research (which is not allowable under exception 10(f)(ii)), researchers will be exempted from this responsibility for the purposes of the research.

Disclosure

83 *Principle 11: Limits on disclosure of personal information.* The research does not involve disclosure of individual-level data to any person, body or other agency outside of the research team. Research findings may be published on the MSD website. This is allowable under exception h(ii) to Privacy Principle 11 on the grounds that the information is used for statistical or research purposes and will not be published in a form that could reasonably be expected to identify the individual concerned.[4]

Unique identifiers

84 *Principle 12: Unique identifiers*. Agencies' unique identifiers cannot assist with the linking of data and will not be included in the raw agency datasets that contribute to the research. The exception to this is birth and death registration numbers which will be included in data requested from the Registrar General of Births, Deaths and Marriages and will be used to match deaths to matching births.

85 An anonymised research number will be given to each individual for the purposes of data analysis. This is necessary for the researchers to carry out their research functions efficiently because there is no common unique identifier that spans all of the service systems that are of interest. The research number will have no application outside of the research and will be deleted together with the individual-level research datasets at the completion of the research (no later than June 2015).

---

[4] The Project will follow Statistics New Zealand guidelines that prevent identification of individuals' data from research outputs http://www.stats.govt.nz/~/media/Statistics/Methods%20and%20Services/microdata-access/datalab-guide.ashx

4 Privacy risk assessment and privacy enhancing responses

Privacy risk	Privacy enhancing response
1 Information is published in a form that **identifies** an individual	Identifying data (names, relationship and address information, unique identifiers and date of birth) will not be retained on files used for data analysis, and will be securely held separate from other data with access restricted only to members of the project team who are undertaking or checking the data linking and IT systems administrators. Only summary tables will be downloaded from MSD's Information and Analysis Platform during data analysis. Individual data will not be downloaded during data analysis. All research findings will be reported at the aggregate level. No individual-level data will be included in reports or briefings.
2 **Unauthorised access by people outside the research team** to individual-level data sets created in linking and analysis	The source data and linked datasets will be stored and analysed on MSD's Information and Analysis Platform, a secure data warehouse access to which is restricted to authorised users. Access to all files created as part of the research project will be further restricted to members of the project team. Identifying data (names, relationship and address information, unique identifiers and date of birth) will not be retained on files used for data analysis, and will be securely held separate from other data with access restricted only to members of the project team who are undertaking or checking the data linking and IT systems administrators. Only summary tables will be downloaded from the IAP during data analysis. Individual data will not be downloaded during data analysis. The project team will meet at the commencement of the project to discuss the importance of privacy issues relating to the Project and data security procedures (such as locking desktops when they leave their desks) and privacy enhancing responses put in place for the Project. MSD team members and systems administrators are permanent employees of MSD and bound by an obligation under the MSD Code of Conduct to keep information secure and respect the privacy of personal information held by the Ministry. For the avoidance of doubt, MSD research team members (and any contracted provider) will also be required to sign a letter that acknowledges the particularly sensitive nature of the linked data and records their agreement to comply with specified processes put in place to protect the data from unauthorised use, misuse and disclosure.

Privacy risk	Privacy enhancing response (continued)
3 **Unauthorised access to or loss** of data in carrying out the sample-based in-depth clerical assessment	Individual data for samples of records will need to be downloaded to spreadsheets for clerical assessment of the quality of the electronic linking. These spreadsheets will be held in password encrypted zip files and stored in a folder that can be viewed only by members of the project team who are involved in this part of the project. They will not include any details of the variables being analysed (eg. the care and protection or birth information of the individual). They will include only details relevant to verifying the electronic linkage (ie. names, date of birth, gender, address history and relationships). Any printouts to enable the checking to proceed will be securely stored, and will be destroyed immediately (by shredding) at the completion of checking..
4 **Misuse** - demand to use the linked data for purposes other than this research once the potential of the linked data is realised	Two documents set out the scope of the research and authorised uses as part of the research: • this Privacy Impact Assessment • the Research Protocol. Other research proposals that seek to draw on the linked data assembled for the Research Project may be developed. In some cases, these may involve other Government agencies or contracted providers. Interagency project governance will be provided through the Vulnerable Children's Board, which is made up of social sector Chief Executives. Where proposals gain consensus approval from the Vulnerable Children's Board and ethics approval (if required) and have a Privacy Impact Assessment in place, provision will be made for de-identified linked datasets to be supplied for the purposes of the research under strict confidentiality agreements and only where those organisations are able to meet high data security standards and comply with the Privacy Act in respect of the use, disclosure and destruction of the data. Where the de-identified linked dataset to be supplied for the purposes of the research includes Ministry of Health data, approval of the Director General of Health will also be required before the data are supplied. Where it includes birth and death information, approval of the Registrar General will be required. All individual-level data assembled for the research will be deleted at the completion of the Project (no later than June 2015). .
5 **Misuse** - ability to detect and obligation to report suspected benefit fraud	MSD researchers have a responsibility under the MSD Code of Conduct to report any suspected fraud. Because this would involve use of the data for a purpose other than research (which is not allowable under exception (f)(ii) to Privacy Principle 10), researchers will be exempted from this responsibility for the purposes of the Project.

6 **Misuse** - unauthorised use by research team members	The normal requirements for MSD staff set out in the MSD Code of Conduct apply to all researchers working on the datasets. For the avoidance of doubt, research team members will also be required to sign a letter that acknowledges the particularly sensitive nature of the linked data and records their agreement to comply with specified processes put in place to protect the data from unauthorised use, misuse and disclosure.
7 Public concern or individual complaint about loss of privacy	Office of the Privacy Commissioner informed and feedback addressed Compliance with the Privacy Act 1993, and other relevant codes and legislation Ethics Committee approval for the research Transparency about objectives and processes Communications statements to be prepared outlining the benefits of the research and the privacy enhancing protections that have been used in the research.

5 Compliance mechanisms

86 This Privacy Impact Assessment has been reviewed by the MSD Privacy Officer. The project manager will review the status of the project and implementation of privacy enhancing responses throughout the project and will consult with the Privacy Officer as necessary.

87 An annual formal audit of data security systems will be undertaken and will form part of the annual report on the project to the approving ethics committee. The audit will include reviewing whether:

- the research team has met at the outset of the project to discuss data security
- the research team members have signed data security letters
- transit of files between agencies adheres to agreed security arrangements, and details of external data transfers have been logged
- access to data files has been restricted to project team members and authorised IT systems administrators
- data security has been reviewed as a standing item at project meetings with specific reference to:
    - ensuring each team member has only had access to files that are necessary for the stage of the Project they are responsible for
    - checking that data sets containing identifying information or linked individual data have been deleted at the earliest opportunity
- research outputs downloaded from the secure data analysis environment include only tables and figures and predictive risk model algorithms and do not include any individual-level data (except for that required for manual inspection of samples of linked records).
- manual inspection of samples of linked records to access accuracy in the linking adheres to agreed security arrangements
- the data have not been used for any other purpose (except where a proposal for additional research has gained approval from the interagency governance group overseeing this research programme and ethics approval (if required) and has a Privacy Impact Assessment in place and required approvals from the Director General of Health and the Registrar General).

6 Conclusions

The risks to individual privacy arising from the Project have been identified, addressed and can be managed to a low level.

The potential risk of negative public perception regarding individual privacy issues arising from the project is very real, given the intended end use of the research findings. This privacy impact assessment process, privacy protection procedures, and the development of communications statements are designed to address these risks.

Risks to privacy law are minimal as the uses of personal information come within the research exceptions to the Privacy Act. Careful procedures are being implemented to prevent loss of the data or use of the data for purposes other than those authorised.

References

Centre for Social Research and Evaluation. (2012). *Children's Contact with MSD Services*. Ministry of Social Development.

Child Youth and Family (2006). *Children at increased risk of death from maltreatment and strategies for prevention.* Ministry of Social Development.

Gilbert, R., Fluke, J., O'Donnell, M., Gonzalez-Izquierdo, A., Brownell, M., Gulliver, P., Janson, S. and Sidebotham, P. (2011). Child maltreatment: variation in trends and policies in six developed countries. *The Lancet* December (online).

MacMillan, H.L., Wathen, C.N., Barlow, J., Fergusson, D.M., and Leventhal, J.M. (2009). Interventions to prevent child maltreatment and associated impairment. *The Lancet* 373.9659 (Jan 17-Jan 23): 250-66.

Putnam-Hornstein, E. & Needell, B. (2011). Predictors of child welfare contact between birth and age five: an examination of California's 2002 birth cohort. *Children and Youth Services Review*, 33(11), 2400-2407.

Vaithianathan, R., Maloney, T., Jiang, N., De Haan, I., Dale, C., Putnam-Hornstein, E., Dare, T. and Thompson, D. (2012). *Vulnerable children: Can administrative data be used to identify children at risk of adverse outcomes?* Report prepared for the Ministry of Social Development.

WHO (2012). Child abuse and neglect by parents and other caregivers, Chapter 3, World *Report on Violence and Health.*