

Case Identification of Mental Health and Related Problems in Children and Young People using the New Zealand Integrated Data Infrastructure

Nicholas Bowden (✉ Nick.Bowden@otago.ac.nz)

University of Otago <https://orcid.org/0000-0003-4589-9956>

Sheree Gibb

University of Otago Wellington

Hiran Thabrew

University of Auckland

Jesse Kokaua

University of Otago

Richard Audas

University of Otago

Sally Merry

University of Auckland

Barry Taylor

University of Otago

Sarah Hetrick

University of Auckland

Research article

Keywords: Integrated Data Infrastructure, IDI, big data, mental health, case identification

Posted Date: July 18th, 2019

DOI: <https://doi.org/10.21203/rs.2.11629/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published on February 27th, 2020. See the published version at <https://doi.org/10.1186/s12911-020-1057-8>.

Abstract

Background In a novel endeavour we aim to develop a clinically relevant case identification method for research into the mental health of children and young people in New Zealand using the Integrated Data Infrastructure (IDI). The IDI is a linked individual-level database containing New Zealand government and survey microdata. **Methods** We draw on diagnostic and pharmaceutical information contained within five secondary care service use and medication dispensing datasets to identify cases of mental health and related problems. A systematic classification and refinement of codes, including restrictions by age, was undertaken to assign cases into 13 different mental health problem categories. This process was carried out by a group of seven professionals (a clinical psychologist, one child and adolescent psychiatrist, two child psychiatrists/paediatricians and two research fellows). The case identification algorithm was applied to the New Zealand youth estimated resident population for the 2014/15 fiscal year. **Results** Over 84,000 unique individuals aged 0-24 with at least one specified mental health or related problem were identified using the case identification algorithm for the 2014/15 fiscal year. The most prevalent mental health problem subgroups were emotional problems (31,299 individuals), substance problems with (16,545), and disruptive behaviours (14,091). Overall the pharmaceutical collection was the largest source of case identification data (59,862). **Conclusion** This study demonstrates the value of utilising IDI data for mental health research, taking us beyond relying solely on diagnostic based incidence, while providing directions for future use, including further linkage of data to the IDI.

Background

Mental health problems are common among children and young people, with a worldwide estimated prevalence of 13.4% affected by any mental disorder (Polanczyk, Salum, Sugaya, Caye, & Rohde, 2015). In New Zealand, school-based survey results indicate 31% of young people experience at least two weeks of low mood, 15.7% report suicidal ideation, and 24% engage in self-harm each year (Fleming et al., 2014). The short-term consequences of childhood and adolescent mental health problems can include interference with education (Schulte-Körne, 2016) and developmental milestones (Frisk, 1999). Longer-term, they may be associated with personal costs, such as reduced employment (Fergusson, Boden, & Horwood, 2007; Mojtabai et al., 2015), poorer quality of life, (Chen et al., 2006) and societal costs such as greater economic burden (Smith & Smith, 2010).

Most information regarding the prevalence and treatment of mental health problems originates from small cross-sectional studies with short-term evaluation, and occasional, expensive longitudinal trials with finite long-term outcomes. To date, there has been limited use of administrative data for mental health research (Davis, Sudlow, & Hotopf, 2016; Harvey et al., 2017; Stewart & Davis, 2016), especially in children and young people (Reid et al., 2015). However, large amounts of administrative data, including information on hospital attendance, community care specialist services, and medication prescriptions, are routinely collected, stored by national health providers and related institutions, and may be valuable

for health research (Cadarette & Wong, 2015; Garland, Gershengorn, Marrie, Reider, & Wilcox, 2015; Hinds, Lix, Smith, Quan, & Sanmartin, 2016). Their use for this purpose is covered in some countries by privacy legislation (New Zealand Statutes, 1993).

The advantages of using administrative data for research include the large, heterogeneous and representative nature of samples, which allows the reflection of real-world practice, ongoing tracking of problems via regular collection of up-to-date data, long observation periods and low cost; while disadvantages include erroneous interpretation of data beyond the scope for which they were intended, the variability of data quality, limited clinical detail, and potential public concern about administrative data being used for research purposes (Mazzali & Duca, 2015). In New Zealand administrative data are readily available, free to use, typically national and, because of the IDI, a large anonymised research database managed by Statistics New Zealand, containing a wide range of administrative and survey data about people and households, linked at the individual level (Statistics New Zealand, 2017b).

Case identification for physical health problems using administrative data, typically utilising International Classification for Diseases (ICD) coding, has been widespread (Aronsky, Haug, Lagor, & Dean, 2005; Elixhauser, Steiner, Harris, & Coffey, 1998; Hebert et al., 1999; Neff et al., 2002; Saczynski et al., 2012; Sung et al., 2016; Tu, Campbell, Chen, Cauch-Dudek, & McAlister, 2007), but less so in mental health (Frayne et al., 2010). A standardised and accessible case identification approach means research can be more comparable, it negates the need to duplicate work, and permits researchers without specialist mental health knowledge to contribute more easily to the field. (Blakely, Collings, & Atkinson, 2003; Kake, Arnold, & Ellis, 2008; Social Investment Agency, 2017).

This paper describes the development of an administrative data-based case identification method for research into the mental health of children and young people in New Zealand. The method utilises secondary care service use and medication dispensing data held within the IDI. Given a large number of individuals with mental health problems are treated in primary care, through alternative therapies, or not at all, the method is not intended to calculate prevalence estimates for mental health problems but can be used to examine health trajectories, comorbidities and a range of other wellbeing outcomes. These uses are especially important in understanding the burden of disease for mental health conditions and the impact they have on the lives of children and young people.

Methods

Integrated Data Infrastructure (IDI)

The IDI is a large anonymised research database managed by Statistics New Zealand, containing a wide range of administrative and survey data[1] about people and households (Statistics New Zealand, 2017b). Cabinet directives dating back to 1997 mandated Statistics New Zealand to undertake cross-agency data integration and the IDI was established in 2011[2]. Data in the IDI is held in a secure environment and can be accessed by approved researchers only for projects that are in the public interest. Since 1993 it has been possible to link health datasets together using the National Health Index (NHI) number, however, until the creation of the IDI, it was not possible to routinely combine health and non-health related data. The IDI provides secure access to linked data at an individual level. It enables more extensive use of government data for research including supporting the evaluation of the long-term impact of health interventions, with the aim of improved public health (Atkinson & Blakely, 2017).

Data Privacy

Statistics New Zealand's 'five safes' framework (Statistics New Zealand, 2017a) is used to ensure data privacy is protected. Only approved researchers can use the IDI for projects that have a statistical purpose and are for the public good. All data are de-identified and only accessible via a secure connection from a secure Datalab. Data and results must be aggregated and confidentialised according to Statistics New Zealand protocols, and all results are checked by Statistics New Zealand prior to their release from the secure environment.

The legal requirements to protect the IDI data include the [Statistics Act 1975](#), the [Privacy Act 1993](#), and the [Tax Administration Act 1994](#) (Statistics New Zealand, 2017a). In addition to legal requirements a number of Statistics New Zealand policies, protocols, and guidelines exist (Statistics New Zealand, 2017c). These include information privacy guidelines, security guidelines, confidentiality policy and [data integration guidelines](#), [microdata access guidelines](#), and [privacy and confidentiality guidelines](#). Regular privacy impact assessments for the IDI also provide a systematic evaluation of the benefits and risks associated with integrating data from a number of sources (Statistics New Zealand, 2017d).

Data

Five datasets are used in this study and are all from administrative sources held within the IDI accessed in June 2019 using the most recent refresh of IDI data. These datasets are each described below, including associated strengths and weaknesses for case identification of mental health and related problems.

Programme for the Integration of Mental Health Data (PRIMHD)

PRIMHD is a national collection of publicly funded specialist mental health service use (PRIMHD activity data) and diagnoses (PRIMHD classification (diagnosis) data). Data is collected from district health boards (DHBs) and non-governmental organisations (NGOs) that provide specialist mental health services and is used to report on what services are being provided, and who is providing the services, for health consumers across New Zealand's mental health sector (Statistics New Zealand, 2015b). A limitation of PRIMHD data is that it only covers publicly funded specialist mental health care which is targeted to those with most serious mental health problems. It does not cover mental health care provided in a primary care setting, or mental health care provided in the private sector.

PRIMHD Classification (Diagnosis) Data

PRIMHD classification (diagnosis) data in the IDI was collected from 1 July 2008 to 31 December 2016 and includes primary and secondary diagnosis codes (ICD-10-AM and DSM-IV). PRIMHD is the only national collection of formal psychiatric diagnoses in New Zealand but it has some limitations. Many people have contact with mental health services but do not have specific diagnoses recorded in PRIMHD. For most (nearly 85% of these non-specific diagnosis codes), this is because clients were seen for only a brief period and there was insufficient time for a diagnosis to be assigned (Ministry of Health, 2017b). This gap in data is reflected in approximately 37% of all records (not clients).

PRIMHD Activity Data

PRIMHD activity data in the IDI was collected between 1 July 2008 and 30 June 2018 and contains (i) an activity type code, i.e. a code that classifies the type of healthcare provided, and/or (ii) a team type code, i.e. a code that identifies which team provided a service. These codes can be used to identify individuals with mental health problems. Utilising PRIMHD activity data, a diagnosis can sometimes be implied by the type of service the client receives, as with substance use, and can help to improve the coverage and quality of diagnosis information.

The National Minimum Dataset (NMDS)

NMDS is a national collection of publicly funded New Zealand hospital admissions, including day patients (stays of either three hours or more but not overnight) and emergency department visits of greater than three hours. Primary and secondary diagnosis codes (ICD-10-AM) are recorded for every hospital event and are used to identify the mental health and related problems. NMDS data within the IDI were collected between 1 January 1988 and 31 December 2017. As there have been several changes to NMDS data collection over the years, most notably prior to 1994, we restrict the potential for NMDS use to 1994 onwards (Statistics New Zealand, 2015c).

A key advantage of NMDS is the ability to utilise secondary diagnoses, i.e. issues deemed material to the individual's care that are not the main reason for their hospital admission. For example, a patient may be admitted to hospital for a non-mental health reason but their mental health may affect treatment or recovery and therefore is recorded as a secondary diagnosis. In some cases these individuals will not have accessed services for mental health, or may have done so only in the primary care setting (data not held in the IDI). In these cases NMDS permits case identification of mental health problems not otherwise possible in the IDI.

Socrates

Socrates is the national database of the Ministry of Health's (MoH's) Disability Support Services clients and service providers. Individuals have data recorded in Socrates when they apply for a needs assessment to access home help or other support services via a Needs Assessment and Service Co-ordination Agency (NASC) throughout New Zealand. A range of disabilities can be recorded on an individual's record and these include some mental health diagnoses. These diagnoses come with the referral for the client, for example from a general practitioner, social worker, treatment and rehabilitation provider, or psychologist. Socrates was established in 2008 and data are robust from 1 January 2010. While Socrates data exists in the IDI prior to 2008 they are sparse and not deemed reliable.

There is uncertainty around the diagnostic detail in Socrates: it is not known who provides the diagnosis and therefore the accuracy of the diagnosis may vary. However, people with mental health problems, in particular neurodevelopmental problems such as Attention Deficit and Hyperactivity Disorder (ADHD) can be identified, some of whom will not have been referred to specialist mental health services. Therefore, Socrates offers an additional source of case identifications.

Pharmaceutical Collection (Pharms)

Pharms contains claim and payment information from pharmacists for government-subsidised medication dispensing throughout New Zealand (Statistics New Zealand, 2015a). Pharms data in the IDI was collected between 1 January 2005 and 31 December 2017. However, as pre-2007 data was collected with less than 90% coverage, only data collected from 2007 is used for research purposes, as per MoH recommendations. 'Chemical codes' assigned to each dispensing are used to identify the specific medication dispensed and can be used as indications for mental health problems.

Mortality Collection

The mortality collection contains information about the underlying causes of all deaths registered in New Zealand (Ministry of Health, 2017a). It uses the ICD-10-AM classification and *World Health Organization Rules and Guidelines for Mortality Coding* (World Health Organisation, 2003). Mortality data is robust and

of high quality. However, due to the timeframe for coronial processes, there is a two-year lag in its availability. For the current analysis, the mortality dataset within the IDI was collected between 1 January 1988 and 31 December 2015 and has been used in this study to identify cases of death by suicide.

Data Summary

Figure 1 displays the periods of time^[3] each of the five datasets are available within the IDI. It also breaks down this coverage by 'available data' and 'best quality data' as per the discussions above. Data was available for each of the datasets within a five-fiscal year period from 1 July 2010 (start of 2010/11 fiscal year) until 30 June 2015 (end of 2014/15 fiscal year). For the purposes of demonstrating an application of the algorithm we decided to use the most recent of these fiscal years, 2014/15. The results are presented in the next section.

Establishing and refining case definitions for mental health problems

Our aim was to create an algorithm for identifying common, clinically relevant mental health problems for New Zealand children and young people (aged 24 and under) using available IDI data, which also included self-harm. The method built on a similar approach taken by the SIA (Social Investment Agency, 2017) but was developed specifically for children and young people, with an increased focus on clinical verification of diagnoses among children and young people. The processes undertaken to establish the case identification method are summarised below in three stages.

In the first stage, a list of 13 mental health (and related) problems of interest was established: anxiety; depression; bipolar disorders; emotional problems^[4]; disruptive behaviours; substance problems; eating problems; psychosis; personality disorders; sleep problems; self-harm; other mental health problems^[5]; and mental health not defined^[6].

The next stage of work involved the systematic classification and refinement of codes used to assign cases into each mental health problem category, by data source. Seven individuals (a clinical psychologist, one child and adolescent psychiatrist, two child psychiatrists/paediatricians and two research fellows) independently assigned diagnostic codes to the 13 mental health problem categories. Disagreements were resolved using a consensus approach. For pharmaceutical data, medications that were considered more likely to be used for the treatment of non-mental health problems than mental health problems were not included. Medications deemed to be used for both anxiety and depression were assigned to the category 'emotional disorders' and medications used for several mental health problems (e.g. Risperidone, which may be used for the treatment of psychosis, disruptive behaviour, obsessive

compulsive disorder, bipolar disorder and emotional dysregulation) to the category 'mental health not defined'. The data sources used to identify each specific mental health problem group are outlined in

Table 1.[7] Some mental health problems were derived from as few as two datasets (e.g. eating problems were identified using NMDS and PRIHMD Diagnosis) and others from as many as four datasets (e.g. anxiety was identified using NMDS, PRIMHD, Pharms, and Socrates).

In the final stage, again using a consensus approach, we used age stratification to further refine our case identification method. Five age strata were employed: 0-4, 5-9, 10-14, 15-19, and 20-24 and case identifications were restricted by age in that a case was not assigned to a problem group (counted) if it was deemed unlikely that individuals below an age threshold would be suffering from that given mental health problem. These restrictions varied according to the data source and type of code used (ICD, DSM, pharmaceutical etc.) (e.g. generalised anxiety disorder had no age restriction whereas social phobias were restricted to age five and up) [8]. This process was particularly relevant for case identifications derived from pharmaceutical dispensing where case identifications are not tied to specific clinical diagnoses and where some medications are more likely to be used for different problems at different ages (e.g. Quetiapine being more likely to be used for behavioural disturbance in children vs. psychosis, bipolar disorder, and depression in older adolescents and young adults)[9].

Data management

Data preparation was done in SAS 7.1 within the IDI environment. There were three main steps. First, event level data (e.g. dispensings for pharms, hospitalisations for NMDS) were extracted, separately for each of the five datasets used in the study, for all individuals in the New Zealand youth population (0-24) for the 2014/15 fiscal year. Second, using the coding system for case identification described, 13 dichotomous mental health problem indicator variables were generated for each individual. Finally, data from each of the five datasets were appended and then collapsed to one set of mental health problem indicators per person. For individuals who had a 'mental health not defined' and another mental health problem group indicator (excluding self-harm) the 'mental health not defined' indicator was set to zero. The resulting data were analysed using StataMP 15. All counts were randomly rounded to base three in line with Statistics New Zealand confidentiality requirements.

Establishing the New Zealand Youth (0-24) Population 2014/15

The New Zealand youth population (0-24) was calculated using existing methods for estimating a resident New Zealand population from the IDI (Gibb, Bycroft, & Matheson-Dunning, 2016; Zhao, Gibb, Jackson, Mehta, & Exeter, 2017). More specifically, this method included people whose presence in New

Zealand was indicated by activity in key datasets. Individuals who had died or moved overseas were excluded. The total resident population generated using this method was within 2% of the official estimated resident population estimate. Case identifications were restricted to people from within this population and 12-month prevalence rates derived accordingly.

Ethics Approval

The University of Otago Human Research Ethics Committee reviewed the study for ethics consideration. The study was reviewed as a 'Minimal Risk Health Research – Audit and Audit related studies' proposal and was approved.

[1] For more details on the data continued within the IDI see http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-data.aspx

[2] For more details on the history of the IDI development see http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/idi-how-it-works.aspx#history

[3] It is important to note that data in the IDI is updated periodically, typically quarterly.

[4] This is a composite group formed because a number of medications exist which are typically good indications of either Anxiety Disorders or Depressive Disorders but not specifically one in particular. There are several diagnostic codes that contribute to this group as well.

[5] This is a composite group, which for the sake of completeness includes all mental health diagnostic codes not otherwise used in the first ten groups.

[6] This is a composite group formed because a number of medications are typically indications of a range of potential mental health problems but not specific disorders and in many cases 'mental health not defined' is a diagnostic code assigned to people with mental health problems that for whatever reason cannot be specified with more detail.

[7] A codebook containing full details of all case identifications of mental health (and related) problems, including age restrictions is available from the first author by request.

[8] Details of the final classifications for pharmaceuticals is contained within the codebook and available upon request.

[9] The SAS code for the case identification algorithm is also available upon request.

Results

The case identification method was employed using data from the 2014/15 fiscal year. Over 84,000 unique individuals aged 0-24 with at least one specified mental health problem including self-harm, other mental health, and mental health not defined were identified (see Table 2), indicating a 12-month prevalence of 5,446 per 100,000 population (equivalent to 5.4%).

The most prevalent mental health problem subgroups were 'emotional problems' with 31,299 individuals (2.0%), followed by substance problems with 16,545 individuals (1.8%), and disruptive behaviours with 14,091 individuals (0.9%).

Overall, pharms identified the most individuals (almost 60,000) and was also the data source used to identify the most individuals in five of the 13 problem groups (anxiety, depression, emotional problems, disruptive behaviours, and sleep problems). PRIMHD was the data source that contributed the second most case identifications (over 35,000) and was the biggest contributor to a further seven of 13 specific mental health problem groups (bipolar disorders, substance problems, eating problems, personality disorders, psychosis, mental health not defined, and other mental health). NMDS was the only data source used to identify cases on non-fatal self-harm, but also contributed to the case identifications in all other mental health problem groups. Socrates was used in only eight of 13 mental health problem groups and while the corresponding case identifications numbers were generally low, it contributed to nearly 600 disruptive behaviour cases.

Discussion

Key findings

Using a range of data from the IDI with mental health information it has been possible to construct an algorithm to identify and classify mental health problems among New Zealand youth. The algorithm identified over 84,000 individuals aged between 0-24 with mental health problems, affecting 5.4 percent of all youth in 2014/15. The algorithm is not designed to estimate prevalence of all diagnosed mental conditions due to an undercount from relying mostly on secondary service use data. However, it will provide a valuable source of data for tracking individuals with mental health problems, at least serious enough to require some level of public health funded intervention. Additionally, it can provide information to help to understand their use of mental health services and pharmaceuticals, and more broadly facilitating research on those affected by mental health problems.

Limitations and Strengths

Ethical issues

Further Research and Potential Uses

Further research is needed to formally validate and refine the described method. This may be initially undertaken using medical record review. Presumably the IDI will improve over time, and the addition of interaction and diagnosis data from the primary healthcare sector would be a significant step forward. The development of a truly robust algorithm is likely to be iterative, but once it has been demonstrated, it could be used to track mental health problems in children and young people over time to better understand pathways to risk and resilience. The IDI algorithm could also be used for evaluating the long-term impact of public mental health interventions and in time, reducing health disparities and inequalities.

Conclusions

We have described how multiple data sources from within the IDI could be used to identify and classify mental health problems according to secondary service use and medication dispensing data among New Zealand children and young people. This algorithm enables improved capabilities for mental health research and evaluation; however, its current limitations should be kept firmly in mind. It could be further strengthened by the inclusion of additional data sources in the IDI, in particular primary care data. Undertaking a validation process through a medical records review would allow for greater confidence in the validity. The creation of the IDI is an excellent step forward in tracking health and well-being in New Zealand, but it is a new resource and has clear limitations, including in this important field, although the future potential is clear.

Declarations

Ethics approval

The University of Otago Human Research Ethics Committee reviewed the study for ethics consideration. The study was reviewed as a 'Minimal Risk Health Research – Audit and Audit related studies' proposal and was approved.

Consent of publication

Not applicable with respect to individual/person level data. Consent for publication of the data presented in this paper, subject to the disclaimer presented, has been given by Statistics New Zealand.

Availability of data and material

The data used in this study is held with the Integrated Data Infrastructure and is managed by Statistics New Zealand. These data are publicly available, although access to these data are restricted. Please see <https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/> for more details.

The SAS code will be made available to interested parties.

Competing Interests

None

Funding

This research was part of the Better Start National Science Challenge, funded by the New Zealand Ministry of Business, Innovation and Employment (MBIE).

Authors' contributions

All listed authors have made substantial contributions to the manuscript, have approved the submitted version, have agreed to be personally accountable for their contributions, and ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature:

NB: Design and conception of work, coding, analysis, writing of manuscript.

SG: Design and conception of work, coding, analysis, writing of manuscript.

HT: Design and conception of work, diagnostic classification, writing of manuscript.

JK: Design and conception of work, coding, analysis, substantial revisions of work.

RA: Design and conception of work, substantial revisions of work.

SM: Design and conception of work, diagnostic classification, writing of manuscript.

BT: Design and conception of work, substantial revisions of work.

SH: Design and conception of work, diagnostic classification, drafted the work.\

Acknowledgements

We'd also like to acknowledge Dr. Grant Christie, University of Auckland, and Dr. Rebecca Mairs, University of Auckland for their contributions.

References

Aronsky, D., Haug, P. J., Lagor, C., & Dean, N. C. (2005). Accuracy of administrative data for identifying patients with pneumonia. *American Journal of Medical Quality*, *20*(6), 319-328.

Atkinson, J., & Blakely, T. (2017). New Zealand's Integrated Data Infrastructure (IDI): Value to date and future opportunities. *International Journal of Population Data Science*, *1*(1).

- Blakely, T. A., Collings, S. C., & Atkinson, J. (2003). Unemployment and suicide. Evidence for a causal association? *Journal of Epidemiology & Community Health*, *57*(8), 594-600.
- Bowden, N., Gibb, S., & Thabrew, H. A., Richard Camp, Justine Taylor, Barry Hetrick, Sarah. (2019). IDI Trends in Antidepressant Dispensing to New Zealand Children and Young People Between 2007/08 and 2015/16. *Manuscript submitted for publication*.
- Cadarette, S. M., & Wong, L. (2015). An introduction to health care administrative data. *The Canadian journal of hospital pharmacy*, *68*(3), 232.
- Chen, H., Cohen, P., Kasen, S., Johnson, J. G., Berenson, K., & Gordon, K. (2006). Impact of adolescent mental disorders and physical illnesses on quality of life 17 years later. *Archives of pediatrics & adolescent medicine*, *160*(1), 93-99.
- Davis, K. A., Sudlow, C. L., & Hotopf, M. (2016). Can mental health diagnoses in administrative data be used for research? A systematic review of the accuracy of routinely collected diagnoses. *BMC psychiatry*, *16*(1), 263.
- Durie, M. (2006). Measuring Māori wellbeing. *New Zealand Treasury Guest Lecture Series*, 1.
- Elixhauser, A., Steiner, C., Harris, D. R., & Coffey, R. M. (1998). Comorbidity measures for use with administrative data. *Medical care*, 8-27.
- European Economic and Social Committee. (2017). *The ethics of Big Data: Balancing economic benefits and ethical questions of Big Data in the EU policy context*.
- Fergusson, D. M., Boden, J. M., & Horwood, L. J. (2007). Recurrence of major depression in adolescence and early adulthood, and later mental health, educational and economic outcomes. *The British Journal of Psychiatry*, *191*(4), 335-342.
- Fleming, T. M., Clark, T., Denny, S., Bullen, P., Crengle, S., Peiris-John, R., et al. (2014). Stability and change in the mental health of New Zealand secondary school students 2007–2012: Results from the national adolescent health surveys. *Australian & New Zealand Journal of Psychiatry*, *48*(5), 472-480.
- Frayne, S. M., Miller, D. R., Sharkansky, E. J., Jackson, V. W., Wang, F., Halanych, J. H., et al. (2010). Using administrative data to identify mental illness: what approach is best? *American Journal of Medical Quality*, *25*(1), 42-50.
- Frisk, M. (1999). A complex background in children and adolescents with psychiatric disorders: Developmental delay, dyslexia, heredity, slow cognitive processing and adverse social factors in a multifactorial entirety. *European child & adolescent psychiatry*, *8*(3), 225-236.
- Garland, A., Gershengorn, H. B., Marrie, R. A., Reider, N., & Wilcox, M. E. (2015). A practical, global perspective on using administrative data to conduct intensive care unit research. *Annals of the American*

Thoracic Society, 12(9), 1373-1386.

Gibb, S., Bycroft, C., & Matheson-Dunning, N. (2016). *Identifying the New Zealand resident population in the Integrated Data Infrastructure (IDI)* (No. 0908350333): Statistics New Zealand.

Harvey, S. B., Deady, M., Wang, M.-J., Mykletun, A., Butterworth, P., Christensen, H., et al. (2017). Is the prevalence of mental illness increasing in Australia? Evidence from national health surveys and administrative data, 2001–2014. *Med J Aust, 206(11), 490-493.*

Hebert, P. L., Geiss, L. S., Tierney, E. F., Engelgau, M. M., Yawn, B. P., & McBean, A. M. (1999). Identifying persons with diabetes using Medicare claims data. *American Journal of Medical Quality, 14(6), 270-277.*

Hinds, A., Lix, L. M., Smith, M., Quan, H., & Sanmartin, C. (2016). Quality of administrative health databases in Canada: a scoping review. *Can J Public Health, 107(1), 56-61.*

Take, T. R., Arnold, R., & Ellis, P. (2008). Estimating the prevalence of schizophrenia among New Zealand Māori: a capture–recapture approach. *Australian and New Zealand Journal of Psychiatry, 42(11), 941-949.*

Mazzali, C., & Duca, P. (2015). Use of administrative data in healthcare research. *Internal and emergency medicine, 10(4), 517-524.*

Ministry of Health. (2017a). *Mortality Collection Data Dictionary*. Wellington.

Ministry of Health. (2017b). *PRIMHD Classification – Summary and Metadata*. Wellington.

Mojtabai, R., Stuart, E. A., Hwang, I., Susukida, R., Eaton, W. W., Sampson, N., et al. (2015). Long-term effects of mental disorders on employment in the National Comorbidity Survey ten-year follow-up. *Social psychiatry and psychiatric epidemiology, 50(11), 1657-1668.*

Neff, J. M., Sharp, V. L., Muldoon, J., Graham, J., Popalisky, J., & Gay, J. C. (2002). Identifying and classifying children with chronic conditions using administrative data with the clinical risk group classification system. *Ambulatory Pediatrics, 2(1), 71-79.*

Privacy Act (1993).

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry, 56(3), 345-365.*

Reid, G., Stewart, S. L., Zaric, G. S., Carter, J. R., Neufeld, R. W., Tobon, J. I., et al. (2015). Defining episodes of care in children's mental health using administrative data. *Administration and Policy in Mental Health and Mental Health Services Research, 42(6), 737-747.*

- Saczynski, J. S., Andrade, S. E., Harrold, L. R., Tjia, J., Cutrona, S. L., Dodd, K. S., et al. (2012). A systematic review of validated methods for identifying heart failure using administrative data. *Pharmacoepidemiology and drug safety*, 21, 129-140.
- Schulte-Körne, G. (2016). Mental health problems in a school setting in children and adolescents. *Deutsches Ärzteblatt International*, 113(11), 183.
- Smith, J. P., & Smith, G. C. (2010). Long-term economic costs of psychological problems during childhood. *Social science & medicine*, 71(1), 110-115.
- Social Investment Agency. (2017). Mental health conditions and addictions service access and/or treatment data definition.
- Statistics New Zealand. (2015a). IDI Data Dictionary: Pharmaceutical data October 2015 edition.
- Statistics New Zealand. (2015b). IDI Data Dictionary: Programme for the integration of mental health data October 2015 edition.
- Statistics New Zealand. (2015c). IDI Data Dictionary: Publicly funded hospital discharges – event and diagnosis / procedure information November 2015 edition.
- Statistics New Zealand. (2017a). How we keep IDI and LBD data safe. from http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe.aspx
- Statistics New Zealand. (2017b). Integrated Data Infrastructure. from http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx
- Statistics New Zealand. (2017c). Legislation, policies, and protocols.
- Statistics New Zealand. (2017d). Privacy impact assessments for the IDI and LBD. from http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure/keep-data-safe/privacy-impact-assessments.aspx
- Stewart, R., & Davis, K. (2016). 'Big data' in mental health research: current status and emerging possibilities. *Social psychiatry and psychiatric epidemiology*, 51(8), 1055-1072.
- Sung, S.-F., Hsieh, C.-Y., Lin, H.-J., Chen, Y.-W., Yang, Y.-H. K., & Li, C.-Y. (2016). Validation of algorithms to identify stroke risk factors in patients with acute ischemic stroke, transient ischemic attack, or intracerebral hemorrhage in an administrative claims database. *International journal of cardiology*, 215, 277-282.
- Tu, K., Campbell, N. R., Chen, Z.-L., Cauch-Dudek, K. J., & McAlister, F. A. (2007). Accuracy of administrative databases in identifying patients with hypertension. *Open medicine*, 1(1), e18.

World Health Organisation. (2003). Intentional self-harm from <http://apps.who.int/classifications/apps/icd/icd10online2003/fr-icd.htm?gx60.htm>

Zhao, J., Gibb, S., Jackson, R., Mehta, S., & Exeter, D. J. (2017). Constructing whole of population cohorts for health and social research using the New Zealand Integrated Data Infrastructure. *Australian and New Zealand Journal of Public Health, in press.*

Disclaimer

The results in this paper are not official statistics. They have been created for research purposes from the Integrated Data Infrastructure (IDI), managed by Statistics New Zealand. The opinions, findings, recommendations, and conclusions expressed in this paper are those of the authors, not Statistics New Zealand.

Access to the anonymised data used in this study was provided by Statistics New Zealand under the security and confidentiality provisions of the Statistics Act 1975. Only people authorised by the Statistics Act 1975 are allowed to see data about a particular person, household, business, or organisation, and the results in this paper have been confidentialised to protect these groups from identification and to keep their data safe.

Careful consideration has been given to the privacy, security, and confidentiality issues associated with using administrative and survey data in the IDI. Further detail can be found in the Privacy impact assessment for the Integrated Data Infrastructure available from www.stats.govt.nz.

Tables

Table 1: Sources of data for each disorder group

Mental Health Problems	NMDS	PRIMHD	Pharms	Socrates	Mortality
Anxiety	☐	☐	☐	☐	
Depression	☐	☐	☐	☐	
Emotional Problems	☐	☐	☐		
Bipolar Disorders	☐	☐		☐	
Substance Problems	☐	☐	☐	☐	
Eating Problems	☐	☐			
Disruptive Behaviours	☐	☐	☐	☐	
Psychosis	☐	☐	☐	☐	
Personality Disorders	☐	☐			
Sleep Problems	☐	☐	☐		
Self-harm	☐				☐
Other Mental Health	☐	☐			
Mental health not defined		☐	☐	☐	

Table 2: Total Individuals by Disorder Group, Data Source and 12-month Population Prevalence Rates

Mental Health Problem	PRIMHD	NMDS	Pharms	Socrates	Mortality	Total	Pop. Rate*
Emotional	450	60	30,930			31,299	2,022
Substance	14,247	3,069	270	S		16,545	1,766 ^{\$}
Disruptive	3,423	180	12,414	591		14,091	910
Anxiety	3,957	1,377	6,633	148		11,040	714
Depression	3,837	1,311	7,116	21		10,911	706
Sleep	15	12	9,552			9,582	619
Psychosis	1,155	855	990	6		2,037	217 ^{\$}
Eating	1,215	294				1,299	84
Personality	447	276		S		591	128 [#]
Bipolar	369	210		S		477	74 [%]
Sub Total (Any Problem)	25,107	5,715	54,417	690	n/a	72,363	4,676
MH not defined	9,588	S	5,442	S		10,863	702
Self-harm		2,769			105	2,874	186
Other MH	945	339				1,254	81
Total (Any Problem)	35,316	7,629	59,862	690	105	84,291	5,446

* Total unique individuals (across all data sources) per 100,000 youth (0-24) population (unless otherwise stated)

S Data suppressed due to number of case identifications being less than 6

^{\$} per 100,000 10-24 year old population

[#] per 100,000 18-24 year old population

[%] per 100,000 15-24 year old population

Figures

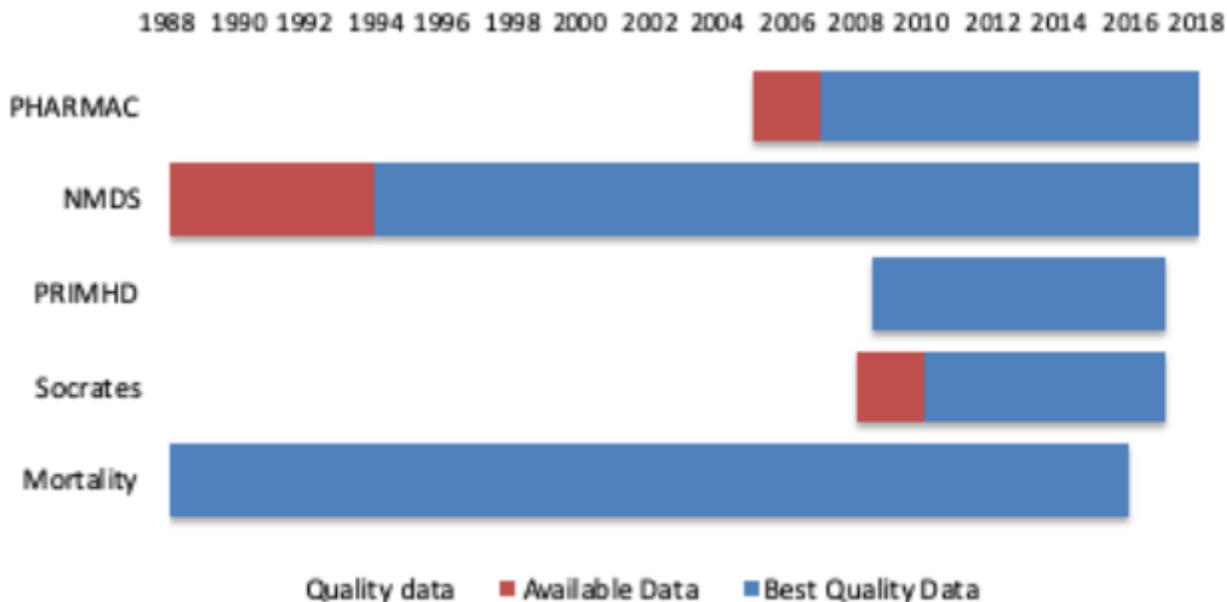


Figure 1

Dataset Coverage within the IDI