Substance and alcohol use data in the Integrated Data Infrastructure (IDI)

October 2019

Acknowledgements

Copyright

Disclaimer

Table of Contents

Introduction

This research was commissioned by the Health Promotion Agency / Te Hiringa Hauora (HPA) to understand and document what data on alcohol use is held within the Integrated Data Infrastructure (IDI). This information will support alcohol and IDI researchers to make better use of the IDI for understanding alcohol-related harm in New Zealand.

The outcome of the research is that some people who use harmful substances (which includes alcohol but is a broader definition) can be defined using available data in the IDI, as can some people who have health conditions as a result of alcohol use in utero. However, there are limitations in how these populations are captured within the IDI due to the reliance on government funded service use information.

 This report sets out a working data definition of these two populations:

1. People who use harmful substances directly (substance use population, of which alcohol is a subset)
2. People who have health conditions as a result of alcohol use in utero (alcohol-use in utero population)

The IDI is a research database managed by Statistics New Zealand (Stats NZ). It includes linked data about individuals from administrative data from government departments and non-governmental organisations (NGOs), and national collections of survey data. More information on the IDI can be found on the [Stats NZ website](#).

This report details the data sources used to identify substance and alcohol use as well as commentary on:

- How this data can be used in the IDI, including appropriate time periods for analysis
- The quality of the datasets used in this report
- Available survey data in the IDI that has not been used in the definitions of this report.

Definition of substance and alcohol use using the IDI

In the IDI there are multiple sources of data that contain information on substance and alcohol use. Broadly, the use of substances and alcohol from available data sources can be categorised into two distinct populations:

1. People who use harmful substances directly (substance use population, of which alcohol is a subset)
2. People who have developmental conditions as a result of alcohol use in utero (alcohol-use in utero population).

The reason the first population (substance use population) is defined as all substance use, rather than just alcohol use, is that many of the codes in the data capture all substance use and don't distinguish between alcohol use and other substance use. People often have alcohol and substance use issues at the same time and so it doesn't make sense clinically to separate the treatment of these two conditions.

This means any analysis cannot be meaningfully restricted to alcohol users only, it must include all substance users.

The reason the second population (alcohol-use in utero population) can be defined as alcohol effects only, rather than all substance effects, is that diagnosis of Foetal Alcohol Spectrum Disorders (FASD) and Foetal Alcohol Syndrome (FAS) are both alcohol specific. FASD and FAS diagnoses are available in the data to identify this population.

The definitions in Table 1 set out alcohol-specific codes and substance use more broadly, so users can understand why alcohol use cannot be separated in the data.

Definition of harmful substance use using the IDI: substance use population

In the IDI we recommend using four administrative data sources to identify a population of people who are currently using harmful substances (or are in treatment for the use of harmful substances):

1. the Programme for the Integration of Mental Health Data (PRIMHD)
2. the Pharmaceutical Collection
3. the National Minimum Dataset (NMDS) – publicly funded hospital discharges
4. the Ministry of Social Development (MSD) incapacity data

This is all administrative data. The IDI also has one survey data source which captures self-identified use of alcohol: the Survey of Family, Income and Employment (SoFIE). This survey data has not been included in this standard definition because of small counts available in survey data, which limits robust analysis. However, for some specific purposes analysts may wish to use survey data, so information on this has been included in Appendix 2.

PRIMHD is a Ministry of Health National Collection of mental health and addiction information relating to service activity and outcomes for healthcare users. This data is collected from District Health Boards and non-government organisations. This dataset includes information on healthcare users who access secondary specialist alcohol and drug services funded by the government.

The Pharmaceutical Collection contains claim and payment information on publically subsidised community dispensed pharmaceuticals. It is jointly owned by the Ministry of Health and PHARMAC.

The National Minimum Dataset (NMDS) captures publicly funded hospital discharges in New Zealand. It is a Ministry of Health National Collection.

The MSD incapacity data collects information on medical certificates for income support.

Table 1 details how these datasets have been used to identify this population.

Table 1: Data sources and codes used to define the current substance use population

Data source	Codes used to define current substance use population	Notes on these codes	Alcohol specific or substance specific?
PRIMHD	Team type: 03 = Alcohol and Drug Team 10 = Alcohol and Drug Kaupapa Māori Team 11 = Co-Existing Problems Team (includes alcohol and drug use) 21 = Children and youth, alcohol and drug services Team 23 = Kaupapa Māori dual diagnosis mental health and alcohol and drug services Team And/or Activity type: T16 = Substance abuse Withdrawal management/detoxification occupied bed nights (medical) T17 = Substance abuse detoxification attendances (social) T18 = Methadone treatment specialist service attendances T19 = Methadone treatment specialist service attendances (consumers of authorised GP's) T20 = Substance abuse residential service occupied bed nights T48 – Co-existing mental health problems with alcohol and/or other drug misuse – residential service occupied bed nights And/or Diagnosis type: As per ICD-10 codes listed below	Team type is a code that identifies which team provided a service to the healthcare user. Activity type is a code that classifies the type of healthcare activity provided. Diagnosis codes are incomplete for NGO organisations and for newly diagnosed individuals.	All team types and activity types in PRIMHD relate to harmful substance use, not specifically to alcohol use. Diagnosis code can be used to refine to alcohol specific service use.
Pharmaceutical Collection	Chemical ID codes: 1432 = Disulfiram 2367 = Calcium carbimide 3793 = Naltrexone hydrochloride		All relate to harmful substance use, not alcohol use specific
NMDS	**ICD-10-AM diagnosis code:** F10 = Alcohol related disorders F11 = Opioid related disorders F12 = Cannabis related disorders F13 = Sedative, hypnotic, or anxiolytic related disorders F14 = Cocaine related disorders F15 = Other stimulant related disorders F16 = Hallucinogen related disorders F18 = Inhalant related disorders	Note - we have not included codes that may relate to historic alcohol use e.g. alcoholic liver disease	These codes can be used to distinguish between alcohol use and other substance use. Both have been included for completeness of the definition.

	F19 = Other psychoactive substance related disorders		
	F55 = Abuse of non-psychoactive substances		
	ICD-9-CMA-II diagnosis code:		
	291 = Alcohol-induced mental disorders		
	292 = Drug-induced mental disorders		
	303.0 = Acute alcoholic intoxication		
	304.0 = Opioid type dependence		
	305.2 = Cannabis abuse		
	305.3 = Hallucinogen abuse		
	305.4 = Sedative, hypnotic or anxiolytic abuse		
	305.5 = Opioid abuse		
	305.6 = Cocaine abuse		
	305.7 = Amphetamine or related acting sympathomimetic abuse		
	305.8 = Antidepressant type abuse		
	305.9 = Other, mixed, or unspecified drug abuse		
MSD incapacity	Incapacity code = 006, 007, 170, 171, 172	These incapacity codes capture medical reasons in a medical certificate relating to substance use. At present little is known about the quality of the diagnostic information in the incapacity dataset.	All relate to harmful substance use, not alcohol use specific

Limitations of this definition include:

- Undercounting: As these data sources are all administrative, they capture treatment service use in relation to harmful substance use, not harmful substance use directly. This means this data only captures a subset of harmful substance use in New Zealand.
- Over counting: Where these data capture treatments rather than diagnoses, there may be an over counting of substance use. For example, stimulant pharmaceuticals can be used to treat substance addiction, but can sometimes be used to treat other, unrelated health conditions.

Other notes to be aware of:

- In using PRIMHD, the Ministry of Health have advised to use the diagnosis data with care because of known quality issues. In particular, diagnosis codes are incomplete for NGO organisations and for newly diagnosed individuals.
- The diagnosis categories provided by this definition are not exhaustive and should be treated as simple high-level inferences made from the IDI data. These should not be treated as definitive medical diagnosis information.
- More details on each of these data sources, including comments on their quality, can be found in Appendix 1.

Definition of alcohol use using the IDI: alcohol-use in utero population

In the IDI we recommend using three administrative data sources to identify a population of people who either have or potentially could have developmental conditions as a result of harmful alcohol-use in utero:

1. SOCRATES - national needs assessment and service coordination information.
2. the Programme for the Integration of Mental Health Data (PRIMHD)
3. the National Minimum Dataset (NMDS) – publicly funded hospital discharges.

These are administrative data sources. As with the substance use population, it is possible to use SoFIE survey data to identify self-reported alcohol use during pregnancy. However, the counts for this are very small so we have not included this use of survey data in this standard definition. Details of SoFIE are in Appendix 2.

The National Needs Assessment and Service Coordination Information (SOCRATES) is used by Ministry-funded Needs Assessment and Service Coordination (NASC) agencies to record information about clients who are eligible for Disability Support Services (DSS).

The NMDS captures publicly funded hospital discharges in New Zealand and PRIMHD captures mental health and addiction service use.

Table 2: Data sources and codes used to define the alcohol-use in utero population

Data source	Codes to be used to define alcohol-use in utero population	Notes on these codes	Alcohol specific or substance specific?
SOCRATES	Any record with an associated FASD or FAS diagnosis	Identified in an individual	Alcohol specific
PRIMHD	Diagnosis type: As per ICD-10 codes listed above	Identified in an individual's birth mother during individuals in utero period (during birth mothers' pregnancy). Birth mother to individual relationship is defined through DIA's Birth, Deaths, and Marriage register in the IDI.	Alcohol specific
NMDS	ICD-10-AM diagnosis code: P04.3 = Newborn affected by maternal use of alcohol Q86.0 = Foetal alcohol syndrome (dysmorphic)	Identified in an individual	Alcohol specific
NMDS	ICD-10-AM diagnosis code: F10 = Alcohol related disorders ICD-9-CMA-II diagnosis code: 291 = Alcohol-induced mental disorders 303.0 = Acute alcoholic intoxication	Identified in an individual's birth mother during individuals in utero period (during birth mothers' pregnancy). Birth mother to individual relationship is defined through DIA's Birth, Deaths,	Alcohol specific

		and Marriage register in the IDI.	

The main limitation with this definition of the alcohol-use in utero population is undercounting. These data sources are administrative, capturing either:

- needs assessments for people eligible for disability support services who have a recorded diagnosis of FASD or FAS;
- babies being treated in hospital with a diagnosis of maternal use of alcohol or FAS symptoms; or
- mothers being treated in hospital or accessing addiction services for alcohol related diagnoses while pregnant.

These scenarios are severe outcomes of harmful alcohol use in pregnancy and will under-represent the total harm caused by alcohol.

The definition may also overcount as it assumes that mothers with an alcohol abuse diagnosis are current drinkers, which may not be accurate. It also assumes that all children exposed to alcohol in pregnancy (to the extent that warrants hospital or addiction service care) will be affected by developmental problems.

The diagnosis categories provided by this definition are not exhaustive and should be treated as simple high-level inferences made from the IDI data. These should not be treated as definitive medical diagnosis information.

Using this data for analysis, including suggested time periods

This definitions in this report aim to capture as many individuals who are users of substances or who suffer from the effects of alcohol use in utero as possible. Analysts can use subsets of the codes provided in Table 1 and Table 2 to answer specific research questions. The limitations of these datasets and the assumptions being made should however be noted.

The administrative data sources in Table 1 and Table 2 are available over different time periods. The common overlap between these data sources is 2010 – 2018.

PRIMHD data for DHBs is available from 1 July 2008. NGO data was included from 2010, but the completeness of this information has changed over time and may affect any time-trends.

The major benefit of using IDI data is the ability to link individuals' data across a wide range of domains. For instance, once an analysis population has been defined using the definitions outlined in this report, it is possible to analyse service use and outcomes across such subject areas as:

- Education participation and outcomes
- Employment and unemployment
- Access to support and services
- Demographic and geographic information
- Interactions with the justice sector

Appendix 1: Quality of data sources used in this report

Data dictionaries for the administrative datasets described in this report are available from the StatsNZ Data in the IDI website. In many cases these data dictionaries set out data quality limitations, as well as providing a useful reference for analysts on the use of the datasets.

While each dataset does have quality issues, the data is still appropriate for most analytical purposes, providing the quality issues are considered when drawing conclusions from the data. Some of these limitations are summarised below.

PRIMHD data quality

PRIMHD was established in July 2008. Prior to PRIMHD, mental health data was collected in the Mental Health Information National Collection (MHINC). This older data remains available for use in reporting and ad hoc queries run by the Ministry of Health Data Services team and is available in the IDI on request to Stats NZ.

Limitations

- It is recommended that care is taken when using diagnosis data from PRIMHD due to known quality and completeness issues.
- A diagnosis is not required to be submitted to PRIMHD within the first 3 months of treatment. This means there may be a large number of clients with no diagnosis recorded, particularly if they received short term treatment.
- NGO's do not generally submit diagnosis data to PRIMHD.
- Some NGOs are behind on their reporting, skewing results. Further, as these NGOs come up to speed, increased NGO reporting will influence apparent trends.
- There are regional differences in the completeness of data for older people. Mental health conditions and addiction services for older people are funded as mental health conditions, and as addiction services in the Northern and Midland regions. In the Southern and Central regions these services are funded as disability support services. Data capture in PRIMHD is unequal across these sources, with the result that older people living in the Southern and Central regions are likely to be underreported.
- Observed variance and trends may be a result of differences in coding practices across service providers and time, for example, coding changes have influenced the number of crisis contact services reported by some DHBs. The Ministry of Health has published a Guide to PRIMHD Activity Collection and Use to address this issue. PRIMHD is a living data collection, which continues to be revised and updated as data reporting processes are improved. For this reason, previously published data may be liable to amendments. In particular, there was notable change made to the coding of team types as part of the HISO review of the PRIMHD Code set. Team type data, extracted before 1/7/2014, should not be compared with the data within the PRIMHD tables.
- It is known that there are referrals in PRIMHD with no referral end date but the referral is no longer active. It is believed that up to 30% of all open referrals in PRIMHD at the end of 2016 had not had any activity in the last 15 months.

Pharmaceutical Collection data quality

The Pharmaceutical Collection supports the management of pharmaceutical subsidies. It is jointly owned by the Ministry of Health and PHARMAC. The Pharmaceutical Collection contains claim and payment information for publicly subsidised community dispensed pharmaceuticals.

Limitations

- This data source does not include prescriptions that were never dispensed. Research has shown that people who report high levels of psychological distress are more likely to face financial barriers to collecting prescriptions than those reporting low levels of psychological distress (Jatrana, Crampton, Carter, & Richardson, 2008). Similarly, patients may pick up prescriptions but not follow through with taking the medication. This may happen for many reasons including losing the medication or actively choosing not to take the medication.
- The data does not capture where the prescription for the dispensed medicine was issued. This means it is not possible to distinguish between primary or secondary care or public or private care using this data source.
- Pharmaceuticals dispensed in hospital settings (with the exception of oncology drugs) are not captured in the Pharmaceutical Collection.
- Data is known to include exact duplicate records due to the credit, resubmit and reversal process. This will not over count the number of people accessing services, but may overinflate medications and costs. The Ministry of Health has developed methodology to deal with this issue.
- Medications for addiction are not recommended in pregnancy. However, there are some cases where these medications have been prescribed in pregnancy. In these cases it is most likely a prescribing doctor and patient have considered and weighed up the risks of medication as an appropriate treatment for alcohol use.
- NHI completeness drops to below 90% prior to 2006. This limits the ability to link dispensing information to an individual.

Needs Assessment and Service Coordination (SOCRATES)

Limitations

- SOCRATES will not capture all children diagnosed with FAS/FASD, only those who are eligible for Disability Support System Services.
- It often takes time to diagnose these children and see them come through in the data. Often children go undiagnosed. Research has suggested the prevalence is around 3% of births so there should be 1,800 children identified each year. Our case study into pregnant mothers showed that across three years 2009 to 2011 there were less than 30 children born who were subsequently diagnosed with FASD.
- Ministry of Health data dictionaries on this dataset do not include any information about data quality of this dataset. The Ministry does not have any information available in their collections website on the quality of this dataset.

National Minimum DataSet (NMDS) quality

The NMDS is a national collection of publicly funded hospital discharge information, including clinical information, for inpatients and day patients.

Limitations

- Privately funded hospital events have been excluded from the version of this dataset in the IDI because of the lack of completeness.
- Some publicly funded hospital events may not have cost associated because the purchase unit costing information was not available.

- While short stay emergency department events are included, coverage of these events has changed over time. Additional information is provided here: https://www.health.govt.nz/publication/factsheet-short-stay-emergency-department-events

Appendix 2: Survey data available in the IDI relating to alcohol use

The Survey of Family, Income and Employment (SoFIE) includes self-identified excessive drinking screening questions, similar to the Alcohol Use Disorders Identification Test (AUDIT).

As at August 2019, this is the only survey information available in the IDI relating to alcohol use.

SoFIE has not been included in the standard definitions of harmful substance use or effects of alcohol use in utero because:

- There is a small number of records relating to alcohol use, especially during pregnancy
- relevant data is over 10 years old
- working with weighted survey data alongside administrative data sources is complex.

If researchers wish to use SoFIE for specific alcohol related research questions it is important to note, due to the data being survey data, all responses are self-reported and are therefore likely to result in an undercount of alcohol use, especially during pregnancy.

Information available in SoFIE about alcohol use

A simple yes/no question asks if the respondent has had a drink of alcohol in the last 12 months. If yes, this is followed up with questions about their drinking patterns in the last four weeks. They are asked:

- how many days in the last four weeks have you had a drink containing alcohol?
- how many drinks containing alcohol did you have on a typical day when you were drinking?
- how many occasions did you have six/eight (six for women, eight for men) or more drinks containing alcohol?
- what alcoholic drink did you drink most often in the last four weeks?

SoFIE was a longitudinal survey, with respondents surveyed each year. It was run over an 8-year period from 2002 - 2010, each year is known as a wave.

The alcohol related questions were only available in the health module to SoFIE, which were run in waves 3, 5 and 7 (2005, 2007, and 2009).

Appendix 3: IDI Disclaimer

The results in this report 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 report are those of the author(s), not Statistics NZ.

Access to the anonymised data used in this study was provided by Statistics NZ 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 report 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.