Novel applications of the Integrated Data Infrastructure for longitudinal analysis for the Māori population

Tori Diamond

A thesis submitted in partial fulfilment of the requirements for
the degree of Master of Science in Statistics
the University of Auckland, 2023

Acknowledgements

There are many people whom I would like to thank for their support throughout my Masters.

Firstly, I want to express my appreciation to my primary and co-supervisors, Matt Edwards and Andrew Sporle. I'm immensely grateful for your unwavering support, guidance and mentorship throughout this project. Your work repeatedly inspires me to continue my research journey, and I hope to one day provide the same inspiration for young Māori researchers.

I would like to express my gratitude to others in this research space who have offered endless support in developing this research. I want to give particular thanks to Nicole Satherley, Tom Elliot, Lara Greaves and Chao Li. Whether it be technical advice, aid in the Datalab or other avenues of mentorship. I'm thankful to be surrounded by a great team of like-minded individuals.

Outside the academic world, I'm surrounded by the most validating and compassionate group of friends and whānau. Your continual motivation has kept me emotionally grounded with my spirits high.

Finally, I'm forever grateful to be surrounded by incredibly powerful women - my Mum, Aunty, and Nana. You have shaped me into who I am today, and I appreciate your ongoing love and support.

Abstract

New Zealand has a strong history of longitudinal research, with studies such as the *Multidisciplinary Health and Development Study*, the *Christchurch Health and Development Study*, the *Pacific Island Families Study*, and *Growing Up in New Zealand*. However, all these studies involve single cohorts, and only *Growing Up in New Zealand* includes a large proportion of Māori in its cohort. New Zealand also has an important research resource in the Integrated Data Infrastructure (IDI), a database of administrative and survey datasets containing a wide range of variables linkable at the individual level. The existing longitudinal studies are not able to link to the range of information within the IDI. Therefore, such longitudinal studies are yet to demonstrate the ability to link official statistics with administrative data in the IDI resource. This thesis aims to extend the utility of linking an official statistics survey with administrative data and the ability to complete longitudinal analysis on this cohort.

Practical examples using various individual, household and geographic variables from the IDI are conveyed, and their effects on two health outcomes for the Te Kupenga 2013 cohort: Ambulatory Sensitive Hospitalisations (ASH) and COVID-19 Vaccinations. Regression analysis displayed that housing and geographic factors do not affect ASH events. However, individual characteristics such as disability and medical discrimination impacted the odds of an ASH event. Further analysis uncovered that measures of trust in fair healthcare and whānau wellbeing impact COVID-19 Vaccinations.

This thesis demonstrates that it is feasible to turn a sample survey into a cohort for longitudinal analysis in the IDI. The process for doing this is described in detail, including data management and analytic code that can be used with Te Kupenga or other datasets with some modifications. Developing this method has highlighted that changes to the structure and function of the IDI resources would simplify similar research in the future. This thesis concludes by outlining these issues and potential solutions and then provides recommendations for improving IDI's capability for longitudinal research.

Disclaimers

Access to the data used in this study was provided by Stats NZ under conditions designed to give effect to the security and confidentiality provisions of the Data and Statistics Act 2022.
The results presented in this study are the work of the author, not Stats NZ or individual data suppliers

These results are not official statistics.
They have been created for research purposes from the Integrated Data Infrastructure (IDI) which is carefully managed by Stats NZ.
For more information about the IDI please visit https://www.stats.govt.nz/integrated-data/.

The results are based in part on tax data supplied by Inland Revenue to Stats NZ under the Tax Administration Act 1994 for statistical purposes.
Any discussion of data limitations or weaknesses is in the context of using the IDI for statistical purposes, and is not related to the data's ability to support Inland Revenue's core operational requirements.

Data in this report have been reported in accordance with Stats NZ's confidentiality rules for microdata use. For weighted counts applicable to the Te Kupenga survey, rounding to the base of 500 has been applied, with weighted counts less than 500 suppressed (S). Random rounding to the base 3 has been applied to all unweighted count data, and counts of less than 6 have been suppressed (S) in these circumstances.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

This thesis creates new robust methods for enhancing official statistics by linking official statistics sample survey data with Census data and data from the Integrated Data Infrastructure (IDI) to create the capability to undertake longitudinal analysis of outcomes for the survey cohort. While the IDI holds a wealth of information about New Zealand's population, the Te Kupenga post-censal survey has the largest representative sample of any official statistics survey and is the only survey with Māori culturally informed variables. This research uses Te Kupenga as a foundational cohort for longitudinal analysis with novel linkage methods applied to outcomes in different time periods. Once the data is linked, various regression analysis approaches are applied to examine the impact of factors that impact real-life outcomes for Māori. This project demonstrates practical examples by using various individual, household and geographic variables from the IDI and their effects on two health outcomes for the Te Kupenga cohort: Ambulatory Sensitive Hospitalisations (ASH) and COVID-19 Vaccinations. Example code is also included for reproducible research in the future. Overall, this project creates novel statistical methods for longitudinal analysis using sample survey data and data linkage using Te Kupenga as an example. However, the methods developed could be applied to any sample survey in the official statistics system co-located with IDI data.

1.1 Linkable administrative data in Aotearoa, New Zealand

The IDI is a research database of linkable datasets that holds data about people and households within Aotearoa, New Zealand (Milne et al., 2019). This database includes information from multiple government agencies, including immigration, health, income, and education. Surveys conducted by Stats NZ, such as the Census, are also included in this database. Although a substantial amount of New Zealand's population is included, all entries are de-identified to ensure confidentiality for individuals. De-identification removes personal information such as name, address, and date of birth in addition to the encryption of individual NHI and IRD identification numbers. Therefore, while the IDI holds a wide range of information, protecting an individual's identity remains a priority. The IDI spine holds information about those who have ever been resident in New Zealand (Milne et al., 2019) and is created through birth records, visa records and tax records (Black, 2016; Virtual Health Information Network, 2020). Datasets within the IDI are firstly linked to the spine. Using the unique identifier within the IDI, individual-specific data can be linked across different datasets, including datasets from different agencies (Milne et al., 2019).

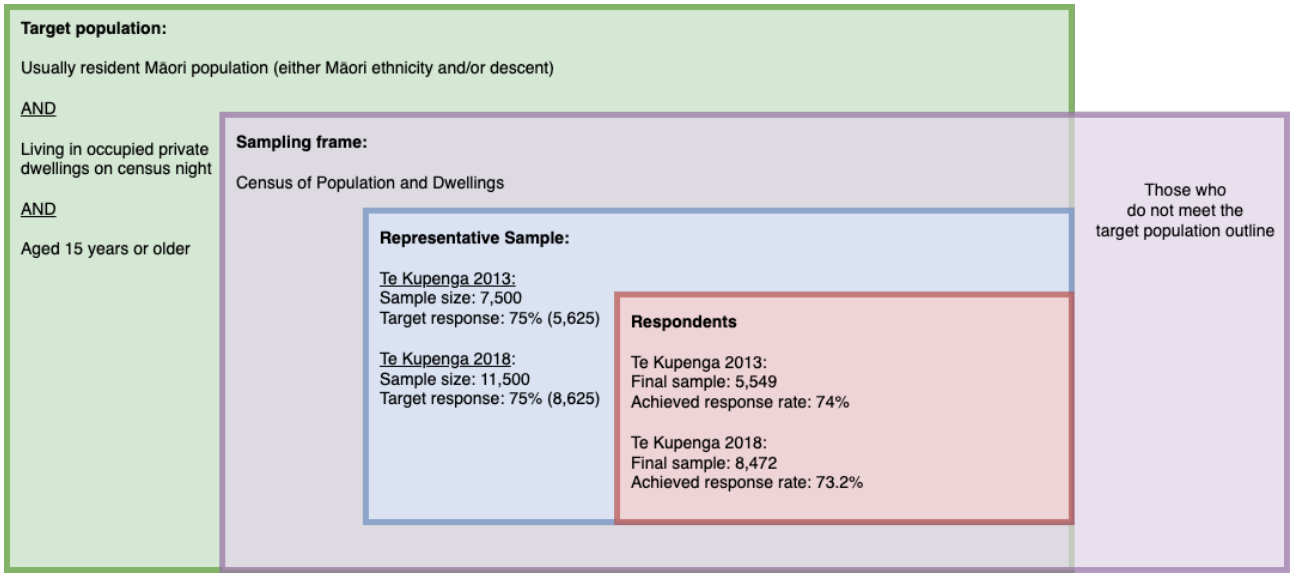
The IDI is used across research within government sectors and institutions to explore complex issues within the population. All research conducted using the IDI is set out to benefit individuals throughout New Zealand (Stats NZ, 2022). Using the IDI enables linking individuals across datasets and government agencies. This allows in-depth information on how a person interacts with government services. For researchers, this provides data for the issues that most affect our population and serves as levers for improvements to policy.

The IDI holds data about individuals who have interacted with government departments within New Zealand. Therefore, the database holds more records from those who interact with the government system at higher rates. Furthermore, Māori are shown to have more data points within the IDI. As a relatively young population and majority seemingly born in NZ, Māori are more likely to have a digital record and higher levels of interaction with state agencies that require a digital record, such as the police (Milne et al., 2019; Greaves et al., 2023).

The IDI contains a large amount of data about Māori but is often created through interactions with agencies or organisations, thus producing data focusing on problems. Such a predominance of 'negative' data can lead to research with a deficit framing, where Māori are described in terms of problems. This creates a system that holds more information **about** Māori, rather than **for** Māori (Greaves et al., 2023). This thesis contributes to addressing this issue by creating robust statistical approaches for longitudinal analysis incorporating survey data into an analysis of outcomes for Māori.

1.2 Te Kupenga

Figure 1.1: Sampling method for 2013 and 2018 Te Kupenga



Modified from Barry Milne's STATS240 slides (Milne, 2023)

Te Kupenga is the only official statistics survey containing tikanga-informed wellbeing variables about the Māori population (defined by either Māori ethnicity or Māori descent) and those aged 15 or over. In New Zealand official statistics, Māori identity is measured through two distinct methods - Māori ethnicity and Māori descent. Māori ethnicity is defined by a person's sense of belonging and identity, while descent is defined by whakapapa (ancestry) (Gleisner et al., 2015). Ethnicity measurements have evolved over multiple Censuses, from blood quantum methods to the ability to identify to more than one ethnic group (Cormack and Robson, 2010; Cormack et al., 2019; Allan, 2001). Alongside this, a person's ethnic identity can also evolve throughout time, influencing their self-reported ethnicity. While this is a fluid measure, Māori descent offers an objective measure considering whakapapa and whānau connections, whether an individual identifies as ethnically Māori or not.

The first version of Te Kupenga was released in 2013, which included a sample of 5,549 respondents (Stats NZ, 2018, 2014). This iteration of the survey included questions on areas surrounding Māori wellbeing including wairuaranga (spirituality), tikanga (Māori customs and practices), te reo Māori (Māori language) and whanaungatanga (social connectedness). Numerous Māori stakeholders participated in the development of the Te Kupenga survey including leaders from iwi and Māori businesses and groups, researchers across New Zealand universities and multiple government agencies (Stats NZ,

2018). The aim of this post censal survey was to collect information specific to Māori wellbeing, which is critical to Te Ao Māori (Māori worldview) and improving outcomes for Māori.

After Te Kupenga 2013, consultation processes between Stats NZ and selected stakeholders meant the 2018 version could address any need for content updates and the possibility of increasing the sample size. Te Kupenga 2018 is the most recent version of this survey, which includes roughly 8,500 adults and an Auckland representative sub-sample and provides further data on Māori cultural wellbeing, with kaitiakitanga as a newly added variable (Stats NZ, 2018).

The Te Kupenga surveys are cross-sectional surveys involving nationally representative samples from the 2013 and 2018 Census of adult Māori (15 years and older) who are selected through the Census population and Dwellings information (see Figure 1.1). Te Kupenga 2018 involved a new and larger sample, so those individuals in the 2013 survey are not necessarily present in the 2018 sample (ie. It is not a repeated sample). Although Te Kupenga is not a direct sub-sample of the Census, there is limited documentation about the sampling frame applied. Te Kupenga 2013 and 2018 involve different sampling methodologies; however, the only published resources available apply to methods used for Te Kupenga 2018 (Stats NZ, 2020a). Therefore, this must be considered when applying a longitudinal lens to this sample, as timeliness and weighting implications must be prioritised.

1.3 Te Kupenga 2013 vs. Te Kupenga 2018

While the Te Kupenga survey offers a wealth of information based on social contexts for Māori, it is essential to examine the differences across both survey iterations. This incorporates viewing both the strengths and limitations in each survey's context and possible influences caused by the different Censuses.

The table 1.1 below displays the number of Māori (by ethnicity and/or descent) recorded for the 2013 Census, 2013 Te Kupenga, 2018 Census and 2018 Te Kupenga samples (diagonals). The surrounding entries reflect the overlap in observations between each sample (for example, the number of Māori recorded in Census 2013 and recorded in Census 2018 is 570,873). These unweighted counts are output from the IDI and randomly rounded to base 3, with any counts under 6 suppressed. This table shows the number of matched individuals included across the Census rounds and Te Kupenga surveys to enable the inclusion of variables across time. Although Te Kupenga assesses similar domains of Te Ao Māori and wellbeing, the two surveys only overlap by 48 observations. Between the two Census, those who fall into the categorisation of identifying as Māori (by descent and/or ethnicity) differ by approximately 200,000.

Table 1.1: Number of observations (Māori by ethnicity and/or descent) in initial joining

	Census2018	Census2013	TK2018	TK2013	Unmatchable
Census2018	926,490	570,873	8,430	4,515	
Census2013	570,873	707,250	6,780	5,457	
TK2018	8,430	6,777	8,472	48	42
TK2013	4,515	5,457	48	5,457	S
Unmatchable			42	S	

Note: observation numbers from Te Kupenga (TK) surveys 2013 and 2018 are unweighted statistics (survey weights were not used in these initial summary statistics) (Li, 2023)
S (*Suppression*) occurs when raw, unweighted counts are < 6 to protect an individuals' privacy and reduce risk of identification

Data collection methods for the 2013 and 2018 Census were carried out with different approaches, therefore displaying consequential differences in output for each Census and the associated Te Kupenga surveys. The 2018 Census used a digital-first approach, creating missing data issues and low response

rates overall (Milne et al., 2019). Therefore, to combat this issue and its potential ongoing impacts, Statistics NZ used methods such as imputation from administrative sources to cover missing data (Jack and Graziadei, 2019). This has resulted in an improved population coverage by the Census, but it has differences in methodology and lower completion rates for many Census variables. As a result of this issue, Census 2013 continues to be used as a data source for research for the Māori population (Nicholson Consulting & Kōtātā Insight, 2021; Te Kāhui Raraunga Charitable Trust (TKR), 2023).

Census 2023 outputs are still being completed by Stats NZ at the time of writing this thesis and are not available for analysis. Stats NZ has outlined that there will not be a Te Kupenga survey following the recent Census, with the next Te Kupenga survey being held in 2028 (Huirama, 2023). This emphasises the need for novel methodological approaches to be created now to build longitudinal capability for the existing Te Kupenga survey data resources. This thesis will provide longitudinal methodology using Te Kupenga as an exemplar, but could be adapted and applied across other official statistics sample surveys.

1.4 Longitudinal Health Research in New Zealand & Utilisation of the IDI

Longitudinal studies are often used to explore past or future outcomes for a population of interest. In health research, longitudinal analysis is widely used to address changes in health outcomes for a sample across a period of time. Furthermore, longitudinal studies can assess for the risk of disease or the effectiveness of a treatment.Fitzmaurice et al. (2011) explain that measuring a repeated sample of individuals through time creates a defining element of longitudinal studies. Changes which occur during this period can be explored on an individual and group level, along with the possible factors that influence such change.

New Zealand currently has high-quality longitudinal studies, including the *Dunedin Multidisciplinary Health and Development Study*, the *Christchurch Health and Development Study*, the *Pacific Island Families Study*, and *Growing Up in New Zealand*. Only *Growing Up in New Zealand* has a sufficiently large number (1500 approximately at recruitment) and a proportion of Māori (24%) in its cohort to enable Māori-specific analysis. However, the cohort is still very young, at around 14 years of age (Morton et al., 2012). The Christchurch and Dunedin Studies include Māori in their cohort; however, the numbers are too small for anything other than simple descriptive analysis (Broughton et al., 2000). The *New Zealand Census Mortality Study* completed whole population data linkage based on 6 linked Census population cohorts (1981, 1986, 1991, 1996, 2001, 2016) and included a specific focus on Māori outcomes but was limited to mortality and cancer outcomes.

This thesis can, therefore, make a significant contribution to improving the capability of New Zealand's official statistics to provide information of value to Māori as it creates a longitudinal study from a nationally-representative sample of Māori with access to a broad range of information, including Māori culturally-informed variables and information from multiple years. This broad range of information includes measures of socio-economic position at the level of the individual, household or geographic area, allowing for a range of different theoretical and causal socio-economic models to be included in the longitudinal analysis (Galobardes, 2006a,b).

1.5 Aims for Thesis

This thesis focuses on demonstrating novel approaches to using IDI data and the Te Kupenga survey for longitudinal analyses of Māori health outcomes. This will explore the ability to transform Te Kupenga into a longitudinal survey by linking to microdata from other datasets in the IDI. Linkage to other datasets enables this work to utilise individual, household and geographic level variables as possible predictors for health outcomes.

Demonstrating different possible longitudinal approaches using IDI data, the thesis starts by exploring outcomes at different population levels. The first outcome, Ambulatory Sensitive Hospitalisations (ASH), will be analysed for the whole population, then for the Te Kupenga sample only. The ASH outcome will be explored on different population levels, such as Census 2013 Māori vs. Non-Māori, Census 2013 Māori only and Te Kupenga 2013. As Te Kupenga is a nationally representative sample of adult Māori, this will be used as the underlying sample from this point onward in the project.

Exploring the specific impacts for the Te Kupenga sample will involve a second health outcome example using COVID-19 vaccinations, which uses linkages across two Census and further sources of microdata. As the sample and outcome periods are more than seven years apart, this section of the project addresses longitudinal approaches while considering the potential impacts on the sample weighting.

While using the Te Kupenga 2013 survey enabled the exploration of weighting methodology, the survey also offers measures specific to Māori wellbeing. Incorporating these variables means the ability to explore factors that are essential outcomes in themselves but may also protect Māori from detrimental health outcomes. Using Te Kupenga as an example of what is possible, the final aim of this thesis is to provide recommendations for improving and continuing this type of analysis in the IDI using an official statistics survey.

1.5.1 National population levels (using ASH)

ASH are hospitalisations that are preventable in a primary health care setting and are often used as a marker of lack of access to timely and adequate health care, including inequitable outcomes by ethnicity or socio-economic position (Sheridan et al., 2023; Singh et al., 2021). Therefore, it is essential to first explore this outcome on a whole population level before examining the results for the Māori population. The first stage of this project will use ASH records (2014 and onward) as the outcome and linking to Census 2013 variables and other microdata to examine differences between Māori and Non-Māori populations.

For the context of my thesis, ASH will be examined using three measures: as a binary variable (ASH event present or not), count of unique ASH events, and total duration of stay specific to an ASH event across the period. The former will be explored nationally, while the latter two measures will focus on Māori only on a national level. An exploratory analysis of the ASH binary is completed at the beginning of the project for Non-Māori vs. Māori from the 2013 Census sample before conducting deeper analyses for the Te Kupenga 2013 sample.

1.5.2 Te Kupenga 2013 sample (using ASH & COVID-19 Vaccinations)

This project aims to emphasise the longitudinal capability of Te Kupenga and will implement this by focusing on the Te Kupenga 2013 specific sample. Variables measured in the Te Kupenga 2013 survey are included to measure impacts on ASH and COVID-19 vaccinations and the individual, household and geographic factors implemented at the national population level. This type of analysis displays the novel methods used to include variables on a detailed level for a more specific sample.

COVID-19 vaccination status from 2020 onward is the second outcome measured for the Te Kupenga 2013 sample. As this outcome period begins at a different time period (seven years later), linkages are performed across IDI datasets to include variables from Census and Te Kupenga 2018. Transitions between 2013 and 2018 variables are included to demonstrate how individuals have changed between the two Census iterations. Furthermore, weighting techniques will be discussed when individuals are missing from the sample (due to death or leaving the country permanently).

These examples are selected as an introductory step to using the Te Kupenga as a longitudinal cohort; therefore, variables from 2013 and 2018 Census and Te Kupenga surveys are chosen as predictors for outcomes at different time periods. Only the Te Kupenga 2013 sample is used for this portion of the project as the focus shifts toward weighting approaches for an outcome multiple years ahead of the predictors used. Linkage across datasets in the IDI creates possible exploration of impacts from not

only Te Kupenga 2013 variables but also Census 2013 and Census 2018 variables for the underlying sample.

Carrying across approaches from the ASH branch of the project creates consistency for two different health outcomes. For COVID-19 vaccinations, this means using the same underlying sample and approach to using individual, household and geographic factors. Regression analysis with such variables will also be completed while considering the implications of weighting.

1.5.3 Developing novel methodology

The project includes two practical examples using IDI microdata with outcomes at varying time frames. By joining Te Kupenga, predictor variables from prior years will be included to assess impacts later in time. This allows for approaches to longitudinal analysis for outcomes that are not in the immediate future. Therefore, timeliness and temporality factors mean that implications on causation will be viewed throughout this project.

This research involves creating steps for using microdata for Māori across time. These steps will include demonstrating code implemented for joining and producing output using IDI health data and Te Kupenga and displaying output resulting from this process. However, any limitations found will also be documented to assist with continuing research in this area. While communicating the methodology process, this project will form the early development of linking detailed administrative data to Te Kupenga, a nationally-representative survey for Māori. Creating novel methodology using Te Kupenga as a longitudinal survey allows areas specific to Māori wellbeing to be considered, impacting Māori health outcomes through time. The exploration of data in this manner aims to provide autonomy and sovereignty for Māori and Māori data to create a strengths-based approach toward understanding what factors enable greater equitable health outcomes.

Chapter 2

Methods

2.1 Creating methodology with intent

Practical Examples (Ambulatory Sensitive Hospitalisations & COVID-19 Vaccinations)

This project incorporates two practical examples using health data from the IDI to demonstrate potential methodological approaches to longitudinal analysis with the Māori weighted social survey, Te Kupenga. The examples of interest are Ambulatory Sensitive Hospitalisations (ASH) and COVID-19 vaccinations, which aim to demonstrate outcomes highly influenced by policy action in Aotearoa, New Zealand.

ASH refer to hospitalisations that have occurred due to conditions responsive to medical interventions (prophylactic or therapeutic) in the primary healthcare setting (Jackson and Tobias, 2001). ASH is used to monitor equitable healthcare, as it measures access to primary care on a community level (Singh et al., 2021; Barker et al., 2016). Fundamentally, a high ASH rate means a significant burden on primary care and hospitals in the community (Jackson and Tobias, 2001; Palapar et al., 2020). International Classification of Diseases 10th Revision (ICD-10) determines an index of ASH conditions (Ministry of Health - Manatū Hauora, 2023, 2022). As this list represents hospitalisations preventable by primary care access, there are difficulties with determining conditions that are directly affected. Therefore, this list of conditions is extensive but only partial. For this study, ASH conditions are restricted to individuals 15 years and older, as Te Kupenga is the primary sample for this project. The complete list of ASH conditions with ICD-10 codes and relevant age groups is outlined in Table A.1.

Similar to ASH, the COVID-19 vaccination rollout demonstrates a strategy driven by policy and equitable outcomes. However, alongside the slow adoption of this vaccination strategy, recent research displayed the disproportionate impacts of COVID-19 outcomes on Māori compared to Non-Māori (Satherley et al., 2023; Steyn et al., 2020). Therefore, vaccination status is included as another practical example in this project.

COVID-19 Vaccination status is recorded by Te Whatu Ora (formally The Ministry of Health) and is available as a vaccination register within the IDI. Although full vaccination typically requires a particular number of doses, eligibility differs between individuals. According to Te Whatu Ora, vaccination eligibility depends on age, timeliness, and whether an individual is immunocompromised (Ministry of Health Manatū Hauora, 2020).

Given that Te Kupenga is a weighted social survey for Māori, variables relating to Māori wellbeing are included to view possible effects on ASH and vaccination outcomes. Combining such health outcome examples allows the exploration of longitudinal capability by merging elements of other administrative data in the IDI. To better understand the potential effects on an individual, household and geographic level, Census and other microdata will be integrated into the dataset. This approach incorporates the

prioritisation of Te Tiriti to address policy amenable factors and their impacts on health outcomes for Māori. Furthermore, methods for survey weighting will be studied when Te Kupenga 2013 is used as the underlying sample across time. This investigation aims to open the possibility of reframing a post-censal survey into a longitudinal sample while considering approaches to re-weighting and missing data (treated as loss-to-follow-up).

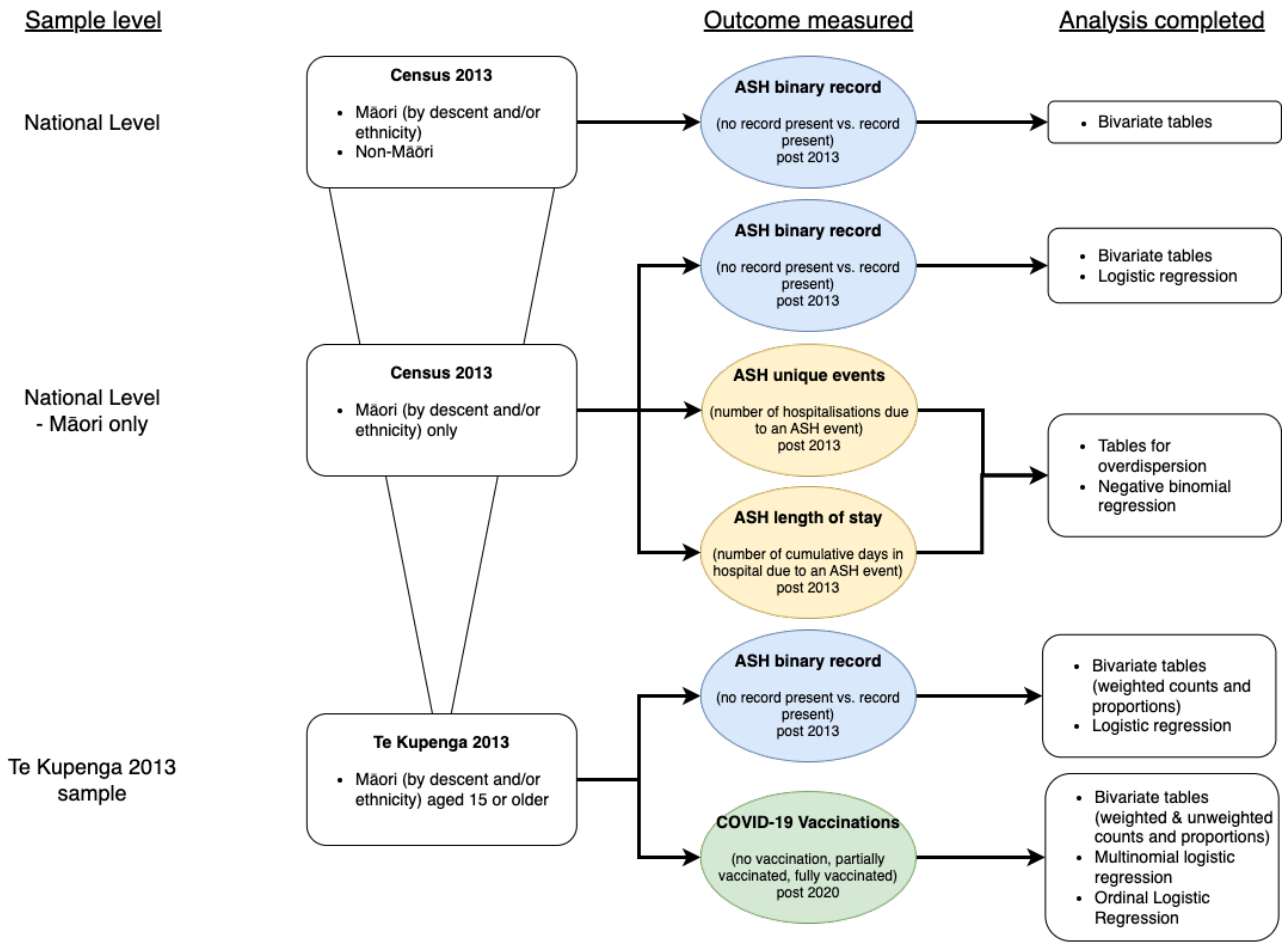
2.2 Steps taken to create methodology

2.2.1 Data Sources

This project focuses on using microdata from within the IDI. The data was sourced from Stats NZ 2013 and 2018 Census, Te Kupenga 2013 and 2018, Inland Revenue Department (IRD), Department of Internal Affairs (DIA), Ministry of Health (MoH), and core data derived by Stats NZ for date of birth, date of death and overseas spell records (Stats NZ, 2023). For the context of this project, multiple variables across datasets were collated to create variables specific to the topic of interest. The IDI undergoes a 'refresh' up to four times per year to incorporate more data and re-do record linkages to the IDI spine (Virtual Health Information Network, 2020). All datasets aside from Te Kupenga and COVID-19 Immunisation Register (CIR) data used the June 2022 IDI refresh, and are sourced from ad hoc tables within the IDI.

The first stages of the project involved reviewing the impacts of ASH on a national scale before focusing on smaller populations stratified by Māori ethnicity, then for the Te Kupenga 2013 sample. It was essential to provide a brief overview of these variables across the whole population, as an ASH event does not necessarily account for the extent of impact on the health system at one given time. For COVID-19 vaccinations, only individuals from Te Kupenga 2013 were included to explore the continuation of the sample across outcomes at different times. A visual representation of the project is displayed in Figure 2.1

Figure 2.1: Overview of outcomes measured and output produces



The analysis for ASH outcomes were carried out using a 'funnelled' approach, beginning on a national level and funnelling into a smaller population to represent Māori. This project's ASH section explored avoidable hospitalisations nationally before a focused analysis for the smaller Te Kupenga 2013 sample. Refer to Figure 2.1 for the complete visual overview of this process.

2.2.2 National level (Census 2013 sample)

The first step for analyses was examining the ASH binary outcome nationally by comparing Māori vs. Non-Māori counts. This analysis used the Census 2013 sample to create frequency tables stratified by Non-Māori vs. Māori (self-identified by descent and/or ethnicity). Understanding the distribution of Non-Māori vs. Māori across ASH events was completed by exploring counts by explanatory variables. This exploratory breakdown was used to assess for any disparity in the outcome on a larger scale.

2.2.3 National level (Census 2013 - Māori only sample)

To assess the impact of ASH on Māori specifically, it was of interest to view how ASH conditions impacted Māori on a further level of detail - through 'Length of Stay' and 'count of unique ASH events'. Bi-variate tables and various types of regression analysis were completed for the three outcomes, depending on the type of outcome (for example, binary vs. discrete count).

2.2.4 Te Kupenga 2013 sample

Once the impacts of ASH were assessed for Māori on the census level, the Te Kupenga 2013 sample was used as the primary sample for ongoing analyses. This decision allows for utilising tikanga-informed

variables measured in Te Kupenga, specific to Māori, to explore potential effects on health outcomes. The appropriate standard errors can be calculated using Te Kupenga's replicate weights to create statistics representative for the Māori population. Using a post-censal survey in a longitudinal format is largely uncommon, as methodological approaches throughout the literature are minimal. Therefore, the focus was shifted to the ASH binary outcome, including Te Kupenga 2013 specific variables and individual, household and geographic factors from Census 2013.

ASH

ASH were explored for the Te Kupenga 2013 sample only using the ASH binary outcome variable. This portion of the study focuses on predictor variables recorded in 2013 to assess their effects on ASH from 2014 onward. These include standard demographic variables used at higher levels of the analysis (sex, age, IRD individual income and highest qualification) while including disability, Te Kupenga 2013 specific variables, and household and geographic variables.

Further tables of weighted counts and proportions (with 95% CI's) were created to explain the potential impacts of predictor variables on a weighted count of the Māori population. Logistic regression analysis was also conducted to explore the impacts of the predictor variables.

COVID-19 Vaccinations

The analysis for COVID-19 vaccinations used the Te Kupenga 2013 sample only. However, the vaccination outcome period occurred seven years after the original Te Kupenga 2013 sample was created. Therefore, this opened the possibility of linking across multiple Census and Te Kupenga datasets in the IDI.

Those within the sample are linked to individual, household and geographic variables from the Census and Te Kupenga 2013 and 2018 surveys. The ability to link between surveys provides an overview of change across time points and possible effects on the outcome later in time. The first assessment of this change is displayed through transition tables for matching variables measured in both 2013 and 2018 Census (such as disability and NZDep). This explores how the sample has changed between the Census iterations before considering direct effects on the outcome.

Bivariate tables between vaccinations and predictor outcomes using weighted counts with 95% confidence intervals for the Te Kupenga 2013 sample are also provided. Weighted proportions by total population and by levels of the outcome (no vaccination, partially and fully vaccinated) with 95% CI's will be explored to establish how the outcome differs on various levels of detail. For variables included in Census 2013 and 2018, only the latter will be included for the bivariate tables and regression analyses to factor in timeliness.

Exploring approaches to analytic output will involve different regression analyses and acknowledging the potential weighting issues due to missingness (see section 4.4). An ordinal logistic regression is provided for such regression output due to the ordered nature of the vaccination outcome (no, partially, and fully vaccinated).

2.2.5 Variable definitions

All predictor variables were included on individual, household and geographic levels and are displayed in table 2.1. This approach was to understand the tiered impacts on the outcomes of interest

Table 2.1: List of variables (outcome and predictors) included - specific to using Te Kupenga 2013 sample

ASH		
Variable level	Source	Variable
Outcome	Publicly funded hospital discharges	ASH Binary
Individual	Census 2013	Age
	Census 2013	Sex
	Census 2013	Disability
	Census 2013	Highest Qualification
	IRD (2013)	IRD Individual income
	Te Kupenga 2013	Connection to tūrangawaewae
	Te Kupenga 2013	Importance of connection to culture
	Te Kupenga 2013	Trust in fair healthcare
	Te Kupenga 2013	Any discrimination while accessing medical services
Household	Te Kupenga 2013	Household crowding
Geographic	Te Kupenga 2013	NZDep2013
COVID-19 Vaccinations		
Variable level	Source	Variable
Outcome	CIR Immunisation Register	COVID-19 Vaccination Status
Individual	Census 2013	Age
	Census 2013	Sex
	Census 2018	Disability
	Census 2018	Highest Qualification
	IRD (2018)	IRD Individual income
	Te Kupenga 2013	Connection to tūrangawaewae
	Te Kupenga 2013	Importance of connection to culture
	Te Kupenga 2013	Trust in fair healthcare
	Te Kupenga 2013	Any discrimination while accessing medical services
Household	Census 2018	Household crowding
	Census 2018	Household income
	Census 2018	Household composition
	Census 2018	Housing quality
	Te Kupenga 2018	How are whānau doing?
	Te Kupenga 2018	How well are whānau getting along?
Geographic	Census 2018	NZDep2018
Censoring	Department of Internal Affairs	Date of death
	Overseas spell	Date left country permanantely

2.2.6 Outcome variables

ASH

For an overview of all the variables used in the ASH analyses for this project, refer to Table 2.1 and Figure 2.4.

ASH Binary The ASH binary variable was derived for whether an individual had never (0) or ever (1) had an ASH record after 2013. Measuring ASH as a binary variable was necessary to provide an

overview of those who are affected on a wider scale.

ASH Length of Stay Length of stay was derived by calculating the total number of days spent in hospital, due to an ASH event after 2013. This variable is presented as a continuous value. Total length of stay was of interest to explore differences in hospitalisation length across Māori vs. non-Māori. This measure is also used as a proxy to assess severity of an individuals' hospitalisation, therefore demonstrating the need for greater access to primary health care.

Count of ASH Unique Events The number of unique ASH related events were derived by the calculating the total discrete count of individual records after 2013. Similar to ASH length of stay, the number of unique ASH events are a proxy for the severity of an individuals' health.

COVID-19 Vaccinations

For an overview of all the variables used in the COVID-19 Vaccinations analyses for this project, refer to Table 2.1 and Figure 2.5

Vaccination status Vaccination status was provided by Ministry of Health records, and categorised into (1) partially vaccinated or (2) fully vaccinated up to 21 February 2023. Those who were without a vaccination record in New Zealand were classified as (0) not vaccinated. Differences between partially vaccinated and fully vaccinated depend on multiple factors such as time between primary dose and/or booster, age of individual and whether someone is considered immunocompromised (Ministry of Health Manatū Hauora, 2020). Vaccination status was only considered for those in the underlying 2013 Te Kupenga sample. Therefore, this included individuals 15 years and older, in order to align with a consistent denominator.

2.2.7 Individual predictor variables

Māori ethnicity/descent

A binary variable for Māori ethnicity was implemented for stratification purposes, by coding those who were either of Māori ethnicity (1), Māori descent (1), both (1) or not (0) in the 2013 Census. This variable was used at the beginning stages of analyses to provide an overview of the three ASH outcome variables, before placing restrictions on the sample by using the Te Kupenga 2013 only.

Sex

For both sections, age in years and sex (1 = male, 2 = female) information from 2013 Census were used.

Age

Age was included from two different sources for ASH vs. COVID vaccinations. This decision was made to factor timeliness and age at the relevant outcome period. For both age variables, The age-band categories were as follows: 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75+. For the purposes of confidentiality with the Te Kupenga sample, 10 year age-bands were derived and kept for consistency across the three outcomes of interest.

Age information was sourced from Census 2013 for the ASH section of the project. This was also included in the model to allow for stratification by age, as ASH conditions impact different age groups. Accuracy needed to be considered when understanding age for vaccinations. Therefore, an age variable was created for age at the beginning of the vaccination outcome period (2020). This age information is extracted from a full dataset of date of birth information, and represented as age at 1 Jan 2020 (general date for beginning of outcome period). This creates a variable with discrete counts,

however this information was collapsed into 10-year age-bands to better represent age groups without risk of identification.

Disability

Disability was represented as Disability status (0 = no disability, 1 = disability present) from Census 2013 and Census 2018 surveys. The former was included for ASH, while the latter was used for vaccinations. This variable (irrespective of survey) is derived from Stats NZ's definition of disability which considers disability as having difficulty or being unable to do one or more of six activities in the area of seeing, hearing, walking or climbing steps, memory, washing or dressing and communicating. These questions are based on the Washington Group question set (Washington Group on Disability Statistics, 2017; Stats NZ, 2017).

Total individual income

Four income variables were created to represent individual income. Census 2013, 2013 IRD gross income, Census 2018 and 2018 IRD gross income. Income recorded in 2018 was used for the vaccination outcome only.

Census individual income (irrespective of year) is a self-reported measure and was re-coded into six categories for this project: Zero income or a loss, $1-$15,000, $15,001-$30,000, $30,001-$50,000, $50,001-$100,000, $100,001 or more. IRD records were used as a second variable for total individual income from April to 31st March (of the corresponding year). This IRD dataset was created from taxation data, and only included records from those who have received taxable income during this period. Therefore, there were no records for individuals who recorded zero income or a loss. This was a five category variable: $1-$15,000, $15,001-$30,000, $30,001-$50,000, $50,001-$100,000, $100,001 or more.

Highest Qualification

Information about education was used from 2013 and 2018 Census records and a four-category variable derived for highest qualification completed. These categories included: no education, school level, post school and degree.

Connection to Tūrangawaewae

"How connected do you feel to your tūrangawaewae?" is a variable included from the 2013 Te Kupenga survey, which measures the extent of an individual's connection to their whenua and sense of belonging as Māori. This is a six-levelled variable ranging from "very strongly connected", through to "not at all connected" and is used on an individual level.

Importance of connection to culture

Connection to culture is another factor considered for individuals within the Te Kupenga 2013 survey, which assesses the question "Thinking about your life as a whole, how important is it for you to be involved in things to do with Māori culture?". Ranging from "very important" to "not at all important" with five possible responses. Don't know values are also included.

Trust in Fair healthcare

Trust in the healthcare institution is highly relevant to this project, as both outcomes assess outcomes for Māori. Within Te Kupenga 2013, the question "Where zero is not at all, and ten is completely, how much do you trust the health system to treat people fairly?" was asked of individuals, with responses ranging on an eleven-point scale. This question does not ask for responses relating to a particular timeframe, but a general assessment of trust in fair treatment by the healthcare system.

Any discrimination while accessing medical services

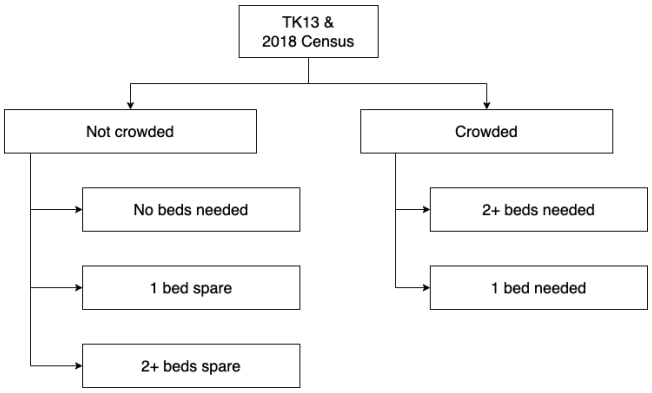
"Have you ever been discriminated against while trying to get medical care?" was assessed in the Te Kupenga 2013 survey and included in this project to establish any effects of self-reported discrimination within the healthcare system. This is a binary variable (1) yes, (0) no, and measures 'any' discrimination in contrast to discrimination within a particular timeframe. Don't know values are also included.

2.2.8 Household predictor variables

Household crowding

Household crowding was included for both sections of the project, the Te Kupenga survey from 2013 and Census 2018. This information was not available for all Census sample from 2013 (see discussion on data quality 4.2) and therefore is restricted to Te Kupenga 2013 only. Household crowding is based on the Canadian National Occupancy Standard (Canada Mortgage and Housing Corporation (CMHC), 2022), which Stats NZ uses to assess household crowding (Stats NZ, 2019b). As Figure 2.2 displays, this collapses the variable into three categories: (0) not crowded (no additional beds needed) , (1) crowded (one or more additional beds needed).

Figure 2.2: Grouping of 2013 and 2018 Crowding variable to create an aggregated household crowding variable



Household income

Household income was used from 2018 Census records, but re-categorised to match the Stats NZ Classification: Census grouped family or household income V2.0.0 (Stats NZ, 2019a) . These eight categories are as follows: $20,000 or Less, 20,001 - $30,000, $30,001 - $50,000, $50,001 - $70,000, $70,001 - $100,000, $100,001 - $150,000, $150,001 or more, Not Stated.

Household composition

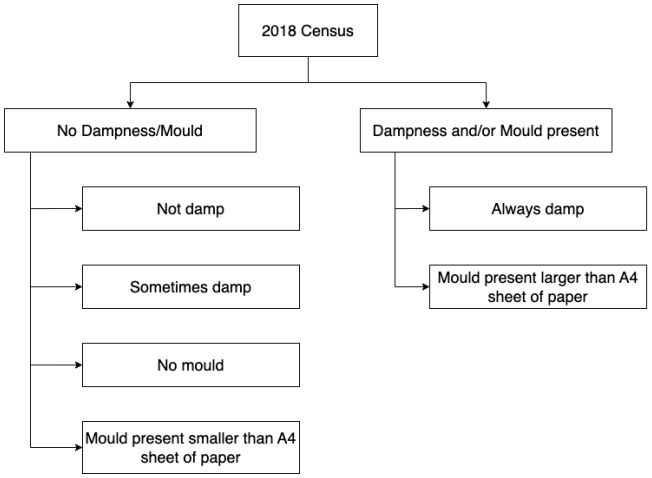
Household composition was obtained through the 2018 Census and collapsed into seven categories: 1 person, 1 couple, 1 couple with children, 1 parent with children, 2 or more families, other multi person household, and missing. These categories were chosen to represent multiple layers of household types, with an emphasis on disaggregating the typical 1 family category which can include couple, couple with one person, couple with children, one parent with children etc.

Household quality

Household quality was derived from 2018 Census variables which include a dampness indicator and mould indicator. This has the values of (0) not damp/some dampness/no mould/mould present smaller

than A4 sheet of paper, (1) always damp/mould larger than A4 sheet of paper. See Figure 2.3 for the aggregation of this variable.

Figure 2.3: Grouping of 2018 Dampness and Mould variables to create an aggregated housing quality variable



How are whānau doing?

Whānau factors are included to consider effects within households. Using the question "Where zero means extremely badly and ten means extremely well, how would you rate how your whānau is doing these days?" from the Te Kupenga 2018 survey, with responses on an eleven-point scale.

How well are whānau getting along?

A further measure of whānau wellbeing, Te Kupenga 2018 asks "In general, how would you rate the way your whānau get along with one another?". Responses range from "very well" to "very badly", with five options. Don't know values are also included.

2.2.9 Geographic predictor variables

Deprivation

Deprivation was included across the project, however, information was sourced from the Te Kupenga survey from 2013 and Census 2018. This information was not available for all Census sample from 2013 (see discussion on data quality 4.2) and therefore is restricted to Te Kupenga 2013 only. This variable included a ten levelled variable ranging from 1 (least deprived) to 10 (most deprived). However, for the purposes of consistency across official statistics research, the deprivation scale was divided into quintiles.

Figure 2.4: The IDI data sources relevant to ASH outcome

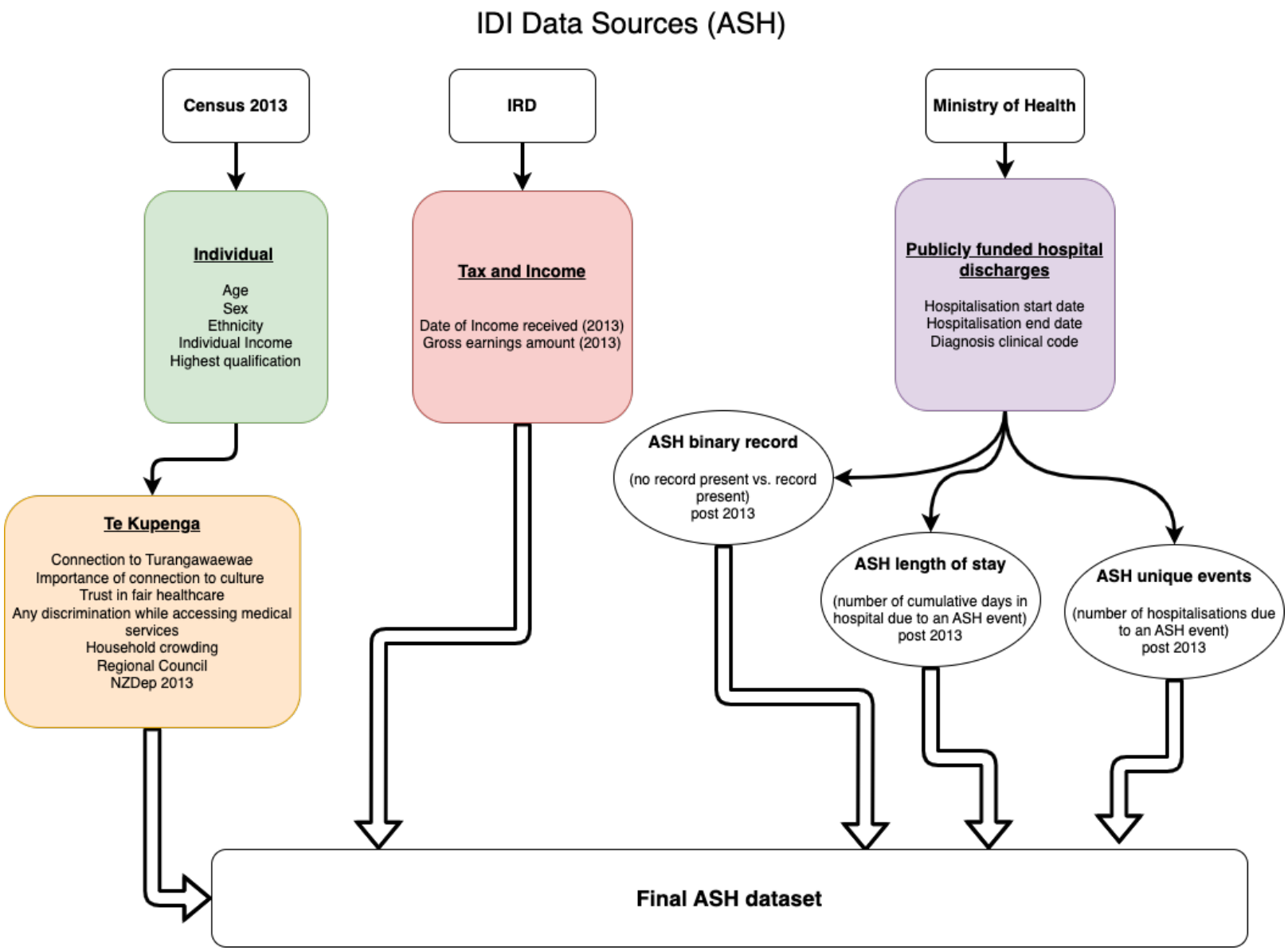
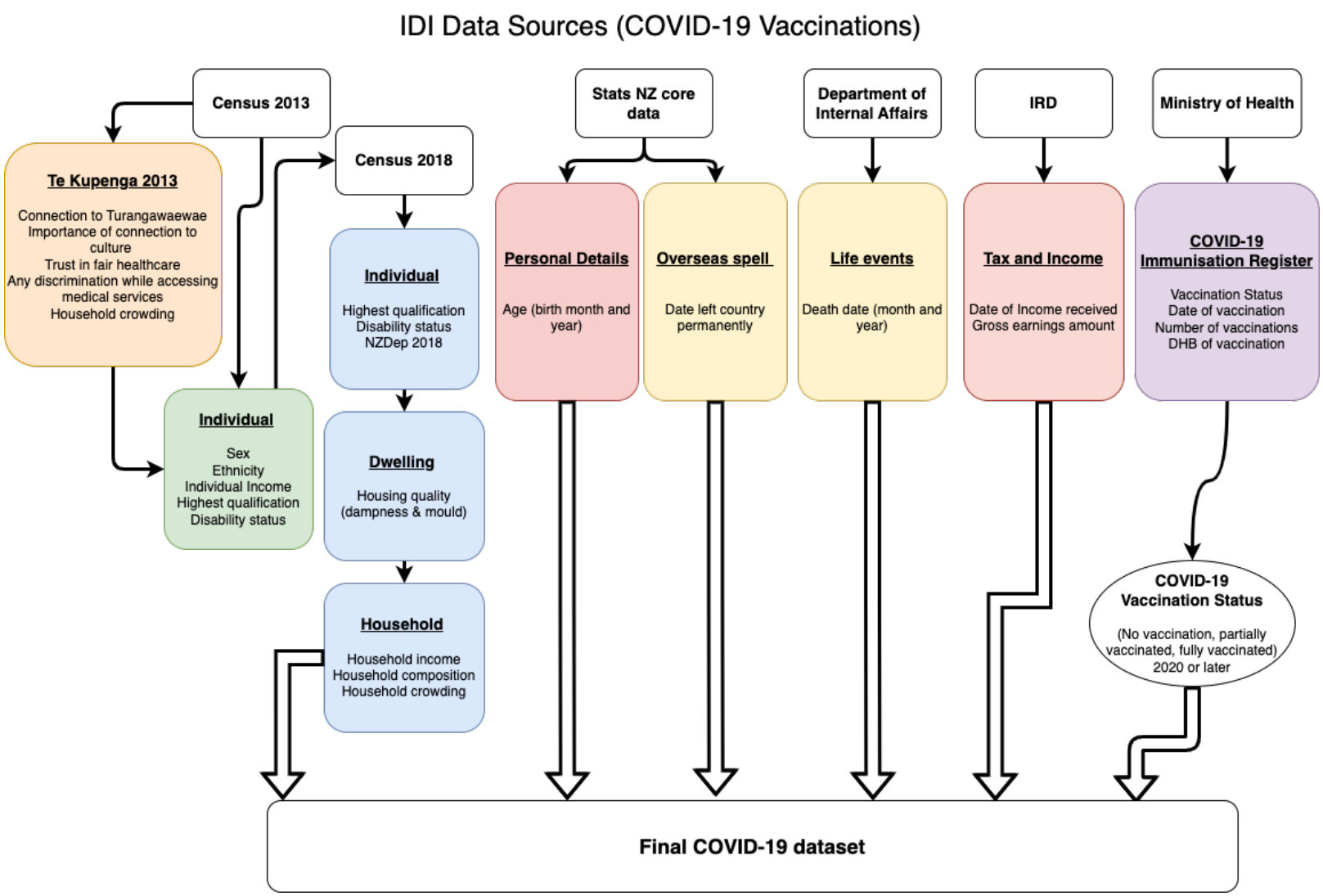


Figure 2.5: The IDI sources relevant to COVID-19 Vaccination outcome



2.2.10 'Censoring' Variables

A variation of censoring was used to include or exclude individuals who left the country permanently, or died between the Te Kupenga 2013 sample creation and the COVID-19 vaccination outcome period (2020 onwards). If someone left the country or died within this seven year period, their observation was flagged in the dataset.

Left country permanently

Overseas spell is included to gather information about whether an individual left the country permanently. Those who left the country permanently between 2014 and 2020 are recorded in this variable (1), while all others have missing values.

Death recorded

DIA records are utilised to gather information about date of death. For anyone who died between 2014 and 2020, their observation was recorded as 1, while all others have missing values.

2.2.11 Weighting Methods

Te Kupenga is a post-censal survey and a nationally representative sample of Māori in the New Zealand population at the time of the Census. This is possible as Te Kupenga sampled private dwellings throughout Aotearoa and conducted through a multi-stage sampling design(Stats NZ, 2014). Jackknife replicate weights are used to protect information regarding Primary Sampling Units (PSU), which may identify the sampling design, including detailed geography which could compromise the non-identifiability of participants. Using the replicate weights supplied with this microdata, approaches to survey weighting are examined in this project to consider possible methods when applying a longitudinal lens.

For the two outcome streams - ASH and vaccinations, different weighting approaches are discussed as each outcome is recorded at different time periods. Firstly, the whole 2013 Te Kupenga sample is kept throughout this analysis when viewing the ASH outcome. This decision was made as the ASH outcome takes place directly after the predictor variables are recorded and Te Kupenga sample was created (post 2013). This is in contrast to vaccinations, which uses an outcome variable recorded in 2020 onward. By implementing the use of censoring variables to view those who left the country permanently and/or died between 2014 and 2020, individuals who fall into this category are flagged for later investigation.

Exploring approaches to analytic output will involve different regression analyses, and acknowledging the potential weighting issues due to missingness. Demographic factors (age, sex and NZDep13) will be investigated for those who died and/or left the country permanently between 2013 and 2020. This is important to consider as these individuals will be missing from the potentially vaccinated population (denominator) and have impacts on the jackknife weights used in Te Kupenga.

2.3 Steps for linking microdata in the IDI

The following steps were implemented to join datasets within the IDI, and create the final datasets for each outcome. This included creating the relevant output used for the results section in this project. Code reflecting these steps can be found in appendix B. This code intends to create reproducible steps for longitudinal analysis in the IDI, using Te Kupenga as the foundational cohort. Additionally, this code may be used as an exemplar for future research linking other social surveys in the IDI.

2.3.1 Initial Joining - National and Te Kupenga 2013 level (ASH)

Coding steps are outlined in appendix B.1

Using Stata:

1. Using snz_uid's for those with ethnicity and/or Māori descent = **MID13** dataset

- Includes snz_uid (unique identifier), if an individual is present in Census 2013, Te Kupenga 2013, is of Māori descent and identifies with Māori ethnicity.

Using SAS:

1. Using MID13 dataset and left joining Census13 dataset (highest education level, total income, age and sex), matching by snz_uid = **ID13** dataset
2. Using ID13 dataset and left joining ASH hospitalisations after 2013= **hosp** dataset

Using SQL for extracting IRD records (for income cross checking):

1. Using snz_uid from census 2013 individual and using left join to attach 2013 gross income records for these individuals = **ird** dataset

- Joining done in SQL due to issues with using SAS
- Restricting income period to April 2012 to 31st March 2013 to match Census 2013 time period
- Each row is for an individual income period, e.g. monthly entries of income = 12 rows

Using Stata for data cleaning of base dataset **ID13**:

1. Cleaning and assigning labels to variables
2. Collapsing age, education & income variables to respective categories

Using Stata for data cleaning of **hosp** dataset:

1. Creating ASH variable to match ASH ICD-10 diagnostic codes (see Table A.1
2. Cleaning and assigning labels to variables
3. Creating ASH binary (*ashbinary*), ASH Unique events (*ashfreq*) & ASH Length of Stay (*ashlos*) variables

Using Stata for joining and cleaning **ird** dataset: Merging 1:1 on snz_uid with **ID13** dataset

1. Collapsing each row of gross income amount (by a sum) for each snz_uid
2. Collapsing income into respective categories to match income variable from Census 2013 and assigning labels to variables

Using Stata for merging base **ID13** with **hosp** and **ird** datasets:

1. Merging **hosp** dataset to **ID13** = **ASH_C13TK13_merge** dataset
2. Merging IRD dataset to **ASH_C13TK13_merge** dataset = **ASH_C13TK13_IRD_merge** dataset

2.3.2 National level output

Using Stata for output tables using **ASH_C13TK13_IRD_merge**

1. National level (Census 2013: Māori compared to Non-Māōri)
   (a) ASH Binary (see table 3.1)
   (b) ASH Binary by sex (see table 3.2)
   (c) ASH Binary by 10 year age-bands (see table 3.3)
   (d) ASH Binary by Highest Qualification (see table 3.4)

(e) ASH Binary by Individual income (Census 2013) (see table 3.5)

2.3.3 National level - Māori only

Using Stata to reduce ID13 sample:

1. Creating whole Māori sample dataset from Census 2013 = **c13Maori-ASH_C13TK13_IRD_merge** dataset
2. Creating Te Kupenga 2013 only sample dataset from Census 2013 = **tk13-ASH_C13TK13_IRD_merge** dataset

Using Stata to produce output using **c13Maori-ASH_C13TK13_IRD_merge** for Census 2013: Māōri only

1. ASH Binary
   (a) Logistic Regression Analysis: ASH Binary - age, sex, highest qualification & IRD Individual income - Odds Ratio (see table 3.6)
2. ASH Length of Stay & count of Unique ASH Events - Tables
   (a) ASH Length of Stay by Sex (Overdispersion tables) (see table 3.7)
   (b) ASH Length of Stay by Age (Overdispersion tables) (see table 3.8)
   (c) count of Unique ASH Events by Sex (Overdispersion tables) (see table 3.9)
   (d) count of Unique ASH Events by Age (Overdispersion tables) (see table 3.10)
3. ASH Length of Stay & count of Unique ASH Events - Negative Binomial Regression Analysis
   (a) Negative Binomial Regression Analysis: ASH Length of Stay - age, sex, highest qualification & IRD Individual Income - Incidence Rate Ratio (see table 3.11)
   (b) Negative Binomial Regression Analysis: count of Unique ASH Events - age, sex, highest qualification & IRD Individual Income - Incidence Rate Ratio (see table 3.12)

2.4 Te Kupenga 2013 sample only

From this point forward, the Te Kupenga 2013 sample is used as the cohort of interest. The following analysis demonstrates the use of tikanga-informed measures included in the survey alongside other IDI datasets to explore the impacts on ASH Events and COVID-19 vaccinations.

2.4.1 ASH outcome

Using SAS

1. to extract Te Kupenga 2013 from IDI adhoc tables, but joining to security concordance table first = **TK13_securitycon** dataset

Using Stata to rejoin Te Kupenga sample and Jackknife weights for Te Kupenga 2013 dataset

1. Rejoining Jackknife weights to Te Kupenga sample = **tk13-ASH_FINAL_merge**

Using Stata to produce output tables and regression analysis using **tk13-ASH_FINAL_merge**

1. Te Kupenga 2013 sample only
   (a) Bivariate tables (Weighted counts with 95% CI's)
      i. ASH Binary by sex

ii. ASH Binary by age

iii. ASH Binary by disability

The following regression analysis will be included in the results section:

1. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table 3.19)
2. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Connection to Tūrangawaewae, Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Error with Jackknife standard errors (see table 3.20)
3. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Odds Ratio (see table 3.21)
4. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), Household Crowding (2013) - Odds Ratio (see table 3.22)
5. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), NZDep Quintiles (2013) - Odds Ratio (see table 3.23)
6. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio (see table 3.24)

The following regression analysis were also completed and are provided in the appendix A

1. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), sex*age, Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table A.12)
2. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Highest Qualification (2013)*IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table A.13)
3. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: NZDep Quintiles (2013) - Odds Ratio (see table A.14)
4. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Household Crowding (2013), NZDep Quintiles (2013) - Odds Ratio (see table A.15)
5. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio (see table A.16)

2.4.2 Initial Joining - Te Kupenga 2013 level (COVID-19 Vaccinations)

Using Stata:

1. Using snz_uid's for those with ethnicity and/or Māori descent = **MID13** dataset

- Includes snz_uid (unique identifier), if an individual is present in Census 2013, Te Kupenga 2013, is of Māori descent and identifies with Māori ethnicity.

Includes snz_uid (unique identifier), if an individual is present in Census 2013, Te Kupenga 2013, is of Māori descent and identifies with Māori ethnicity.

Using SAS:

1. Using Census13 (individual variables, household ID, dwelling ID) and left joining **MID13** dataset, matching by snz_uid = **ID13ind** dataset
2. Using Census13 household dataset and left joining **ID13ind** dataset, matching by snz_cen_hhld_uid (housing ID) = **ID13hhld** dataset
3. Using **ID13hhld** and left joining Census18 individual and dwelling datasets, matching by snz_cen_dwell_uid and snz_uid = **ID13_Cen18dwl** dataset
4. Using **ID13_Cen18dwl** and left joining Census18 household dataset, matching by snz_cen_hhld_uid = **ID13_Cen18** dataset

Date of death

1. Using **ID13ind** dataset and left joining date of death from 2014 onwards, matching on snz_uid = **ID13_Cen18_dth** dataset

Overseas spell

1. Using **ID13ind** dataset and left joining date of leaving country permanently, matching on snz_uid = **ID13_Cen18_ovsp** dataset

Date of Birth

1. Using **ID13ind** dataset and left joining full date of birth, matching on snz_uid = **ID13_Cen18_dob** dataset
2. Using **ID13ind** dataset and left joining year of birth, month of birth and date of birth proxy, matching on snz_uid = **ID13_Cen18_pddob** dataset

Using SQL for extracting IRD records (for income cross checking):

1. Using snz uid from census 2013 individual and using left join to attach 2013 gross income records for these individuals = **ird** dataset

- Joining done in SQL due to issues with using SAS
- Restricting income period to April 2017 to 31st March 2018 to match Census 2018 time period
- Each row is for an individual income period, e.g. monthly entries of income = 12 rows

Using SQL for extracting **COVID vaccination datasets**:

1. Extracting Event dataset (Vaccination records & DHB) = **ciract** dataset
2. Extracting Status dataset (Vaccination status and date occurred) = **cirstat** dataset

Using Stata for joining/merging/cleaning datasets:

Using Stata for joining datasets for deaths, overseas spell, DOB, COVID-19 Vaccinations to Census dataset:

1. Merging 1:1 on snz_uid with **ID13_Cen18** dataset

Using Stata for data cleaning of dataset **ID13_Cen18** = **ID13_Cen18_CIR** dataset

1. Cleaning and assigning labels to variables (2013 and 2018)
2. Collapsing disability, qualification, income variables etc to respective categories
3. Creating vaccination status categories (0 = no vaccination, 1 = partially vaccinated, 2 = fully vaccinated)
4. Creating age variable to align with age at beginning of outcome period (2020).

Using Stata for creating variable for left country permanently and/or died between 2013 and 2020.

1. Creating variable lc_before = 1 if left country before start of vaccination period
2. Creating variable dth_before = 1 if died before start of vaccination period

Using SAS:

1. to extract Te Kupenga 2013 from IDI adhoc tables, but joining to security concordance table first = **TK13_securitycon** dataset

Using Stata to rejoin Te Kupenga sample and Jackknife weights for Te Kupenga 2013 dataset

1. Rejoining Jackknife weights to Te Kupenga sample = **CIR_TK13merge** dataset

Using Stata for Te Kupenga variable cleaning of dataset **CIR_TK13merge** dataset = **ID13_Cen18_CIR_TK13n** dataset

1. Cleaning and assigning labels to variables (2013 and 2018)

Using Stata to produce output tables and regressions using **ID13_Cen18_CIR_TK13merge** dataset:

1. Te Kupenga 2013 sample only
   (a) Transition tables of unweighted counts (2013 to 2018)
   (a) Bivariate tables (Weighted counts with 95% CI's)
      i. ASH Binary by sex
      ii. ASH Binary by age
      iii. ASH Binary by disability

The following regression analysis will be included in the results section:

1. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table 3.34)
2. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Disability (2018) - Coefficients (see table 3.36)
3. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Coefficients (see table 3.37)

The following regression analysis will be included in the appendix:

1. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Sex*Age, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table A.22)
2. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability

(2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table A.23)

3. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018) - Coefficients (see table A.24)

4. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018), NZDep Quintiles (2018) - Coefficients (see table A.25)

5. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), NZDep Quintiles (2018) - Coefficients (see table A.26)

6. Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018) - Relative Risk Ratio (see table A.27)

7. Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Relative Risk Ratio (see table A.28)

Chapter 3

Results

The following chapter discusses results found when implementing the methodology created during this project. The structure will follow the process outlined in the methods section (Figure 2.1), by viewing results across each sample level. Any issues found throughout the methodological development process will be investigated in the discussion section.

Firstly, the national level (Māori compared to Non-Māori), national level with Māori only and Te Kupenga 2013 sample. The first national level will include the ASH binary outcome only by viewing bi-variate tables. The national level with Māori only will assess the outcomes: ASH binary (using bi-variate tables and logistic regression), unique events and length of stay (using overdispersion tables and negative binomial regression).

Finally, the Te Kupenga 2013 sample is used for the ASH binary and COVID-19 vaccination outcomes. At this stage, ASH binary will be viewed with bi-variate tables of weighted counts, proportions and logistic regression. COVID-19 vaccinations will be inspected with bi-variate tables of weighted and unweighted counts and proportions and ordinal logistic regression.

3.1 National level - Māori compared to Non-Māori (Census 2013)

3.1.1 Bi-variate Tables (ASH Binary)

Each table below display the counts of ASH events vs. No ASH events (from 2014 onward) for Māori and Non-Māori across Census 2013 variables such as sex, age, highest qualification and individual income. These results display any areas that impact the distribution of ASH on a national scale, when stratified by Māori vs. Non-Māori.

The following tables will be displayed in this section:

1. National level (Census 2013: Māori compared to Non-Māōri)
   (a) ASH Binary (see table 3.1)
   (b) ASH Binary by sex (see table 3.2)
   (c) ASH Binary by 10 year age-bands (see table 3.3)
   (d) ASH Binary by Highest Qualification (see table 3.4)
   (e) ASH Binary by Individual income (Census 2013) (see table 3.5)

Table 3.1: ASH Binary by Māori vs. Non-Māori

ASH Binary Indicator	**Māori (by Descent/Ethnicity)**	**Non-Māori**	**Total**
ASH Event	163,215	736,371	899,589
	(0.23)	(0.20)	(0.21)
No ASH Event	544,035	2,909,574	3,453,609
	(0.77)	(0.80)	(0.79)
Total	707,250	3,645,945	4,353,198
	(1.00)	(1.00)	(1.00)

Note: Bracketed values are the proportion of outcome for given category

On a national scale, there are a lower proportion of ASH events in comparison to no ASH events. However, it can be seen that Māori had a higher proportion of ASH events, given the total Māori population within the 2013 Census (0.23 vs. 0.20 for Non-Māori). Therefore, this provides a motivation for viewing ASH outcomes with greater detail, applying stratification by Māori and Non-Māori.

Table 3.2: ASH by sex for Māori vs. Non-Māori

ASH Binary Indicator	**Sex**	**Māori**	**Non-Māori**	**Total**
ASH Event				
	Male	70,629	345,837	416,466
		(0.43)	(0.47)	(0.46)
	Female	92,586	390,534	483,120
		(0.57)	(0.53)	(0.54)
	Total	163,215	736,371	899,586
		(1.00)	(1.00)	(1.00)
No ASH Event				
	Male	270,339	1,432,662	1,703,001
		(0.50)	(0.49)	(0.49)
	Female	273,696	1,476,915	1,750,611
		(0.50)	(0.51)	(0.51)
	Total	544,035	2,909,574	3,453,609
		(1.00)	(1.00)	(1.00)
Total		**707,250**	**3,645,945**	**4,353,195**

Note: Bracketed values are the proportion of outcome for given category

ASH events for Māori vs. Non-Māori by sex display that for those with an ASH event, the proportion of Māori females is higher than for Non-Māori. In contrast, for Non-Māori males, there is a lower proportion of ASH events in comparison to Māori males.

Table 3.3: ASH by Age for Māori vs. Non-Māori

ASH Binary Indicator	**10 Year Age-bands (Census 2013)**	**Māori**	**Non-Māori**	**Total**
ASH Event				
	15-24	21,777	50,445	125,880
		(0.18)	(0.08)	(0.27)
	25-34	15,618	50,715	84,318
		(0.13)	(0.08)	(0.18)
	35-44	18,378	65,910	87,321
		(0.15)	(0.10)	(0.18)
	45-54	23,103	94,872	80,907
		(0.19)	(0.14)	(0.17)
	55-64	21,384	118,146	53,763
		(0.17)	(0.18)	(0.11)
	65-74	14,682	134,187	27,588
		(0.12)	(0.20)	(0.06)
	75 and above	7,992	146,340	12,861
		(0.07)	(0.22)	(0.03)
	Total	122,937	660,612	472,641
		(1.00)	(1.00)	(1.00)
No ASH Event				
	15-24	104,103	431,928	482,373
		(0.30)	(0.18)	(0.16)
	25-34	68,697	401,589	452,304
		(0.20)	(0.17)	(0.15)
	35-44	68,943	430,563	496,473
		(0.20)	(0.18)	(0.16)
	45-54	57,804	436,857	531,729
		(0.17)	(0.19)	(0.18)
	55-64	32,379	340,863	459,009
		(0.09)	(0.15)	(0.15)
	65-74	12,909	201,609	335,793
		(0.04)	(0.09)	(0.11)
	75 and above	4,866	106,050	252,390
		(0.01)	(0.05)	(0.08)
	Total	349,704	2,349,459	3,010,071
		(1.00)	(1.00)	(1.00)
Total		**472,641**	**3,010,071**	**3,482,712**

Note: Bracketed values are the proportion of outcome for given category

Table 3.4: ASH by Highest Qualification (Census 2013) for Māori vs. Non-Māori

Population	**Highest Qualification (Census 2013)**	**Māori**	**Non-Māori**	**Total**
ASH Event				
	No Qualification	44,640	176,802	137,088
		(0.41)	(0.30)	(0.32)
	School-level	33,495	199,260	159,426
		(0.31)	(0.33)	(0.27)
	Post-school	21,687	137,382	88,776
		(0.20)	(0.23)	(0.15)
	Degree Level	8,643	84,573	45,144
		(0.08)	(0.14)	(0.08)
	Total	108,462	598,020	430,434
		(1.00)	(1.00)	(0.72)
No ASH Event				
	No Qualification	92,448	314,487	491,292
		(0.29)	(0.16)	(0.19)
	School-level	125,931	739,737	938,997
		(0.39)	(0.38)	(0.37)
	Post-school	67,086	446,367	583,749
		(0.21)	(0.23)	(0.23)
	Degree Level	36,501	471,591	556,161
		(0.11)	(0.24)	(0.22)
	Total	321,969	1,972,182	2,570,202
		(1.00)	(1.00)	(1.00)
Total		**430,431**	**2,570,202**	**3,000,636**

Note: Bracketed values are the proportion of outcome for given category

Table 3.4 displays that for those who have had an ASH event, there are a greater proportion for those with no qualification, particularly for Māori.

Table 3.5: ASH by individual income (Census 2013) for Māori vs. Non-Māori

Population	Individual Income (Census 2013)	Māori	Non-Māori	Total
ASH Event				
	Loss/No Income	9,897	34,905	47,469
		(0.09)	(0.06)	(0.11)
	$1 - $15,000	30,780	131,448	108,105
		(0.28)	(0.21)	(0.25)
	$15,001 - $30,000	32,013	207,360	102,066
		(0.29)	(0.34)	(0.23)
	$30,001 - $50,000	21,294	119,919	93,915
		(0.19)	(0.19)	(0.22)
	$50,001 - $100,000	14,595	99,600	71,493
		(0.13)	(0.16)	(0.16)
	$100,001 and above	2,259	25,242	12,924
		(0.02)	(0.04)	(0.03)
	Total	110,838	618,480	435,969
		(1.00)	(1.00)	(1.00)
No ASH Event				
	Loss/No Income	37,572	177,177	212,085
		(0.12)	(0.09)	(0.08)
	$1 - $15,000	77,325	379,917	511,368
		(0.24)	(0.19)	(0.20)
	$15,001 - $30,000	70,050	395,013	602,373
		(0.22)	(0.20)	(0.23)
	$30,001 - $50,000	72,621	438,441	558,363
		(0.22)	(0.22)	(0.21)
	$50,001 - $100,000	56,898	462,486	562,086
		(0.18)	(0.23)	(0.21)
	$100,001 and above	10,662	143,004	168,249
		(0.03)	(0.07)	(0.06)
	Total	325,131	1,996,041	2,614,518
		(1.00)	(1.00)	(1.00)
Total		**435,969**	**2,614,521**	**3,050,487**

Note: Bracketed values are the proportion of outcome for given category

3.2 National level - Māori only

The overview of ASH events on a national level highlight the impacts of ASH on Māori. This is displayed across sex, age, highest qualification and individual income. Therefore, it is important to continue this analysis focusing on the Māori population only to emphasise which factors greatest impact ASH outcomes. Creating an analysis for Māori only provides a population level view, before focusing on longitudinal methods for the Te Kupenga sample.

The following output will be included in the upcoming sections:

1. ASH Binary
   (a) Logistic Regression Analysis: ASH Binary - age, sex, highest qualification & IRD Individual income - Odds Ratio (see table 3.6)
2. ASH Length of Stay & count of Unique ASH Events - Tables

(a) ASH Length of Stay by Sex (Overdispersion tables) (see table 3.7)

(b) ASH Length of Stay by Age (Overdispersion tables) (see table 3.8)

(c) count of Unique ASH Events by Sex (Overdispersion tables) (see table 3.9)

(d) count of Unique ASH Events by Age (Overdispersion tables) (see table 3.10)

3. ASH Length of Stay & count of Unique ASH Events - Negative Binomial Regression Analysis

   (a) Negative Binomial Regression Analysis: ASH Length of Stay - age, sex, highest qualification & IRD Individual Income - Incidence Rate Ratio (see table 3.11)

   (b) Negative Binomial Regression Analysis: count of Unique ASH Events - age, sex, highest qualification & IRD Individual Income - Incidence Rate Ratio (see table 3.12)

The following output will also be included in the appendix A

1. ASH Length of Stay & count of Unique ASH Events - Tables

   (a) ASH Length of Stay by Highest Qualification (Overdispersion tables)

   (b) ASH Length of Stay by Individual Income (Census 2013) (Overdispersion tables)

   (c) count of Unique ASH Events by Highest Qualification (Overdispersion tables)

   (d) count of Unique ASH Events by Individual Income (Census 2013) (Overdispersion tables)

3.2.1 Regression models (ASH Binary)

The following regression model is completed using ASH binary indicator as the outcome variable. This output includes the parameter `\_cons`. `\_cons` is the expected value of the odds for an ASH event, when all other explanatory variables are set to the reference population (Rothman, 2021). Within this analysis, the baseline outcome is no ASH event. For this level of analysis, the sample is including all Māori from the Census 2013 and the explanatory variables used include sex, age, highest qualification and IRD individual income (all from 2013). IRD individual income is used from this point onward (including for the Te Kupenga sample), as there are large discrepancies found between self-reported income and income records sourced from administrative data - such as IRD (see Chapter 4.2).

Table 3.6: Māori Census 2013: Logistic Regression for ASH binary using Highest Qualification, IRD individual income, Sex and Age - Odds Ratio

No ASH event (vs. ASH event)

	Odds Ratio	95% CI Lower	95% CI Upper	Std. Error	t	p
sex						
Male	1	(base)				
Female	1.238	1.218	1.259	0.010	25.630	0.000***
age						
15-24	1	(base)				
25-34	1.227	1.194	1.261	0.017	14.850	0.000***
35-44	1.528	1.487	1.569	0.021	31.000	0.000***
45-54	2.261	2.201	2.322	0.031	60.150	0.000***
55-64	3.639	3.536	3.744	0.053	88.170	0.000***
65-74	5.621	5.427	5.822	0.101	96.450	0.000***
75 and above	8.408	7.989	8.848	0.219	81.740	0.000***
Highest Qualification						
No Qualification	1	(base)				
School-level Qualification	0.708	0.695	0.722	0.007	-34.620	0.000***
Post-school Qualification	0.779	0.762	0.796	0.009	-22.600	0.000***
Degree Qualification	0.613	0.595	0.632	0.009	-32.080	0.000***
IRD Individual Income						
$1 - $15,000	1	(base)				
$15,001 - $30,000	0.897	0.878	0.917	0.010	-9.880	0.000***
$30,001 - $50,000	0.641	0.626	0.656	0.008	-37.160	0.000***
$50,001 - $100,000	0.586	0.571	0.601	0.008	-40.990	0.000***
$100,001 and above	0.478	0.451	0.506	0.014	-25.410	0.000***
_cons	0.279	0.272	0.285	0.003	-102.280	0.000***

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Log Likelihood = -190532.96, Number of obs = 354,216, LR $chi^2(14)$ = 30,767.32, Pseudo R^2 = 0.0747, Prob > chi^2 = 0.000

Viewing the results from the above table 3.6 , this displays that for the Māori population within the Census 2013, the odds of having an ASH event for females is 1.238 times that of males. Furthermore, for those aged 75 and above, the odds of having an ASH event are 8.408 times of those who are aged 15-24. However, for the highest qualification and IRD income variables, those who have higher completed qualifications and higher income display lower odds of an ASH event incomparison to the baseline (no qualification and $1 - $15,000). In this example, all results are significant ($p < 0.01$).

3.2.2 ASH Length of Stay & count of Unique ASH Events

As the outcomes for ASH Length of Stay and count of Unique ASH events are both discrete count variables, the following tables display the process of checking for possible overdispersion. Overdispersion can occur when the variance in the response variable is greater than the assumed variance within the model. This creates issues as the model assumptions are no longer met and overdispersion needs to be factored into the model. (Casella and Berger, 2001)

Poisson regression is commonly used when the response variable is a discrete count. The model assumes that the variance is equal to the mean (a one-parameter model). This is a restricting assumption when the data are overdispersed, i.e., when the variance of the data is larger than the model assumptions allow for. An alternative approach is to use negative binomial regression which is a two-parameter model for counts that allows the variance to be greater than or equal to the mean, thus providing more flexibility when we have overdispersion.

The following section provides tables to check for overdispersion across ASH Length of Stay & count of Unique ASH events for Sex and Age variables. However, additional tables are provided in the appendix A.1.1

Overdispersion for ASH Length of Stay

Table 3.7: Checking for over-dispersion in Māori Census 2013 only sample:
ASH Length of Stay by Sex

Sex	Mean	Variance	N
Male	2.4	368.5	340,968
Female	2.5	394.7	366,285
Total	2.4	382.1	707,250

Table 3.8: Checking for over-dispersion in Māori Census 2013 only sample:
ASH Length of Stay by Age-bands

Age-Bands	Mean	Variance	N
15-24	1	117.9	125,880
25-34	1.4	194.8	84,318
35-44	2.1	362.9	87,321
45-54	3.3	321.5	80,907
55-64	6.3	915.9	53,763
65-74	11.3	1915.4	27,588
75 and above	18.7	4768.1	12,861
Total	3.3	547.5	472,638

Across both tables 3.7 and 3.8, the mean for explanatory variables are shown to be lower than the variance of each corresponding row. This gives evidence of overdispersion, providing the rationale for a Negative Binomial Regression to be implemented in the analysis of the ASH Length of Stay outcome.

Overdispersion for count of Unique ASH Events

Table 3.9: Checking for over-dispersion in Māori Census 2013 only sample: ASH Unique Events by Sex

Sex	Mean	Variance	N
Male	0.5	4	340,968
Female	0.6	4.5	366,285
Total	0.6	4.3	707,250

Table 3.10: Checking for over-dispersion in Māori Census 2013 only sample: ASH Unique Events by Age-bands

Age-bands	Mean	Variance	N
15-24	0.3	1.6	125,880
25-34	0.4	3.1	84,318
35-44	0.5	4.1	87,321
45-54	0.8	6.4	80,907
55-64	1.4	11.4	53,763
65-74	2.1	14.9	27,588
75 and above	2.5	14.2	12,861
Total	0.7	5.7	472,638

Similarly to ASH length of stay outcome, there is evidence of overdispersion for the count of Unique ASH events across sex and age explanatory variables (see tables 3.9, 3.10). Furthermore, this formulates the motivation toward implementing Negative Binomial Regression for count of Unique ASH events.

3.2.3 Negative Binomial Regression models

Negative Binomial Regression will be implemented for ASH Length of Stay and count of Unique ASH Events due to the overdispersion found in tables 3.7, 3.8, 3.9, 3.10. This regression output will produce Incidence Rate Ratio (IRR) - a commonly used method for measuring outcomes in epidemiology. Incidence rate ratio is a measure used to compare two incidence rates of events occurring at the sample period of time (Rothman, 2021) When categorical explanatory variables are included, the IRR is the ratio of events in one category in comparison to the baseline category.

The following regression models include `\_cons`, `alpha` and `/lnalpha` parameters. `\_cons` is the negative binomial regression estimate which is included when all included variables are zero. `alpha` is the dispersion parameter estimate, displaying that the model is greater than zero (by significance) then overdispersion is present and a negative binomial model is best suited. However, if `alpha` is zero, then a poisson model is preferred. `/lnalpha` is the logged value of the dispersion parameter `alpha`

ASH Length of Stay

Table 3.11: Māori Census 2013: Negative Binomial Regression for ASH Length of Stay using Sex, Age, Highest Qualification, IRD individual income - Incidence Rate Ratio

	IRR	95% CI Lower	95% CI Upper	Std. Error	t	P> |t|
sex						
Male	1	(base)				
Female	1.022	0.999	1.045	0.012	1.850	0.064*
age						
15-24	1	(base)				
25-34	1.758	1.698	1.820	0.031	32.020	0.000***
35-44	2.839	2.741	2.941	0.051	58.050	0.000***
45-54	4.668	4.499	4.844	0.088	81.800	0.000***
55-64	8.235	7.898	8.585	0.175	99.050	0.000***
65-74	13.559	12.869	14.287	0.362	97.760	0.000***
75 and above	20.993	19.480	22.624	0.801	79.770	0.000***
Highest Qualification (2013)						
No Qualification	1	(base)				
School-level Qualification	0.701	0.682	0.721	0.010	-25.360	0.000***
Post-school Qualification	0.691	0.670	0.713	0.011	-23.390	0.000***
Degree Qualification	0.548	0.526	0.571	0.011	-28.890	0.000***
IRD Individual Income (2013)						
$1 - $15,000	1	(base)				
$15,001 - $30,000	0.637	0.617	0.657	0.010	-28.830	0.000***
$30,001 - $50,000	0.389	0.377	0.402	0.006	-57.040	0.000***
$50,001 - $100,000	0.296	0.286	0.307	0.005	-66.780	0.000***
$100,001 and above	0.238	0.221	0.256	0.009	-38.410	0.000***
_cons	1.727	1.670	1.785	0.029	32.310	0.000***
/lnalpha	2.324	2.316	2.332	0.004		
alpha	10.21784	10.13823	10.29808	0.0407788		

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Dispersion = mean, Log Likelihood = -465492.75, Number of obs = 354,216, Population size = 29,082.00, Prob> chi^2 = 0.000, Pseudo R^2 = 0.0303

The above output 3.11 displays that the estimate of dispersion is greater than 0 (`alpha` = 10.21784), therefore the negative binomial regression model is preferred over a Poisson regression. For females in comparison to males, when all other variables are held constant in the model, their expected rate for length of stay is expected to be 1.022 times greater. Furthermore, for those who are ages 75 and above, when all other variables are held constant in the model, are expected to have a rate 20.993 times greater for length of stay in comparison to those aged 15-24 (the baseline). For highest qualification and IRD individual income, when comparing to their baseline, all categories display a decrease in rate while holding all other variables constant in the model.

Count of ASH Unique Events

Table 3.12: Maori Census 2013: Negative Binomial Logistic Regression for count of ASH Unique Events using Sex, Age, Highest Qualification, IRD individual income - Incidence Rate Ratio

	IRR	95% CI Lower	95% CI Upper	Std. Error	t	P>|t|
sex						
Male	1	(base)				
Female	1.181	1.161	1.201	0.010	19.720	0.000***
age						
15-24	1	(base)				
25-34	1.433	1.395	1.472	0.020	26.370	0.000***
35-44	2.038	1.984	2.093	0.028	52.430	0.000***
45-54	3.156	3.072	3.242	0.043	83.740	0.000***
55-64	5.116	4.968	5.269	0.077	108.720	0.000***
65-74	6.971	6.727	7.225	0.127	106.500	0.000***
75 and above	7.789	7.413	8.184	0.196	81.400	0.000***
Highest Qualification (2013)						
No Qualification	1	(base)				
School-level Qualification	0.705	0.691	0.719	0.007	-34.830	0.000***
Post-school Qualification	0.769	0.753	0.786	0.009	-23.410	0.000***
Degree Qualification	0.588	0.571	0.606	0.009	-34.620	0.000***
IRD Individual Income (2013)						
$1 - $15,000	1	(base)				
$15,001 - $30,000	0.808	0.791	0.826	0.009	-19.170	0.000***
$30,001 - $50,000	0.535	0.523	0.548	0.006	-51.700	0.000***
$50,001 - $100,000	0.458	0.446	0.470	0.006	-59.270	0.000***
$100,001 and above	0.366	0.346	0.387	0.010	-35.110	0.000***
_cons	0.458	0.447	0.470	0.006	-61.390	0.000***
/lnalpha	1.338	1.328	1.349	0.005		
alpha	3.812645	3.773384	3.852315	0.0201354		

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Dispersion = mean, Log Likelihood = -356206.85, Number of obs = 354,216, Population size = 33,579.00, Prob > chi^2 =0.000, Pseudo R^2 = 0.045

For the above regression output 3.12, it can be seen that the females also have a higher rate for the number of unique ASH events in comparison to males (1.181 times greater) while all variables are held constant. The age variable follows a similar trend to the negative binomial regression for Length of Stay 3.11, where all age-bands over 24 years have a greater rate of unique ASH events when all other variables are held constant. Similarly, when all variables are held constant, highest qualification and IRD individual income produce lower rates of unique events for those categories higher than their baselines. Evidence for overdispersion is also provided here, with a parameter of `alpha`= 3.812645 (greater than 0). Additionally, all variables within this model are significant at the $p < 0.01$ level.

3.3 Te Kupenga 2013 sample

3.3.1 ASH Binary

ASH will first be explored for the Te Kupenga 2013 sample through tables of weighted counts, cell proportions and proportions - conditioned on ASH events (with 95% CI's). Corresponding plots for Sex and Age will also be provided as a visual demonstration of this output. Exemplars of this output are provided below using individual, household and geographic variables, with further tables included in Appendix A. The following output will only be representative of the Te Kupenga 2013 sample, which only includes Māori aged 15 and older. For the Te Kupenga 2013 sample, Te Kupenga specific variables and household level variables are incorporated to view an added level of complexity for the analysis.

The following tables of weighted counts, cell proportions and proportions - conditioned on ASH events (with 95% CI's) are provided in this section:

1. ASH binary by sex (see table 3.13)
2. ASH binary by Age (2013 Census) (see table 3.14)
3. ASH binary by Disability (2013 Census) (see table 3.15)
4. ASH binary by Any Medical Discrimination (TK 2013) (see table 3.16)
5. ASH binary by Household Crowding (2013 Census) (see table 3.17)
6. ASH binary by NZDep (2013 Census) (see table 3.18)

All weighted counts rounded to base 500, with raw counts < 500 suppressed.

Sex (Census 2013)

Table 3.13: ASH binary by sex (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	Sex	Weighted counts	Cell proportion	Conditional proportions
ASH Event				
	Male	48,000	0.092	0.390
		(43,000 , 53,000)	(0.083 , 0.102)	(0.36 , 0.42)
	Female	75,000	0.144	0.610
		(70,500 , 80,000)	(0.135 , 0.153)	(0.58 , 0.64)
	Total	123,000	0.236	1.000
		(113,500 , 133,000)	(0.218 , 0.255)	
No ASH Event				
	Male	201,000	0.387	0.507
		(196,500 , 206,000)	(0.377 , 0.396)	(0.498 , 0.516)
	Female	196,000	0.377	0.493
		(191,000 , 201,000)	(0.367 , 0.386)	(0.484 , 0.502)
	Total	397,000	0.764	1.000
		(387,500 , 407,000)	(0.744 , 0.782)	
Total		520,000	1.000	
		(501,000 , 540,000)		

Note: Bracketed values are the 95% CI for given category

The above table 3.13 displays that across the whole Te Kupenga 2013 sample, weighted counts demonstrate that there is a higher proportion of no ASH events in comparison to ASH events. However, females have a higher proportion of ASH events and males have a higher proportion of no ASH events across the weighted counts. However, when conditioning on ASH event given sex, females have higher proportions of ASH events in comparison to males. The reverse is shown when conditioning on no ASH event given sex.

Figure 3.1: ASH binary by sex (weighted counts) with 95% CI's

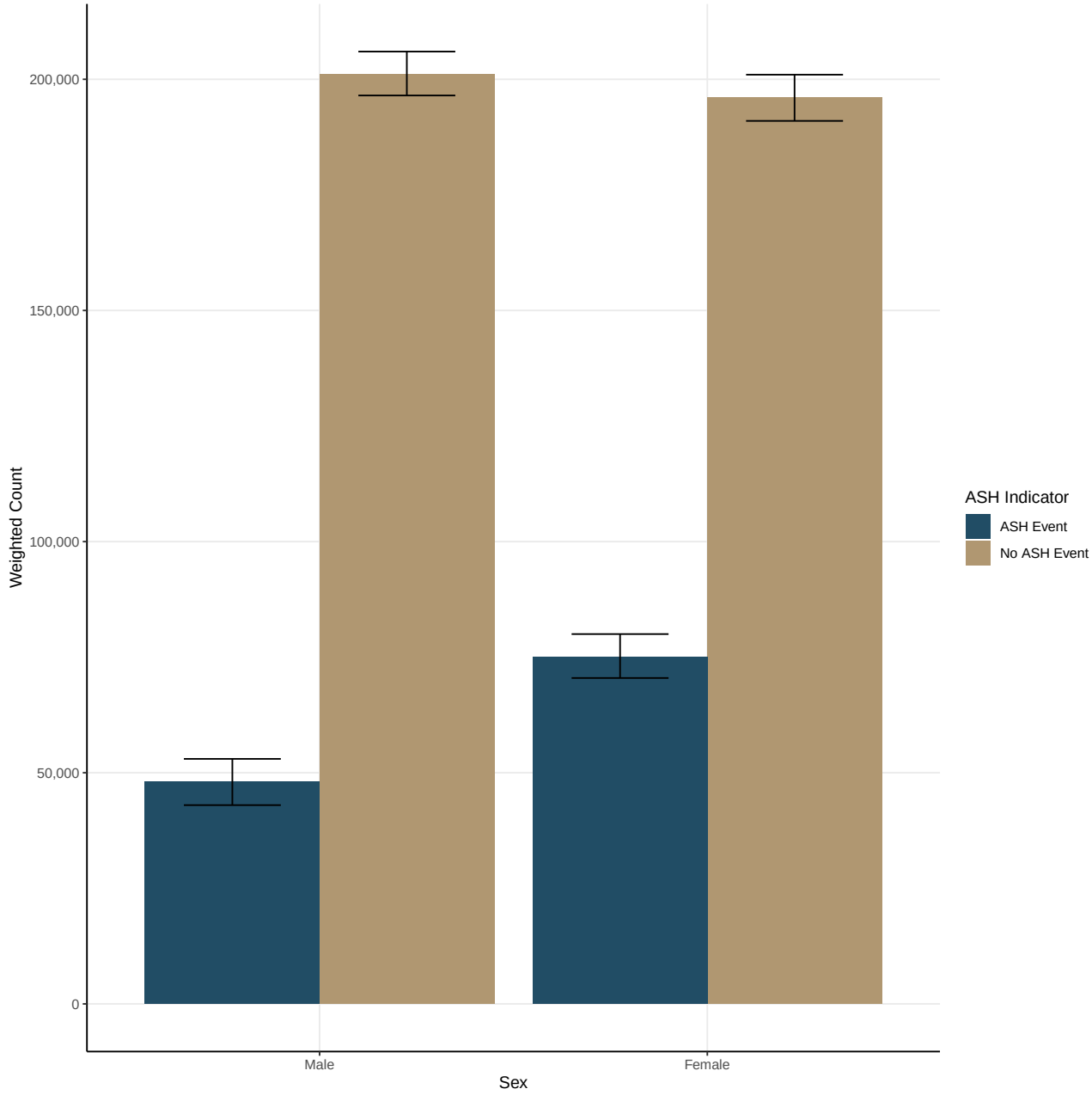
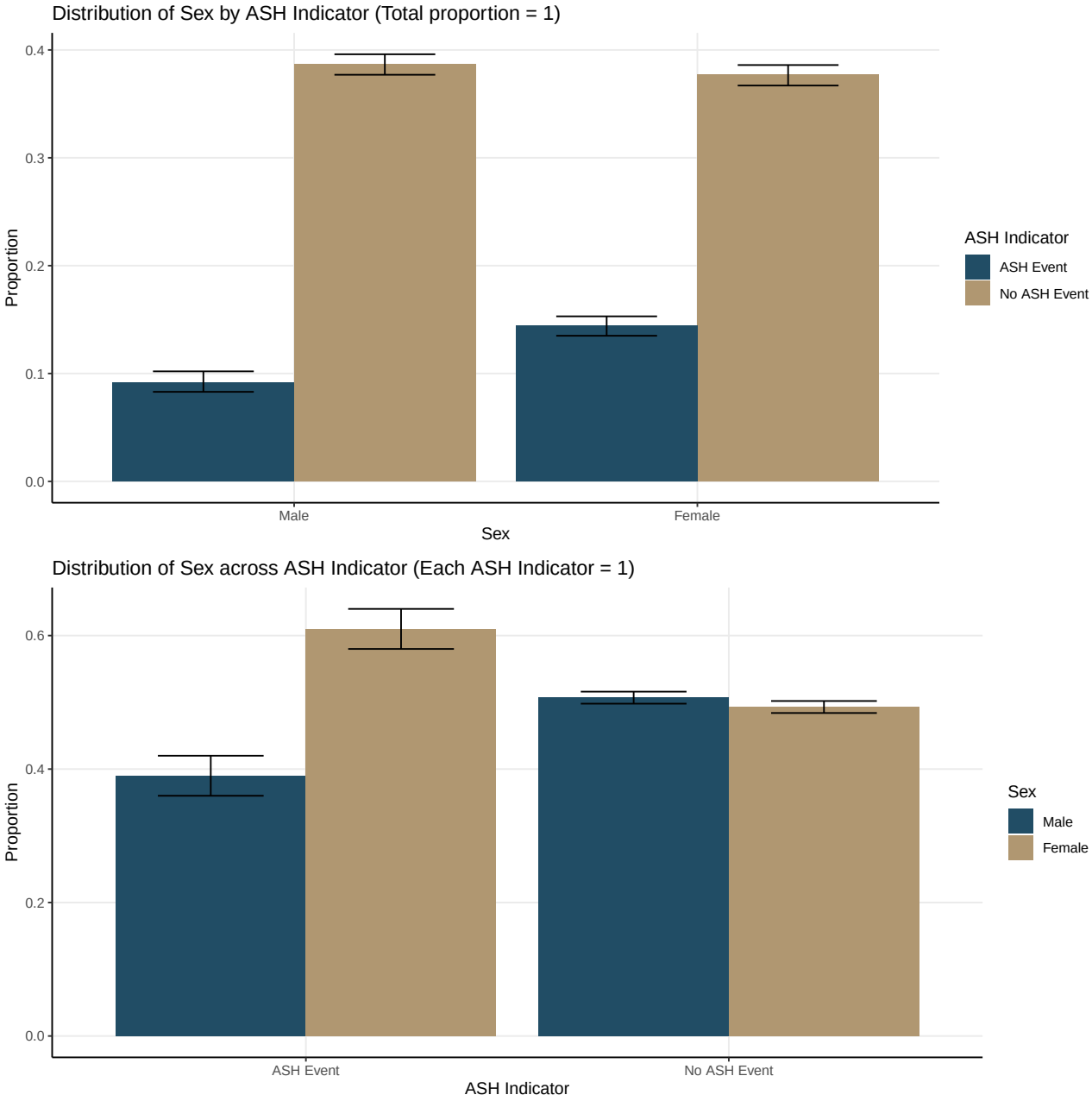


Figure 3.2: ASH binary by sex (proportions)



Age (Census 2013)

Table 3.14: ASH binary by Age (2013 Census) (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	Age-bands (2013 Census)	Weighted counts	Cell proportion	Conditional proportions
ASH Event				
	15-24	25,500	0.049	0.208
		(22,500 , 28,500)	(0.043 , 0.055)	(0.186 , 0.229)
	25-34	15,500	0.03	0.125
		(13,000 , 17,500)	(0.025 , 0.034)	(0.108 , 0.142)
	35-44	17,500	0.034	0.142
		(14,500 , 20,500)	(0.028 , 0.039)	(0.121 , 0.164)
	45-54	22,000	0.042	0.178
		(19,000 , 25,000)	(0.036 , 0.048)	(0.155 , 0.201)
	55-64	21,000	0.04	0.171
		(18,500 , 23,500)	(0.036 , 0.045)	(0.153 , 0.188)
	65-74	15,500	0.03	0.125
		(13,500 , 17,500)	(0.026 , 0.034)	(0.108 , 0.142)
	75 and above	6,500	0.012	0.051
		(4,500 , 8,000)	(0.009 , 0.015)	(0.039 , 0.064)
	Missing	S	S	S
		S	S	S
	Total	123,500	0.237	1.000
		(105,500 , 140,500)		
No ASH Event				
	15-24	122,000	0.235	0.308
		(117,000 , 127,000)	(0.225 , 0.244)	(0.296 , 0.319)
	25-34	78,000	0.15	0.196
		(72,000 , 83,500)	(0.139 , 0.16)	(0.183 , 0.21)
	35-44	78,000	0.15	0.197
		(73,000 , 83,000)	(0.14 , 0.16)	(0.185 , 0.209)
	45-54	65,000	0.125	0.163
		(59,500 , 70,000)	(0.115 , 0.134)	(0.15 , 0.176)
	55-64	34,500	0.067	0.087
		(32,000 , 37,500)	(0.061 , 0.072)	(0.08 , 0.094)
	65-74	15,000	0.029	0.038
		(12,500 , 17,500)	(0.024 , 0.034)	(0.032 , 0.044)
	75 and above	4,000	0.008	0.011
		(3,000 , 5,500)	(0.006 , 0.01)	(0.008 , 0.013)
	Missing	S	S	S
		S	S	S
	Total	396,500	0.764	1.000
		(369,000 , 424,000)	(0.71 , 0.814)	
Total		520,000	1.000	
		(474,500 , 564,500)		

Note: Bracketed values are the 95% CI for given category

The above table 3.14 displays higher weighted counts for no ASH event. However, when conditioning on the ASH outcome given age, ASH events are higher for those aged 15-24, in comparison to other

age groups within the sample. A similar result is found for those aged 45-64.

Figure 3.3: ASH binary by 10 year age-bands (weighted counts) with 95% CI's

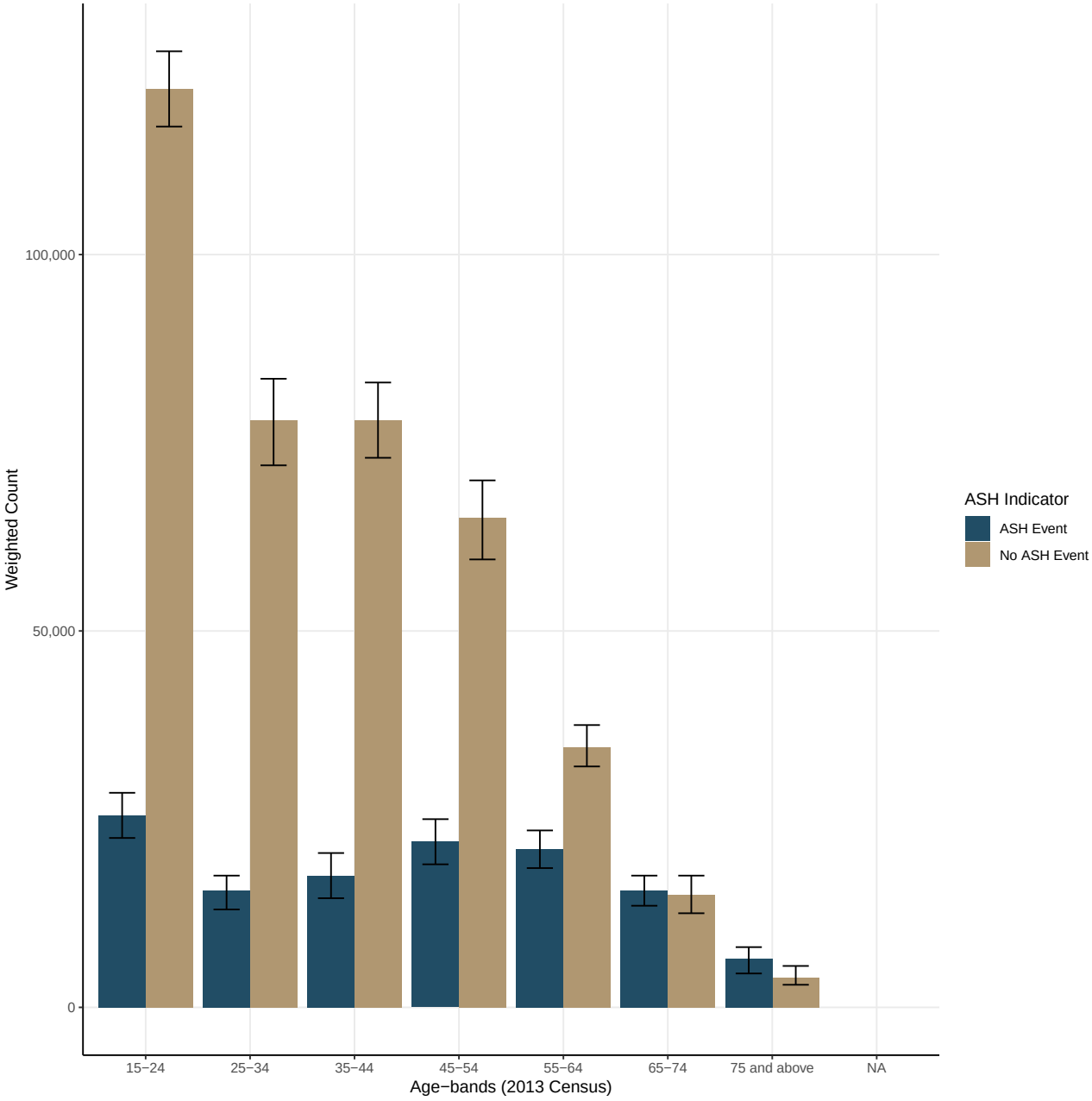
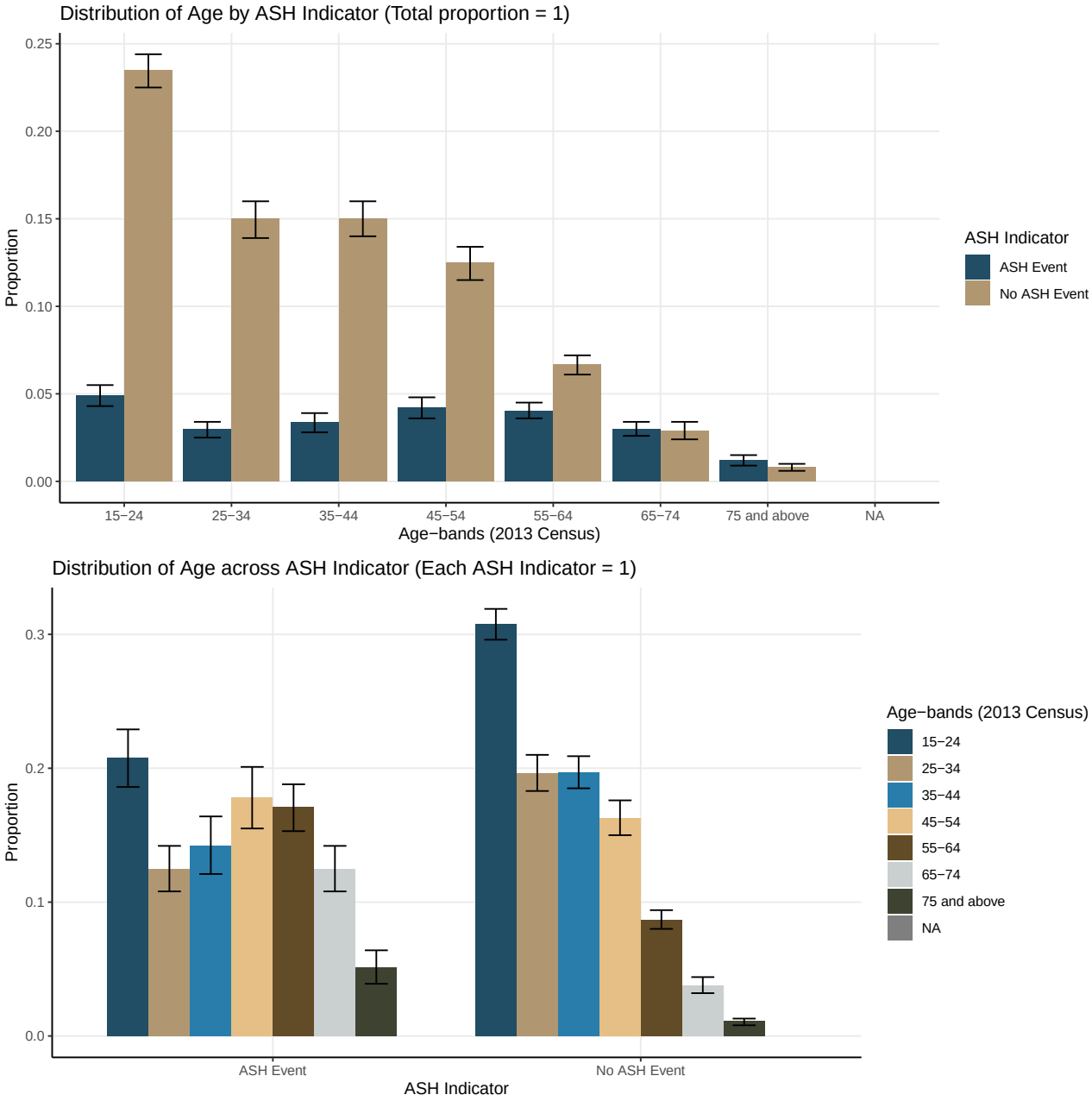


Figure 3.4: ASH binary by 10 year age-bands (proportions)



Disability (2013 Census)

Table 3.15: ASH binary by Disability (2013 Census) (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	Disability (2013)	Weighted counts	Cell proportion	Conditional proportions
ASH Event				
	Disability	43,500	0.084	0.354
		(40,000 , 47,500)	(0.077 , 0.091)	(0.33 , 0.379)
	No Disability	76,000	0.146	0.615
		(70,500 , 81,000)	(0.135 , 0.156)	(0.589 , 0.64)
	Missing	4,000	0.007	0.031
		(2,500 , 5,500)	(0.005 , 0.01)	(0.019 , 0.043)
	Total	123,500	0.237	1.000
		(113,000 , 134,000)	(0.217 , 0.257)	
No ASH Event				
	Disability	57,500	0.110	0.144
		(52,500 , 62,000)	(0.101 , 0.119)	(0.133 , 0.156)
	No Disability	331,000	0.636	0.833
		(323,500 , 338,000)	(0.622 , 0.649)	(0.821 , 0.845)
	Missing	9,000	0.017	0.022
		(7,000 , 11,000)	(0.013 , 0.021)	(0.017 , 0.027)
	Total	397,500	0.763	1.000
		(383,000 , 411,000)	(0.736 , 0.789)	
Total		521,000	1.000	
		(496,000 , 545,000)		

Note: Bracketed values are the 95% CI for given category

Table 3.15 displays that when conditioning on ASH outcome, there are a higher proportion of individuals without a disability, who have not had an ASH event. The same can be found for those who have had an ASH event. However, across the whole population, those who have a disability are closely distributed between having an ASH event and no ASH event.

Any Medical Discrimination (TK 2013)

Table 3.16: ASH binary by Any Medical Discrimination (TK 2013) (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	Any Medical Discrimination (TK 13)	Weighted count	Cell proportion	Conditional proportions
ASH Event				
	No	111,000	0.213	0.900
		(104,500 , 117,500)	(0.201 , 0.225)	(0.883 , 0.917)
	Yes	10,500	0.020	0.086
		(8,500 , 12,500)	(0.016 , 0.024)	(0.07 , 0.103)
	Missing	1,500	0.003	0.014
		(1,000 , 2,500)	(0.002 , 0.005)	(0.007 , 0.021)
	Total	123,000	0.236	1.000
		(114,000 , 132,500)	(0.219 , 0.254)	
No ASH Event				
	No	374,000	0.718	0.941
		(366,500 , 381,000)	(0.705 , 0.732)	(0.933 , 0.949)
	Yes	15,500	0.030	0.039
		(12,500 , 18,000)	(0.024 , 0.035)	(0.032 , 0.046)
	Missing	8,000	0.015	0.020
		(6,000 , 10,000)	(0.011 , 0.019)	(0.015 , 0.025)
	Total	397,500	0.763	1.000
		(385,000 , 409,000)	(0.74 , 0.786)	
Total		520,500	1.000	
		(499,000 , 541,500)		

Note: Bracketed values are the 95% CI for given category

The above table displays that within the ASH event group, there is a higher proportion of individuals who have not experienced any medical discrimination, however, a similar proportion is shown for those who have not had an ASH event.

Household Crowding (2013 Census)

Table 3.17: ASH binary by Household Crowding (2013 Census) (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	Household Crowding (2013)	Weighted count	Cell proportion	Conditional proportions
ASH Event				
	Crowded	17,500	0.034	0.142
		(14,500 , 20,500)	(0.028 , 0.039)	(0.12 , 0.164)
	Not crowded	104,500	0.201	0.847
		(98,000 , 110,500)	(0.189 , 0.212)	(0.824 , 0.869)
	Missing	1,500	0.003	0.012
		(500 , 2,000)	(0.001 , 0.004)	(0.006 , 0.017)
	Total	123,500	0.238	1.000
		(113,000 , 133,000)	(0.218 , 0.255)	
No ASH Event				
	Crowded	51,000	0.098	0.129
		(45,500 , 57,000)	(0.088 , 0.109)	(0.115 , 0.143)
	Not crowded	342,500	0.659	0.863
		(335,000 , 350,500)	(0.644 , 0.673)	(0.849 , 0.877)
	Missing	3,000	0.006	0.008
		(2,000 , 4,500)	(0.004 , 0.008)	(0.005 , 0.011)
	Total	396,500	0.763	1.000
		(382,500 , 412,000)	(0.736 , 0.79)	
Total		520,000	1.000	
		(495,500 , 545,000)		

Note: Bracketed values are the 95% CI for given category

Household crowding displays that there is a lower proportion of individuals in a crowded house (conditioning on ASH).

NZDep (2013 Census)

Table 3.18: ASH binary by NZDep (2013 Census) (weighted counts, total proportions and conditional proportions) with 95% CI's

ASH Indicator	NZDep (TK 13)	Weighted count	Cell proportion	Conditional proportions
ASH Event				
	Dep 1&2	10,000	0.020	0.083
		(8,000 , 12,500)	(0.015 , 0.024)	(0.065 , 0.101)
	Dep 3&4	17,000	0.033	0.138
		(13,500 , 20,500)	(0.026 , 0.039)	(0.114 , 0.161)
	Dep 5&6	20,000	0.039	0.163
		(17,000 , 23,000)	(0.033 , 0.044)	(0.142 , 0.184)
	Dep 7&8	28,500	0.055	0.233
		(25,000 , 32,000)	(0.049 , 0.062)	(0.209 , 0.256)
	Dep 9&10	47,000	0.090	0.381
		(44,000 , 50,000)	(0.084 , 0.096)	(0.357 , 0.406)
	Missing	S	S	S
		S	S	S
	Total	122,500	0.237	1.000
		(107,500 , 138,000)	(0.207 , 0.265)	
No ASH Event				
	Dep 1&2	49,500	0.095	0.125
		(45,000 , 54,500)	(0.086 , 0.104)	(0.113 , 0.136)
	Dep 3&4	60,500	0.116	0.152
		(55,000 , 65,500)	(0.106 , 0.126)	(0.139 , 0.165)
	Dep 5&6	73,500	0.141	0.185
		(67,000 , 79,500)	(0.129 , 0.153)	(0.17 , 0.2)
	Dep 7&8	96,000	0.184	0.241
		(90,000 , 101,500)	(0.173 , 0.195)	(0.227 , 0.255)
	Dep 9&10	118,000	0.227	0.297
		(112,500 , 123,500)	(0.217 , 0.237)	(0.285 , 0.31)
	Missing	S	S	S
		S	S	S
	Total	397,500	0.763	1.000
		(369,500 , 424,500)	(0.711 , 0.815)	
Total		520,000	1.000	
		(477,000 , 562,500)		

Note: Bracketed values are the 95% CI for given category

For individuals in a higher quintile of NZDep (Dep9&10), there are larger proportions of ASH events when compared to those in the lower quintiles of deprivation. However, this trend is also seen for those without an ASH event.

3.3.2 Regression Analysis (ASH Binary)

The following output is of logistic regression output using ASH binary as the outcome variable. No ASH event is the baseline for comparison, therefore all output is for the odds of having an ASH event in comparison to no ASH event.

The following regression models include `\_cons`. `\_cons` is the expected value of the odds for an ASH event, when all other explanatory variables are set to the reference population. Also to note: the

columns `RSE \%` and `Output rule` are required when outputting regression output from the IDI, using the Te Kupenga survey. `RSE\%` is the relative sampling error, calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage. The following rules are required for releasing any output using Te Kupenga in the IDI (Stats NZ, 2021);(Stats NZ, 2020b):

1. Suppress estimates with a relative sampling error (RSE) of 100 percent or greater.
2. Identify estimates with an RSE between 30 percent and less than 50 percent with one hash symbol (#).
3. Identify estimates with an RSE between 50 percent and less than 100 percent with two hash symbols (##).

The following regression analysis will be included in this results section:

1. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table 3.19)
2. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Connection to Tūrangawaewae, Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Error with Jackknife standard errors (see table 3.20)
3. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Odds Ratio (see table 3.21)
4. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), Household Crowding (2013) - Odds Ratio (see table 3.22)
5. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), NZDep Quintiles (2013) - Odds Ratio (see table 3.23)
6. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio (see table 3.24)

The following regression analysis were also completed and are provided in the appendix A

1. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), sex*age, Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table A.12)
2. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Highest Qualification (2013)*IRD Individual Income (2013), Disability (2013) - Odds Ratio (see table A.13)
3. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: NZDep Quintiles (2013) - Odds Ratio (see table A.14)
4. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Household Crowding (2013), NZDep Quintiles (2013) - Odds Ratio (see table A.15)

5. Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio (see table A.16)

Table 3.19: Te Kupenga 2013 Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.652	1.359	2.009	0.163	5.100	0.000	***	9.850	
Age (2013)									
15-24	1.000	(base)							
25-34	1.019	0.775	1.338	0.140	0.130	0.894		13.742	
35-44	1.243	0.926	1.670	0.185	1.460	0.146		14.870	
45-54	1.677	1.243	2.262	0.253	3.420	0.001	**	15.098	
55-64	2.936	2.151	4.007	0.460	6.870	0.000	***	15.678	
65-74	5.132	3.623	7.270	0.901	9.320	0.000	***	17.553	
75 and above	7.000	3.774	12.982	2.179	6.250	0.000	***	31.133	#
Highest Qualification (2013)									
No Qualification	1.450	1.077	1.952	0.217	2.480	0.015	**	14.975	
School-level Qualification	1.159	0.872	1.541	0.166	1.030	0.305		14.345	
Post-school Qualification	1.118	0.818	1.528	0.176	0.710	0.482		15.750	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.151	1.218	3.797	0.616	2.670	0.009	**	28.643	
$15,001 - $30,000	1.528	0.837	2.790	0.464	1.400	0.165		30.346	#
$30,001 - $50,000	1.712	0.930	3.149	0.526	1.750	0.083	*	30.731	#
$50,001 - $100,000	1.439	0.777	2.666	0.447	1.170	0.245		31.079	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.592	2.162	3.106	0.237	10.430	0.000	***	9.128	
_cons	0.061	0.032	0.116	0.020	-8.540	0.000	***	32.841	#

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 390,500 (rounded to base 500), Replications = 100, Design df = 99, $F(15, 85) = 30.97$, Prob > $F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% < 50% ; Output rule = ## if *RSE*% 50% < 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Upon viewing results from the logistic regression output in 3.19, it can be seen that the odds of females having an ASH event are 1.652 times higher than that of males. For age bands above 24 years old, all have higher odds of an ASH event (all greater than 1), in comparison to those who are 15-24 years old. However, this cannot be concluded for the 25-34 year olds, as the P value is greater than the 0.1 level of significance. For all those with highest qualification lower than a degree qualification, the odds of an ASH event are higher than those with a degree. IRD individual income shows that for all those with income less than $100,000, the odds of an ASH event are also greater than if individual income is above $100,000, however, this can only be concluded for those with income between $1 - $15,000 and $30,001 - $50,000, as these categories are significant at the 0.1 level. Disability also displays a large impact on ASH events, as those with a disability have 2.592 times greater odds of an ASH event in comparison without a disability. All of these findings occur when all other variables are held constant in the model.

This output shows similar results overall, when comparing to ASH Length of Stay and count of Unique ASH events (3.11, 3.12). All rows across this regression are displayed, without the need for suppression as all meet the requirements for output rules (RSE $< 30\%$).

Table 3.20: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Connection to Tūrangawaewae, Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Error with Jackknife standard errors

```
* all individual var
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.conTWW
                                                    i.cultureimp thealthfair i.meddiscr, base

Jackknife replications (100)
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx 50
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx 100
insufficient observations to compute jackknife standard errors
no results will be saved
r(2000);
```

The above table 3.20 is a demonstration of errors that occurred when including Te Kupenga 2013 variables, particularly Connection to Tūrangawaewae, into the analysis. This created errors when using the Jackknife survey weights, as there were insufficient observations available. Therefore, from this point the Connection to Tūrangawaewae variable is dropped from the regression analysis when considering Te Kupenga variables.

Table 3.21: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.606	1.316	1.959	0.161	4.730	0.000	***	10.014	
Age (2013)									
15-24	1.000	(base)							
25-34	0.954	0.724	1.257	0.133	-0.340	0.736		13.889	
35-44	1.149	0.850	1.553	0.175	0.910	0.364		15.195	
45-54	1.564	1.148	2.133	0.244	2.870	0.005	**	15.613	
55-64	2.722	1.976	3.749	0.439	6.200	0.000	***	16.142	
65-74	5.096	3.541	7.332	0.935	8.880	0.000	***	18.340	
75 and above	7.366	3.934	13.789	2.328	6.320	0.000	***	31.603	#
Highest Qualification (2013)									
No Qualification	1.434	1.064	1.932	0.216	2.400	0.018	**	15.029	
School-level Qualification	1.158	0.870	1.541	0.167	1.020	0.310		14.401	
Post-school Qualification	1.122	0.815	1.545	0.181	0.710	0.476		16.121	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.070	1.168	3.669	0.597	2.520	0.013	**	28.850	
$15,001 - $30,000	1.434	0.779	2.640	0.441	1.170	0.244		30.752	#
$30,001 - $50,000	1.642	0.889	3.033	0.508	1.600	0.112		30.919	#
$50,001 - $100,000	1.420	0.758	2.662	0.450	1.110	0.270		31.659	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.542	2.116	3.055	0.235	10.080	0.000	***	9.260	
Importance of Culture (TK 2013)									
Very	1.000	(base)							
Quite	1.020	0.818	1.274	0.114	0.180	0.856		11.163	

Somewhat	0.984	0.767	1.264	0.124	-0.120	0.901		12.585	
A little	0.724	0.549	0.955	0.101	-2.320	0.023	**	13.950	
None	0.922	0.702	1.211	0.127	-0.590	0.557		13.738	
DK	S	S	S	S	S	S		S	S
Any Medical Discrimination (TK 2013)									
No	1.000	(base)							
Yes	1.473	1.041	2.084	0.258	2.210	0.029	**	17.492	
Trust in Fair Healthcare (TK 2013)	0.969	0.935	1.004	0.017	-1.750	0.083	*	1.801	
_cons	0.086	0.043	0.171	0.030	-7.110	0.000	***	34.494	#

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 383,500 (rounded to base 500), Replications = 100, Design df = 99, $F(22, 78) = 21.95$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

The above regression analysis 3.21 includes the Te Kupenga 2013 specific variables and their potential impacts on ASH events. Interestingly, the importance of culture is seen to impact ASH events, but only when culture is considered to be 'a little important', in comparison to culture being very important (baseline). Considering the ordered nature of the importance of culture variable, we may expect to see the same trend for those who consider 'no importance of culture' in comparison to 'culture is very important', however, this is not the case. Nonetheless, for those who view culture with 'a little' importance, there are lower odds of an ASH event. Furthermore, if an individual has experienced any medical discrimination, the odds of an ASH event is 1.473 times greater than those who have not experienced medical discrimination. A one unit change in the measure of Trust in Fair Healthcare multiplies the odds of an ASH event by 0.969. This may imply that if individuals have greater trust in fair healthcare, they have a decreased odds in having an ASH event. However, this finding is only significant at the 10% level. All aspects explained are shown to occur when other variables are held constant.

Table 3.22: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), Household Crowding (2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.616	1.323	1.973	0.163	4.760	0.000	***	10.076	
Age (2013)									
15-24	1.000	(base)							
25-34	0.957	0.726	1.261	0.133	-0.320	0.752		13.927	
35-44	1.151	0.848	1.563	0.177	0.910	0.364		15.414	
45-54	1.564	1.143	2.141	0.247	2.830	0.006	**	15.821	
55-64	2.774	2.004	3.840	0.455	6.220	0.000	***	16.393	
65-74	5.056	3.459	7.390	0.967	8.470	0.000	***	19.130	
75 and above	7.373	3.805	14.287	2.458	5.990	0.000	***	33.338	#
Highest Qualification (2013)									
No Qualification	1.434	1.059	1.941	0.219	2.360	0.020	**	15.265	
School-level Qualification	1.153	0.866	1.535	0.166	0.990	0.327		14.426	
Post-school Qualification	1.122	0.811	1.551	0.183	0.700	0.485		16.351	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.037	1.153	3.600	0.584	2.480	0.015	**	28.686	
$15,001 - $30,000	1.427	0.775	2.627	0.439	1.160	0.250		30.752	#
$30,001 - $50,000	1.602	0.870	2.950	0.493	1.530	0.129		30.766	#
$50,001 - $100,000	1.417	0.756	2.656	0.449	1.100	0.274		31.676	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.577	2.127	3.122	0.249	9.780	0.000	***	9.676	
Importance of Culture (TK 2013)									
Very	1.000	(base)							
Quite	1.004	0.801	1.259	0.114	0.030	0.973		11.393	

Somewhat	0.971	0.755	1.250	0.123	-0.230	0.819		12.706	
A little	0.715	0.540	0.945	0.101	-2.380	0.019	**	14.094	
None	0.915	0.696	1.204	0.127	-0.640	0.524		13.830	
DK	S	S	S	S	S	S		S	S
Any Medical Discrimination (TK 2013)									
No	1.000	(base)							
Yes	1.436	1.010	2.043	0.255	2.040	0.044	**	17.751	
Trust in Fair Healthcare (TK 2013)	0.971	0.936	1.008	0.018	-1.570	0.121		1.864	
Household Crowding (2013)									
Not crowded	1.000	(base)							
Crowded	1.094	0.820	1.459	0.159	0.620	0.539		14.530	
_cons	0.085	0.043	0.169	0.029	-7.130	0.000	***	34.527	#

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 381,000 (rounded to base 500), Replications = 100, Design df = 99, $F(23, 77) = 20.73$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Incorporating household crowding into the model in 3.22 shows that there is no effect on the odds of ASH events. This is while all other variables are held constant in the model.

Table 3.23: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), NZDep Quintiles (2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.600	1.308	1.957	0.162	4.630	0.000	***	10.157	
Age (2013)									
15-24	1.000	(base)							
25-34	0.960	0.729	1.264	0.133	-0.300	0.767		13.878	
35-44	1.144	0.847	1.546	0.174	0.890	0.377		15.169	
45-54	1.568	1.152	2.136	0.244	2.890	0.005	**	15.563	
55-64	2.714	1.968	3.741	0.439	6.170	0.000	***	16.180	
65-74	5.100	3.542	7.345	0.937	8.860	0.000	***	18.381	
75 and above	7.162	3.794	13.518	2.293	6.150	0.000	***	32.016	#
Highest Qualification (2013)									
No Qualification	1.426	1.054	1.928	0.217	2.330	0.022	**	15.225	
School-level Qualification	1.158	0.869	1.544	0.168	1.010	0.314		14.500	
Post-school Qualification	1.117	0.813	1.535	0.179	0.690	0.490		16.009	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.045	1.143	3.659	0.600	2.440	0.017	**	29.324	
$15,001 - $30,000	1.422	0.768	2.632	0.441	1.130	0.260		31.045	#
$30,001 - $50,000	1.644	0.881	3.066	0.516	1.580	0.117		31.419	#
$50,001 - $100,000	1.419	0.756	2.660	0.450	1.100	0.273		31.691	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.539	2.108	3.058	0.238	9.930	0.000	***	9.380	
Importance of Culture (TK 2013)									
Very	1.000	(base)		(base)					
Quite	1.025	0.823	1.278	0.114	0.230	0.822		11.095	

Somewhat	0.999	0.783	1.273	0.122	-0.010	0.990		12.228	
A little	0.736	0.557	0.972	0.103	-2.180	0.031	**	14.048	
None	0.948	0.724	1.243	0.129	-0.390	0.698		13.631	
DK	S	S	S	S	S	S		S	S
Any Medical Discrimination (TK 2013)									
No	1.000	(base)							
Yes	1.467	1.036	2.077	0.257	2.190	0.031	**	17.527	
Trust in Fair Healthcare (TK 2013)	0.969	0.935	1.004	0.017	-1.740	0.085	*	1.806	
NZDep Quintiles (2013)									
Dep 1&2	1.000	(base)							
Dep 3&4	1.195	0.787	1.815	0.252	0.850	0.400		21.071	
Dep 5&6	1.067	0.732	1.557	0.203	0.340	0.733		19.021	
Dep 7&8	0.919	0.633	1.333	0.172	-0.450	0.653		18.760	
Dep 9&10	1.187	0.859	1.641	0.194	1.050	0.295		16.314	
_cons	0.080	0.039	0.163	0.029	-7.010	0.000	***	36.078	#

Note: $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 383,500 (rounded to base 500), Replications = 100, Design df = 99, $F(26, 74) = 18.87$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

The regression output from 3.23 shows that when considering geographic factors such as NZDep alongside Te Kupenga specific variables, it can be seen that NZDep does not impact ASH events to the same extent. For example, no quintiles of NZDep impact ASH events, however any medical discrimination and Trust in Fair Healthcare continue to display an effect on the odds of ASH events.

Table 3.24: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife t Std. Err.		$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.598	1.326	1.925	0.150	4.990	0.000	***	9.392	
Age (2013)									
15-24	1.000	(base)							
25-34	0.930	0.718	1.206	0.122	-0.550	0.582		13.064	
35-44	1.157	0.884	1.516	0.157	1.070	0.285		13.607	
45-54	1.634	1.227	2.176	0.236	3.400	0.001	**	14.434	
55-64	2.660	2.009	3.522	0.376	6.910	0.000	***	14.149	
65-74	4.960	3.515	6.998	0.861	9.230	0.000	***	17.354	
75 and above	6.508	3.886	10.900	1.692	7.210	0.000	***	25.991	
IRD Individual Income (2013)									
$1 - $15,000	2.374	1.327	4.250	0.697	2.950	0.004	**	29.342	
$15,001 - $30,000	1.758	0.967	3.194	0.529	1.870	0.064	*	30.098	#
$30,001 - $50,000	1.868	1.014	3.441	0.575	2.030	0.045	**	30.784	#
$50,001 - $100,000	1.527	0.825	2.827	0.474	1.360	0.175		31.030	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.510	2.105	2.994	0.223	10.370	0.000	***	8.875	
Any Medical Discrimination (TK 2013)									
No	1.000	(base)							
Yes	1.669	1.205	2.311	0.274	3.120	0.002	**	16.401	
_cons	0.071	0.038	0.132	0.022	-8.410	0.000	***	31.490	#

Note: $^{*}\, p < 0.1$; $^{**}\, p < 0.05$; $^{***}\, p < 0.01$

Number of obs = 4,500 (rounded to base 500), Population size = 411,500 (rounded to base 500), Replications = 100, Design df = 99, $F(13, 87) = 34.16$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% < 50% ; Output rule = ## if *RSE*% 50% < 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Throughout the exploration of possible factors that affect ASH events, variables which may have the greatest impact are included in the regression output above 3.24. This model does not include any household or geographic variables, due to the minimal effect found in previous models. Overall, females are shown to have 1.598 times greater odds of an ASH event than males. For those aged 45 and above, there are greater odds of an ASH event in comparison to those aged 15-24 (significant at the 0.001 level). These odds are particularly greater for those aged 75 and above. IRD individual income showed that those who earned between \$1 - \$50,000 have greater odds of an ASH event in comparison to those earning above \$100,000 (baseline). An individuals' health and their experiences also impact ASH events, as those with a disability had 2.510 times greater odds of an ASH event in comparison to those without a disability. Furthermore, experiencing medical discrimination also increased your odds of an ASH event by 1.669 times in comparison to those who had not experienced discrimination in a medical setting.

3.3.3 COVID-19 Vaccinations

COVID-19 Vaccinations will first be explored for the Te Kupenga 2013 sample through changes across Census and tables of weighted counts and proportions. As this outcome occurs from 2020 onwards, the Te Kupenga 2013 is linked between 2013 and 2018 Census and Te Kupenga surveys. Therefore, it is important to explore how this sample changes across time, for variables that can be compared through time. This approach is demonstrated by exploring variables included in both 2013 and 2018 Census and Te Kupenga surveys and how respondents have transitioned across the time period.

Additionally, tables of weighted counts, cell proportions and proportions - conditioned on Vaccination status (with 95% CI's) will be displayed. Exemplars of this output are provided below using individual, household and geographic variables, with further tables included in Appendix A. This output will be a weighted representation of the Te Kupenga 2013 sample, which only includes Māori aged 15 and older. For the Te Kupenga 2013 sample, Te Kupenga specific variables and household level variables are incorporated to view an added level of complexity for the analysis.

3.3.4 Census 2013 to 2018 Transition

The following tables of 2013 to 2018 unweighted responses are included in this section:

1. 2013 Census Disability and 2018 Census Disability (see table 3.25)
2. 2013 Census Household crowding and 2018 Census Household crowding (see table 3.26)
3. 2013 Census NZDep and 2018 Census NZDep (see table 3.27)

The following tables of 2013 to 2018 unweighted responses are included in the appendix A:

1. 2013 Census Highest Qualification and 2018 Census Highest Qualification
2. 2013 Census Individual Income and 2018 Census Individual Income
3. 2013 IRD Individual Income and 2018 IRD Census Individual Income
4. 2013 Census Household Composition and 2018 Census Household Composition
5. 2013 Census Total Household Income and 2018 Census Total Household Income

All unweighted counts are randomly rounded to base 3, with all counts < 6 suppressed.

Disability

Table 3.25: Unweighted counts of 2013 Census Disability and 2018 Census Disability

	2018 Variable			
	Have Disability	No Disability	missing	Total
2013 Variable				
Have Disability	186	555	393	1134
No Disability	135	2895	1152	4182
missing	15	66	63	144
Total	336	3516	1608	5460

Household crowding

Table 3.26: Unweighted counts of 2013 Census Household crowding and 2018 Census Household crowding

	2018 Variable			
	Crowded	Not crowded	missing	Total
2013 Variable				
Crowded	189	261	231	681
Not crowded	390	3246	1080	4716
missing	9	30	21	60
Total	588	3537	1332	5457

NZDep

Table 3.27: Unweighted counts of 2013 Census NZDep and 2018 Census NZDep

	2018 Variable						
	Dep 1&2	Dep 3&4	Dep 5&6	Dep 7&8	Dep 9&10	missing	Total
2013 Variable							
Dep 1&2	198	129	60	42	24	48	501
Dep 3&4	126	135	135	81	57	78	612
Dep 5&6	63	141	252	204	102	117	879
Dep 7&8	57	90	210	447	345	195	1344
Dep 9&10	27	75	120	327	1266	303	2118
missing	S	S	S	S	S	S	S
Total	471	570	777	1101	1794	741	5454

Transition tables provided above 3.25 3.26 3.27 display the changes between Census measured variables for 2013 and 2018. Disability shows that there were higher counts of individuals with a disability in 2013. Interestingly, there are many individuals who were counted as having a disability in 2013, but did not have a disability in 2018 Census. Household crowding shows that for those not in a crowded household in 2013, many remained in this category during 2018 Census. However, 390 were recorded as later being in a crowded household during 2018. NZDep displays that majority of individuals in lower quintiles (Dep 1&2, Dep 3&4, Dep 5&6) remained in the same category between 2013 and 2018 Censuses. Additionally, majority of those in the Dep 9&10 quintle also remained in the identical category for Census 2018.

3.3.5 Vaccination Summary Tables (Weighted Counts)

The following tables of weighted counts, cell proportions and proportions - conditioned on Vaccination status (with 95% CI's) are provided in this section:

1. Vaccination Status by Sex (see table 3.28)
2. Vaccination Status by Age (2020) (see table 3.29)
3. Vaccination Status by Disability (2018 Census) (see table 3.30)
4. Vaccination Status by How well are whānau getting along? (TK 2018) (see table 3.31)
5. Vaccination Status by Household Crowding (2018 Census) (see table 3.32)
6. Vaccination Status by NZDep (2018 Census) (see table 3.33)

All weighted counts rounded to base 500, with raw counts < 500 suppressed according to the IDI output requirements (Stats NZ, 2020b)

Sex

Table 3.28: Vaccination Status by Sex (weighted counts, total proportions and conditional proportions) with 95% CI's

Vaccination Status	Sex	Weighted counts	Cell proportion	Conditional proportions
No Vaccination				
	Male	56,500	0.1	0.5
		(51,500 , 61,500)	(0.1 , 0.1)	(0.5 , 0.6)
	Female	52,000	0.1	0.5
		(5,000 , 56,000)	(0.1 , 0.1)	(0.4 , 0.5)
	Total	108,500	0.2	1.0
		(56,500 , 117,500)	(0.2 , 0.2)	
Partially Vaccinated				
	Male	3,000	0.0	0.4
		(2,000 , 4,000)	(0 , 0)	(0.3 , 0.5)
	Female	4,000	0.0	0.6
		(2,500 , 5,000)	(0 , 0)	(0.5 , 0.7)
	Total	7,000	0.0	1.0
		(4,500 , 9,000)	(0 , 0)	
Fully Vaccinated				
	Male	190,000	0.4	0.5
		(184,500 , 195,000)	(0.4 , 0.4)	(0.5 , 0.5)
	Female	215,000	0.4	0.5
		(210,500 , 219,500)	(0.4 , 0.4)	(0.5 , 0.5)
	Total	405,000	0.8	1.0
		(395,000 , 414,500)	(0.8 , 0.8)	
Total		520,500	1.0	
		(456,000 , 541,000)		

Note: Bracketed values are the 95% CI for given category

Age (Personal Detail 2020)

Table 3.29: Vaccination Status by Age (Personal Detail 2020) (weighted counts, total proportions and conditional proportions) with 95% CI's **(Continued on the following two pages)**

Vaccination Status	Age-band	Weighted counts	Cell proportion	Conditional proportions
No Vaccination	15-24	10,500	0.0	0.1
		(8,000 , 12,500)	(0 , 0)	(0.1 , 0.1)
	25-34	29,500	0.1	0.3
		(25,500 , 33,500)	(0 , 0.1)	(0.2 , 0.3)
	35-44	14,000	0.0	0.1
		(11,500 , 16,500)	(0 , 0)	(0.1 , 0.2)
	45-54	16,000	0.0	0.1
		(13,000 , 19,000)	(0 , 0)	(0.1 , 0.2)
	55-64	16,000	0.0	0.1
		(1,500 , 18,500)	(0 , 0)	(0.1 , 0.2)
	65-74	9,000	0.0	0.1
		(7,500 , 11,000)	(0 , 0)	(0.1 , 0.1)
	75 and above	13,000	0.0	0.1
		(11,000 , 15,500)	(0 , 0)	(0.1 , 0.1)
	Missing	S	S	S
		S	S	S
	Total	108,000	0.1	0.9
		(78,000 , 126,500)	(0 , 0.1)	
Partially Vaccinated	15-24	500	0.0	0.1
		(0 , 1,000)	(0 , 0)	(0 , 0.1)
	25-34	3,000	0.0	0.4
		(1,500 , 4,000)	(0 , 0)	(0.3 , 0.5)
	35-44	1,500	0.0	0.2
		(1,000 , 2,500)	(0 , 0)	(0.1 , 0.3)
	45-54	1,000	0.0	0.1
		(500 , 1,500)	(0 , 0)	(0.1 , 0.2)
	55-64	1,000	0.0	0.1
		(0 , 1,500)	(0 , 0)	(0 , 0.2)
	65-74	S	S	S
		S	S	S
	75 and above	S	S	S
		S	S	S
	Missing	S	S	S
		S	S	S
	Total	7,000	0.0	0.9
		(3,000 , 10,500)	(0 , 0)	
Fully Vaccinated	15-24	41,000	0.1	0.1
		(36,500 , 45,500)	(0.1 , 0.1)	(0.1 , 0.1)
	25-34	92,500	0.2	0.2
		(86,500 , 98,500)	(0.2 , 0.2)	(0.2 , 0.2)
	35-44	73,000	0.1	0.2
		(68,000 , 78,000)	(0.1 , 0.1)	(0.2 , 0.2)
	45-54	80,500	0.2	0.2
		(75,500 , 85,000)	(0.1 , 0.2)	(0.2 , 0.2)
	55-64	64,500	0.1	0.2

		(59,500 , 69,500)	(0.1 , 0.1)	(0.1 , 0.2)
	65-74	34,500	0.1	0.1
		(31,500 , 38,000)	(0.1 , 0.1)	(0.1 , 0.1)
	75 and above	19,000	0.0	0.0
		(16,000 , 21,500)	(0 , 0)	(0 , 0.1)
	Missing	S	S	S
		S	S	S
	Total	405,000	0.8	1.0
		(373,500 , 436,000)	(0.7 , 0.8)	
Total		520,000	0.9	
		(454,500 , 573,000)		

Note: Bracketed values are the 95% CI for given category

Disability (2018)

Table 3.30: Vaccination Status by Disability (2018) (weighted counts, total proportions and conditional proportions) with 95% CI's

Vaccination Status	Disability Status (Census 2018)	Weighted counts	Cell proportion	Conditional proportions
No Vaccination				
	Disability	5,000	0.0	0.0
		(500 , 6,500)	(0 , 0)	(0 , 0.1)
	No Disability	41,000	0.1	0.4
		(36,500 , 45,500)	(0.1 , 0.1)	(0.3 , 0.4)
	Missing	62,500	0.1	0.6
		(57,500 , 67,000)	(0.1 , 0.1)	(0.5 , 0.6)
	Total	108,500	0.2	1.0
		(94,500 , 119,000)	(0.2 , 0.2)	
Partially Vaccinated				
	Disability	S	S	S
		S	S	S
	No Disability	4,000	0.0	0.6
		(3,000 , 5,500)	(0 , 0)	(0.5 , 0.7)
	Missing	2,500	0.0	0.4
		(1,500 , 3,500)	(0 , 0)	(0.3 , 0.5)
	Total	6,500	0.0	1.0
		(4,500 , 9,000)	(0 , 0)	
Fully Vaccinated				
	Disability	23,500	0.0	0.1
		(20,000 , 27,000)	(0 , 0.1)	(0.1 , 0.1)
	No Disability	292,000	0.6	0.7
		(283,500 , 300,000)	(0.5 , 0.6)	(0.7 , 0.7)
	Missing	89,500	0.2	0.2
		(83,500 , 95,500)	(0.2 , 0.2)	(0.2 , 0.2)
	Total	405,000	0.8	1.0
		(387,000 , 422,500)	(0.7 , 0.9)	
Total		520,000	1.0	
		(486,000 , 550,500)		

Note: Bracketed values are the 95% CI for given category

How well are whānau getting along? (TK 2018)

Table 3.31: Vaccination Status by How well are whānau getting along? (TK 2018) (weighted counts, total proportions and conditional proportions) with 95% CI's **(Continued on the following two pages)**

Vaccination Status	Whanau getting along (TK 18)	Weighted counts	Cell proportion	Conditional proportions
No Vaccination	Very well	47,000	0.1	0.4
		(42,500 , 51,000)	(0.1 , 0.1)	(0.4 , 0.5)
	Well	43,000	0.1	0.4
		(38,500 , 47,500)	(0.1 , 0.1)	(0.4 , 0.4)
	Badly	2,000	0.0	0.0
		(0 , 2,500)	(0 , 0)	(0 , 0)
	Neither well/bad	15,000	0.0	0.1
		(12,500 , 17,500)	(0 , 0)	(0.1 , 0.2)
	Very Badly	S	S	S
		S	S	S
	DK	S	S	S
		S	S	S
	Missing	1,500	0.0	0.0
		(1,000 , 2,500)	(0 , 0)	(0 , 0)
	Total	108,500	0.2	0.9
		(94,500 , 121,000)	(0.2 , 0.2)	
Partially Vaccinated	Very well	2,500	0.0	0.4
		(1,500 , 3,500)	(0 , 0)	(0.2 , 0.5)
	Well	3,000	0.0	0.4
		(2,000 , 4,500)	(0 , 0)	(0.3 , 0.6)
	Neither well/bad	1,000	0.0	0.1
		(500 , 1,500)	(0 , 0)	(0 , 0.2)
	Badly	S	S	S
		S	S	S
	Very Badly	S	S	S
		S	S	S
	DK	S	S	S
		S	S	S
	Missing	S	S	S
		S	S	S
	Total	6,500	0.0	0.9
		(4,000 , 9,500)	(0 , 0)	
Fully Vaccinated	Very well	183,000	0.4	0.5
		(175,500 , 191,000)	(0.3 , 0.4)	(0.4 , 0.5)
	Well	155,500	0.3	0.4
		(147,500 , 163,000)	(0.3 , 0.3)	(0.4 , 0.4)
	Neither well/bad	48,500	0.1	0.1
		(43,500 , 54,000)	(0.1 , 0.1)	(0.1 , 0.1)
	Badly	10,000	0.0	0.0
		(8,000 , 12,500)	(0 , 0)	(0 , 0)
	Very Badly	2,500	0.0	0.0
		(1,500 , 3,500)	(0 , 0)	(0 , 0)
	DK	S	S	S
		S	S	S

	Missing	4,500 (3,000 , 6,000)	0.0 (0 , 0)	0.0 (0 , 0)
	Total	404,000 (379,000 , 430,000)	0.8 (0.7 , 0.8)	1.0
Total		519,000 (477,500 , 560,500)	1.0	

Note: Bracketed values are the 95% CI for given category

Household Crowding (Census 2018)

Table 3.32: Vaccination Status by Household Crowding (2018) (weighted counts, total proportions and conditional proportions) with 95% CI's

Vaccination Status	Household Crowding (2018)	Weighted counts	Cell proportion	Conditional proportions
No Vaccination				
	Crowded	10,000	0.0	0.1
		(7,500 , 12,500)	(0 , 0)	(0.1 , 0.1)
	Not crowded	41,500	0.1	0.4
		(37,000 , 46,500)	(0.1 , 0.1)	(0.3 , 0.4)
	Missing	57,000	0.1	0.5
		(52,000 , 62,000)	(0.1 , 0.1)	(0.5 , 0.6)
	Total	108,500	0.2	1.0
		(96,500 , 121,000)	(0.2 , 0.2)	
Partially Vaccinated				
	Crowded	1,000	0.0	0.1
		(500 , 1,500)	(0 , 0)	(0.1 , 0.2)
	Not crowded	4,000	0.0	0.6
		(2,500 , 5,500)	(0 , 0)	(0.5 , 0.7)
	Missing	2,000	0.0	0.3
		(1,000 , 2,500)	(0 , 0)	(0.2 , 0.4)
	Total	7,000	0.0	1.0
		(4,000 , 9,500)	(0 , 0)	
Fully Vaccinated				
	Crowded	45,500	0.1	0.1
		(41,500 , 50,000)	(0.1 , 0.1)	(0.1 , 0.1)
	Not crowded	289,500	0.6	0.7
		(28,000 , 297,000)	(0.5 , 0.6)	(0.7 , 0.7)
	Missing	69,500	0.1	0.2
		(64,500 , 75,000)	(0.1 , 0.1)	(0.2 , 0.2)
	Total	404,500	0.8	1.0
		(134,000 , 422,000)	(0.7 , 0.8)	
Total		520,000	1.0	
		(234,500 , 552,500)		

Note: Bracketed values are the 95% CI for given category

NZDep (Census 2018)

Table 3.33: Vaccination Status by NZDep (Census 2018) (weighted counts, total proportions and conditional proportions) with 95% CI's

Vaccination Status	NZDep (Census 2018)	Weighted counts	Cell proportion	Conditional proportions
No Vaccination	Dep 1&2	5,500	0.0	0.1
		(3,500 , 7,500)	(0 , 0)	(0 , 0.1)
	Dep 3&4	8,500	0.0	0.1
		(6,500 , 10,500)	(0 , 0)	(0.1 , 0.1)
	Dep 5&6	10,000	0.0	0.1
		(7,500 , 12,000)	(0 , 0)	(0.1 , 0.1)
	Dep 7&8	10,000	0.0	0.1
		(8,000 , 12,500)	(0 , 0)	(0.1 , 0.1)
	Dep 9&10	29,000	0.1	0.3
		(26,000 , 32,500)	(0 , 0.1)	(0.2 , 0.3)
	Missing	45,500	0.1	0.4
		(41,000 , 50,000)	(0.1 , 0.1)	(0.4 , 0.5)
	Total	108,500	0.2	1.1
		(92,500 , 125,000)	(0.1 , 0.2)	
Partially Vaccinated	Dep 1&2	500	0.0	0.1
		(0 , 1,000)	(0 , 0)	(0 , 0.2)
	Dep 3&4	500	0.0	0.1
		(0 , 1,000)	(0 , 0)	(0 , 0.2)
	Dep 5&6	500	0.0	0.1
		(0 , 1,000)	(0 , 0)	(0 , 0.2)
	Dep 7&8	1,500	0.0	0.2
		(500 , 2,500)	(0 , 0)	(0.1 , 0.3)
	Dep 9&10	3,000	0.0	0.4
		(1,500 , 4,000)	(0 , 0)	(0.3 , 0.5)
	Missing	1,000	0.0	0.1
		(500 , 1,500)	(0 , 0)	(0 , 0.2)
	Total	7,000	0.0	1.0
		(2,500 , 11,000)	(0 , 0)	
Fully Vaccinated	Dep 1&2	49,500	0.1	0.1
		(44,500 , 54,500)	(0.1 , 0.1)	(0.1 , 0.1)
	Dep 3&4	54,000	0.1	0.1
		(48,000 , 60,500)	(0.1 , 0.1)	(0.1 , 0.1)
	Dep 5&6	67,000	0.1	0.2
		(62,500 , 71,500)	(0.1 , 0.1)	(0.2 , 0.2)
	Dep 7&8	89,000	0.2	0.2
		(83,000 , 95,000)	(0.2 , 0.2)	(0.2 , 0.2)
	Dep 9&10	118,500	0.2	0.3
		(112,500 , 125,000)	(0.2 , 0.2)	(0.3 , 0.3)
	Missing	26,500	0.1	0.1
		(22,500 , 30,000)	(0 , 0.1)	(0.1 , 0.1)
	Total	404,500	0.8	1.0
		(373,000 , 436,500)	(0.7 , 0.8)	
Total		520,000	1.0	
		(468,000 , 572,500)		

Note: Bracketed values are the 95% CI for given category

Tables 3.28 3.29 3.30 3.31 3.32 3.33 convey the distribution of explanatory variables given Vaccination status. Between males and females, there are a higher proportion of fully vaccinated females compared to males. For distribution of age, many rows are suppressed, likely due to low counts particularly for partially vaccinated. For those not vaccinated, there are a larger proportion of those aged 25-34, in comparison to the surrounding age-bands, however, this trend is also displayed for fully vaccinated. Disability indicators revealed that for those not vaccinated, there are a higher proportion of those without a disability. Likewise, those fully vaccinated display the same trend. When measuring how well whānau get along, for fully vaccinated there are a higher proportion of those where whānau get along 'very well'. Household crowding shows that for fully vaccinated, there are a higher proportion of individuals not in a crowded house (likewise for non-vaccinated). NZDep shows that when conditioning on no vaccination and fully vaccinated (individually), those in higher levels of deprivation (Dep 9&10) have a higher proportion compared to those in lower quintiles of deprivation.

3.3.6 Regression Analysis (COVID-19 Vaccinations)

The following output is of logistic regression output using COVID-19 Vaccinations as the outcome variable. The first set of output uses Multinomial Logistic Regression and results in Relative Risk Ratio in comparison to the baseline. However, vaccinations are an ordered outcome variable and are therefore used in an Ordered Logistic Regression. This method is preferred as this takes the ordered nature of the outcome variable into consideration throughout the analysis. Nonetheless, two Multinomial Logistic Regression will be provided in appendix A for reference.

Ordinal Logistic Regression uses the Vaccination outcome as a 3-levelled ordered variable (from no vaccination, partially vaccinated, through to fully vaccinated). The Ordinal Logistic regressions will produce coefficients for each parameter, that can be transformed into proportional odds ratios (Agresti, 2013). Due to the ordered nature of the vaccination outcome, the Ordinal Logistic Regressions provide an improved view of variables that may effect vaccinations. Interpreting ordinal logistic regression will involve providing the original coefficients, alongside their transformed odds ratio for easy of interpretation. This process is completed by exponentiation of the coefficient.

Ordinal Logistic Regression was implemented to view the impacts of the ordered response variable for vaccinations. Within this output, the regression models include `/cut1` and `/cut2` parameters, which demonstrate where the latent variable is cut, demonstrating the three cut off points within the data. However, these cut points are not applied to interpretation of the results, and are often used for other purposes in Stata (Statistical Methods and Data Analytics, 2023).

As required by Stats NZ, `RSE \%` and `Output rule` are included as columns for each regression to demonstrate the relative sampling error when using the Te Kupenga survey. `RSE\%` is presented as a percentage, with the following rules displayed in the `Output rule` column (Stats NZ, 2021);(Stats NZ, 2020b):

1. Suppress estimates with a relative sampling error (RSE) of 100 percent or greater.
2. Identify estimates with an RSE between 30 percent and less than 50 percent with one hash symbol (#).
3. Identify estimates with an RSE between 50 percent and less than 100 percent with two hash symbols (##).

The following regression analysis will be included in this results section:

1. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table 3.34)
2. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Disability (2018) - Coefficients (see table 3.36)

3. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Coefficients (see table 3.37)

The following regression analysis will be included in the appendix:

1. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Sex*Age, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table A.22)
2. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients (see table A.23)
3. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018) - Coefficients (see table A.24)
4. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018), NZDep Quintiles (2018) - Coefficients (see table A.25)
5. Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), NZDep Quintiles (2018) - Coefficients (see table A.26)
6. Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018) - Relative Risk Ratio (see table A.27)
7. Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Relative Risk Ratio (see table A.28)

Table 3.34: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	0.418	0.214	0.621	0.102	4.080	0.000	***	24.526	
IRD Individual Income (2018)									
$1 - $15,000	-2.644	-3.731	-1.557	0.548	-4.830	0.000	***	20.716	
$15,001 - $30,000	-2.414	-3.575	-1.254	0.585	-4.130	0.000	***	24.218	
$30,001 - $50,000	-2.058	-3.214	-0.903	0.582	-3.540	0.001	**	28.288	
$50,001 - $100,000	-1.553	-2.735	-0.370	0.596	-2.600	0.011	**	38.388	#
$100,001 and above	0.000	(base)							
Trust in Fair Healthcare (TK 2013)	0.048	0.008	0.088	0.020	2.400	0.018	**	41.593	#
How are whānau doing? (TK 2018)	0.070	0.024	0.117	0.024	2.980	0.004	**	33.563	#
/cut1	-3.036	-4.207	-1.864	0.590				19.447	
/cut2	-2.903	-4.073	-1.733	0.590				20.313	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 4,500 (rounded to base 500), Population size = 414,500 (rounded to base 500), Replications = 100, Design df = 99, $F(7, 93) = 19.5$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Table 3.35: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Odds Ratios

	Odds Ratio	95% CI Lower	95% CI Upper
Sex			
Male	1.000	(base)	
Female	1.518	1.239	1.860
IRD Individual Income (2018)			
$1 - $15,000	0.071	0.024	0.211
$15,001 - $30,000	0.089	0.028	0.285
$30,001 - $50,000	0.128	0.040	0.405
$50,001 - $100,000	0.212	0.065	0.691
$100,001 and above	1.000	(base)	
Trust in Fair Healthcare (TK 2013)	1.049	1.008	1.092
How are whānau doing? (TK 2018)	1.073	1.024	1.124

Tables 3.34 and 3.35 display the coefficients and corresponding odds ratios when including Sex, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013) and How are whānau doing? (TK 2018) to explore the impacts on vaccination.

All variables within this model are significant. Trust in fair healthcare and how are whānau doing display a relative sampling error of approximately 40%, therefore, estimates may not be entirely reliable. However, when exponentiating the coefficients to interpret their odds ratio (displayed in Table 3.35) the following is displayed: In contrast to males, females display odds 1.518 (95% CI: 1.239, 1.860) times greater for the combined partially and fully vaccinated categories in comparison to being non-vaccinated while all other variables are held constant. Alternatively, the odds of partially and fully vaccinated combined categories versus non-vaccinated is 1.518 (95% CI: 1.239, 1.860) times greater than non-vaccinated.

For all IRD individual income categories (2018) less than $100,000, every category displayed lower odds for combined partially and fully vaccinated categories, in comparison to being non vaccinated, while all other variables are held constant. This can also be said for fully vaccinated versus partially and non-vaccinated categories combined, which has lower odds when all other variables are held constant.

Viewing variables relevant to Te Kupenga surveys, Trust in Fair healthcare and how whānau are doing displayed greater odds of combined partially and fully vaccinated categories in comparison to being non-vaccinated (can also be viewed as greater odds of being fully vaccinated versus combined partially vaccinated and non-vaccinated). Specifically, for every one unit increase in trust in fair healthcare, odds of fully vaccinated versus the combined partially vaccinated and non-vaccinated is 1.049 times greater (95% CI: 1.008, 1.092), given all other variables are held constant. Furthermore, the odds of the combined partially and fully vaccinated categories versus non-vaccinated is 1.049 times greater (95% CI: 1.008, 1.092), with all other variables held constant. How are whānau doing displays that for one unit increase, the odds of being fully vaccinated versus combined partially vaccinated and

non-vaccinated is 1.073 times greater (95% CI: 1.024, 1.124), when all other variables are held constant. To elaborate on this, the odds of combined partially and fully vaccinated categories versus non-vaccinated is 1.073 times greater (95% CI: 1.024, 1.124), with all other variables held constant.

Table 3.36: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Disability (2018) - Coefficients

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	S	S	S	S	S	S		S	S
Age (2020)									
15-24	0.000	(base)							
25-34	-0.384	-0.836	0.068	0.228	-1.680	0.095	*	59.388	##
35-44	S	S	S	S	S	S		S	S
45-54	S	S	S	S	S	S		S	S
55-64	S	S	S	S	S	S		S	S
65-74	0.354	-0.209	0.916	0.283	1.250	0.215		80.168	##
75 and above	-0.456	-1.049	0.137	0.299	-1.530	0.130		65.501	##
Disability (2018)									
No Disability	0.000	(base)							
Have Disability	S	S	S	S	S	S		S	S
/cut1	-2.136	-2.544	-1.728	0.205				9.620	
/cut2	-2.015	-2.425	-1.606	0.206				10.239	

Note: $^{*}\, p < 0.1$; $^{**}\, p < 0.05$; $^{***}\, p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 358,500 (rounded to base 500), Replications = 100, Design df = 99, $F(8, 92) = 2.27$, Prob $> F = 0.0289$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

For this regression output, the variables sex and disability are suppressed, implying that the relative sampling error is greater than 100%. This means that these variables cannot be included within the model as the estimates are no longer reliable. The only value for this output that meets the 0.05 significance level is for the ages 25-34, which means we expect a decrease in the log odds of being in a higher level of vaccination, given all other variables are held constant. If interpreting ths using odds ratios, then for those ages 25-34, the odds of being fully vaccinated versus combined partially and non-vaccinated is 1.424 times greater (95% CI: 0.434, 1.071). Furthermore, for odds of combined partially vaccinated and fully vaccinated versus non-vaccinated is 1.424 times greater (95% CI: 0.434, 1.071). This assumption is given all other variables in the model are held constant.

Table 3.37: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Coefficients

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
NZDep Quintiles (2018)									
Dep 1&2	0.000	(base)							
Dep 3&4	-0.290	-0.824	0.244	0.269	-1.080	0.285		92.952	##
Dep 5&6	S	S	S	S	S	S		S	S
Dep 7&8	S	S	S	S	S	S		S	S
Dep 9&10	-0.782	-1.250	-0.314	0.236	-3.310	0.001	**	30.168	#
/cut1	-2.393	-2.840	-1.946	0.225				9.417	
/cut2	-2.265	-2.719	-1.811	0.229				10.098	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 4,500 (rounded to base 500), Population size = 436,500 (rounded to base 500), Replications = 100, Design df = 99, $F(4, 96) = 8.69$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ 30% $<$ 50% ; Output rule = ## if $RSE\%$ 50% $<$ 100%; Output rule = S if $RSE\% \geq 100\%$

This example uses NZDep Quintiles from 2018 only, as an explanatory variable for vaccination level. However, across all quintiles of deprivation aside from Dep 9&10, each category is either suppressed or non-significant. For those with Deprivation level 9&10, the odds of the partially and fully vaccinated categories of vaccination are 0.458 (95% CI: 0.287, 0.731) times lower than non vaccination, while all other variables are held constant (calculation: $\exp -0.782 = 0.458$. transforming the coefficient into an odds ratio). Likewise, the odds of fully vaccinated versus the combined partially and non-vaccinated are 0.458 (95% CI: 0.287, 0.731) times lower, when all other variables are held constant.

Chapter 4

Discussion

The aims of this thesis are to develop longitudinal methods using Te Kupenga as an example, and the discussion will outline the results found across this analysis and the methodology created. This discussion will provide an overview of the results found in this study and any limitations discovered throughout. This overview outlines the advantages and disadvantages of using IDI datasets, linkage processes and confidentiality trade-offs. The challenges to data linkage and data quality will be examined alongside the factors that need to be considered when re-weighting the Te Kupenga sample. Furthermore, recommendations are provided in 5 for future approaches needed to continue developing longitudinal methodology with Te Kupenga.

4.1 Overview of results

This project creates novel methodological approaches to using the Te Kupenga 2013 sample for longitudinal analysis. Application of this required providing an overview of ASH on a national level. This initial stage uncovered the differences in ASH events between Māori and Non-Māori by sex, age, highest qualification and individual income using Census 2013 variables. It was found that although the proportion of ASH events is lower on a national population level, the proportion of ASH events is higher for Māori than Non-Māori. When conditioning on ASH events, Māori females display a greater proportion of ASH events in contrast to Māori males. Furthermore, ASH events were higher for Māori and Non-Māori with lower levels of qualification. This trend is also found for individuals with lower incomes compared to those with higher incomes. These findings informed the motivation to explore the impacts of ASH for Māori only on a national level.

ASH events were examined through logistic regression for Māori only (using the Census 2013 sample). It was found that Māori females have higher odds of an ASH event than Māori males; Māori aged 25 and above display increased odds of an ASH event compared to those aged 15-24 (when all other variables are kept constant). Using IRD individual income records, Māori with income greater than \$15,000 displayed decreased odds of ASH events compared to those who earned $1-15,000. This trend continued for those with higher levels of qualification (school level and above), who displayed decreased odds in ASH events compared to Māori with no qualification. For those with an ASH event, ASH length of stay and the count of unique ASH events were also investigated for Māori from the Census 2013 sample. Once overdispersion was explored for both outcomes, negative binomial regression models were fit to explore the implications of sex, age, highest qualification and IRD individual income. Similar to the results for ASH events, Māori females displayed greater rates in the length of stay and count of unique ASH events in comparison to Māori males, and those aged above 25 followed the same pattern, with greater expected rates in comparison to 15-24-year-olds. Highest qualification and IRD individual income displayed the opposite for both outcomes, with lower expected rates for those above school-level qualification and earning above \$15,000 compared to the baseline categories. All conclusions found for parameters occurred when all other variables are held constant.

As outlined in the aims of this thesis, it was of interest to explore longitudinal analysis for the Te Kupenga sample, through the use of individual, housing and geographic level variables. When investigating the impacts of ASH events for this sample using logistic regression analysis (applying Jackknife weighting) (table 3.24), it was found that housing and geographic variables (household crowding and NZDep from 2013) do not have an effect on ASH events. However, variables relating to individual circumstances such as sex, age, IRD individual income, disability and any medical discrimination were found to impact the odds of an ASH event. For example, females had higher odds of an ASH event than males, and those aged 25 and above also had greater odds of an ASH event. Furthermore, those with a disability had 2.510 greater odds of an ASH event in comparison to those without a disability. Experience of medical discrimination was measured in Te Kupenga 2013, and it was found that those who had experienced medical discrimination had 1.669 greater odds of an ASH event than those who had not experienced medical discrimination. These findings occurred when all other variables within the model were held constant.

Continuing longitudinal analysis for the Te Kupenga 2013 sample involved considering COVID-19 vaccinations from 2020 onward by incorporating variables from the 2013 and 2018 Census and Te Kupenga surveys. To demonstrate changes across time, tables of transition were created for variables measured in both the 2013 and 2018 Census before exploring weighted counts for these individuals, given vaccination status. When applying an ordered logistic regression to this cohort, geographic measures (NZDep in 2018) only affected vaccination status when applied without other variables. However, sex, IRD individual income (2018), trust in fair healthcare (Te Kupenga 2013) and how whānau are doing (Te Kupenga 2018) displayed impacts on vaccination status 3.34. Examining variables specific to Te Kupenga surveys, a one unit increase in the measure for trust in healthcare is associated with an increase in the odds of being at least partially vaccinated by 1.049. Furthermore, the odds of being at least partially vaccinated is 1.049 times greater than non-vaccinated. How whānau are doing displayed the same trend, with a one unit increase in how well whānau are doing, producing odds of being partially vaccinated and above to be 1.073 times greater. This can also be interpreted as the odds of being partially vaccinated and above being 1.073 times greater than non-vaccinated. These findings occur when all variables are held constant in the model. However, despite these findings, causality cannot be applied as multiple variables within this model displayed a relative sampling error of approximately 40%; therefore, the parameters may not be reliable. This may result from loss to follow-up and potential implications when observations are missing from the sample.

4.2 Data Linkage & Quality of IDI data

This section of this discussion will explore the quality of IDI data and their implications for data linkage. When using datasets in the IDI, it is essential to distinguish the difference between the original survey dataset created (the survey conducted in the population) and the datasets as available in the IDI. As this section explains, although the Te Kupenga sample was created using a sub-sample of Māori from the Census, the variables available are inconsistent across both surveys. This issue is found across individual, household and geographic level variables.

4.2.1 Data Linkage

Data linkage is a crucial stage of working in the IDI as minor errors can cause considerable implications at later stages of data analysis. Therefore, it is essential to consider the issues found at two stages: firstly, when issues are found within datasets outside of a researcher's control and secondly, when performing bespoke linkages as an IDI user.

Linking death information and vaccination status to the Te Kupenga sample uncovered issues in the dataset entries. When joining DIA deaths to the Te Kupenga sample, duplicate records were found for individuals. While duplicates are possible for datasets with multiple time points (for example, an individual has more than one vaccination record), this scenario is implausible as a person should only

have a singular death record. This situation implies issues with the quality of death data and linkage to the IDI spine. Another example of linkage errors outside of the IDI user's control is displayed with vaccination records. While the general population's vaccination roll-out began in 2021, many observations were found for 2020. Although this does not reflect the main vaccination period, this may reflect vaccinations for those in border, MIQ and front-line work throughout the COVID-19 pandemic. Due to the risk of dropping many observations from the Te Kupenga sample, this project used a vaccination outcome period beginning in 2020. However, issues with vaccination records persisted as the first COVID-19 vaccination recorded in the dataset occurred before 2000, and death entries were found before vaccination records. Given that COVID-19 vaccinations began in 2020 and an individual cannot be vaccinated after death, this finding represents an impossible scenario. Although these issues were found for only a small number of observations, likely reflecting data entry errors (Kvalsvig et al., 2019), this can have enduring and possibly cumulative impacts if unnoticed and these observations are kept in the sample. This issue points toward implementing robust 'sense checking' processes made across IDI data and their variables to ensure commitment to data quality is maintained. While StatsNZ needs more resources to undertake all the required checking, they could produce guidelines for robust sense checking for IDI users.

It is also possible to generate errors due to data linkage – routinely through the IDI refresh process or by an IDI user when performing bespoke data linkage. Within the IDI, there are fully integrated datasets and ad-hoc datasets. Fully integrated datasets include information that needs to be regularly updated through the IDI refresh. In contrast, ad-hoc datasets are typically one-off datasets that can be linked to others using a unique identifier. Te Kupenga is an example of the latter, as this survey sample is linked to the Census by snz_uid and Census ID. As each refresh randomises the unique identifiers, this requires an extra step when linking Te Kupenga to the Census to link individuals across datasets correctly. This creates the potential for linkage error and linkage bias, which may be inconsistent between refreshes for longer time periods and specific users. Linkage error and linkage bias have been documented in the Census/ Mortality data linkage work (Fawcett et al., 2002) and is a recognised issue when working with IDI data (Kvalsvig et al., 2019). Linkage errors and biases could be an issue for using linked Te Kupenga data in the IDI, but methods to address these are well beyond the remit of this thesis. These issues are the subject of a PhD in statistics by Eileen Li as part of the Te Rorou research project (https://terourou.org/), with her thesis due to be submitted in mid-2024.

As demonstrated in this project, there were numerous circumstances where coding needed to be completed across multiple software programs. This was different from the original intention for coding in the IDI; however, these decisions were needed to extract the appropriate IDI datasets without errors. For example, most of the datasets were extracted using SAS; however, IRD income and COVID-19 vaccinations required using SQL. The primary reason for this issue was due to the length of the dataset IRD income and COVID-19 names. Although it requires minimal effort to export datasets from SQL, the added step is time-consuming and adds to the coding complexities required, and therefore, the potential for error, when working with IDI data.

4.2.2 Self-reporting vs. administrative data

Income was considered as a predictor of interest for each health outcome. While Census 2013 includes a variable for individual income, it is a self-reported measure of income. To consider the most accurate measure of individual income, this was compared to an administrative data source, individual income reported by IRD during 2013. Comparisons are firstly made on a national population level in Table 4.1 and for the Te Kupenga 2013 sample in Table 4.2.

Table 4.1: National level:
IRD Individual Income (2013) vs. Census Individual Income (2013)

		Census 2013 Individual Income							
		Loss/No Income	$1 - $15,000	$15,001 - $30,000	$30,001 - $50,000	$50,001 - $100,000	$100,001 >	Missing	Total
IRD Individual Income (2013)									
	$1 - $15,000	48,075	329,592	100,746	31,146	15,792	4,686	56,004	586,038
	$15,001 - $30,000	9,822	160,926	422,214	112,632	34,800	8,565	64,425	813,390
	$30,001 - $50,000	1,758	10,227	87,015	338,757	46,092	3,522	14,667	502,038
	$50,001 - $100,000	765	1,866	5,379	87,819	435,144	20,454	8,265	559,692
	$100,001 >	99	153	252	726	20,388	105,387	1,167	128,169
	Missing	199,032	116,712	88,830	81,195	81,357	38,559	1,158,180	1,763,871
	Total	259,551	619,473	704,436	652,278	633,579	181,170	1,302,711	4,353,198

Table 4.2: Te Kupenga 2013 sample only (unweighted counts):
IRD Individual Income (2013) vs. Census Individual Income (2013)

		Census 2013 Individual Income							
		Loss/No Income	$1 - $15,000	$15,001 - $30,000	$30,001 - $50,000	$50,001 - $100,000	$100,001 >	Missing	Total
IRD Individual Income (2013)									
	$1 - $15,000	117	702	192	57	12	9	93	1182
	$15,001 - $30,000	18	339	693	156	24	S	120	1350
	$30,001 - $50,000	S	30	183	654	72	6	33	978
	$50,001 - $100,000	S	9	12	225	636	24	15	921
	$100,001 >	S	S	S	S	21	78	S	99
	Missing	408	144	102	87	66	36	66	909
	Total	543	1224	1182	1179	831	153	327	5439

Across both Tables 4.1 and 4.2, there are large discrepancies between self-reported individual income and administrative data income records. Therefore, it can be concluded that self-reported income is not always an accurate measure. To combat this issue, IRD income was used for the majority of the analysis within this project. However, this decision may come with limitations. This includes creating another step of linking IRD income data to our sample dataset. This added the complexity of requiring SQL as the variable name was too long for SAS to read. Also, IRD-based income is not routinely available, and access to IRD records requires an additional permission step to gain access to a restricted dataset within the IDI database. Furthermore, when comparing Census Individual income, IRD contains no record of income loss since gross income is collected for a time period; therefore, conclusions cannot be made for those who receive a loss or no income. The implications of this discrepancy mean that distinctions between self-reported and administrative measures need to be considered in future IDI-based research. A possible solution for this is providing disclaimers that outline the potential biases within the dataset.

4.2.3 Individual Information

Age-band issues

Investigating vaccination rates requires detailed information on age because vaccination eligibility has a precise age threshold, which changes over time. In this research, we are measuring outcomes over a period of time. The IDI has multiple potential sources of age information; therefore, determining the best option for use in this study was the issue.

Differences in COVID-19 vaccinations by age group uncovered an issue with the timeliness of data provided in the IDI and coverage of data sources available. Timeliness is essential for vaccination data, as using an incorrect time point can shift vaccination eligibility. Data quality is another important feature, as the IDI has multiple sources for individual age, each with strengths and limitations. This section will focus on the process of deciding a suitable time point to record age and outline the coverage of the available datasets used for age.

An appropriate time point for age was decided between age at the time of vaccination and age at the beginning of the vaccination outcome period. Age at the time of vaccination provides a detailed account of age-dependant vaccination eligibility. An example of this is age at the first vaccination event. However, this approach creates difficulties for making comparisons across individuals, as age at vaccination differs between individuals and eligibility changes over time. In contrast, age at the beginning of the outcome period (vaccinations from 2020 onward) creates a consistent cutoff date for everyone within the sample. However, age at the beginning of the outcome period does not account for changes in age during the span of the COVID-19 vaccination time period (approximately three years in this analysis). However, as this project aims to create methodology for longitudinal approaches, it is crucial to consider the impacts of age as a predictor before the outcome period. This approach allows for within-group comparisons and ensures that age as a predictor variable occurs before the outcome.

The second issue covers the issue of deciding the appropriate source of age information. These sources of age include using Census records, the Ministry of Health (MoH) (full date of birth table), and the Core data (Personal detail table). This will focus on the strengths and limitations of each option and provide reasons for using Core data as the most appropriate choice. Additionally, this issue is discussed through dataset coverage and deciding on a suitable time point to record age.

Age information details varied between the three IDI datasets available to this project. Census 2013 IDI dataset provides discrete age at the time of the Census, while the Census 2018 IDI dataset provides birth month and year (due to potential confidentiality issues). Therefore, age needs to be derived in creating an age response for individuals during vaccinations. For example, if an individual was 20 during the 2018 Census, their derived age will be 22 in 2020. Although this is a simple calculation, the resulting age is only sometimes accurate, even to the year.

Additionally, not all individuals are included in the 2018 Census (particularly from Te Kupenga 2013); age may only be derived using age at the 2013 Census or Te Kupenga survey. This inconsistency across both Censuses implies difficulties in gathering a consistent source of age information and requires further coding commitments. The second source for age information is using the Ministry of Health's date of birth table. This provides greater detail on specific dates of birth - date, month, and year of birth- compared to using the Census dataset. However, this dataset is pulled from Ministry of Health records and is limited to only those within MoH data. During this project, the sample was linked to this dataset to gain specific date of birth information; however, many individuals were missing from this resource. Therefore, the coverage limitations outweighed the benefits of the potential accuracy of age calculation. The final source for age information uses Stats NZ core data - the personal detail table. This resource provides information across the whole population within the IDI, creating greater coverage for the sample. However, this table is limited to the month and year of birth only. Despite the limited specificity, the personal detail table provides coverage for all individuals, and a general date of birth can be applied across the sample for consistency. For this study, the personal detail table was used as the primary age source and for those born after June, age is derived by subtracting 1 from age at the year 2020.

Considering age in this longitudinal analysis uncovers the decision-making process for appropriate time points for age cutoff and sources of age information. Age is highly impactful as it is vital to age-dependent outcomes, such as vaccinations. Outside of these circumstances, time point cutoff and sources for age may be less critical (for example, age of death). However, vaccination age requires high levels of precision as age affects vaccination eligibility criteria. This project uses age at the beginning of the outcome period, with age information sourced from the personal detail table. To create accurate age bands, Stats NZ granted permission to use date of birth records for the project. However, without such access, this project would rely on deriving age from Census 2013 and 2018 records, potentially compromising accuracy. Therefore, accessing more accurate age data requires an additional step for researchers to use IDI data. Future research in the IDI requires trade-off decisions between coverage and accuracy, depending on the sample size and level of detail needed for age information. Thus, standard approach documentation is required when using age variables in IDI datasets, with emphasis on using specific methods depending on the outcome of interest.

4.2.4 Housing & Geographic Information

Upon linking individuals between the Census and Te Kupenga 2013 and 2018 IDI datasets, inconsistencies were discovered between these IDI datasets in two ways. Firstly, variables are not available within the same timeframe, between the Te Kupenga 2013 sample and Census 2013 dataset for the same population. Secondly, time comparisons are inconsistent, creating difficulty with comparisons within the whole population and when using 2013 Te Kupenga.

This project found issues when using household crowding and NZDep variables from 2013. Household crowding and NZDep are important variables to include in this analysis as they display the impacts of socio-economic inequities within the population (Galobardes, 2006a,b). Household crowding is a vital measurement as a policy amenable factor, which may lead to inequitable health outcomes for Māori if not addressed by the Crown. NZDep is a measure of deprivation that considers elements of area-based social deprivation within the New Zealand context (Atkinson et al., 2019). Therefore, it is important to include crowding and NZDep as variables within the longitudinal analysis for Māori. However, when using the IDI datasets, household crowding and NZDep are only found in the 2013 Te Kupenga IDI variables and are not present in Census 2013 dataset. Although ensuring consistency across datasets and surveys is intuitive, this is not demonstrated within the IDI, therefore creating issues when viewing the impacts of household crowding and deprivation on a national level for comparing Māori and Non-Māori.

For this longitudinal analysis, it was important to consider variables that are measured consistently across each Census and Te Kupenga survey. However, not all variables included in Census 2018 appear

in Census 2013, and likewise for Te Kupenga surveys. While a common issue found in longitudinal surveys, this created difficulties with comparisons throughout different time periods. Therefore, variables available only in the Te Kupenga 2013 dataset are included only for the Te Kupenga 2013 sample. Nonetheless, for variables found in 2013 and 2018 Census, comparisons were made by examining how respondents changed and transitioned through time. This limitation adds to the difficulty of creating longitudinal analysis with IDI data, created for original surveys.

For future use of IDI data, improved validation processes need to be included when adding datasets to the IDI database to ensure that the available data accurately represents the original intention of the survey. Additionally, requiring validated processes for adding data into the IDI acts as a safeguard for ensuring standard measures included in the Census (such as household crowding and NZDep), are also included across linked datasets.

4.3 Confidentiality - Trade-offs with output checking processes

Releasing output from the IDI involves an output-checking process conducted by a team at Stats NZ. IDI users are required to follow the output checking guide (Stats NZ, 2020b). Different rules are applied to the output depending on the datasets used. For this project, unweighted counts under 6 are suppressed, random rounding to base 3 is applied, while weighted counts applicable to the Te Kupenga survey are to base 500, with all counts under 500 suppressed.

While the output-checking process is essential, precision is forfeited, particularly when using weighted counts from Te Kupenga, and row suppression occurs due to low counts. Furthermore, the output must meet further confidentiality rules (Stats NZ, 2020b) when calculating proportions with their corresponding 95% CI's in the IDI. Throughout this project, weighted counts and conditional proportions are displayed for ASH and COVID-19 Vaccinations. However, the level of confidentiality checks applied to cleared files differs between the two outcomes. Proportions and standard errors (used to calculate 95% CI's) for ASH events in the Te Kupenga 2013 sample were calculated from unweighted counts in Stata and cleared for release with precision to 3 decimal places. The same approach was applied to COVID-19 vaccinations for the Te Kupenga sample; however, values released were restricted to 1 decimal place (see table3.28). Therefore, conflicting approaches within the output-checking process are illustrated. Furthermore, this presented issues when displaying proportions for COVID-19 vaccinations with 95% CI's as the level of precision decreased. Proportions can also be calculated directly from the weighted counts released from the IDI output; however, precision remains a concern as these are calculated from rounded counts, not raw counts.

The regression analysis completed also displays issues encountered with suppressed parameters. When a parameter's relative sampling error exceeds 100%, the row is suppressed as a further requirement for using Te Kupenga surveys. Due to reduced sample sizes and observations dropped, the relative sampling error may be inflated, and the model is less reliable. Parameter suppression is a concern when addressing potential causal effects on the outcome - an issue found throughout this study. For future longitudinal methods using Te Kupenga, suppressed values need to be further explored alongside their implications on Jackknife weighting.

4.4 Re-weighting considerations

Te Kupenga is a unique social survey, as it is not a direct sub-sample of the Census. Unlike other surveys such as the General Social Survey that are sub-samples of the Census, there are limited details about the variable distributions included in the survey. Furthermore, finite documentation is released about the actual population included in the sample. Classification of Māori ethnicity and descent has impacted the population of individuals included in the Te Kupenga survey, as Te Kupenga's sampling frame is unpublished and researchers are blind to this detail. Therefore, this creates implications on re-weighting with missing data and observations due to loss to follow-up.

Potential effects of loss to follow-up need to be considered because the original sample weights used for the remaining sample population are no longer weighted to reflect a representative sample. This research uses Te Kupenga 2013 sample as the foundational cohort to analyse the impacts of ASH and COVID-19 Vaccinations. When viewing ASH outcomes, these response variables are recorded from 2014 onward. Thus, the outcome reporting time period is very close to that of the explanatory variables used in this analysis. Therefore, in the context of ASH outcomes, the sample does not require re-weighting due to loss to follow-up. This is important to consider for implications of causality in a longitudinal analysis, as the number of missing observations can create biases within the model (Hernan and Robins, 2020). If dealing with Census data, there are longitudinal methods for dealing with loss to follow-up, as this includes a cohort without weighting considerations. This approach cannot be applied to a nationally representative survey such as Te Kupenga, which requires updated weighting to transform the residual survey population into a representative sample. Suppose this approach were to be generalised to other sample surveys. In that case, researchers need access to information about the weighting methodology of such surveys to create longitudinal analysis across extended periods where a loss to follow-up could impact the analysis.

The COVID-19 Vaccinations analysis demonstrated the issues when using an outcome measured at a different time period to when the data for the original sample was collected. Due to the extent and non-randomness of loss to follow-up (the number of people who left the country or died within the seven years between the sample creation and outcome period), re-weighting approaches need to be considered; however, documentation for existing re-weighting strategies has not been published for the Te Kupenga 2013 sample. Furthermore, causation cannot be applied to this analysis as the sample may be biased due to the number and non-randomness of observations that need to be dropped from the sample due to loss to follow-up. Common epidemiological approaches to attrition and loss to follow-up include multiple imputation and booster sampling to adjust and refresh cohort measurements (Deng et al., 2013; Asendorpf et al., 2014). However, at this early stage of developing longitudinal analysis methods for use with Te Kupenga data, this is not possible due to a lack of sampling frame and weighting documentation that could inform such actions. Longitudinal analysis methods can create robust causal inferences, but this requires robust datasets. This thesis has demonstrated that linkage and longitudinal approaches with Te Kupenga are possible. However, with longer time periods, there can be issues of loss to follow-up – and there is no readily available documentation for researchers to assess this impact or inform possible remediation processes accurately. The approaches developed in this thesis could apply to other official statistics sample surveys, especially those such as the General Social Survey, which have much more detailed and accessible information about their sampling frame and weighting process.

4.4.1 Missing Data due to Loss to follow-up

Creating novel methods for using Te Kupenga 2013 as a longitudinal survey means investigating those who may be dropped from the sample between 2013 and the outcome period of interest. Using the examples in this project, ASH records directly follow from the sample as this outcome is recorded from 2014 onward. In contrast, COVID-19 vaccinations are recorded from 2020 onward, creating a difference of seven years between the original sample and the outcome period. Therefore, it is essential to consider the impacts of the observations missing from the sample across this time. Missing data heavily impacts outcome measurement and analysis, due to forfeiting precision and accuracy. If individuals are dropped from the sample, a smaller sample size is recorded, and overall precision lowers.

Additionally, missing data impacts accuracy, depending on the nature of missingness (e.g. missing at random, missing not at random or missing completely at random) and informs what potential responses could be used to address the issue. In this example, missing data is demonstrated by those who have died or left the country permanently before the outcome in 2020. As this project aims to create a longitudinal study analysis, missing observations may be viewed as a loss to follow-up.

The tables below outline the missing observations from the Te Kupenga 2013 sample due to death or permanent leave of the country. Characteristics about these individuals are represented by sex, age and NZDep18, using unweighted and weighted counts. Note: Tables 4.3, 4.5 and 4.7 all display unweighted counts and are randomly rounded to base 3. Therefore, total values are different across each table due to random rounding rules. However, Tables 4.4, 4.6, 4.8 each display weighted counts with 95% CI's, rounded to base 500.

Table 4.3: Missing observations by sex (Unweighted counts)

Sex	Count (unweighted)
Male	186
Female	180
Total	366

Table 4.4: Missing observations by sex (Weighted counts)

Sex	Count (Weighted)
Male	17,500 (14,500, 20,500)
Female	15,500 (13,000, 18,000)
Total	33,000 (27,500, 38,500)

Table 4.5: Missing observations by 10 year age-bands (Unweighted counts)

10 year age-bands	Count (unweighted)
15-24	21
25-34	66
35-44	24
45-54	27
55-64	66
65-74	63
75 and above	99
Total	366

Table 4.6: Missing observations by 10 year age-bands (Weighted counts)

10 year age-bands	Count (Weighted)
15-24	2,500
	(1,500, 3,500)
25-34	7,500
	(5,500, 9,500)
35-44	2,000
	(1,500, 3,000)
45-54	3,500
	(2,000, 5,000)
55-64	6,000
	(4,500, 7,500)
65-74	4,500
	(3,500, 5,500)
75 and above	7,000
	(5,500, 8,500)
Total	33,000
	(24,000, 42,500)

Table 4.7: Missing observations by NZDep18 (Unweighted counts)

NZDep18	Count (unweighted)
Dep 1&2	9
Dep 3&4	18
Dep 5&6	18
Dep 7&8	18
Dep 9&10	60
missing	246
Total	369

Table 4.8: Missing observations by NZDep18 (Weighted counts)

NZDep18	Count (Weighted)
Dep 1&2	1,000
	(500, 2,000)
Dep 3&4	1,500
	(1,000, 2,500)
Dep 5&6	2,000
	(1,000, 2,500)
Dep 7&8	1,000
	(500, 2,000)
Dep 9&10	5,500
	(3,500, 7,000)
missing	22,000
	(19,000, 24,500)
Total	33,000
	(25,500, 40,500)

Te Kupenga 2013 uses Jackknife weighting to create a representative survey of the Māori population.

Portraying missing data through unweighted and weighted counts allows for impacts to be quantified for the Te Kupenga 2013 sample and the Māori population. The tables above demonstrate the differences in missing observations across sex, age and deprivation characteristics. Tables 4.3 and 4.4 displays that differences between males and females are similar; however, this impact may be more profound when weighting is applied. Age-band differences (see tables 4.5, 4.6) are found to be greater for the ages 75 years and older, which may represent mortality - an age-dependant outcome. An age group that displays higher counts compared to the surrounding age bands is 25-34, which may represent higher rates of overseas migration in this age group. Viewing differences across deprivation reveals that for those in higher deprivation, the 9& 10 quintile, there are higher recorded counts of missing observations compared to other quintiles. Nonetheless, while the missing sample shows differences within each characteristic, the total count of those missing is approximately only 360 individuals.

These results demonstrate that the loss-to-followup using Te Kupenga is not randomly distributed across the sample. It is demonstrably associated with variables included in the dataset, so it may be missing at random, with missingness being able to be addressed via imputation. However, it may also be associated with variables external to the available data, especially as there is no publicly available information on the population it represents. This would make the missing data missing not at random and not resolvable via imputation. One solution could be reweighting the remaining sample, but as discussed previously, researchers do not have access to the information required to make that possible.

It is possible to quantify the characteristics of missing individuals between 2013 and 2020 from the Te Kupenga 2013 sample before dropping these observations. However, creating a new sample by dropping these individuals impacts the Jackknife weighting methods used to create a representative sample. As discussed previously, there is finite documentation released on the weighting methodology, including the sampling frame used for this study. Therefore, reweighting to an updated sample is not possible, given the limited resources available to create updated Jackknife weights. Furthermore, using a sample with missing data means the ability to determine causality is compromised, as the Jackknife weights no longer represent the complete Te Kupenga 2013 sample. This issue emphasises the importance of releasing methodological processes for weighting, indicating the need for improved documentation by StatsNZ.

Given the appropriate information is provided, a possible methodological approach to future re- weighting may consist of:

1. Outcome tables without those who died and/or left the country, but without reweighting
    (a) Tables about those who are missing from sample (characteristics about these individuals: age and gender)
2. Outcome tables: coding those dropped from sample as missing Jackknife weights
    (a) Code those dropped as missing. Vaccination would be yes, no, missing
3. Outcome tables: coding missing Jackknife weights then redistributing their Jackknife weights to similar individuals)

Chapter 5

Recommendations

As outlined in the aims section of this thesis, this project intends to develop robust processes for creating longitudinal analysis capability using sample surveys in the IDI. Although novel methods are created throughout this research, improvements to the IDI could be made to improve and streamline processes for future IDI researchers wanting to apply these methods. The recommendations suggested below align with each step of this process developed in this thesis. Recommendations are categorised by the actions taken when using IDI data, including Information Access, Data Management tools, Data analytic approaches and Data reporting. Within the categories, bullet points provide a concise outline and are described in more detail in the text below.

5.1 Information Access

- How surveys are created: Improved access documentation for Te Kupenga and other surveys. Te Kupenga was created as a Māori social survey and further iterations to be completed (Huirama, 2023), therefore, Te Kupenga documentation needs to be accessible
- How surveys are weighted: Limited documentation about Te Kupenga weighting methods available

Improvements to survey documentation and weighting are recommended to allow a greater understanding of the structure of social statistics survey data sets in New Zealand and how they're created. This project uses Te Kupenga as a foundational cohort due to its value as the only Māori social survey. While Te Kupenga holds valuable information about the Māori population, relevant documentation must reflect this importance in official statistics. Te Kupenga was first created in 2013 as a post-censal survey with further iterations created or planned (Huirama, 2023)). Therefore, resources surrounding its methodology require improved accessibility for researchers. Although the Te Kupenga sample is a representative sample of a sampling frame consisting of individuals in the Census aged 15 or over who report Māori descent or Māori ethnicity. However, no readily available information about this sampling frame was created from both Māori identifiers. Stats NZ does publish information about the Māori ethnic population and the Māori descent population, but not the population defined in the same way as the Te Kupenga sampling frame. Therefore, users of linked Te Kupenga data cannot assess any potential linkage or missing data biases against the base population sampling frame.

Furthermore, there needs to be more documentation about the Jackknife weighting methodology for Te Kupenga 2013. While Stats NZ holds concerns about confidentiality risks when publishing details of the sampling process, this creates difficulty when utilising Te Kupenga for longitudinal purposes. Throughout this thesis, issues with missing data were discussed when individuals were dropped from the cohort. These issues create knock-on effects for any potential re-weighting of the remaining sample, alongside a need for more information about the original sampling frame. Although re-weighting is possible for a new sample, documentation about the original sampling frame process is essential

to creating appropriate adjusted weights. While maintaining the confidentiality of individuals is a priority, weighting methodology and sampling frame information could be available within the secure IDI environment where confidentiality is protected. This revision allows researchers to apply suitable techniques when using Te Kupenga and other sample surveys without compromising the privacy of individuals in the dataset.

During the course of this thesis, the manager of Stats NZ surveys was contacted for additional information about the methodological processes used to create Te Kupenga and any relevant documentation. However, no written material was available to be provided on this request. Although a formal request can be made under the Official Information Act, this should not be necessary, as general descriptors about sampling and weighting strategy should be routinely available from Stats NZ. This information is required to provide enough detail to understand how the Te Kupenga sample was created and enable the re-weighting of this nationally representative sample.

5.2 Data management tools

- Stats NZ creating standard but modifiable code, which is applicable to official sample surveys.
- Stats NZ creating standard linking code with census to standard derived variables: For example, from individual census level, to household level measures.
- Stats NZ creating standard linked file between Census datasets (e.g. 2013 and 2018) and between each Te Kupenga datasets and the Census from which it was sampled.
- Stats NZ creating consistency in variables across datasets: For example, this project found that Household Crowding and NZDep only available in Te Kupenga 2013, but not Census 2013, despite Te Kupenga survey variables derived from the Census.
- Stats NZ required to outline the limitations of software available in the IDI: Software consistency and their relevant restrictions

Resources for data management are recommended when implementing novel methods, as outlined in this thesis. A set of standard linking code, linked files and modifiable code libraries by Stats NZ would greatly benefit working with IDI datasets in the future. Due to the complex level of coding needed for linking datasets within the IDI, a resource of standard linking code would reduce coding inconsistencies across users. An example includes creating standard linking code from individual census-level measures to household-level measures - a common approach to using Census data. As researchers often require multiple levels of Census variables (e.g. individual to household), offering standard linking code reduces the need for bespoke data linkage and lowers the risk of differing output between researchers due to incorrect methodology or coding errors. This approach also allows Stats NZ to maintain greater control of output produced in the IDI by conducting functionality tests on the standard code resource.

Providing standard files that link between datasets would also be beneficial for future longitudinal research when using Te Kupenga in the IDI. For example, creating a standard file linking Census 2013 and Te Kupenga 2013 datasets and between Census 2018 and Te Kupenga 2018. This reduces potential issues when conducting bespoke linkages. An additional example is demonstrated when using Te Kupenga and COVID-19 Vaccination data, which requires the added step of linking from the IDI ad-hoc tables to the 'security concordance' table to ensure consistency of snz_uid's across refresh iterations. If this step is neglected, individuals may be incorrectly linked across datasets in the IDI. As a regularly completed linkage across datasets, standard linking files would decrease the likelihood of such potential errors. Furthermore, this resource allows for Stats NZ to complete further checks across datasets to establish variable consistencies. As discussed earlier in this thesis, household crowding and NZDep variables were found for Te Kupenga 2013 only, but not Census 2013. If a standard linking file between Census 2013 and Te Kupenga were to exist, inconsistencies would be discovered and addressed before being released in the IDI.

This project undertakes complex steps when extracting and merging IDI datasets, proving difficult when working with statistical software available in the IDI environment. This included challenges with extracting datasets (mainly using SAS), given issues with lengths of filenames in the IDI. Furthermore, utilising software such as R as operationalised in the IDI environment proves burdensome when working with large datasets - for example, the Te Kupenga dataset with $>$ 5,000 rows and $>$ 100 columns with Jackknife replicate weights, as this slows R substantially. Although not at the fault of Stats NZ, and instead a restriction with the available software, the possibility of encountering these issues needs to be realised by Stats NZ. Outlining such limitations will decrease the additional and potentially unavoidable steps needed when working with IDI data.

5.3 Data analytic approaches

- Stats NZ creating standard data quality measures: As novel approaches are created by researchers, Stats NZ could validate these methods and be accessible for future reference.
- Stats NZ creating a list of recommended statistical approaches useful for particular projects (for example, utilising negative binomial regression vs. poisson regression, or ordered logistic regression vs. multinomial logistic regression)

The following stage of the process relates to recommendations surrounding analytic tools and approaches by implementing data quality measures and statistical approaches by Stats NZ. As researchers develop novel approaches to using IDI data, Stats NZ does not maintain engagement with researchers throughout all the stages of the research project; therefore, methods and approaches applied by IDI users are not validated by Stats NZ. Similar to the data management stage - user error may occur during analysis by not using the appropriate statistical techniques. However, Stats NZ has the opportunity to combat this issue by developing a validation stage of output checking. For a particular scenario, the appropriate analysis may be suggested at the beginning of a research project once outcome variables are decided. When this situation occurs, researchers can consult Stats NZ about their options. Furthermore, this information can be archived for future reference and made accessible to researchers for reuse, gradually building up a body of 'best practice' approaches to using IDI resources.

Researchers who thoroughly analyse datasets in the IDI would benefit from recommended statistical approaches when completing their projects. Across both outcomes for this project, statistical checks were implemented to check for the appropriate type of analysis. The Count of Unique ASH Events and Length of Stay displayed evidence of overdispersion; therefore, negative binomial logistic regressions were completed. While COVID-19 vaccinations are an ordered outcome, an ordinal logistic regression was optimal. If such decisions had not been implemented, results may have been misinterpreted. However, this project has proven to review health outcomes appropriately while adjusting for overdispersion and ordering within the response variables. Without prior statistical knowledge, IDI users may not have an understanding of this area and are led into using incorrect approaches. Therefore, having recommended approaches routinely archived and made available by Stats NZ would assist researchers and create reliable methods for conducting an appropriate statistical analysis.

5.4 Data reporting

- Stats NZ storing and archiving: once output is produced by researchers, this should be archived and validated by Stats NZ
- Stats NZ utilising a published code library for researcher use

The final stage of utilising Te Kupenga for longitudinal analysis within the IDI involves methods for storing and archiving output produced from IDI projects. Users of the IDI produce output for research projects and ongoing research in areas related to New Zealand's administrative data. Like the recommendations for analytic tools and approaches, Stats NZ needs to create improved approaches at

the final stages of output production. When IDI-based researchers create output, a system that stores and archives code, methodology and processes used in projects could improve future IDI research. A validation process is also recommended throughout this stage to ensure that the correct approaches are utilised. This also creates the opportunity for Stats NZ to demonstrate reliable methodology for reuse in future projects.

The IDI user commons is a recently established resource by Stats NZ that allows researchers to share code and recommendations when working with particular datasets in the IDI. Stats NZ could better utilise this resource to create a code library for standard but modifiable code applicable to different IDI surveys. Where new methods are created, the standard code researchers implement should be published by Stats NZ as a guide for using IDI data. The IDI search tool (Elliott et al., 2022) is a resource for searching through the list of IDI datasets available and can be utilised outside the IDI environment. This demonstrates an opportunity for a similar library searchable by variable name, where the base code is stored and can be accessed internally or externally to the IDI. The current IDI user commons acts as a voluntary code repository, but the IDI search engine has displayed the impact and usefulness of a standard user approach. Therefore, if Stats NZ adopts the same approach to archiving user code, this adds to commitments for producing methods for the public good and improves researcher capability.

5.5 Overview of recommendations

Information Access

- How surveys are created: Improved access documentation for Te Kupenga and other surveys. Te Kupenga was created as a Māori social survey and further iterations to be completed (Huirama, 2023), therefore, Te Kupenga documentation needs to be accessible
- How surveys are weighted: Limited documentation about Te Kupenga weighting methods available

Data management tools

- Stats NZ creating standard but modifiable code, which is applicable to official sample surveys.
- Stats NZ creating standard linking code with census to standard derived variables: For example, from individual census level, to household level measures.
- Stats NZ creating standard linked file between Census datasets (e.g. 2013 and 2018) and between each Te Kupenga datasets and the Census from which it was sampled.
- Stats NZ creating consistency in variables across datasets: For example, this project found that Household Crowding and NZDep only available in Te Kupenga 2013, but not Census 2013, despite Te Kupenga survey variables derived from the Census.
- Stats NZ required to outline the limitations of software available in the IDI: Software consistency and their relevant restrictions

Data analytic approaches

- Stats NZ creating standard data quality measures: As novel approaches are created by researchers, Stats NZ could validate these methods and be accessible for future reference.
- Stats NZ creating a list of recommended statistical approaches useful for particular projects (for example, utilising negative binomial regression vs. poisson regression, or ordered logistic regression vs. multinomial logistic regression)

Data reporting

- Stats NZ storing and archiving: once output is produced by researchers, this should be archived and validated by Stats NZ
- Stats NZ utilising a published code library for researcher use

Bibliography

Agresti, A. (2013). *Categorical data analysis*. Wiley Series in Probability and Statistics Series. John Wiley & Sons, Incorporated.

Allan, J. (2001). Review of the measurement of ethnicity classification and issues. https://www.stats.govt.nz/methods/review-of-the-measurement-of-ethnicity-classifications-and-issues/.

Asendorpf, J. B., van de Schoot, R., Denissen, J. J. A., and Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, 38(5):453–460.

Atkinson, J., Salmond, C., and Crampton, P. (2019). NZDep2018 Index of Deprivation.

Barker, C., Crengle, S., Bramley, D., Bartholomew, K., Walsh, M., and Wignall, J. (2016). Pathways to ambulatory sensitive hospitalisations for Māori in the Auckland and Waitemata regions. *New Zealand Medical Journal*, 129(1444):22.

Black, A. (2016). *The IDI prototype spine's creation and coverage*. Statistics New Zealand, Wellington, New Zealand. OCLC: 993615665.

Broughton, J., Fergusson, D. M., Rimene, C., Horwood, J. L., and Sporle, A. (2000). Nga tini aho o te ao hou. the many strands of contemporary maori society: Maori ethnicity and identity in the christchurch health and development study. Dunedin, NZ: The Ngai Tahu Maori Health Research Unit (NTMHRU).

Canada Mortgage and Housing Corporation (CMHC) (2022). National Occupancy Standard. https://www.cmhc-schl.gc.ca/professionals/industry-innovation-and-leadership/industry-expertise/affordable-housing/provincial-territorial-agreements/investment-in-affordable-housing/national-occupancy-standard.

Casella, G. and Berger, R. L. (2001). *Statistical Inference*. Duxbury, 2nd edition.

Cormack, D., Reid, P., and Kukutai, T. (2019). Indigenous data and health: critical approaches to 'race'/ethnicity and indigenous data governance. *Public health*, 172:116–118. ID: 272854.

Cormack, D. and Robson, C. (2010). Classification and output of multiple ethnicities: issues for monitoring māori health. https://www.fmhs.auckland.ac.nz/assets/fmhs/Te%20Kupenga%20Hauora%20M%C4%81ori/docs/classification.pdf.

Deng, Y., Hillygus, D. S., Reiter, J. P., Si, Y., and Zheng, S. (2013). Handling Attrition in Longitudinal Studies: The Case for Refreshment Samples. *Statistical Science*, 28(2).

Elliott, T., Milne, B., Li, E., Simpson, C., and Sporle, A. (2022). Idi search app. https://terourou.org/idisearch/.

Fawcett, J., Blakely, T., and Atkinson, J. (2002). *Weighting the 81, 86, 91 & 96 census-mortality cohorts to adjust for linkage bias*. Dept. of Public Health, Wellington School of Medicine and Health Sciences, Wellington, N.Z. OCLC: 156174150.

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2011). *Applied Longitudinal Analysis*, volume 998 of *Wiley Series in Probability and Statistics Series*. John Wiley & Sons, Incorporated, 2 edition.

Galobardes, B. (2006a). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health*, 60(1):7–12.

Galobardes, B. (2006b). Indicators of socioeconomic position (part 2). *Journal of Epidemiology & Community Health*, 60(2):95–101.

Gleisner, F., Downey, A., and McNally, J. (2015). *Enduring census information requirements for and about Māori*. https://www.stats.govt.nz/assets/Research/Enduring-census-information-requirements-for-and-about-Maori/enduring-census-information-requirements-for-and-about-maori.pdf.

Greaves, L. M., Latimera, C. L., Emerald Muriwai, C. M., Li, E., Sporle, A., Clark, T. C., and Milne, B. J. (2023). Māori and the integrated data infrastructure: an assessment of the data system and suggestions to realise māori data aspirations [te māori me te integrated data infrastructure: he aromatawai i te pūnaha raraunga me ngā marohitanga e poipoia ai ngā wawata raraunga māori]. *Journal of the Royal Society of New Zealand*, 0(0):1–17.

Hernan, M. A. and Robins, J. M. (2020). *Causal Inference: What If*. Chapman & HallCRC.

Huirama, M. (2023). Revisiting Te Kupenga - A National Survey of Māori Wellbeing. StatsNZ.

Jack, M. and Graziadei, C. (2019). *Report of the independent review of New Zealand's 2018 census*. https://www.stats.govt.nz/reports/report-of-the-independent-review-of-new-zealands-2018-census.

Jackson, G. and Tobias, M. (2001). Potentially avoidable hospitalisations in New Zealand, 1989–98. *Australian and New Zealand Journal of Public Health*, 25(3):212–221.

Kvalsvig, A., Gibb, S., and Teng, A. (2019). Linkage error and linkage bias: A guide for IDI users. https://vhin.co.nz/wp-content/uploads/2019/11/Linkage-error-and-linkage-bias.pdf.

Li, C. (2023). Number of observations (māori by ethnicity and/or descent) in initial joining.

Milne, B. (2023). Analysis lecture 1 - stats240, university of auckland.

Milne, B. J., Atkinson, J., Blakely, T., Day, H., Douwes, J., Gibb, S., Nicolson, M., Shackleton, N., Sporle, A., and Teng, A. (2019). Data Resource Profile: The New Zealand Integrated Data Infrastructure (IDI). *International Journal of Epidemiology*, 48(3):677–677e. _eprint: https://academic.oup.com/ije/article-pdf/48/3/677/29005185/dyz014.pdf.

Ministry of Health - Manatū Hauora (2022). National minimum dataset (hospital events). https://www.tewhatuora.govt.nz/our-health-system/data-and-statistics/nz-health-statistics/national-collections-and-surveys/collections/national-minimum-dataset-hospital-events.

Ministry of Health - Manatū Hauora (2023). Ambulatory sensitive (avoidable) hospital admissions. https://www.health.govt.nz/new-zealand-health-system/accountability-and-funding/planning-and-performance-data/ambulatory-sensitive-avoidable-hospital-admissions.

Ministry of Health Manatū Hauora (2020). *Immunisation handbook 2020*. Ministry of Health Manatū Hauora, Wellington, New Zealand, 2020 edition edition. OCLC: 1230484041.

Morton, S. M. B., Carr, P. E. A., Grant, C. C., Robinson, E. M., Bandara, D. K., Bird, A., Ivory, V. C., Kingi, T. K. R., Liang, R., Marks, E. J., Perese, L. M., Peterson, E. R., Pryor, J. E., Reese, E., Schmidt, J. M., Waldie, K. E., and Wall, C. (2012). Cohort Profile: Growing Up in New Zealand. *International Journal of Epidemiology*, 42(1):65–75. _eprint: https://academic.oup.com/ije/article-pdf/42/1/65/18481828/dyr206.pdf.

Nicholson Consulting & Kōtātā Insight (2021). He ara poutama mō te reo māori. Wellington, New Zealand.

Palapar, L., Wilkinson-Meyers, L., Lumley, T., and Kerse, N. (2020). GP- and practice-related variation in ambulatory sensitive hospitalisations of older primary care patients. *BMC Family Practice*, 21(1):217.

Rothman, K. J. (2021). *Modern epidemiology*. Philadelphia : Wolters Kluwer 2021, Philadelphia, 4th ed.. edition.

Satherley, N., Diamond, T., and Sporle, A. (2023). More than just living in a deprived area: An equity-focused analysis of policy amenable factors associated with māori covid-19 outcomes. iNZight Analytics Ltd. The National Hauora Coalition. Tāmaki Makaurau.

Sheridan, N., Love, T., Kenealy, T., The Primary Care Models Study Group, Aguirre-Duarte, N., Arroll, B., Atmore, C., Carryer, J., Crampton, P., Dowell, A., Fishman, T., Gauld, R., Harwood, M., Hoare, K., Jackson, G., Jansen, R. M., Kerse, N., Lampshire, D., McBain, L., MacRae, J., Mills, J., Øvretveit, J., Percival, T., Perera, R., Roland, M., Ryan, D., Schmidt-Busby, J., Stokes, T., Stubbe, M., Hewitt, S., Watt, D., and Peck, C. (2023). Is there equity of patient health outcomes across models of general practice in Aotearoa New Zealand? A national cross-sectional study. *International Journal for Equity in Health*, 22(1):79.

Singh, H., Papaconstantinou, D., and Jackson, G. (2021). Ambulatory Sensitive Hospitalisations in Counties Manukau Health 2020 update. https://www.countiesmanukau.health.nz/assets/About-CMH/Performance-and-planning/health-status/2021_-ambulatory-sensitive-hospitalisations_adults.pdf.

Statistical Methods and Data Analytics (2023). Ordered logistic regression — stata data analysis examples. https://stats.oarc.ucla.edu/stata/dae/ordered-logistic-regression/.

Stats NZ (2014). Te kupenga 2013 (english). https://www.stats.govt.nz/information-releases/te-kupenga-2013-english.

Stats NZ (2017). *Improving New Zealand disability data*. https://www.stats.govt.nz/assets/Reports/Improving-New-Zealand-disability-data/improving-new-zealand-disability-data.pdf.

Stats NZ (2018). *Differences between Te Kupenga 2013 and 2018 surveys*. https://www.stats.govt.nz/methods/differences-between-te-kupenga-2013-and-2018-surveys.

Stats NZ (2019a). Census grouped family or household income v2.0.0. https://aria.stats.govt.nz/aria/#ClassificationView:uri=http://stats.govt.nz/cms/ClassificationVersion/EC7A1NgYznVggNM7.

Stats NZ (2019b). *Data sources, editing, and imputation in the 2018 Census*. https://www.stats.govt.nz/methods/data-sources-editing-and-imputation-in-the-2018-census/.

Stats NZ (2020a). *Assessment of potential bias in the Te Kupenga sample frame: 2018*. https://www.stats.govt.nz/assets/Uploads/Assessment-of-potential-bias-in-the-Te-Kupenga-sample-frame-2018-.pdf.

Stats NZ (2020b). *Microdata output guide (fifth edition)*. https://www.stats.govt.nz/assets/Methods/Microdata-Output-Guide-2020-v5-Sept22update.pdf.

Stats NZ (2021). Te Kupenga data collection 2018. https://datainfoplus.stats.govt.nz/Item/nz.govt.stats/c606e95d-3003-478b-9ece-685a8de8357d/9.

Stats NZ (2022). Integrated data infrastructure. https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/#about.

Stats NZ (2023). Data in the idi. `https://www.stats.govt.nz/integrated-data/integrated-data-infrastructure/data-in-the-idi/`.

Steyn, N., Binny, R. N., Hannah, K., Hendy, S. C., James, A., Kukutai, T., Lustig, A., McLeod, M., Plank, M. J., Ridings, K., and Sporle, A. (2020). Estimated inequities in covid-19 infection fatality rates by ethnicity for aotearoa new zealand. *The New Zealand Medical Journal (Online)*, 133(1521):28.

Te Kāhui Raraunga Charitable Trust (TKR) (2023). Te whata. `https://tewhata.io/`.

Virtual Health Information Network (2020). Understanding linkage in the idi. `https://vhin.co.nz/guides/understanding-linkage-in-the-idi/`.

Washington Group on Disability Statistics (2017). Washington group short set on functioning (wg-ss). `https://www.washingtongroup-disability.com/question-sets/wg-short-set-on-functioning-wg-ss/`.

Appendix A

Appendix: Tables

Table A.1: Ambulatory Sensitive Hospitalisations - ICD 10 Codes **(Continued on the following five pages)**

ASH Chapter	ASH Condition	Diagnosis Code	Diagnosis Description	Applicable Ages
Cardiovascular	Angina and chest pain	R072	Precordial pain	15+
		R073	Other chest pain	15+
		R074	Chest pain, unspecified	15+
		I20	Angina pectoris	15+
	Congestive heart failure	I50	Heart failure	15+
		J81	Pulmonary oedema	15+
	Hypertensive disease	I10	Essential (primary) hypertension	15+
		I11	Hypertensive heart disease	15+
		I12	Hypertensive kidney disease	15+
		I13	Hypertensive heart and kidney disease	15+
		I15	Secondary hypertension	15+
		I674	Hypertensive encephalopathy	15+
	Myocardial infarction	I21	Acute myocardial infarction	15+
		I22	Subsequent myocardial infarction	15+
		I23	Certain current complications following acute myocardial infarction	15+
		I241	Dressler's syndrome	15+
	Other ischaemic heart disease	I240	Coronary thrombosis not resulting in myocardial infarction	15+
		I248	Other forms of acute ischaemic heart disease	15+
		I249	Acute ischaemic heart disease, unspecified	15+
		I25	Chronic ischaemic heart disease	15+
	Rheumatic fever/heart disease	I00	Rheumatic fever without mention of heart involvement	All
		I01	Rheumatic fever with heart involvement	All
		I02	Rheumatic chorea	All
		I05	Rheumatic mitral valve diseases	All
		I06	Rheumatic aortic valve diseases	All
		I07	Rheumatic tricuspid valve diseases	All
		I08	Multiple valve diseases	All

		I09	Other rheumatic heart diseases	All
Dental	Dental conditions	K02	Dental caries	All
		K04	Diseases of pulp and periapical tissues	All
		K05	Gingivitis and periodontal diseases	All
Dermatological	Cellulitis	L01	Impetigo	All
		L02	Cutaneous abscess, furuncle and carbuncle	All
		L03	Cellulitis	All
		L04	Acute lymphadenitis	All
		L08	Other local infections of skin and subcutaneous tissue	All
		H000	Hordeolum and other deep inflammation of eyelid	All
		H010	Blepharitis	All
		J340	Abscess, furuncle and carbuncle of nose	All
		L980	Pyogenic granuloma	All
	Dermatitis and eczema	L20	Atopic dermatitis	All
		L21	Seborrhoeic dermatitis	All
		L22	Diaper [napkin] dermatitis	All
		L23	Allergic contact dermatitis	All
		L24	Irritant contact dermatitis	All
		L25	Unspecified contact dermatitis	All
		L26	Exfoliative dermatitis	All
		L27	Dermatitis due to substances taken internally	All
		L28	Lichen simplex chronicus and prurigo	All
		L29	Pruritus	All
		L30	Other dermatitis	All
Gastrointestinal	Constipation	K590	Constipation	All
	Gastroenteritis/dehydration	A02	Other salmonella infections	All
		A03	Shigellosis	All
		A04	Other bacterial intestinal infections	All
		A05	Other bacterial food-borne intoxications, not elsewhere classified	All
		A06	Amoebiasis	All
		A07	Other protozoal intestinal diseases	All
		A08	Viral and other specified intestinal infections	All
		A09	Other gastroenteritis and colitis of infectious and unspecified origin	All

		R11	Nausea and vomiting	All
		K529	Noninfective gastroenteritis and colitis, unspecified	All
	GORD (Gastro-oesophageal reflux disease)	K21	Gastro-oesophageal reflux disease	All
	Nutrition deficiency and anaemia	D50	Iron deficiency anaemia	All
		D51	Vitamin B12 deficiency anaemia	All
		D52	Folate deficiency anaemia	All
		D53	Other nutritional anaemias	All
		E40	Kwashiorkor	All
		E41	Nutritional marasmus	All
		E42	Marasmic kwashiorkor	All
		E43	Unspecified severe protein-energy malnutrition	All
		E44	Protein-energy malnutrition of moderate and mild degree	All
		E45	Retarded development following protein-energy malnutrition	All
		E46	Unspecified protein-energy malnutrition	All
		E50	Vitamin A deficiency	All
		E51	Thiamine deficiency	All
		E52	Niacin deficiency [pellagra]	All
		E53	Deficiency of other B group vitamins	All
		E54	Ascorbic acid deficiency	All
		E55	Vitamin D deficiency	All
		E56	Other vitamin deficiencies	All
		E58	Dietary calcium deficiency	All
		E59	Dietary selenium deficiency	All
		E60	Dietary zinc deficiency	All
		E61	Deficiency of other nutrient elements	All
		E63	Other nutritional deficiencies	All
		M833	Adult osteomalacia due to malnutrition	15+
	Peptic ulcer	K25	Gastric ulcer	15+
		K26	Duodenal ulcer	15+
		K27	Peptic ulcer, site unspecified	15+
		K28	Gastrojejunal ulcer	15+
Respiratory	Asthma	J45	Asthma	All

		J46	Status asthmaticus	All
	Bronchiectasis	J47	Bronchiectasis	15+
	COPD	J44	Chronic obstructive pulmonary disease	15+
	Pneumonia	J13	Pneumonia due to Streptococcus pneumoniae	All
		J14	Pneumonia due to Haemophilus influenzae	All
		J15	Bacterial pneumonia, not elsewhere classified	All
		J16	Pneumonia due to other infectious organisms, not elsewhere classified	All
		J18	Pneumonia, organism unspecified	All
	Upper and ENT respiratory infections	J00	Acute nasopharyngitis [common cold]	All
		J01	Acute sinusitis	All
		J02	Acute pharyngitis	All
		J03	Acute tonsillitis	All
		J04	Acute laryngitis and tracheitis	All
		J06	Acute upper respiratory infections of multiple and unspecified sites	All
		H65	Nonsuppurative otitis media	All
		H66	Suppurative and unspecified otitis media	All
		H67	Otitis media in diseases classified elsewhere	All
Other	Cervical cancer	C53	Malignant neoplasm of cervix uteri	15+
	Diabetes	E10	Type 1 diabetes mellitus	15+
		E11	Type 2 diabetes mellitus	15+
		E13	Other specified diabetes mellitus	15+
		E14	Unspecified diabetes mellitus	15+
		E162	Hypoglycaemia, unspecified	15+
	Epilepsy	G40	Epilepsy	15+
		G41	Status epilepticus	15+
		O15	Eclampsia	15+
		R560	Febrile convulsions	15+
		R568	Other and unspecified convulsions	15+
	Kidney/urinary infection	N10	Acute tubulo-interstitial nephritis	5+
		N12	Tubulo-interstitial nephritis, not specified as acute or chronic	5+
		N136	Pyonephrosis	5+

		N309	Cystitis, unspecified	5+
		N390	Urinary tract infection, site not specified	5+
	Sexually transmitted infections	A50	Congenital syphilis	15+
		A51	Early syphilis	15+
		A52	Late syphilis	15+
		A53	Other and unspecified syphilis	15+
		A54	Gonococcal infection	15+
		A55	Chlamydial lymphogranuloma (venereum)	15+
		A56	Other sexually transmitted chlamydial diseases	15+
		A57	Chancroid	15+
		A58	Granuloma inguinale	15+
		A59	Trichomoniasis	15+
		A60	Anogenital herpesviral [herpes simplex] infection	15+
		A63	Other predominantly sexually transmitted diseases, not elsewhere classified	15+
		A64	Unspecified sexually transmitted disease	15+
		M023	Reiter's disease	15+
		N341	Nonspecific urethritis	15+
	Stroke	I61	Intracerebral haemorrhage	15+
		I63	Cerebral infarction	15+
		I64	Stroke, not specified as haemorrhage or infarction	15+
		I65	Occlusion and stenosis of precerebral arteries, not resulting in cerebral infarction	15+
		I66	Occlusion and stenosis of cerebral arteries, not resulting in cerebral infarction	15+

A.1 National level - Māori only

A.1.1 Overdispersion tables - ASH Length of Stay & ASH count of Unique Events

Table A.2: Checking for over-dispersion in Māori Census 2013 only sample: ASH Length of Stay by Highest Qualification

Highest Qualification	Mean	Variance	N
No Qualification	4.9	787	137,085
School-level	2.2	410.3	159,426
Post-school	2.5	233.1	88,776
Degree Level	1.7	124.7	45,144
Total	3.1	465.4	430,434

Table A.3: Checking for over-dispersion in Māori Census 2013 only sample: ASH Length of Stay by Individual Income

Individual Income (Census 2013)	Mean	Variance	N
$1 - $15,000	4.1	734.2	110,901
$15,001 - $30,000	5.7	1002	114,519
$30,001 - $50,000	2.1	258.8	81,618
$50,001 - $100,000	1.9	106.6	72,066
$100,001 and above	1.4	59.2	9,708
Total	3.7	582.6	388,815

Table A.4: Checking for over-dispersion in Māori Census 2013 only sample: ASH count of Unique ASH Events by Highest Qualification

Highest Qualification	Mean	Variance	N
No Qualification	1	8.5	137,085
School-level	0.5	3.6	159,426
Post-school	0.6	4.5	88,776
Degree Level	0.4	2.3	45,144
Total	0.7	5.2	430,434

Table A.5: Checking for over-dispersion in Māori Census 2013 only sample: ASH count of Unique ASH Events by Individual Income

Individual Income (Census 2013)	Mean	Variance	N
$1 - $15,000	0.9	8.5	110,901
$15,001 - $30,000	1.1	8.5	114,519
$30,001 - $50,000	0.6	3.5	81,618
$50,001 - $100,000	0.5	2.9	72,066
$100,001 and above	0.4	1.7	9,708
Total	0.8	6.3	388,815

A.2 Te Kupenga 2013 sample

A.2.1 ASH Binary

Highest Qualification

Table A.6: ASH binary by Census 2013 Highest Qualification (weighted counts) with 95% CI's

ASH indicator	Highest Qualification	Count	Lower 95% CI	Upper 95% CI
ASH Event				
	No Qualifications	41000	36500	45000
	School-leaver	38000	33500	42500
	Post-school	22500	19500	25500
	Degree Qualification	10500	8500	13000
	missing	11500	9000	13500
	Total	123500	107000	139500
No ASH Event				
	No Qualifications	99500	93500	105500
	School-leaver	148500	141500	156000
	Post-school	79500	73500	86000
	Degree Qualification	43500	39500	47500
	missing	26000	22500	29500
	Total	397000	370500	424500
Grand Total		520500	477500	564000

Figure A.1: ASH binary by Highest Qualification (weighted counts) with 95% CI's

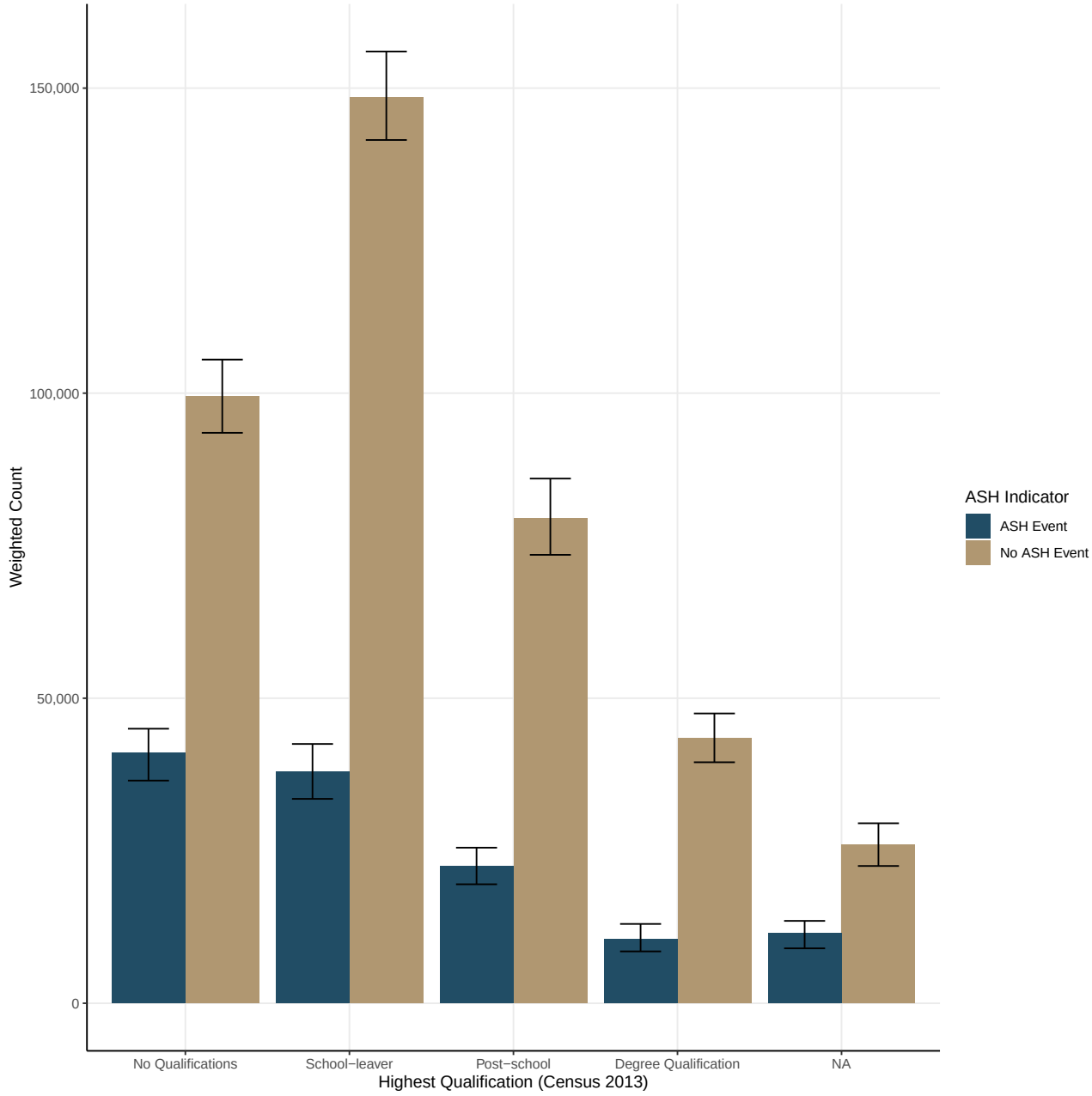


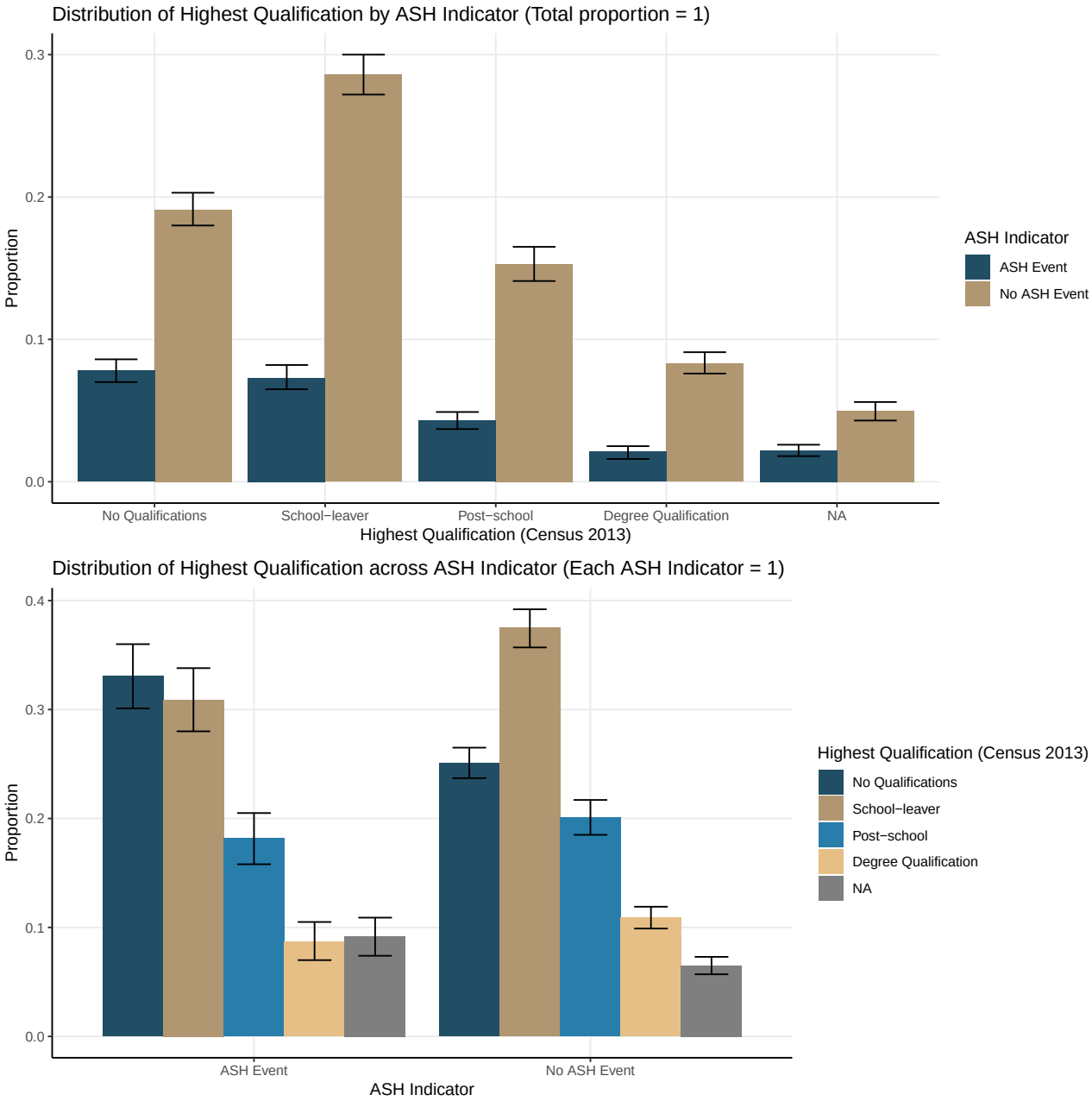
Table A.7: ASH binary by by Census 2013 Highest Qualification (cell proportions)

Highest Qualification	ASH Event	No ASH Event	Grand Total
No Qualifications	0.079	0.191	0.270
School-leaver	0.073	0.285	0.358
Post-school	0.043	0.153	0.196
Degree Qualification	0.020	0.084	0.104
missing	0.022	0.050	0.072
Grand Total	0.237	0.763	1.000

Table A.8: ASH binary by by Census 2013 Highest Qualification (proportions)

Highest Qualification	ASH Event	No ASH Event	Grand Total
No Qualifications	0.332	0.251	0.270
School-leaver	0.308	0.374	0.358
Post-school	0.182	0.200	0.196
Degree Qualification	0.085	0.110	0.104
missing	0.093	0.065	0.072
Grand Total	1.000	1.000	1.000

Figure A.2: ASH binary by Highest Qualification (proportions)



IRD individual income

Table A.9: ASH binary by 2013 IRD Individual income (weighted counts) with 95% CI's

ASH indicator	Individual income	Count	Lower 95% CI	Upper 95% CI
ASH Event				
	$1 - $15	31500	28000	35000
	$15,001 - $30,000	35500	32000	39000
	$30,001 - $50,000	19500	17000	22000
	$50,001 - $100,000	19000	16500	21500
	$100,001 and above	1500	500	2000
	missing	16500	13500	19500
	Total	123500	107500	139000
No ASH Event				
	$1 - $15	83500	77000	89500
	$15,001 - $30,000	83500	76500	90000
	$30,001 - $50,000	70500	65000	75500
	$50,001 - $100,000	73500	68000	79000
	$100,001 and above	10000	8000	12000
	missing	76500	71000	82000
	Total	397500	365500	428000
Grand Total		521000	473000	567000

Figure A.3: ASH binary by 2013 IRD Individual Income (weighted counts) with 95% CI's

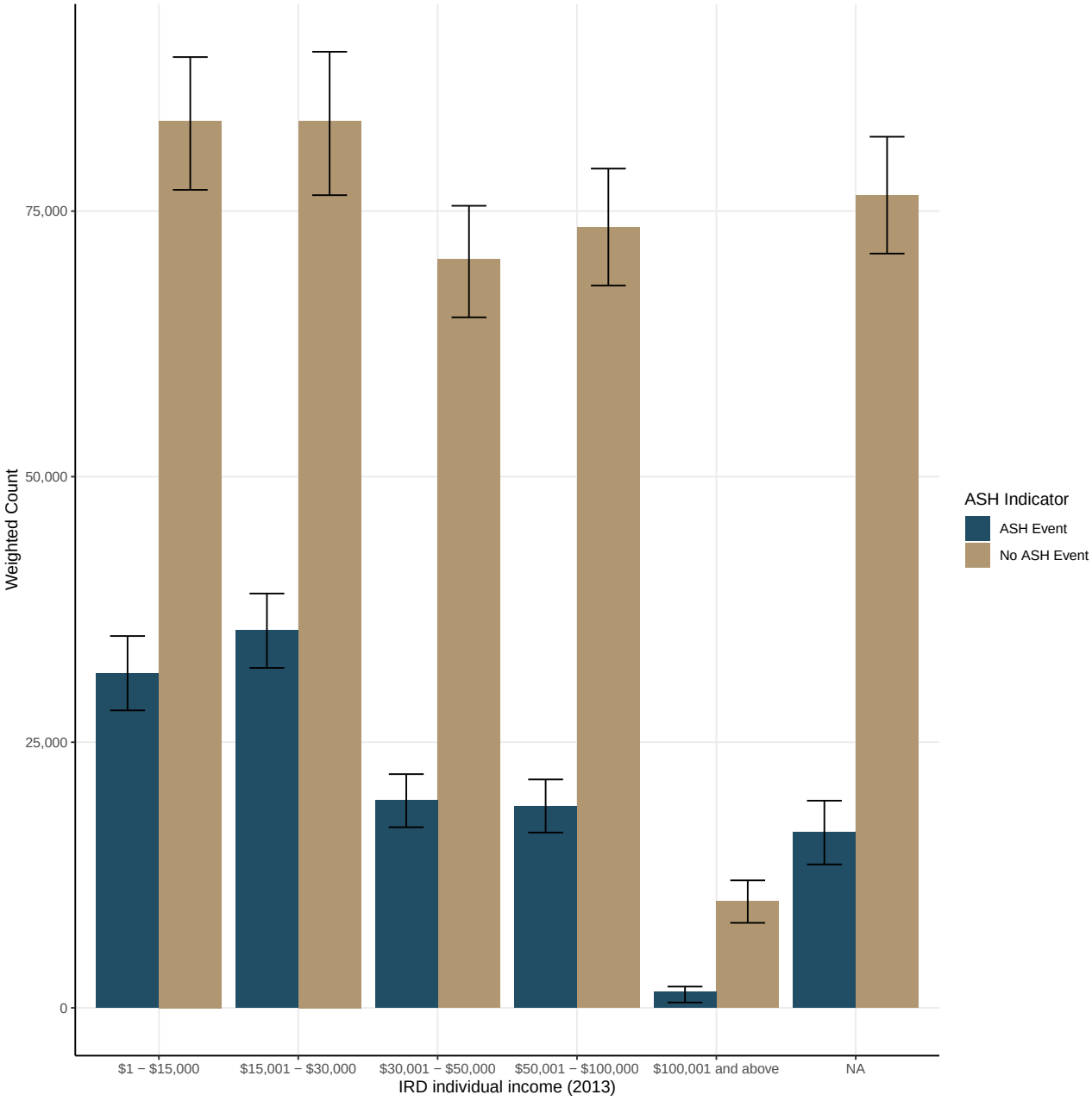


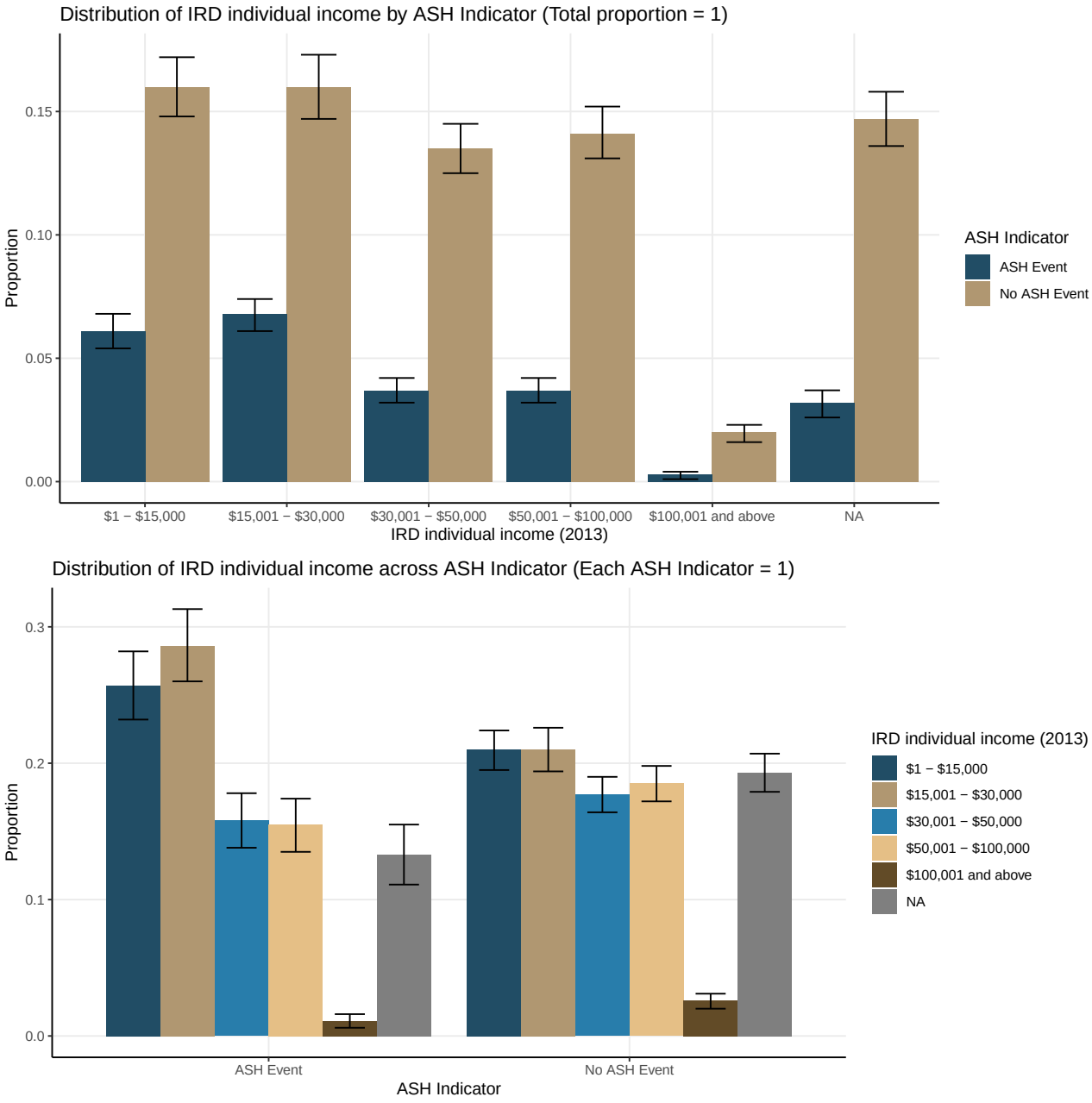
Table A.10: ASH binary by 2013 IRD individual income (cell proportions)

Individual income	ASH Event	No ASH Event	Grand Total
$1 - $15	0.060	0.160	0.221
$15,001 - $30,000	0.068	0.160	0.228
$30,001 - $50,000	0.037	0.135	0.173
$50,001 - $100,000	0.036	0.141	0.178
$100,001 and above	0.003	0.019	0.022
missing	0.032	0.147	0.179
Grand Total	0.237	0.763	1.000

Table A.11: ASH binary by 2013 IRD individual income (proportions)

Individual income	ASH Event	No ASH Event	Grand Total
$1 - $15	0.255	0.210	0.221
$15,001 - $30,000	0.287	0.210	0.228
$30,001 - $50,000	0.158	0.177	0.173
$50,001 - $100,000	0.154	0.185	0.178
$100,001 and above	0.012	0.025	0.022
missing	0.134	0.192	0.179
Grand Total	1.000	1.000	1.000

Figure A.4: ASH binary by 2013 IRD Individual income (proportions)



Household Crowding

Figure A.5: ASH binary by 2013 Household Crowding (weighted counts) with 95% CI's

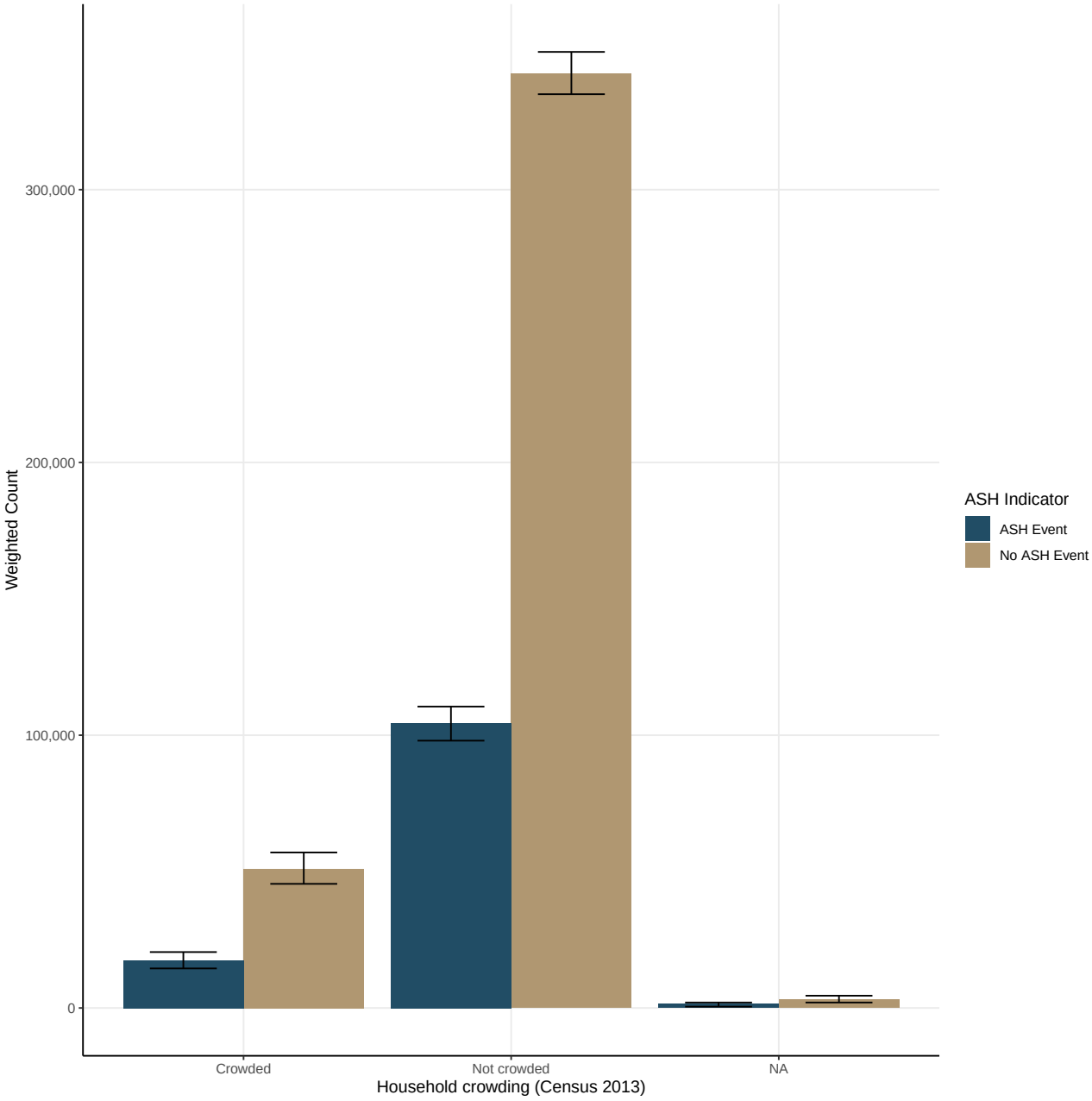
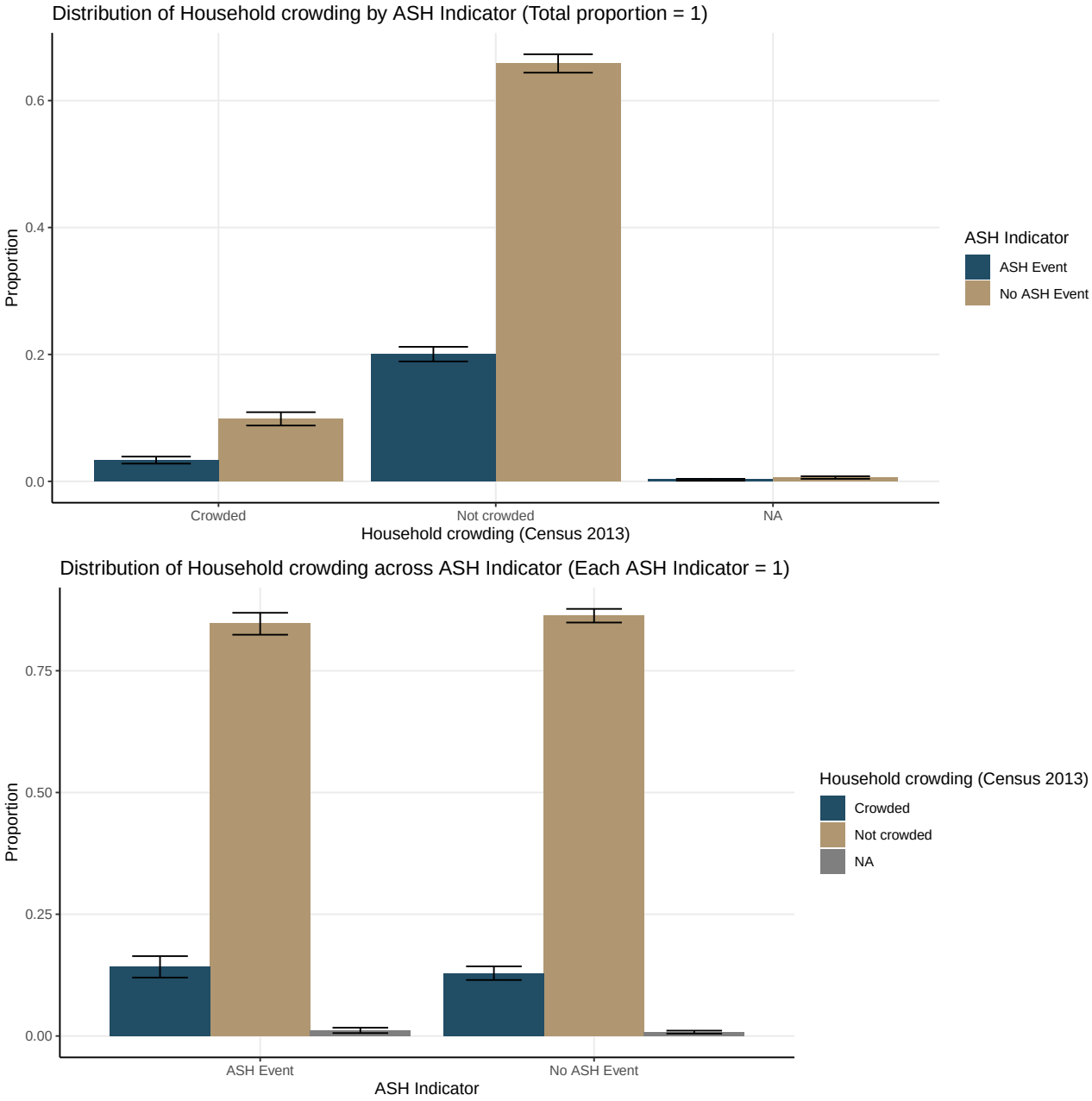


Figure A.6: ASH binary by 2013 Household Crowding (proportions)



NZDep

Figure A.7: ASH binary by 2013 NZDep (weighted counts) with 95% CI's

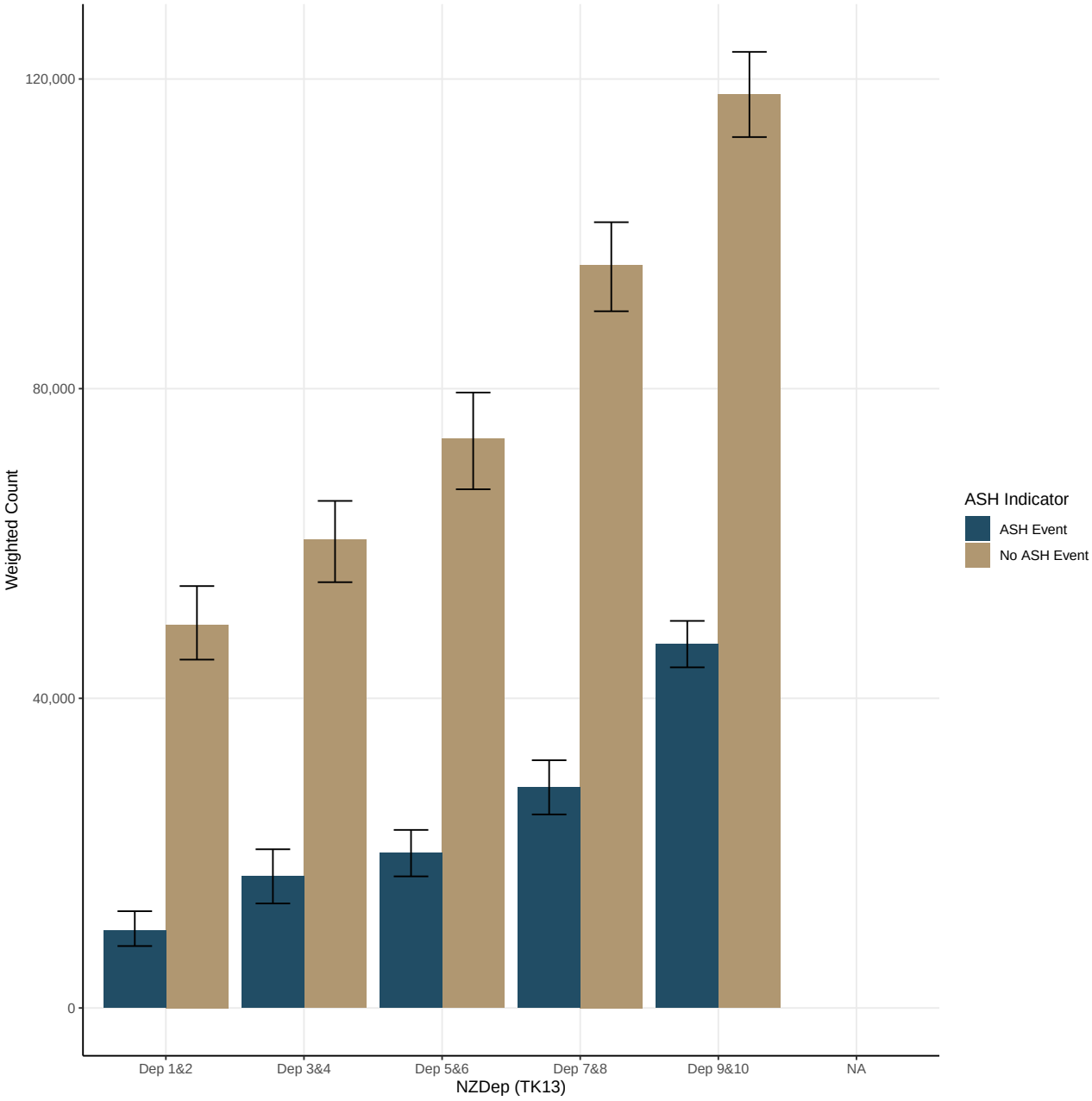
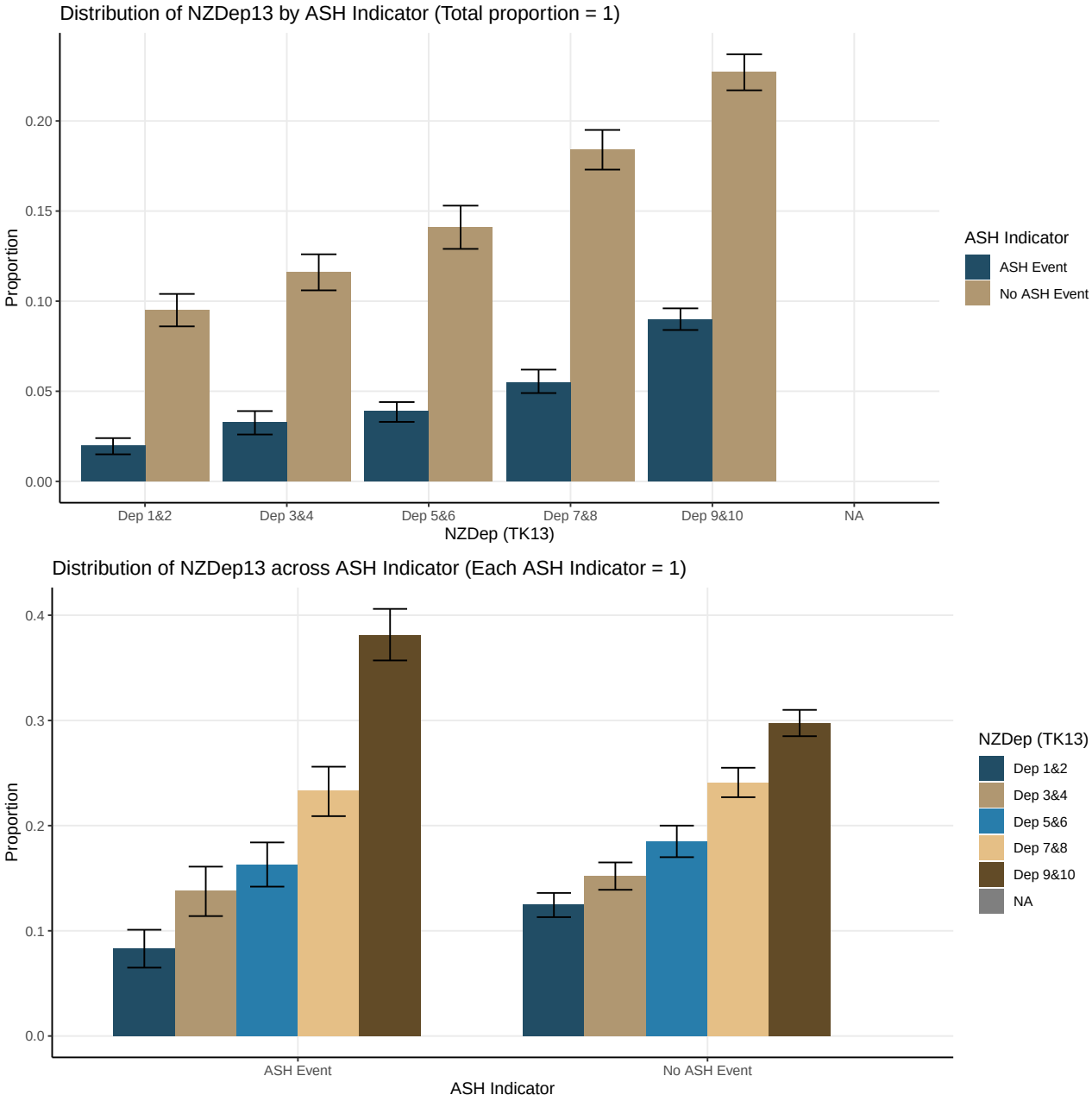


Figure A.8: ASH binary by 2013 NZDep (proportions)



Additional Regression Analysis - ASH Binary

Table A.12: Te Kupenga 2013 Logistic Regression for ASH binary: sex, age, sex*age, Highest Qualification (2013), IRD Individual Income (2013), Disability (2013) - Odds Ratio

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	2.581	1.623	4.105	0.604	4.060	0.000	***	23.381	
Age (2013)									
15-24	1.000	(base)							
25-34	1.025	0.634	1.657	0.248	0.100	0.918		24.196	
35-44	1.664	1.017	2.722	0.413	2.050	0.043	**	24.802	
45-54	2.396	1.521	3.772	0.548	3.820	0.000	***	22.881	
55-64	4.991	3.059	8.143	1.231	6.520	0.000	***	24.668	
65-74	8.296	4.721	14.576	2.357	7.450	0.000	***	28.408	
75 and above	8.753	3.602	21.272	3.917	4.850	0.000	***	44.753	#
Sex*Age									
Female#25-34	0.937	0.513	1.712	0.285	-0.210	0.832		30.361	#
Female#35-44	0.614	0.344	1.094	0.179	-1.670	0.097	*	29.138	
Female#45-54	0.545	0.303	0.980	0.161	-2.050	0.043	**	29.595	
Female#55-64	0.387	0.211	0.709	0.118	-3.110	0.002	**	30.575	#
Female#65-74	0.422	0.220	0.811	0.139	-2.620	0.010	**	32.928	#
Female#75 and above	0.690	0.197	2.417	0.436	-0.590	0.558		63.176	##
Highest Qualification (2013)									
No Qualification	1.453	1.078	1.959	0.219	2.480	0.015	**	15.058	
School-level Qualification	1.154	0.869	1.533	0.165	1.000	0.320		14.314	
Post-school Qualification	1.103	0.807	1.508	0.174	0.620	0.535		15.748	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.255	1.268	4.011	0.654	2.800	0.006	**	29.023	
$15,001 - $30,000	1.615	0.879	2.967	0.495	1.560	0.121		30.653	#
$30,001 - $50,000	1.839	0.994	3.402	0.570	1.970	0.052	*	30.998	#
$50,001 - $100,000	1.517	0.813	2.832	0.477	1.330	0.188		31.457	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.602	2.176	3.111	0.234	10.620	0.000	***	9.008	
_cons	0.044	0.021	0.091	0.016	-8.590	0.000	***	36.364	#

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 390,500 (rounded to base 500), Replications = 100, Design df = 99, $F(21, 79) = 22.74$, Prob > $F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage.

The above regression analysis A.12 includes the use of an interaction term between Sex and Age, to assess whether the Odds ratio for sex also depends on age for individuals. Here, the interaction term is significant across all categories except for the lower age-bands for females. However, for the sex variable in isolation to the interaction term, females have higher odds of an ASH event in comparison to males. Furthermore, all age-bands are considered to have greater odds of an ASH event in comparison to the 15-24 baseline. For the interaction term, the effect the categories of females who are 34 and older, the effect is smaller than males who are in the corresponding category. It can be noted that disability continues to display large effects on ASH events, with 2.602 times greater odds than those without a disability, while all other variables are kept constant. Additionally, upon including an interaction term, there is no change across the other explanatory variables in the model.

Table A.13: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Highest Qualification (2013)*IRD Individual Income (2013), Disability (2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife t Std. Err.		$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.651	1.356	2.011	0.164	5.050	0.000	***	9.940	
Age (2013)							***		
15-24	1.000	(base)					***		
25-34	1.051	0.800	1.381	0.145	0.360	0.719		13.759	
35-44	1.279	0.946	1.730	0.195	1.620	0.109		15.216	
45-54	1.737	1.290	2.339	0.261	3.680	0.000	***	15.005	
55-64	3.034	2.192	4.199	0.497	6.780	0.000	***	16.380	
65-74	5.289	3.720	7.520	0.938	9.390	0.000	***	17.734	
75 and above	7.254	3.892	13.518	2.276	6.320	0.000	***	31.373	#
Highest Qualification (2013)									
No Qualification	S	S	S	S	S	S		S	S
School-level Qualification	1.088	0.249	4.746	0.808	0.110	0.910		74.225	##
Post-school Qualification	0.746	0.151	3.688	0.601	-0.360	0.717		80.560	##
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	1.322	0.486	3.602	0.668	0.550	0.581		50.497	##
$15,001 - $30,000	1.357	0.503	3.658	0.678	0.610	0.543		49.988	#
$30,001 - $50,000	1.348	0.551	3.298	0.608	0.660	0.509		45.073	#
$50,001 - $100,000	1.328	0.577	3.056	0.558	0.680	0.501		42.005	#
$100,001 and above	1.000	(base)							
Highest Qualification*IRD Individual Income									
No Qualification#$1 - $15,000	S	S	S	S	S	S		S	S
No Qualification#$15,001 - $30,000	S	S	S	S	S	S		S	S
No Qualification#$30,001 - $50,000	S	S	S	S	S	S		S	S
No Qualification#$50,001 - $100,000	S	S	S	S	S	S		S	S

School-level Qualification#$1 - $15,000	1.627	0.321	8.252	1.332	0.600	0.553		81.821	##
School-level Qualification#$15,001 - $30,000	1.043	0.215	5.047	0.829	0.050	0.958		79.477	##
School-level Qualification#$30,001 - $50,000	0.971	0.200	4.703	0.772	-0.040	0.971		79.510	##
School-level Qualification#$50,001 - $100,000	0.936	0.189	4.632	0.754	-0.080	0.934		80.610	##
Post-school Qualification#$1 - $15,000	1.908	0.299	12.183	1.783	0.690	0.491		93.441	##
Post-school Qualification#$15,001 - $30,000	1.316	0.239	7.257	1.132	0.320	0.751		86.060	##
Post-school Qualification#$30,001 - $50,000	2.087	0.380	11.473	1.793	0.860	0.394		85.886	##
Post-school Qualification#$50,001 - $100,000	1.336	0.242	7.383	1.151	0.340	0.737		86.152	##
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.590	2.159	3.107	0.238	10.370	0.000	***	9.174	
_cons	0.070	0.032	0.153	0.028	-6.720	0.000	***	39.644	#

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 390,500 (rounded to base 500), Replications = 100, Design df = 99, $F(27, 73) = 15.05$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ 30% $<$ 50% ; Output rule = ## if $RSE\%$ 50% $<$ 100%; Output rule = S if $RSE\% \geq 100\%$

Table A.14: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: NZDep Quintiles (2013) - Odds Ratio

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
NZDep Quintiles (2013)									
Dep 1&2	1.000	(base)							
Dep 3&4	1.365	0.952	1.958	0.248	1.710	0.090	*	18.172	
Dep 5&6	1.329	0.956	1.847	0.221	1.710	0.090	*	16.595	
Dep 7&8	1.451	1.074	1.959	0.220	2.450	0.016	**	15.153	
Dep 9&10	1.928	1.474	2.522	0.261	4.850	0.000	***	13.527	
_cons	0.206	0.159	0.267	0.027	-12.120	0.000	***	13.015	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 5,500 (rounded to base 500), Population size = 520,000 (rounded to base 500), Replications = 100, Design df = 99, $F(4, 96) = 10.14$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ 30% $<$ 50% ; Output rule = ## if $RSE\%$ 50% $<$ 100%; Output rule = S if $RSE\% \geq 100\%$

Table A.15: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Household Crowding (2013), NZDep Quintiles (2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.651	1.354	2.014	0.165	5.010	0.000	***	10.009	
Age (2013)									
15-24	1.000	(base)							
25-34	1.021	0.777	1.344	0.141	0.150	0.878		13.813	
35-44	1.237	0.917	1.669	0.187	1.410	0.162		15.094	
45-54	1.670	1.231	2.266	0.257	3.340	0.001	**	15.367	
55-64	2.972	2.164	4.082	0.475	6.810	0.000	***	15.996	
65-74	5.109	3.553	7.346	0.935	8.910	0.000	***	18.300	
75 and above	6.903	3.547	13.434	2.316	5.760	0.000	***	33.555	#
Highest Qualification (2013)									
No Qualification	1.427	1.052	1.936	0.219	2.310	0.023	**	15.377	
School-level Qualification	1.153	0.865	1.536	0.167	0.980	0.329		14.464	
Post-school Qualification	1.107	0.808	1.518	0.176	0.640	0.524		15.903	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.069	1.165	3.676	0.599	2.510	0.014	**	28.954	
$15,001 - $30,000	1.488	0.811	2.730	0.455	1.300	0.196		30.578	#
$30,001 - $50,000	1.652	0.891	3.060	0.513	1.610	0.110		31.083	#
$50,001 - $100,000	1.426	0.768	2.649	0.445	1.140	0.258		31.207	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.594	2.145	3.137	0.248	9.950	0.000	***	9.578	
Household Crowding (2013)									
Not crowded	1.000	(base)							
Crowded	1.078	0.810	1.434	0.155	0.520	0.604		14.401	

NZDep Quintiles (2013)									
Dep 1&2	1.000	(base)							
Dep 3&4	1.167	0.778	1.751	0.239	0.750	0.453		20.459	
Dep 5&6	1.078	0.740	1.571	0.205	0.400	0.693		18.972	
Dep 7&8	0.942	0.651	1.363	0.176	-0.320	0.748		18.635	
Dep 9&10	1.228	0.890	1.696	0.200	1.270	0.208		16.247	
_cons	0.057	0.029	0.111	0.019	-8.510	0.000	***	33.738	#

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 387,000 (rounded to base 500), Replications = 100, Design df = 99, $F(20, 80) = 23.06$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Table A.16: Te Kupenga 2013 (with Jackknife weights) Logistic Regression for ASH binary: Sex, Age (2013), Highest Qualification (2013), IRD Individual Income (2013), Disability (2013), Any Medical Discrimination (TK 2013) - Odds Ratio **(Continued on the following two pages)**

	Odds Ratio	95% CI Lower	95% CI Upper	Jackknife t Std. Err.		$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	1.000	(base)							
Female	1.643	1.351	1.999	0.162	5.030	0.000	***	9.874	
Age (2013)									
15-24	1.000	(base)							
25-34	0.976	0.740	1.287	0.136	-0.170	0.862		13.956	
35-44	1.192	0.882	1.611	0.181	1.160	0.250		15.175	
45-54	1.627	1.197	2.211	0.252	3.140	0.002	**	15.472	
55-64	2.814	2.054	3.853	0.446	6.530	0.000	***	15.850	
65-74	5.098	3.557	7.307	0.925	8.980	0.000	***	18.145	
75 and above	7.029	3.759	13.145	2.218	6.180	0.000	***	31.547	#
Highest Qualification (2013)									
No Qualification	1.430	1.064	1.923	0.213	2.400	0.018	**	14.916	
School-level Qualification	1.141	0.858	1.518	0.164	0.920	0.360		14.376	
Post-school Qualification	1.102	0.804	1.509	0.175	0.610	0.543		15.860	
Degree Qualification	1.000	(base)							
IRD Individual Income (2013)									
$1 - $15,000	2.105	1.189	3.726	0.606	2.580	0.011	**	28.795	
$15,001 - $30,000	1.496	0.812	2.756	0.461	1.310	0.194		30.792	#
$30,001 - $50,000	1.696	0.918	3.132	0.524	1.710	0.091	*	30.921	#
$50,001 - $100,000	1.440	0.773	2.682	0.451	1.160	0.248		31.357	#
$100,001 and above	1.000	(base)							
Disability (2013)									
No Disability	1.000	(base)							
Have Disability	2.541	2.115	3.052	0.235	10.090	0.000	***	9.239	
Any Medical Discrimination (TK 2013)									
No	1.000	(base)							
Yes	1.627	1.165	2.272	0.274	2.890	0.005	**	16.832	
_cons	0.063	0.033	0.121	0.021	-8.390	0.000	***	33.016	#

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 4,000 (rounded to base 500), Population size = 384,500 (rounded to base 500), Replications = 100, Design df = 99, $F(16, 84) = 28.56$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

A.2.2 COVID-19 Vaccinations

Census 2013 to 2018 Transition tables

Table A.17: Unweighted Counts of 2013 Census Highest Qualification and 2018 Census Highest Qualification

	2018 Variable					
	No Qualification	School-level Qualification	Post-school Qualification	Degree Qualification	Missing	Total
2013 Variable						
No Qualification	849	165	252	9	288	1563
School-level Qualification	90	891	486	120	264	1851
Post-school Qualification	57	108	699	111	114	1089
Degree Qualification	S	12	30	450	42	534
Missing	108	42	114	18	135	417
Total	1104	1218	1581	708	843	5454

Table A.18: Unweighted Counts of 2013 Census Individual Income and 2018 Census Individual Income

	2018 Variable							
	Loss/No Income	$1 - $15,000	$15,001 - $30,000	$30,001 - $50,000	$50,001 - $100,000	$100,001 and above	Missing	Total
2013 Variable								
Loss/No Income	66	153	132	93	9	S	93	546
$1 - $15,000	42	306	393	201	87	S	192	1221
$15,001 - $30,000	21	138	435	276	117	6	186	1179
$30,001 - $50,000	21	84	180	381	357	18	138	1179
$50,001 - $100,000	6	36	51	69	519	87	69	837
$100,001 and above	S	6	9	6	24	93	18	156
Missing	18	63	99	54	21	S	72	327
Total	174	786	1299	1080	1134	204	768	5445

Table A.19: Unweighted Counts of 2013 IRD Individual Income and 2018 IRD Individual Income

	2018 Variable						
	$1 - $15,000	$15,001 - $30,000	$30,001 - $50,000	$50,001 - $100,000	$100,001 and above	Missing	Total
2013 Variable							
$1 - $15,000	294	444	207	93	S	144	1182
$15,001 - $30,000	159	654	288	126	S	129	1356
$30,001 - $50,000	69	132	348	336	9	87	981
$50,001 - $100,000	33	60	78	606	90	63	930
$100,001 and above	S	9	9	6	69	6	99
Missing	171	162	111	39	6	423	912
Total	726	1461	1041	1206	174	852	5460

Table A.20: Unweighted Counts of 2013 Census Household Composition and 2018 Census Household Composition

	2018 Variable							
	1 person	1 Couple	1 Parent w Children	2> Families	Couple w Children	Other Multiperson	Missing	Total
2013 Variable								
1 person	213	33	36	6	36	18	150	492
1 Couple	51	570	30	42	174	36	192	1095
1 Parent w Children	69	45	399	81	147	51	315	1107
2¿ Families	12	42	69	108	117	12	135	495
Couple w Children	42	171	129	144	1026	63	393	1968
Other Multiperson	45	48	21	18	27	60	78	297
Missing	S	S	S	S	S	S	S	S
Total	432	909	684	399	1527	240	1263	5454

Table A.21: Unweighted Counts of 2013 Census Total Household Income and 2018 Total Household Income

	2018 Variable								
	$20,000 or less	$20,001 - $30,000	$30,001 - $50,000	$50,001 - $70,000	$70,001 - $100,000	$100,001 - $150,000	$150,001 and above	Missing	Total
2013 Variable									
$20,000 or less	84	54	48	30	33	27	12	168	456
$20,001 - $30,000	36	45	84	45	36	15	9	111	381
$30,001 - $50,000	54	54	141	132	129	63	30	216	819
$50,001 - $70,000	27	24	69	120	165	126	54	171	756
$70,001 - $100,000	18	21	48	84	186	273	123	180	933
$100,001 - $150,000	6	12	33	36	87	231	192	117	714
$150,001 and above	6	S	18	24	27	75	240	105	495
Missing	60	39	87	93	102	120	75	339	915
Total	291	249	528	564	765	930	735	1407	5469

Additional Regression Analysis - COVID-19 Vaccinations

Table A.22: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Sex*Age, IRD Individual Income (2018), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients **(Continued on the following two pages)**

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	1.142	0.485	1.800	0.331	3.450	0.001	**	29.001	
Age (2020)									
15-24	0.000	(base)							
25-34	S	S	S	S	S	S		S	S
35-44	S	S	S	S	S	S		S	S
45-54	S	S	S	S	S	S		S	S
55-64	0.428	-0.279	1.135	0.356	1.200	0.233		83.249	##
65-74	1.165	0.475	1.855	0.348	3.350	0.001	**	29.857	
75 and above	S	S	S	S	S	S		S	S
Sex*Age									
Female#25-34	-0.878	-1.704	-0.053	0.416	-2.110	0.037	**	47.353	#
Female#35-44	-0.628	-1.493	0.237	0.436	-1.440	0.153		69.451	##
Female#45-54	-0.732	-1.589	0.124	0.432	-1.700	0.093	*	58.956	##
Female#55-64	-0.789	-1.760	0.181	0.489	-1.610	0.110		61.981	##
Female#65-74	-1.432	-2.414	-0.450	0.495	-2.890	0.005	**	34.566	#
Female#75 and above	S	S	S	S	S	S		S	S
IRD Individual Income (2018)									
$1 - $15,000	-2.562	-3.681	-1.443	0.564	-4.540	0.000	***	22.010	
$15,001 - $30,000	-2.342	-3.536	-1.147	0.602	-3.890	0.000	***	25.703	
$30,001 - $50,000	-1.939	-3.126	-0.752	0.598	-3.240	0.002	**	30.848	#
$50,001 - $100,000	-1.494	-2.696	-0.292	0.606	-2.470	0.015	**	40.540	#
$100,001 and above	0.000	(base)							
Trust in Fair Healthcare (TK 2013)	0.056	0.015	0.096	0.020	2.750	0.007	**	36.361	#
How are whānau doing? (TK 2018)	0.071	0.023	0.119	0.024	2.930	0.004	**	34.084	#
/cut1	-2.786	-4.154	-1.419	0.689				24.736	

/cut2	-2.652	-4.019	-1.285	0.689	25.980

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 4,500 (rounded to base 500), Population size = 414,500 (rounded to base 500), Replications = 100, Design df = 99, $F(19, 81) = 9.37$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

Table A.23: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018) - Coefficients **(Continued on the following two pages)**

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	0.450	0.090	0.811	0.182	2.480	0.015	**	40.369	#
Age (2020)									
15-24	0.000	(base)							
25-34	S	S	S	S	S	S		S	S
35-44	S	S	S	S	S	S		S	S
45-54	S	S	S	S	S	S		S	S
55-64	0.550	-0.357	1.457	0.457	1.200	0.231		83.054	##
65-74	0.858	-0.052	1.769	0.459	1.870	0.064	*	53.477	##
75 and above	S	S	S	S	S	S		S	S
Highest Qualification (2018)									
No Qualification	-0.550	-1.147	0.046	0.300	-1.830	0.070	*	54.590	##
School-level Qualification	-0.348	-0.950	0.254	0.303	-1.150	0.254		87.191	##
Post-school Qualification	-0.308	-0.904	0.289	0.300	-1.020	0.309		97.699	##
Degree Qualification	0.000	(base)							
IRD Individual Income (2018)									
$1 - $15,000	-1.701	-3.093	-0.308	0.702	-2.420	0.017	**	41.260	#
$15,001 - $30,000	-1.468	-2.931	-0.005	0.737	-1.990	0.049	**	50.221	##
$30,001 - $50,000	-1.182	-2.581	0.217	0.705	-1.680	0.097	*	59.662	##
$50,001 - $100,000	S	S	S	S	S	S		S	S
$100,001 and above	0.000	(base)							
Disability (2018)									
No Disability	0.000	(base)							
Have Disability	S	S	S	S	S	S		S	S
Connection to Tūrangawaewae (TK 2013)									

Very Strong	0.000	(base)							
Strong	S	S	S	S	S	S		S	S
Somewhat	S	S	S	S	S	S		S	S
Weak	S	S	S	S	S	S		S	S
Very Weak	S	S	S	S	S	S		S	S
None	S	S	S	S	S	S		S	S
Importance of Culture (TK 2013)									
Very	0.000	(base)							
Quite	S	S	S	S	S	S		S	S
Somewhat	S	S	S	S	S	S		S	S
A little	S	S	S	S	S	S		S	S
None	1.004	-0.824	2.831	0.921	1.090	0.279		91.779	##
DK	15.847	12.828	18.865	1.521	10.420	0.000	***	9.600	
Any Medical Discrimination? (TK 2013)									
No	0.000	(base)							
Yes	S	S	S	S	S	S		S	S
Trust in Fair Healthcare (TK 2013)	0.051	-0.021	0.124	0.036	1.400	0.164		71.259	##
How are whānau doing? (TK 2018)	0.107	0.025	0.189	0.041	2.590	0.011	**	38.670	#
/cut1	-2.328	-4.139	-0.518	0.912				39.187	#
/cut2	-2.183	-3.995	-0.371	0.913				41.836	#
Note: 2 observations completely determined. Standard errors questionable.									

Note: $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Number of obs = 2,000 (rounded to base 500), Population size = 162,000 (rounded to base 500), Replications = 100, Design df = 99, $F(28, 72) = 6.01$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Table A.24: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018) - Coefficients **(Continued on the following two pages)**

	Coef.	95% CI Lower	95% CI Upper	Jackknife t Std. Err.		$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	0.341	0.076	0.606	0.134	2.560	0.012	**	39.116	#
Age (2020)									
15-24	0.000	(base)							
25-34	-0.700	-1.281	-0.119	0.293	-2.390	0.019	**	41.810	#
35-44	-0.409	-0.990	0.172	0.293	-1.400	0.165		71.564	##
45-54	S	S	S	S	S	S		S	S
55-64	S	S	S	S	S	S		S	S
65-74	S	S	S	S	S	S		S	S
75 and above	-0.562	-1.385	0.261	0.415	-1.360	0.178		73.753	##
Highest Qualification (2018)									
No Qualification	-0.375	-0.908	0.158	0.269	-1.400	0.166		71.612	##
School-level Qualification	-0.342	-0.886	0.201	0.274	-1.250	0.214		79.990	##
Post-school Qualification	-0.343	-0.810	0.125	0.236	-1.450	0.149		68.834	##
Degree Qualification	0.000	(base)							
IRD Individual Income (2018)									
$1 - $15,000	-2.465	-3.934	-0.997	0.740	-3.330	0.001	**	30.016	#
$15,001 - $30,000	-2.340	-3.909	-0.770	0.791	-2.960	0.004	**	33.811	#
$30,001 - $50,000	-2.368	-3.868	-0.867	0.756	-3.130	0.002	**	31.943	#
$50,001 - $100,000	-1.876	-3.372	-0.380	0.754	-2.490	0.014	**	40.185	#
$100,001 and above	0.000	(base)							
Disability (2018)									
No Disability	0.000	(base)							
Have Disability	S	S	S	S	S	S		S	S
How are whānau doing? (TK 2018)	0.077	0.006	0.147	0.036	2.160	0.033	**	46.299	#

How well are whānau getting along? (TK 2018)									
Very Well	0.000	(base)							
Well	S	S	S	S	S	S		S	S
Neither well/bad	-0.342	-0.826	0.141	0.244	-1.400	0.163		71.190	##
Badly	S	S	S	S	S	S		S	S
Very Badly	S	S	S	S	S	S		S	S
DK	16.034	7.712	24.355	4.194	3.820	0.000	***	26.156	
Household crowding (2018)									
Not crowded	0.000	(base)							
Crowded	S	S	S	S	S	S		S	S
Total Household Income (2018)									
$20,000 or less	-0.834	-1.625	-0.044	0.398	-2.090	0.039	**	47.749	#
$20,001 - $30,000	-0.781	-1.441	-0.121	0.333	-2.350	0.021	**	42.600	#
$30,001 - $50,000	-0.542	-1.264	0.180	0.364	-1.490	0.140		67.166	##
$50,001 - $70,000	S	S	S	S	S	S		S	S
$70,001 - $100,000	S	S	S	S	S	S		S	S
$100,001 - $150,000	S	S	S	S	S	S		S	S
$150,001 and above	0.000	(base)							
Household Composition (2018)									
1 person	0.000	(base)							
1 Couple	0.525	-0.006	1.055	0.267	1.960	0.052	*	50.939	##
Couple w Children	S	S	S	S	S	S		S	S
1 Parent w Children	S	S	S	S	S	S		S	S
2¿ Families	S	S	S	S	S	S		S	S
Other Multiperson	0.535	-0.323	1.393	0.432	1.240	0.219		80.887	##
Housing Quality (2018)									
No/sometimes damp/mould	0.000	(base)							
Always damp/mould	S	S	S	S	S	S		S	S
/cut1	-4.400	-6.226	-2.573	0.920				20.920	
/cut2	-4.262	-6.097	-2.426	0.925				21.709	

Note: 4 observations completely determined. Standard errors questionable.

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 3,000 (rounded to base 500), Population size = 290,500 (rounded to base 500), Replications = 100, Design df = 99, $F(34, 66) = 3.8$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

Table A.25: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), Connection to Tūrangawaewae (TK 2013), Importance of Culture (TK 2013), Any Medical Discrimination (TK 2013), Trust in Fair Healthcare (TK 2013), How are whānau doing? (TK 2018), How well are whānau getting along? (TK 2018), Household crowding (2018), Total Household income (2018), Household Composition (2018), Housing Quality (2018), NZDep Quintiles (2018) - Coefficients **(Continued on the following three pages)**

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	0.633	0.241	1.025	0.198	3.210	0.002	**	31.198	#
Age (2020)									
15-24	0.000	(base)							
25-34	S	S	S	S	S	S		S	S
35-44	S	S	S	S	S	S		S	S
45-54	S	S	S	S	S	S		S	S
55-64	S	S	S	S	S	S		S	S
65-74	0.763	-0.477	2.002	0.625	1.220	0.225		81.903	##
75 and above	S	S	S	S	S	S		S	S
Highest Qualification (2018)									
No Qualification	-0.397	-1.119	0.325	0.364	-1.090	0.278		91.729	##
School-level Qualification	S	S	S	S	S	S		S	S
Post-school Qualification	-0.437	-1.150	0.277	0.360	-1.210	0.228		82.391	##
Degree Qualification	0.000	(base)							
IRD Individual Income (2018)									
$1 - $15,000	-2.030	-4.283	0.224	1.136	-1.790	0.077	*	55.959	##
$15,001 - $30,000	-1.668	-3.898	0.562	1.124	-1.480	0.141		67.361	##
$30,001 - $50,000	-1.680	-3.790	0.429	1.063	-1.580	0.117		63.269	##
$50,001 - $100,000	-1.345	-3.552	0.861	1.112	-1.210	0.229		82.660	##
$100,001 and above	0.000	(base)							
Disability (2018)									
No Disability	0.000	(base)							
Have Disability	-0.338	-0.923	0.247	0.295	-1.150	0.254		87.201	##

Connection to Tūrangawaewae (TK 2013)									
Very Strong	0.000	(base)							
Strong	S	S	S	S	S	S		S	S
Somewhat	0.318	-0.126	0.763	0.224	1.420	0.159		70.419	##
Weak	S	S	S	S	S	S		S	S
Very Weak	S	S	S	S	S	S		S	S
None	S	S	S	S	S	S		S	S
Importance of Culture (TK 2013)									
Very	0.000	(base)							
Quite	-0.370	-0.837	0.098	0.235	-1.570	0.120		63.718	##
Somewhat	-0.401	-0.964	0.162	0.284	-1.410	0.161		70.776	##
A little	S	S	S	S	S	S		S	S
None	S	S	S	S	S	S		S	S
DK	16.669	7.552	25.785	4.595	3.630	0.000	***	27.564	
Any Medical Discrimination? (TK 2013)									
No	0.000	(base)							
Yes	-0.499	-1.160	0.162	0.333	-1.500	0.138		66.804	##
Trust in Fair Healthcare (TK 2013)	0.059	-0.024	0.143	0.042	1.410	0.162		70.924	##
How are whānau doing? (TK 2018)	0.088	-0.009	0.186	0.049	1.800	0.075	*	55.537	##
How well are whānau getting along? (TK 2018)									
Very Well	0.000	(base)							
Well	S	S	S	S	S	S		S	S
Neither well/bad	S	S	S	S	S	S		S	S
Badly	0.898	-0.436	2.231	0.672	1.340	0.185		74.858	##
Very Badly	S	S	S	S	S	S		S	S
DK	16.085	6.876	25.295	4.642	3.470	0.001	**	28.856	
Household crowding (2018)									
Not crowded	0.000	(base)							
Crowded	0.054	-0.595	0.703	0.327	0.160	0.870		610.299	S
Total Household Income (2018)									
$20,000 or less	-1.394	-2.562	-0.225	0.589	-2.370	0.020	**	42.254	#
$20,001 - $30,000	-1.329	-2.341	-0.316	0.510	-2.600	0.011	**	38.424	#

$30,001 - $50,000	-0.813	-1.760	0.133	0.477	-1.700	0.091	*	58.665	##
$50,001 - $70,000	-0.898	-1.710	-0.086	0.409	-2.190	0.031	**	45.595	#
$70,001 - $100,000	-0.782	-1.672	0.108	0.448	-1.740	0.084	*	57.338	##
$100,001 - $150,000	-0.541	-1.252	0.171	0.359	-1.510	0.135		66.327	##
$150,001 and above	0.000	(base)							
Household Composition (2018)									
1 person	0.000	(base)							
1 Couple	0.454	-0.254	1.161	0.357	1.270	0.206		78.600	##
Couple w Children	S	S	S	S	S	S		S	S
1 Parent w Children	S	S	S	S	S	S		S	S
2¿ Families	S	S	S	S	S	S		S	S
Other Multiperson	S	S	S	S	S	S		S	S
Housing Quality (2018)									
No/sometimes damp/mould	0.000	(base)							
Always damp/mould	0.305	-0.204	0.813	0.256	1.190	0.237		84.074	##
NZDep Quintiles (2018)									
Dep 1&2	0.000	(base)							
Dep 3&4	-1.122	-2.427	0.182	0.657	-1.710	0.091	*	58.558	##
Dep 5&6	-0.805	-2.200	0.590	0.703	-1.140	0.255		87.386	##
Dep 7&8	-0.941	-2.320	0.438	0.695	-1.350	0.179		73.838	##
Dep 9&10	-1.192	-2.499	0.116	0.659	-1.810	0.074	*	55.303	##
/cut1	-4.471	-7.719	-1.223	1.637				36.615	#
/cut2	-4.316	-7.573	-1.059	1.641				38.030	#

Note: 3 observations completely determined. Standard errors questionable.

Note: $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Number of obs = 1,500 (rounded to base 500), Population size = 147,000 (rounded to base 500), Replications = 100, Design df = 99, $F(50, 50) = 2.9$, Prob $> F = 0.0001$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Table A.26: Te Kupenga 2013 (with Jackknife weights) Ordered Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018), NZDep Quintiles (2018) - Coefficients **(Continued on the following two pages)**

	Coef.	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
Sex									
Male	0.000	(base)							
Female	0.315	0.076	0.553	0.120	2.620	0.010	**	38.214	#
Age (2020)									
15-24	0.000	(base)							
25-34	-0.658	-1.180	-0.137	0.263	-2.500	0.014	**	39.939	#
35-44	-0.484	-1.031	0.063	0.276	-1.760	0.082	*	56.957	##
45-54	-0.418	-1.025	0.189	0.306	-1.370	0.175		73.229	##
55-64	S	S	S	S	S	S		S	S
65-74	S	S	S	S	S	S		S	S
75 and above	-0.489	-1.162	0.183	0.339	-1.440	0.152		69.289	##
Highest Qualification (2018)									
No Qualification	-0.479	-0.974	0.017	0.250	-1.920	0.058	*	52.189	##
School-level Qualification	-0.360	-0.878	0.159	0.261	-1.380	0.172		72.644	##
Post-school Qualification	-0.371	-0.825	0.082	0.229	-1.620	0.107		61.547	##
Degree Qualification	0.000	(base)							
IRD Individual Income (2018)									
$1 - $15,000	-1.922	-3.072	-0.771	0.580	-3.310	0.001	**	30.168	#
$15,001 - $30,000	-1.879	-3.104	-0.654	0.617	-3.040	0.003	**	32.847	#
$30,001 - $50,000	-1.649	-2.843	-0.454	0.602	-2.740	0.007	**	36.515	#
$50,001 - $100,000	-1.084	-2.294	0.125	0.610	-1.780	0.078	*	56.222	##
$100,001 and above	0.000	(base)							
Disability (2018)									
No Disability	0.000	(base)							
Have Disability	S	S	S	S	S	S		S	S
NZDep Quintiles (2018)									
Dep 1&2	0.000	(base)							
Dep 3&4	S	S	S	S	S	S		S	S
Dep 5&6	S	S	S	S	S	S		S	S

Dep 7&8	S	S	S	S	S	S		S	S
Dep 9&10	-0.628	-1.329	0.072	0.353	-1.780	0.078	*	56.166	##
/cut1	-4.568	-6.106	-3.031	0.775				16.963	
/cut2	-4.438	-5.982	-2.894	0.778				17.536	

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 3,500 (rounded to base 500), Population size = 318,000 (rounded to base 500), Replications = 100, Design df = 99, $F(19, 81) = 3.83$, Prob $> F = 0$

$RSE\%$ (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and $RSE\%$

Output rule = # if $RSE\%$ $30\% < 50\%$; Output rule = ## if $RSE\%$ $50\% < 100\%$; Output rule = S if $RSE\% \geq 100\%$

Table A.27: Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): Sex, Age (2020), Highest Qualification (2018), IRD Individual Income (2018), Disability (2018) - Relative Risk Ratio **(Continued on the following two pages)**

	RRR	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
No Vaccination									
Sex									
Male	1.000	(base)							
Female	0.698	0.548	0.889	0.085	-2.950	0.004	**	12.17	
Age (2020)									
15-24	1.000	(base)							
25-34	1.844	1.065	3.194	0.511	2.210	0.029	**	27.69	
35-44	1.550	0.867	2.773	0.454	1.500	0.138		29.31	
45-54	1.573	0.834	2.968	0.503	1.420	0.160		31.99	#
55-64	1.286	0.705	2.347	0.390	0.830	0.409		30.32	#
65-74	0.805	0.409	1.586	0.275	-0.630	0.527		34.17	#
75 and above	1.660	0.805	3.422	0.605	1.390	0.167		36.45	#
Highest Qualification (2018)									
No Qualification	1.711	0.971	3.017	0.489	1.880	0.063	*	28.58	
School-level Qualification	1.434	0.822	2.500	0.402	1.290	0.201		28.02	
Post-school Qualification	1.538	0.926	2.553	0.393	1.680	0.095	*	25.55	
Degree Qualification	1.000	(base)							
IRD Individual Income (2018)									
$1 - $15,000	9.802	1.794	53.547	8.388	2.670	0.009	**	85.58	##
$15,001 - $30,000	10.325	1.789	59.577	9.120	2.640	0.010	**	88.33	##
$30,001 - $50,000	8.201	1.440	46.711	7.191	2.400	0.018	**	87.68	##
$50,001 - $100,000	4.395	0.758	25.495	3.894	1.670	0.098	*	88.60	##
$100,001 and above	1.000	(base)							
Disability (2018)									
No Disability	1.000	(base)							
Have Disability	0.957	0.664	1.381	0.177	-0.240	0.814		18.46	
_cons	0.009	0.002	0.059	0.009	-5.070	0.000	***	91.99	##
Partially Vaccinated									

Sex								
Male	1.000	(base)						
Female	1.512	0.712	3.210	0.574	1.090	0.279	37.95	#
Age (2020)								
15-24	1.000	(base)						
25-34	2.987	0.635	14.059	2.332	1.400	0.164	78.07	##
35-44	2.794	0.599	13.027	2.168	1.320	0.189	77.59	##
45-54	0.895	0.150	5.327	0.804	-0.120	0.902	89.92	##
55-64	0.858	0.132	5.573	0.809	-0.160	0.871	94.32	##
65-74	0.829	0.115	5.990	0.826	-0.190	0.852	99.64	##
75 and above	S	S	S	S	S	S	S	S
Highest Qualification (2018)								
No Qualification	2.699	0.736	9.898	1.768	1.520	0.133	65.48	##
School-level Qualification	1.888	0.549	6.494	1.175	1.020	0.310	62.26	##
Post-school Qualification	1.212	0.339	4.340	0.779	0.300	0.765	64.28	##
Degree Qualification	1.000	(base)						
IRD Individual Income (2018)								
$1 - $15,000	S	S	S	S	S	S	S	S
$15,001 - $30,000	S	S	S	S	S	S	S	S
$30,001 - $50,000	S	S	S	S	S	S	S	S
$50,001 - $100,000	S	S	S	S	S	S	S	S
$100,001 and above	1.000	(base)						
Disability (2018)								
No Disability	1.000	(base)						
Have Disability	S	S	S	S	S	S	S	S
_cons	S	S	S	S	S	S	S	S
Fully Vaccinated (base outcome)								

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

Number of obs = 3,500 (rounded to base 500), Population size = 318,000 (rounded to base 500), Replications = 100, Design df = 99, $F(30, 70) = 3.28$, Prob $> F = 0$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% $<$ 50% ; Output rule = ## if *RSE*% 50% $<$ 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Table A.28: Te Kupenga 2013 (with Jackknife weights) Multinomial Logistic Regression for Vaccinations (3-levels): NZDep Quintiles (2018) - Relative Risk Ratio

	RRR	95% CI Lower	95% CI Upper	Jackknife Std. Err.	t	$P > t$	Sig.level	RSE %	Output rule
No Vaccination									
NZDep Quintiles (2018)									
Dep 1&2	1.000			(base)					
Dep 3&4	1.388	0.791	2.434	0.393	1.160	0.250		28.318	
Dep 5&6	1.367	0.781	2.395	0.386	1.110	0.271		28.241	
Dep 7&8	1.171	0.717	1.913	0.290	0.640	0.525		24.734	
Dep 9&10	2.246	1.385	3.643	0.547	3.320	0.001	**	24.375	
_cons	0.091	0.057	0.144	0.021	-10.320	0.000	***	23.253	
Partially Vaccinated									
NZDep Quintiles (2018)									
Dep 1&2	1.000			(base)					
Dep 3&4	0.941	0.262	3.386	0.607	-0.090	0.925		64.526	##
Dep 5&6	0.695	0.185	2.615	0.464	-0.550	0.587		66.815	##
Dep 7&8	1.230	0.378	4.004	0.732	0.350	0.729		59.502	##
Dep 9&10	1.756	0.665	4.641	0.860	1.150	0.253		48.978	#
_cons	0.013	0.005	0.033	0.006	-9.590	0.000	***	45.011	#
Fully Vaccinated (base outcome)									

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Number of obs = 4,500 (rounded to base 500), Population size = 436,500 (rounded to base 500), Replications = 100, Design df = 99, $F(8, 92) = 4.47$, Prob > $F = 0.0001$

RSE% (Relative Sampling Error) is calculated by dividing the standard error of the estimate by the estimate itself, then multiplying by 100 to create a percentage

The *Output rule* corresponds to the requirements for releasing output related to the Te Kupenga survey and *RSE*%

Output rule = # if *RSE*% 30% < 50% ; Output rule = ## if *RSE*% 50% < 100%; Output rule = *S* if *RSE*% $\geq$ 100%

Appendix B

Appendix: Code

B.1 Initial Joining (ASH)

```
* Data Extraction from IDI

* Set up libraries;
libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
libname dat ODBC dsn=idi_clean_202206_srvprd schema=data;
libname mta ODBC dsn=idi_clean_202206_srvprd schema=metadata;
libname ird ODBC dsn=idi_clean_202206_srvprd schema=ir_clean;
libname moe ODBC dsn=idi_clean_202206_srvprd schema=moe_clean;
libname moh ODBC dsn=idi_clean_202206_srvprd schema=moh_clean;
libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data";

* Observation base - refresh 202206;

proc sql;
 create table base as
     select e.*, f.*
     from (select coalesce(a.snz_uid, b.snz_uid) as snz_uid,
                  coalesce(a.snz_cen_uid18, c.snz_cen_uid18) as snz_cen_uid18,
                  coalesce(b.snz_cen_uid13, d.snz_cen_uid13) as snz_cen_uid13,
                  a.c18, a.d18c, a.e18c,
                  b.c13, b.d13c, b.e13c,
                  c.t18, c.d18t, c.e18t,
                  d.t13, d.d13t, d.e13t
                  from (select x.snz_uid, x.snz_cen_uid as snz_cen_uid18
                              ,case when x.snz_uid is not null then 1 else 0 end as c18
                              ,case when x.cen_ind_maori_dscnt_output_code='1' then 1 else 0
                              ↪  end as d18c
                              ,case when x.cen_ind_ethgr_maori_ind_code in ('1', '2') then 1
                              ↪  else 0 end as e18c
                        from cen.census_individual_2018 x
                            where x.cen_ind_maori_dscnt_output_code='1' or
                            ↪  x.cen_ind_ethgr_maori_ind_code in ('1', '2')
                        ) a
                  full join (select y.snz_uid, y.snz_cen_uid as snz_cen_uid13
                                   ,case when y.snz_uid is not null then 1 else 0 end as
                                   ↪  c13
                                  ,case when y.cen_ind_recode_maori_dscnt_code='1' then 1
                                  ↪  else 0 end as d13c
```

```
                                    ,case when y.cen_ind_maori_eth_ind_code in ('1', '2')
                                    ↪  then 1 else 0 end as e13c
                                from cen.census_individual_2013 y
                                    where y.cen_ind_recode_maori_dscnt_code='1' or
                                    ↪  y.cen_ind_maori_eth_ind_code in ('1', '2')
                                ) b
                    on a.snz_uid = b.snz_uid
                    full join (select z.snz_cen_uid as snz_cen_uid18
                                    ,case when z.snz_cen_uid is not null then 1 else 0
                                    ↪  end as t18
                                ,case when z.qDEMMaoriDescent='1' then 1 else 0 end as
                                ↪  d18t
                                ,case when z.DVEthTR_Maori='2' then 1 else 0 end as e18t
                                from tkp.TK_2018 z
                                ) c
                    on a.snz_cen_uid18 = c.snz_cen_uid18
                    full join (select u.snz_cen_uid as snz_cen_uid13
                                    ,case when u.snz_cen_uid is not null then 1 else 0
                                    ↪  end as t13
                                ,case when u.qDEMMaoriDescent=1 then 1 else 0 end as d13t
                                ,case when u.DVEthTR_Maori=2 then 1 else 0 end as e13t
                                from tkp.TK_2013 u
                                ) d
                    on b.snz_cen_uid13 = d.snz_cen_uid13
            ) e
        left join (select v.snz_uid, v.snz_spine_ind, v.snz_deceased_year_nbr,
        ↪  v.snz_deceased_month_nbr
                    from dat.personal_detail v
                    where v.snz_spine_ind=1
                    ) f
        on e.snz_uid = f.snz_uid;
quit;


proc export data=base dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/base.csv";
run;
```

```
/*
Data Cleaning - Census13/TK13
*/

*Setting Working Directory
cd "I:\MAA2021-43\Tori\Thesis"

*Saving Log files:
cmdlog using "I:\MAA2021-43\Tori\Thesis\ASH_Project\LogFiles\ASH-commands.txt", append

log using "I:\MAA2021-43\Tori\Thesis\ASH_Project\LogFiles\ASH-results.txt", append text


**********************************************

* Using Base Census ID's and keeping if Maori (i.e. in the census or in TK13)
use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OriginalData\base.dta"

```

```
19
20  keep if snz_uid!=. & (c13!=. | t13!=.)
21
22  codebook snz_uid
23  codebook snz_uid if c13==1
24  codebook snz_uid if t13==1
25
26  keep snz_uid snz_cen_uid13 c13 t13 d13c e13c
27
28  save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\MID13.dta", replace
29  export delimited using "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\MID13.csv",
    ↪  replace
30
31  clear
```

```
1   * Set up libraries;
2   libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
3   libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
4   libname moh ODBC dsn=idi_clean_202206_srvprd schema=moh_clean;
5   libname ird ODBC dsn=idi_clean_202206_srvprd schema=ir_clean;
6   libname dia ODBC dsn=idi_clean_202206_srvprd  schema=dia_clean;
7   libname data ODBC dsn=idi_clean_202206_srvprd  schema=data;
8   libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data";
9
10
11  proc import out=MID13 dbms=csv
12  file="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/MID13.csv";
13  run;
14
15
16
17  * Joining to Cen13;
18
19  proc sql;
20     create table ID13 as
21         select x.snz_uid, x.cen_ind_ttl_inc_code, x.cen_ind_highest_qual_code,
           ↪  x.cen_ind_sex_code, x.cen_ind_age_code, y.*
22         from cen.census_individual_2013 x
23             left join MID13 y
24             on x.snz_uid = y.snz_uid;
25  quit;
26
27  proc export data=ID13 dbms=csv
28  outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/ID13.csv";
29  run;
30
31
32
33  * Hospitalisation;
34
35  proc sql;
36     create table hosp as
37         select x.*, y.*, w.*
38         from ID13 x
39         left join (select a.snz_uid, a.moh_evt_evst_date, a.moh_evt_even_date,
           ↪  a.moh_evt_event_id_nbr, year(a.moh_evt_evst_date) as sty,
           ↪  year(a.moh_evt_even_date) as eny,
40                                    a.moh_evt_los_nbr
```

```
                    from moh.pub_fund_hosp_discharges_event a
                where year(a.moh_evt_evst_date)>2013
                ) y
        on x.snz_uid = y.snz_uid
        left join (select d.moh_dia_event_id_nbr, d.moh_dia_diag_sequence_code,
        ↪  d.moh_dia_clinical_code, d.moh_dia_clinical_sys_code,
                                d.moh_dia_submitted_system_code,
                                ↪  d.moh_dia_diagnosis_type_code
                                 from moh.pub_fund_hosp_discharges_diag d
                ) w
        on y.moh_evt_event_id_nbr = w.moh_dia_event_id_nbr;
quit;


proc export data=hosp dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/hosp.csv";
run;



* IRD (not working in SAS, need to use sql instead);

proc sql;
 create table ird as
     select a.*, b.snz_uid
      , b.ir_ems_return_period_date
      , b.ir_ems_gross_earnings_amt
     from ID13 a
     left join ird.ird_ems b
     on a.snz_uid = b.snz_uid
     where year(b.ir_ems_return_period_date)=2013;
quit;

proc export data=ird dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/ird.csv";
run;
```

```
/* IRD income Export for 2013 (April 2012 - March 31 2013) */

SELECT c.snz_uid, b.[ir_ems_return_period_date], b.[ir_ems_gross_earnings_amt]
  from [IDI_Clean_202206].[cen_clean].[census_individual_2013] as c
  left join (select a.[snz_uid], a.[ir_ems_return_period_date],
  ↪  a.[ir_ems_gross_earnings_amt]
            from [IDI_Clean_202206].[ir_clean].[ird_ems] as a
                        where (year(a.[ir_ems_return_period_date])=2013 and
                        ↪  month(a.[ir_ems_return_period_date])<=3) or
                        ↪  (year(a.[ir_ems_return_period_date])=2012 and
                        ↪  month(a.[ir_ems_return_period_date])>=4)
                        ) as b
  on c.snz_uid=b.snz_uid
```

```
/*
Data Cleaning - Census13/TK13
*/


***** Data cleaning - Census/TK13 indicators, Education, Income, Gender & Age

```

```
8
9  * Importing Census13 ID's which were joined using SAS
10 import delimited "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OriginalData\ID13.csv"
11
12 ** snz_uid
13 label variable snz_uid "snz uid"
14
15 ** Census Indicator labels
16 label variable snz_cen_uid13 "MCensus 2013 ID"
17
18 * c13 variable label, 1 = Recorded in Census 2013
19 label variable c13 "Maori and present in Census 2013 - Indicator variable"
20 label define c13lbl 1 "MCensus13 ID"
21 label values c13 c13lbl
22
23 codebook c13
24
25 * d13c variable label, 0 = Non-Maori Descent in Census 2013, 1 = Maori Descent in Census
   ↪  2013
26 label variable d13c "Maori Descent Census 2013 - Indicator variable"
27 label define d13clbl 0 "Non-Maori Descent 2013" 1 "Maori Descent 2013"
28 label values d13c d13clbl
29
30 codebook d13c
31
32 * e13c variable label, 0 = Non-Maori Ethnicity 2013, 1 = Maori Ethnicity in Census 2013
33 label variable e13c "Maori Ethnicity Census 2013 - Indicator variable"
34 label define e13clbl 0 "Non-Maori Ethnicity 2013" 1 "Maori Ethnicity 2013"
35 label values e13c e13clbl
36
37 codebook e13c
38
39
40 * t13 variable label, 1 = Recorded in TK13
41 label variable t13 "Present in Te Kupenga 2013 - Indicator variable"
42 label define t13lbl 1 "TK13 Record"
43 label values t13 t13lbl
44
45 codebook t13
46
47
48 ** Education cleaning
49
50 tab cen_ind_highest_qual_code, missing
51 tab cen_ind_highest_qual_code d13c, missing
52 tab cen_ind_highest_qual_code e13c, missing
53
54 * Coding to education level, 0=no qual, 1=school, 2=post school, 3=uni
55 gen edu=0 if cen_ind_highest_qual_code==0
56 replace edu=1 if inrange(cen_ind_highest_qual_code,1,4)
57 replace edu=2 if inrange(cen_ind_highest_qual_code,5,10)
58 replace edu=3 if inrange(cen_ind_highest_qual_code,11,14)
59
60 * Creating labels for edu categories
61 label variable edu "Highest Qualification Categories - Census 2013 (0-3)"
62 label define edulbl 0 "No Qualification" 1 "School-level Qualification" 2 "Post-school
   ↪  Qualification" 3 "Degree Qualification"
63 label values edu edulbl
64
```

```
codebook edu

tab edu, missing

** Income cleaning
tab cen_ind_ttl_inc_code, missing

* Coding to income level, 0=loss/no income, 1=below 15k, 2=below 30k, 3=below 50k, 4=below 100k, 5=above 100k
gen cinc=0 if inrange(cen_ind_ttl_inc_code,11,12)
replace cinc=1 if inrange(cen_ind_ttl_inc_code,13,15)
replace cinc=2 if inrange(cen_ind_ttl_inc_code,16,18)
replace cinc=3 if inrange(cen_ind_ttl_inc_code,19,21)
replace cinc=4 if inrange(cen_ind_ttl_inc_code,22,24)
replace cinc=5 if inrange(cen_ind_ttl_inc_code,25,26)

* Creating labels for income levels
label variable cinc "Census 2013 Individual Income Categories (0-5)"
label define inclbl 0 "Loss/No Income" 1 "$1 - $15,000" 2 "$15,001 - $30,000" 3 "$30,001 - $50,000" 4 "$50,001 - $100,000" 5 "$100,001 and above"
label values cinc inclbl
note cinc: 2013 Census Individual Income from April 2012 - 31st March 2013

codebook cinc

tab cinc, missing

** Sex variable cleaning, 1=male 2=female
tab cen_ind_sex_code, missing

* Rename cen_ind_sex_code sex
gen sex=cen_ind_sex_code

* Creating labels for two sex categories
label variable sex "Sex Category - Census 2013 (1-2)"
label define sexlbl 1 "Male" 2 "Female"
label values sex sexlbl

codebook sex

tab sex, missing


** Age variable cleaning

tab cen_ind_age_code, missing

* Coding to 10-year age-bands, 1 = 15-24, 2 = 25-34, 3 = 35-44, 4 = 45-54, 5 = 55-64, 6 = 65-74, 7 = 75> (excludes <15 as not counted in Census)

gen age=1 if inrange(cen_ind_age_code,15,24)
replace age=2 if inrange(cen_ind_age_code,25,34)
replace age=3 if inrange(cen_ind_age_code,35,44)
replace age=4 if inrange(cen_ind_age_code,45,54)
replace age=5 if inrange(cen_ind_age_code,55,64)
replace age=6 if inrange(cen_ind_age_code,65,74)
replace age=7 if inrange(cen_ind_age_code,75,113)

tab age, missing
```

```
121
122 * Creating labels for age-bands
123 label variable age "10-Year Age-Bands - Census 2013 (1-7)"
124 label define agelbl 1 "15-24" 2 "25-34" 3 "35-44" 4 "45-54" 5 "55-64" 6 "65-74" 7 "75 and
    ↪  above"
125 label values age agelbl
126 note age: 2013 Census Age-Bands excluding 15<
127
128 codebook age
129
130 tab age, missing
131
132
133 ** dropping original variables
134 drop cen_ind_ttl_inc_code-cen_ind_age_code
135
136
137 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ID13.dta", replace
138
139 clear
140
141 ***********************************************
142 /*
143 Data Cleaning - ASH
144 */
145
146
147 ** Hospitalisations
148
149 import delimited "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OriginalData\hosp.csv"
150
151 * dropping census variables as will rejoin these later
152 drop cen_ind_ttl_inc_code-t13
153
154 * reformatting dates for stata readability
155 gen estd=date(moh_evt_evst_date,"DMY")
156 format estd %td
157
158
159
160
161 ** ICD recoding for Ambulatory Sensitive Hospitalisations (ASH) (age15+)
162
163 * ICD
164 gen icd1=substr(moh_dia_clinical_code,1,1)
165 gen icd_23=substr(moh_dia_clinical_code,2,2)
166 gen icd_45=substr(moh_dia_clinical_code,4,2)
167 gen icd_4=substr(moh_dia_clinical_code,4,1)
168 gen icd_5=substr(moh_dia_clinical_code,5,1)
169
170 destring icd_23, gen(icd23)
171 destring icd_4, gen(icd4)
172 destring icd_5, gen(icd5)
173 destring icd_45, gen(icd45)
174
175 * Ambulatory Sensitive Hospitalisations (ASH)
176
177 gen ash=0
178
```

```
*specifying only principle and secondary diagnoses
keep if inlist(moh_dia_diagnosis_type_code,"A","B")

*each entry has both ICD9 and ICD10, so keeping only ICD10 entries
keep if inrange(moh_dia_clinical_sys_code,10,15)

*keeping entry only if system code and clinical code match
keep if moh_dia_submitted_system_code==moh_dia_clinical_sys_code

******* HOWEVER, doing the above steps exludes all those who do not have a hospitalisation, therefore, will add these individuals back in later and give ASH=0

replace ash=1 if icd1=="R" & icd23==7 & inlist(icd4,2,3,4)
replace ash=1 if icd1=="R" & icd23==56 & inlist(icd4,0,8)
replace ash=1 if icd1=="R" & icd23==11

replace ash=1 if icd1=="I" & inlist(icd23,20,50,10,11,12,13,15,21,22,23,25,0,1,2,5,6,7,8,9,61,62,64,65,66)
replace ash=1 if icd1=="I" & icd23==67 & icd4==4
replace ash=1 if icd1=="I" & icd23==24 & inlist(icd4,1,0,8,9)


replace ash=1 if icd1=="J" & icd23==81
replace ash=1 if icd1=="J" & icd23==34
replace ash=1 if icd1=="J" & inrange(icd23,45,47)
replace ash=1 if icd1=="J" & inlist(icd23,13,14,15,16,18)
replace ash=1 if icd1=="J" & inlist(icd23,0,1,2,3,4,6)


replace ash=1 if icd1=="K" & inlist(icd23,2,4,5,59,21)
replace ash=1 if icd1=="K" & icd23==52 & icd4==9
replace ash=1 if icd1=="K" & inrange(icd23,25,28)


replace ash=1 if icd1=="L" & inlist(icd23,1,2,3,4,8,98)
replace ash=1 if icd1=="L" & inrange(icd23,20,30)

replace ash=1 if icd1=="H" & inlist(icd23,0,1)
replace ash=1 if icd1=="H" & inrange(icd23,65,67)


replace ash=1 if icd1=="A" & inrange(icd23,2,9)
replace ash=1 if icd1=="A" & inrange(icd23,50,60)
replace ash=1 if icd1=="A" & inrange(icd23,63,64)


replace ash=1 if icd1=="D" & inrange(icd23,50,53)

replace ash=1 if icd1=="E" & inrange(icd23,40,46)
replace ash=1 if icd1=="E" & inrange(icd23,50,56)
replace ash=1 if icd1=="E" & inrange(icd23,58,61)
replace ash=1 if icd1=="E" & icd23==63
replace ash=1 if icd1=="E" & inlist(icd23,10,11,13,14)
replace ash=1 if icd1=="E" & icd23==16 & icd4==2


replace ash=1 if icd1=="M" & icd23==83 & icd4==3
replace ash=1 if icd1=="M" & icd23==2 & icd4==3

```

```
236
237 replace ash=1 if icd1=="C" & icd23==53
238
239 replace ash=1 if icd1=="G" & inrange(icd23,40,41)
240
241 replace ash=1 if icd1=="O" & icd23==15
242
243 replace ash=1 if icd1=="N" & inlist(icd23,10,12)
244 replace ash=1 if icd1=="N" & icd23==13 & icd4==6
245 replace ash=1 if icd1=="N" & icd23==30 & icd4==9
246 replace ash=1 if icd1=="N" & icd23==39 & icd4==0
247 replace ash=1 if icd1=="N" & icd23==34 & icd4==1
248
249
250 tab ash, missing
251
252
253 * Creating labels for ASH related variables
254
255 label variable estd "Health Event Start Date (Formatted)"
256
257 * event dataset
258
259 label variable moh_evt_evst_date "Health Event Start Date"
260 label variable sty "Health Event Start Year"
261
262 label variable moh_evt_even_date "Health Event Discharge Date"
263 label variable eny "Health Event Discharge Year"
264
265 label variable moh_evt_event_id_nbr "Health Event Ref no."
266
267 label variable moh_evt_los_nbr "Health Event LOS (days)"
268
269
270 * diagnoses dataset
271
272 label variable moh_dia_event_id_nbr "Health Diag Ref no."
273
274 label variable moh_dia_diag_sequence_code "Health Diag Sequence Clin Code"
275
276 label variable moh_dia_clinical_code "Health Diag Clin Code"
277
278 label variable moh_dia_clinical_sys_code "Health Diag Clin Sys Code"
279
280 label variable moh_dia_submitted_system_code "Health Diag Clin Submitted Sys Code"
281
282 label variable moh_dia_diagnosis_type_code "Health Diag Diagnosis Priority"
283
284
285 * ASH event indicator
286 label variable ash "ASH Indicator"
287 label define ashbinarylbl 0 "No ASH Event" 1 "ASH Event"
288 label values ash ashbinarylbl
289
290 codebook ash
291
292
293
294 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH.dta", replace
```

```
295
296 clear
297
298 /*
299 ASH - cleaning & collapsing variables
300 */
301
302 ** Collapsing for ASH binary, frequency and length of stay
303
304 use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH.dta"
305
306 * collapsing to create entries with ASH outcome yes or no - 1st binary step
307
308 sort snz_uid
309 collapse (sum) ash (min) estd sty, by(snz_uid)
310
311 codebook snz_uid
312 tab ash, missing
313
314 gen ashbinary=0
315 replace ashbinary=1 if ash>0
316
317 * ASH event indicator
318 label variable ashbinary "ASH Binary Indicator"
319 *label define ashbinarylbl 0 "No ASH Event" 1 "ASH Event"
320 label values ashbinary ashbinarylbl
321
322 label variable estd "First Health Event Start Date"
323 label variable sty "First Health Event Start Year"
324
325
326 tab ashbinary, missing
327
328
329 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_binary.dta", replace
330
331 clear
332
333 *collapsing to create entries with freq no. of days - 2nd step no. of unique ASH events
334
335 use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH.dta"
336
337 duplicates drop moh_evt_event_id_nbr if ash==1, force
338 sort snz_uid
339 collapse (sum) ash, by(snz_uid)
340
341 codebook snz_uid
342 tab ash, missing
343
344 gen ashfreq=ash
345 tab ashfreq, missing
346 drop ash
347
348 * ASH frequency
349 label variable ashfreq "No. of Unique ASH Events"
350
351
352
353 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_freq.dta", replace
```

```
354
355 clear
356
357
358 * collapsing to create entries with length of stay - 3rd step, total length of stay during
    ↪  ASH event
359
360 use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH.dta"
361
362 replace moh_evt_los_nbr=0 if ash==0
363 replace moh_evt_los_nbr=0.5 if ash==1 & moh_evt_los_nbr==0
364
365
366 duplicates drop moh_evt_event_id_nbr if ash==1, force
367 sort snz_uid
368 collapse (sum) moh_evt_los_nbr, by(snz_uid)
369
370 codebook snz_uid
371 tab moh_evt_los_nbr, missing
372
373 gen ashlos=moh_evt_los_nbr
374 tab ashlos, missing
375
376 drop moh_evt_los_nbr
377
378 * ASH Length of Stay (LOS)
379 label variable ashlos "ASH Cumulative Length of Stay (LOS)"
380
381
382 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_los.dta", replace
383
384 clear
385
386 ***********************************************
387
388 /*
389 Data Cleaning - IRD Income Data
390 */
391
392 **IRD joining for cross checking income, through stata menu. generating numeric var from
    ↪  string var v3
393
394 import delimited
    ↪  "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OriginalData\irdApr12Mar13.csv",
    ↪  encoding(UTF-8)
395
396 tab v3 if v3=="0"
397 replace v3="0" if v3=="NULL"
398 gen ird=real(v3)
399
400 rename v1 snz_uid
401
402 sort snz_uid
403 collapse (sum) ird, by(snz_uid)
404
405 replace ird=. if ird==0
406
407 sum ird
408
```

```
codebook ird

** Coding to income level, 0=loss/no income, 1=below 15k, 2=below 30k, 3=below 50k, 4=below 100k, 5=above 100k
gen iinc=0 if ird==0
replace iinc=1 if ird>0 & ird<=15000
replace iinc=2 if ird>15000 & ird<=30000
replace iinc=3 if ird>30000 & ird<=50000
replace iinc=4 if ird>50000 & ird<=100000
replace iinc=5 if ird>100000
replace iinc=. if ird==.

tab iinc, missing

* Creating Variable labels
label variable ird "Raw Figures - IRD 2013 Individual Income"

* Creating labels for income levels
label variable iinc "IRD 2013 Individual Income Categories (0-5)"
label define inclbl 0 "Loss/No Income" 1 "$1 - $15,000" 2 "$15,001 - $30,000" 3 "$30,001 - $50,000" 4 "$50,001 - $100,000" 5 "$100,001 and above"
label values iinc inclbl
note iinc: 2013 IRD Individual Income from April 2012 - 31st March 2013

codebook iinc

tab iinc, missing


save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ird.dta", replace

clear

**********************************************
```

```
/*
Data Merging/Joining - ASH to C13TK13 & IRD
*/

** Merging main census dataset (ID13) with ASH to create main Cen13_ASH dataset

use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ID13.dta"


codebook snz_uid

merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_binary.dta"
drop ash
replace ashbinary=0 if ashbinary==.

tab ashbinary, missing

drop _merge


* merging to ASH freq dataset
```

```
23
24  merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_freq.dta"
25  replace ashfreq=0 if ashfreq==.
26
27  tab ashfreq, missing
28  tab ashfreq if ashfreq!=0
29
30  drop _merge
31
32
33  *merging to ASH LOS
34
35  merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_los.dta"
36  replace ashlos=0 if ashlos==.
37
38  tab ashlos, missing
39  tab ashlos if ashlos!=0
40
41  drop _merge
42
43
44  * Saving as merged dataset
45  save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_merge.dta", replace
46
47
48
49  ****** Merging IRD to dataset
50
51  *Merging ird data to dataset with ASH and Census info
52
53  use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ird.dta"
54
55
56  merge 1:1 snz_uid using
    ↪  "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_merge.dta"
57  drop _merge
58
59
60  tab iinc cinc, missing
61
62  save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta",
    ↪  replace
63
64  clear
65
```

B.2 National level Output (ASH)

```
1  /*
2  ASH - Output: frequency tables
3  */
4
5
6  ****** Creating frequency distributions for ASH by Maori vs non-Maori
7  use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta"
8
```

```
* ASH Maori vs non-Maori
tab ashbinary mbinary, co
tab mbinary ashbinary, co
table (mbinary) (ashbinary)

* ASH Maori vs non-Maori by age
bysort mbinary: tab ashbinary age, co

* ASH Maori vs non-Maori by sex
bysort mbinary: tab ashbinary sex, co

* ASH Maori vs non-Maori by education
bysort mbinary: tab ashbinary edu, co

* ASH Maori vs non-Maori by income
bysort mbinary: tab ashbinary cinc, co

* Education Maori vs non-Maori
tab mbinary edu, co

* Income Maori vs non-Maori
tab mbinary cinc, co

* cross-checking Census13 income with IRD income
tab iinc cinc, missing

clear


****** Creating smaller datasets with Census13 (Maori only)
use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta"

** Whole Maori sample
tab t13, missing
tab t13 if e13c==1 | d13c==1

keep if e13c==1 | d13c==1

codebook snz_uid

save
↪  "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\c13Maori-ASH_C13TK13_IRD_merge.dta",
↪  replace


clear

```

B.3 National level - Māori only (ASH)

```
/*
ASH - National level (Maori only)
*/

****** Creating smaller datasets with Census13 (Maori only)
use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta"

```

```
** Whole Maori sample
tab t13, missing
tab t13 if e13c==1 | d13c==1

keep if e13c==1 | d13c==1

codebook snz_uid

save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\c13Maori-ASH_C13TK13_IRD_merge.dta", replace

clear

/*
ASH - Regression analysis (Maori only)
*/

**********************************************

*** Maori Census13 sample
use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\c13Maori-ASH_C13TK13_IRD_merge.dta"


* ASH binary (logistic regression),
*i. for categorical, change base reference level using ib..., e.g. education 3 as base = ib3.edu
logistic ashbinary i.edu i.iinc i.sex i.age, coef base
logistic ashbinary i.edu i.iinc i.sex i.age, base


* ASH freq (negative binomial), first checking statistics, mean and variance (for overdispersion evidence)
tabstat ashfreq, by(edu) stats(mean v n)
tabstat ashfreq, by(sex) stats(mean v n)
tabstat ashfreq, by(age) stats(mean v n)
tabstat ashfreq, by(iinc) stats(mean v n)

* ASH LOS (negative binomial), first checking statistics, mean and variance (for overdispersion evidence)
tabstat ashlos, by(edu) stats(mean v n)
tabstat ashlos, by(sex) stats(mean v n)
tabstat ashlos, by(age) stats(mean v n)
tabstat ashlos, by(iinc) stats(mean v n)

* ASH freq and LOS, Negative binomial regressions

nbreg ashfreq i.edu i.iinc i.sex i.age, base
* using incident rate ratios
nbreg ashfreq i.edu i.iinc i.sex i.age, irr base

nbreg ashlos i.edu i.iinc i.sex i.age, base
* using incident rate ratios
nbreg ashlos i.edu i.iinc i.sex i.age, irr base


******* INTERACTION MODEL

```

```
* Regression incl interaction terms for whole Maori Census13 sample (interaction between edu
↪  and income)

* ASH binary (logistic regression),
*i. for categorical, change base reference level using ib...,
*e.g. education 3 as base = ib3.edu. ## = full interaction
logistic ashbinary i.edu##i.iinc i.sex i.age, coef base
logistic ashbinary i.edu##i.iinc i.sex i.age, base


* ASH freq (negative binomial)
nbreg ashfreq i.edu##i.iinc i.sex i.age, base
* using incident rate ratios
nbreg ashfreq i.edu##i.iinc i.sex i.age, irr base

* ASH LOS (negative binomial)
nbreg ashlos i.edu##i.iinc i.sex i.age, base
** using incident rate ratios
nbreg ashlos i.edu##i.iinc i.sex i.age, irr base

clear
```

B.4 Te Kupenga 2013 sample (ASH)

```
* Set up libraries;
libname security ODBC dsn=idi_clean_202206_srvprd  schema=security;
libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis";

*Extracting TK13 from ADHOC, but joining to security concordance table first;

proc sql;
        create table TK13_securitycon as
                select e.*, b.*
                        from (select a.*
                                        from tkp.TK_2013 a
                                       ) b
                        left join (select c.snz_uid, coalesce(d.snz_cen_uid) as
                        ↪  snz_cen_uid_sc
                                                from cen.census_individual_2013 c
                                                left join security.concordance d
                                                on c.snz_uid = d.snz_uid) e
                on e.snz_cen_uid_sc = b.snz_cen_uid;
quit;

proc export data=TK13_securitycon dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/TK13_securitycon.csv";
run;
```

```
/*
Data Merging - TK Variables
*/

** USING TK13 DATASET, JOINED TO SECURITY CONCORDANCE TABLE

import delimited "I:\MAA2021-43\Tori\Thesis\TK13_securitycon.csv"
```

```
8  codebook snz_uid
9
10 merge 1:1 snz_uid using
   ↪  "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta"
11 codebook snz_uid
12 tab t13 _merge, missing
13
14 drop _merge
15
16 codebook snz_uid
17
18 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta",
   ↪  replace
19
20 clear
21
22 ************************************************
23
24 /*
25 TK13 - Creating smaller dataset and merging disability
26 */
27
28
29 use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ASH_C13TK13_IRD_merge.dta"
30
31
32 ****** Creating smaller dataset with TK13 sample only
33
34 tab t13, missing
35 tab t13 if e13c==1 | d13c==1
36
37
38 *keeping only TK13 sample
39 keep if t13==1
40
41 codebook snz_uid
42
43 * merging in disability for TK13 sample
44
45 merge 1:1 snz_uid using
   ↪  "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\ID13_tk13_disability.dta"
46
47 order dsblty13, after(iinc)
48 drop _merge
49
50 save "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\tk13-ASH_FINAL_merge.dta",
   ↪  replace
51
52 clear
53
54 ************************************************
55
56 /*
57 TK13 - Cleaning
58 */
59
60
61 * TK variables
62 * ASH connection to turangawaewae
```

```
63
64  tab ashbinary qtttconnectturangawaewae, missing
65
66  gen conTWW=.
67  replace conTWW=11 if qtttconnectturangawaewae==11
68  replace conTWW=12 if qtttconnectturangawaewae==12
69  replace conTWW=13 if qtttconnectturangawaewae==13
70  replace conTWW=14 if qtttconnectturangawaewae==14
71  replace conTWW=15 if qtttconnectturangawaewae==15
72  replace conTWW=16 if qtttconnectturangawaewae==16
73  replace conTWW=88 if qtttconnectturangawaewae==88
74
75
76  *cleaning turangawaewae variable
77  label variable conTWW "Connection to Turangawaewae"
78  label define conTWWlbl 11 "Very Strong" 12 "Strong" 13 "Somewhat" 14 "Weak" 15 "Very Weak"
    ↪  16 "None" 88 "DK"
79  label values conTWW conTWWlbl
80
81  codebook conTWW
82
83  * ASH connection to importance of culture
84  tab ashbinary qvapimpofculture, missing
85
86  gen cultureimp=.
87  replace cultureimp=11 if qvapimpofculture==11
88  replace cultureimp=12 if qvapimpofculture==12
89  replace cultureimp=13 if qvapimpofculture==13
90  replace cultureimp=14 if qvapimpofculture==14
91  replace cultureimp=15 if qvapimpofculture==15
92  replace cultureimp=88 if qvapimpofculture==88
93
94
95  *cleaning importance of culture variable
96  label variable cultureimp "Importance of culture"
97  label define cultureimplbl 11 "Very" 12 "Quite" 13 "Somewhat" 14 "A little" 15 "None" 88
    ↪  "DK"
98  label values cultureimp cultureimplbl
99
100 codebook cultureimp
101
102 order conTWW-cultureimp, before(estd)
103
104
105 * Trust in healthcare system to treat fairly
106 codebook qcdtinsttrust_health
107 gen thealthfair=qcdtinsttrust_health
108
109 label variable thealthfair "Trust of Fair Healthcare"
110
111 codebook thealthfair
112
113 * Any discrimination while trying to get medical care
114 codebook qcdtanydiscrim_medical
115 gen meddiscr=.
116 replace meddiscr=0 if qcdtanydiscrim_medical==12
117 replace meddiscr=1 if qcdtanydiscrim_medical==11
118
119 label variable meddiscr "Any discrimination during medical care"
```

```
label define meddiscrlbl 0 "No" 1 "Yes"
label values meddiscr meddiscrlbl

codebook meddiscr

tab ashbinary meddiscr, missing


order thealthfair-meddiscr, after(cultureimp)


** Household Crowding, collapse into crowded (one or more bedrooms needed) vs. not

gen crowding13=0 if inlist(houcrowddv,3,4,5)
replace crowding13=1 if inlist(houcrowddv,1,2)

codebook crowding13

label variable crowding13 "CAN Crowding Cen13"
label define crowding13lbl 0 "Not crowded" 1 "Crowded"
label values crowding13 crowding13lbl

tab crowding13,missing


** NZDEP 2013
codebook nzdep2013

* Creating labels for nzdep levels
label variable nzdep2013 "NZDep TK13"
note nzdep2013: TK13 NZDep Index (1=least, 10=most)


** NZDEP 2013 QUINTILES
codebook nzdep2013

gen q_nzdep2013=1 if inlist(nzdep2013,1,2)
replace q_nzdep2013=2 if inlist(nzdep2013,3,4)
replace q_nzdep2013=3 if inlist(nzdep2013,5,6)
replace q_nzdep2013=4 if inlist(nzdep2013,7,8)
replace q_nzdep2013=5 if inlist(nzdep2013,9,10)

* Creating labels for nzdep quintile levels
label variable q_nzdep2013 "NZDep TK13 Quintiles"
label define q_nzdep2013lbl 1 "Dep 1&2" 2 "Dep 3&4" 3 "Dep 5&6" 4 "Dep 7&8" 5 "Dep 9&10"
label values q_nzdep2013 q_nzdep2013lbl
note q_nzdep2013: TK13 NZDep Quintiles (1=least, 5=most)


codebook q_nzdep2013


clear


**********************************************

/*
TK13 Analysis - Frequency Tables
```

```
*/


** Tables for TK13 only
use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\tk13-ASH_FINAL_merge.dta"


tab ashbinary, missing
tab ashbinary age, missing
tab ashbinary sex, missing

*comparing census income vs. ird income
tab cinc iinc, missing


** TK13 JK weight set up
svyset _n [pweight=finalwgt], jkrweight(finalwgt_*) vce(jackknife) singleunit(missing)

* using svy: includes weights, aweight=finalwgt only uses finalwgt without applying to
↪  population size (only within sample count)
tab ashbinary, missing
svy: tab ashbinary, missing
svy: tab ashbinary, missing count format(%14.3gc)
tab ashbinary [aweight=finalwgt]

*sex
        *cell proportions
svy: tab ashbinary sex, missing cell se format(%7.4f)
        *row proportions
svy: tab ashbinary sex, missing row se format(%7.4f)
        *table counts
svy: tab ashbinary sex, missing count se format(%14.3g)


*age
        *cell proportions
svy: tab ashbinary age, missing cell se format(%7.4f)
        *row proportions
svy: tab ashbinary age, missing row se format(%7.4f)
        *table counts
svy: tab ashbinary age, missing count se format(%14.3g)


*education
        *cell proportions
svy: tab ashbinary edu, missing cell se format(%7.4f)
        *row proportions
svy: tab ashbinary edu, missing row se format(%7.4f)
        *table counts
svy: tab ashbinary edu, missing count se format(%14.3g)


*ird income
        *cell proportions
svy: tab ashbinary iinc, missing cell se format(%7.4f)
        *row proportions
svy: tab ashbinary iinc, missing row se format(%7.4f)
        *table counts
svy: tab ashbinary iinc, missing count se format(%14.3g)
```

```
237
238
239 *disability
240         *cell proportions
241 svy: tab ashbinary dsblty13, missing cell se format(%7.4f)
242         *row proportions
243 svy: tab ashbinary dsblty13, missing row se format(%7.4f)
244         *table counts
245 svy: tab ashbinary dsblty13, missing count se format(%14.3g)
246
247
248 *connection to turangawaewae
249         *cell proportions
250 svy: tab ashbinary conTWW, missing cell se format(%7.4f)
251         *row proportions
252 svy: tab ashbinary conTWW, missing row se format(%7.4f)
253         *table counts
254 svy: tab ashbinary conTWW, missing count se format(%14.3g)
255
256
257 *culture importance
258         *cell proportions
259 svy: tab ashbinary cultureimp, missing cell se format(%7.4f)
260         *row proportions
261 svy: tab ashbinary cultureimp, missing row se format(%7.4f)
262         *table counts
263 svy: tab ashbinary cultureimp, missing count se format(%14.3g)
264
265
266 * Trust in healthcare system to treat fairly
267         *cell proportions
268 svy: tab ashbinary thealthfair, missing cell se format(%7.4f)
269         *row proportions
270 svy: tab ashbinary thealthfair, missing row se format(%7.4f)
271         *table counts
272 svy: tab ashbinary thealthfair, missing count se format(%14.3g)
273
274
275 * Any discrimination while trying to get medical care
276         *cell proportions
277 svy: tab ashbinary meddiscr, missing cell se format(%7.4f)
278         *row proportions
279 svy: tab ashbinary meddiscr, missing row se format(%7.4f)
280         *table counts
281 svy: tab ashbinary meddiscr, missing count se format(%14.3g)
282
283
284 ** Household Crowding,
285         *cell proportions
286 svy: tab ashbinary crowding13, missing cell se format(%7.4f)
287         *row proportions
288 svy: tab ashbinary crowding13, missing row se format(%7.4f)
289         *table counts
290 svy: tab ashbinary crowding13, missing count se format(%14.3g)
291
292
293 *demographic/geographic measures
294 *NZDep 2013 (quintiles)
295         *cell proportions
```

```
svy: tab ashbinary q_nzdep2013, missing cell se format(%7.4f)
        *row proportions
svy: tab ashbinary q_nzdep2013, missing row se format(%7.4f)
        *table counts
svy: tab ashbinary q_nzdep2013, missing count se format(%14.3g)


/*
TK13 Analysis - Regressions
*/

use "I:\MAA2021-43\Tori\Thesis\ASH_Project\Data\OutputData\tk13-ASH_FINAL_merge.dta"

** TK13 JK weight set up
svyset _n [pweight=finalwgt], jkrweight(finalwgt_*) vce(jackknife) singleunit(missing)

*individual

svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13, base

* all individual var
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.conTWW i.cultureimp
    ↪  thealthfair i.meddiscr, base

* all individual var (without connection to turangawaewae)
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.cultureimp thealthfair
    ↪  i.meddiscr, base

*individual (without connection to turangawaewae) and crowding
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.cultureimp thealthfair
    ↪  i.meddiscr i.crowding13, base

*individual (without connection to turangawaewae) and nzdep
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.cultureimp thealthfair
    ↪  i.meddiscr i.q_nzdep2013, base

svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 thealthfair i.meddiscr,
    ↪  base

* interaction between sex and age
svy: logistic ashbinary i.sex##i.age ib(3).edu ib(5).iinc i.dsblty13, base

*interaction between edu and income
svy: logistic ashbinary i.sex i.age ib(3).edu##ib(5).iinc i.dsblty13, base

*geographic
svy: logistic ashbinary i.q_nzdep2013, base

*hhld crowding and nzdep variables
svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.crowding13
    ↪  i.q_nzdep2013, base

svy: logistic ashbinary i.sex i.age ib(3).edu ib(5).iinc i.dsblty13 i.meddiscr, base


clear

cmdlog close
```

```
log close
```

B.5 Initial Joining (COVID-19 Vaccinations)

```
* Data Extraction from IDI

* Set up libraries;
libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
libname dat ODBC dsn=idi_clean_202206_srvprd schema=data;
libname mta ODBC dsn=idi_clean_202206_srvprd schema=metadata;
libname ird ODBC dsn=idi_clean_202206_srvprd schema=ir_clean;
libname moe ODBC dsn=idi_clean_202206_srvprd schema=moe_clean;
libname moh ODBC dsn=idi_clean_202206_srvprd schema=moh_clean;
libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data";

* Observation base - refresh 202206;

proc sql;
 create table base as
     select e.*, f.*
     from (select coalesce(a.snz_uid, b.snz_uid) as snz_uid,
                  coalesce(a.snz_cen_uid18, c.snz_cen_uid18) as snz_cen_uid18,
                  coalesce(b.snz_cen_uid13, d.snz_cen_uid13) as snz_cen_uid13,
                  a.c18, a.d18c, a.e18c,
                  b.c13, b.d13c, b.e13c,
                  c.t18, c.d18t, c.e18t,
                  d.t13, d.d13t, d.e13t
                  from (select x.snz_uid, x.snz_cen_uid as snz_cen_uid18
                              ,case when x.snz_uid is not null then 1 else 0 end as c18
                              ,case when x.cen_ind_maori_dscnt_output_code='1' then 1 else 0
                              ↪  end as d18c
                              ,case when x.cen_ind_ethgr_maori_ind_code in ('1', '2') then 1
                              ↪  else 0 end as e18c
                        from cen.census_individual_2018 x
                            where x.cen_ind_maori_dscnt_output_code='1' or
                            ↪  x.cen_ind_ethgr_maori_ind_code in ('1', '2')
                        ) a
                  full join (select y.snz_uid, y.snz_cen_uid as snz_cen_uid13
                                   ,case when y.snz_uid is not null then 1 else 0 end as
                                   ↪  c13
                                  ,case when y.cen_ind_recode_maori_dscnt_code='1' then 1
                                  ↪  else 0 end as d13c
                                  ,case when y.cen_ind_maori_eth_ind_code in ('1', '2')
                                  ↪  then 1 else 0 end as e13c
                             from cen.census_individual_2013 y
                                 where y.cen_ind_recode_maori_dscnt_code='1' or
                                 ↪  y.cen_ind_maori_eth_ind_code in ('1', '2')
                             ) b
                  on a.snz_uid = b.snz_uid
                  full join (select z.snz_cen_uid as snz_cen_uid18
                                   ,case when z.snz_cen_uid is not null then 1 else 0
                                   ↪  end as t18
                                  ,case when z.qDEMMaoriDescent='1' then 1 else 0 end as
                                  ↪  d18t
                                  ,case when z.DVEthTR_Maori='2' then 1 else 0 end as e18t
                             from tkp.TK_2018 z
```

```
                        ) c
                on a.snz_cen_uid18 = c.snz_cen_uid18
                full join (select u.snz_cen_uid as snz_cen_uid13
                                ,case when u.snz_cen_uid is not null then 1 else 0
                                ↪  end as t13
                            ,case when u.qDEMMaoriDescent=1 then 1 else 0 end as d13t
                            ,case when u.DVEthTR_Maori=2 then 1 else 0 end as e13t
                        from tkp.TK_2013 u
                        ) d
                on b.snz_cen_uid13 = d.snz_cen_uid13
        ) e
    left join (select v.snz_uid, v.snz_spine_ind, v.snz_deceased_year_nbr,
    ↪  v.snz_deceased_month_nbr
                from dat.personal_detail v
                where v.snz_spine_ind=1
                ) f
    on e.snz_uid = f.snz_uid;
quit;


proc export data=base dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/ASH_Project/Data/OriginalData/base.csv";
run;
```

```
/*
Data Joining/Linking - Census13/Census18
*/

*Setting Working Directory
cd "I:\MAA2021-43\Tori\Thesis\COVID_Project"

*Saving Log files:
cmdlog using "I:\MAA2021-43\Tori\Thesis\COVID_Project\LogFiles\Linking-commands.txt", append

log using "I:\MAA2021-43\Tori\Thesis\COVID_Project\LogFiles\Linking-results.txt", append
↪  text


**********************************************

* Using Base Census ID's and keeping if Maori (i.e. in the census or in TK13)
use "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\base.dta"


keep if snz_uid!=. & (c13!=. | t13!=.)

codebook snz_uid
codebook snz_uid if c13==1
codebook snz_uid if t13==1

keep snz_uid snz_cen_uid13 c13 t13 d13c e13c

save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\MID13.dta", replace
export delimited using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\MID13.csv", replace

clear
```

```
1
2  * Set up libraries;
3  libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
4  libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
5  libname data ODBC dsn=idi_clean_202206_srvprd  schema=data;
6  libname dia ODBC dsn=idi_clean_202206_srvprd  schema=dia_clean;
7  libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data";
8
9  proc import out=MID13 dbms=csv
10 file="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/MID13.csv";
11 run;
12
13
14
15 * Joining to Cen13 individual variables;
16
17 proc sql;
18    create table ID13ind as
19        select x.snz_uid,
20                                  x.snz_cen_uid as snz_cen_uid13,
21                                  x.snz_cen_hhld_uid as snz_cen_hhld_uid13,
22                                  x.snz_cen_fam_uid as snz_cen_fam_uid13,
23                                  x.snz_cen_extfam_uid as snz_cen_extfam_uid13,
24                                  x.snz_cen_dwell_uid as snz_cen_dwell_uid13,
25                                    x.cen_ind_sex_code as cen_ind_sex_code13,
26                                  x.cen_ind_age_code as cen_ind_age_code13,
27                                  x.cen_ind_dsblty_ind_code as cen_ind_dsblty_ind_code13,
28                                  x.cen_ind_ttl_inc_code as cen_ind_ttl_inc_code13,
29                                  x.cen_ind_highest_qual_code as cen_ind_highest_qual_code13,
                                    ↪  y.*
30        from cen.census_individual_2013 x
31            left join MID13 y
32            on x.snz_uid = y.snz_uid;
33 quit;
34
35 proc export data=ID13ind dbms=csv
36 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13ind.csv";
37 run;
38
39
40 * Joining to Cen13 hhld variables;
41
42
43 proc sql;
44  create table ID13hhld as
45      select a.*,
46                          b.snz_cen_hhld_uid as snz_cen_hhld_uid13,
47                          b.cen_hhd_hhld_comp_code as cen_hhd_hhld_comp_code13,
48                          b.cen_hhd_ttl_inc_hhld_code as cen_hhd_ttl_inc_hhld_code13
49      from ID13ind a
50      left join cen.census_household_2013 b
51      on a.snz_cen_hhld_uid13 = b.snz_cen_hhld_uid;
52 quit;
53
54 proc export data=ID13hhld dbms=csv
55 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13hhld.csv";
56 run;
57
```

```
58
59
60  * Joining Cen13 with indv and hhld variables, to Census 2018;
61
62
63  proc sql;
64   create table ID13_Cen18dwl as
65           select c.*, d.*
66          from ID13hhld c
67          left join (select b.cen_dwl_damp_code, b.cen_dwl_mould_code,
68                           a.snz_uid,
69                           a.snz_cen_uid as snz_cen_uid18,
70                           a.snz_cen_hhld_uid as snz_cen_hhld_uid18,
71                           a.snz_cen_fam_uid as snz_cen_fam_uid18,
72                           a.snz_cen_extfam_uid as snz_cen_extfam_uid18,
73                           a.ur_snz_cen_dwell_uid as ur_snz_cen_dwell_uid18,
74                           a.cen_ind_yrs_at_ur_code as cen_ind_yrs_at_ur_code18,
75                           a.cen_ind_dsblty_ind_code as cen_ind_dsblty_ind_code18,
76                           a.cen_ind_NZDep2018,
77                           a.cen_ind_age_code as cen_ind_age_code18,
78                           a.cen_ind_birth_month_nbr as cen_ind_birth_month_nbr18,
79                           a.cen_ind_birth_year_nbr as cen_ind_birth_year_nbr18,
80                           a.cen_ind_ttl_inc_code as cen_ind_ttl_inc_code18,
81                           a.cen_ind_hst_qual_code as cen_ind_hst_qual_code18
82                      from cen.census_individual_2018 a
83                           left join cen.census_dwelling_2018 b
84                           on a.ur_snz_cen_dwell_uid = b.snz_cen_dwell_uid
85                                  ) d
86          on c.snz_uid = d.snz_uid;
87  quit;
88
89  proc export data=ID13_Cen18dwl dbms=csv
90  outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18dwl.csv";
91  run;
92
93
94
95  *Join to hhd census;
96
97  proc sql;
98   create table ID13_Cen18 as
99       select a.*,
100                          b.snz_cen_hhld_uid as snz_cen_hhld_uid18,
101                          b.cen_hhd_composn_code as cen_hhd_composn_code18,
102                          b.cen_hhd_total_hhld_income_code as
                             ↪  cen_hhd_total_hhld_income_code18,
103                          b.cen_hhd_jen_hhold_income as cen_hhd_jen_hhold_income18,
104              b.cen_hhd_can_crowding_code as cen_hhd_can_crowding_code18
105       from ID13_Cen18dwl a
106       left join cen.census_household_2018 b
107       on a.snz_cen_hhld_uid18 = b.snz_cen_hhld_uid;
108 quit;
109
110 proc export data=ID13_Cen18 dbms=csv
111 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18.csv";
112 run;
113
114
115
```

```
116
117 *Joining ID13/Cen18 to Dia Death;
118
119 proc sql;
120    create table dth as
121        select x.snz_uid, y.*
122        from ID13ind x
123        left join (select a.dia_dth_death_month_nbr, a.dia_dth_death_year_nbr, a.snz_uid
124                   from dia.deaths a
125                   where a.dia_dth_death_year_nbr>2013
126                   ) y
127        on x.snz_uid = y.snz_uid;
128 quit;
129
130 proc export data=dth dbms=csv
131 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18_dth.csv";
132 run;
133
134
135
136 *Joining ID13/Cen18 to Overseas spell;
137
138 proc sql;
139    create table ovsp as
140        select x.snz_uid, y.*
141        from ID13ind x
142        left join (select a.*
143                   from data.person_overseas_spell a
144                   where a.pos_last_departure_ind = 'y') y
145        on x.snz_uid = y.snz_uid;
146 quit;
147
148 proc export data=ovsp dbms=csv
149 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18_ovsp.csv";
150 run;
151
152
153 *Joining ID13/Cen18 to Full DOB records;
154
155 proc sql;
156    create table dob as
157        select x.snz_uid, y.*
158        from ID13ind x
159        left join (select a.*
160                   from data.full_birth_date a
161                   ) y
162        on x.snz_uid = y.snz_uid;
163 quit;
164
165 proc export data=dob dbms=csv
166 outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18_dob.csv";
167 run;
168
169
170 *Joining ID13/Cen18 to Personal Detail DOB records;
171
172 proc sql;
173    create table pddob as
174        select x.snz_uid, y.*
```

```
        from ID13ind x
        left join (select a.snz_uid, a.snz_birth_year_nbr, a.snz_birth_month_nbr,
        ↪  a.snz_birth_date_proxy
                   from data.personal_detail a
                   ) y
        on x.snz_uid = y.snz_uid;
quit;

proc export data=pddob dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/COVID_Project/Data/ID13_Cen18_pddob.csv";
run;
```

```
/* IRD income Export for 2018 (April 2017 - March 31 2018) */

SELECT c.snz_uid, b.[ir_ems_return_period_date], b.[ir_ems_gross_earnings_amt]
  from [IDI_Clean_202206].[cen_clean].[census_individual_2013] as c
  left join (select a.[snz_uid], a.[ir_ems_return_period_date],
  ↪  a.[ir_ems_gross_earnings_amt]
            from [IDI_Clean_202206].[ir_clean].[ird_ems] as a
                        where (year(a.[ir_ems_return_period_date])=2018 and
                        ↪  month(a.[ir_ems_return_period_date])<=3) or
                        ↪  (year(a.[ir_ems_return_period_date])=2017 and
                        ↪  month(a.[ir_ems_return_period_date])>=4)
                        ) as b
  on c.snz_uid=b.snz_uid
```

```


 /* ID from Cen13 & Extracting Covid data (Status);*/


SELECT e.*, b.*
  from (select c.snz_uid, d.[snz_moh_uid]
               from [IDI_Clean_202206].[cen_clean].[census_individual_2013] c
               left join [IDI_Clean_202206].[security].[concordance] d
               on c.snz_uid=d.snz_uid) e
  left join (select a.[snz_moh_uid], a.[vaccination_status], a.[vacc_total_dose_cnt],
  ↪  a.[booster_dose_cnt],
                       a.[first_vacc_activity_date], a.[fully_vaccinated_activity_date],
                       ↪  a.[first_booster_activity_date]
                     from
                     ↪  [IDI_Adhoc].[clean_read_MOH_CIR].[moh_cir_vaccination_status_20230221]
                     ↪  a
                   ) b
  on e.snz_moh_uid=b.snz_moh_uid



/* ID from Cen13 & Extracting Covid data (Activity);*/


SELECT e.*, b.*
  from (select c.snz_uid, d.[snz_moh_uid]
               from [IDI_Clean_202206].[cen_clean].[census_individual_2013] c
               left join [IDI_Clean_202206].[security].[concordance] d
               on c.snz_uid=d.snz_uid) e
  left join (select a.[snz_moh_uid], a.[activity_date], a.[dhb_of_service]
```

```
                        from
                        ↪  [IDI_Adhoc].[clean_read_MOH_CIR].[moh_cir_vaccination_activity_20230221]
                        ↪  a
                    ) b
  on e.snz_moh_uid=b.snz_moh_uid


```

```
/*
Data Merging/Joining - Census13 to Census18, Mortality, Overseas spell \& Vaccinations
*/


import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.csv"

codebook snz_uid

save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace

clear

****** Merging mortality with main Census dataset (MORTALITY FOR 2013 ONWARDS)

import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_dth.csv"

codebook snz_uid

duplicates drop snz_uid, force

merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"
drop _merge

save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace

clear


***** Merging ovsp to main Census dataset
import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_ovsp.csv"

codebook snz_uid

merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"
drop _merge

save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace

clear


******** Merging DOB records

import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_pddob.csv"

codebook snz_uid

merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"
drop _merge
```

```
51
52  gen pdbirth=date(snz_birth_date_proxy,"DMY")
53  format pdbirth %td
54  drop snz_birth_year_nbr-snz_birth_date_proxy
55
56  * since pd birth doesn't include date, if month of birth is past june, then subtracting 1
    ↪  from age
57  gen pd_age=2020-year(pdbirth)
58  replace pd_age=pd_age-1 if month(pdbirth)>6
59  tab pd_age,missing
60
61
62  save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace
63
64  clear
65
66  ******** Merging COVID-19 Vaccinations
67
68  * COVID Activity (Event data) - Vaccination record event
69
70  import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ciract.csv", encoding(UTF-8)
71
72  drop v2
73  gen actd=date(v4,"YMD")
74  format actd %td
75
76  rename v1 snz_uid
77  rename v5 dhb
78  sort snz_uid actd
79
80  drop if actd==.
81  collapse (max) actd (last) dhb, by(snz_uid)
82
83  rename actd r_actd
84  rename dhb r_dhb
85
86  codebook snz_uid
87
88  * Merging with ID13 Census18 dataset
89
90  merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"
91  drop _merge
92
93
94  save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace
95
96
97  clear
98
99
100 * COVID Status (Individual data) - Vaccination Status
101
102 import delimited "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\cirstat.csv", encoding(UTF-8)
    ↪
103
104 drop v2
105
106 rename v1 snz_uid
107 codebook snz_uid
```

```
destring v3, generate (snz_moh_uid) force
drop v3

rename v4 vstat
rename v5 totdose
rename v6 bdose
rename v7 firv_date
rename v8 fullv_date
rename v9 firb_date

* reformatting variables
gen totdose2=real(totdose) if totdose!="NULL"
gen bdose2=real(bdose) if bdose!="NULL"

gen firv_date2=date(firv_date,"YMD")
format firv_date2 %td

gen fullv_date2=date(fullv_date,"YMD")
format fullv_date2 %td

gen firb_date2=date(firb_date,"YMD")
format firb_date2 %td

rename totdose2 totdose
rename bdose2 bdose
rename firv_date2 firv_date
rename fullv_date2 fullv_date
rename firb_date2 firb_date

* Merge to dataset with Event Data
merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"
drop _merge


save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta", replace

clear

**********************************************


```

```
/*
Data Cleaning - Census13/Census18
*/


* Using saved fully linked dataset
use "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18.dta"

* Checking amount of obsv lost between 2013 and 2018 Census
codebook snz_uid if snz_cen_uid18!=.
codebook snz_uid if snz_cen_uid18==.

codebook snz_uid if t13!=.


* Ordering variables
```

```
17
18  order snz_cen_uid13-cen_hhd_can_crowding_code18, after(snz_moh_uid)
19  order snz_cen_uid18-ur_snz_cen_dwell_uid18, after(snz_cen_dwell_uid13)
20
21  order snz_cen_uid18, after(snz_cen_uid13)
22  order snz_cen_hhld_uid18, after(snz_cen_hhld_uid13)
23  order snz_cen_fam_uid18, after(snz_cen_fam_uid13)
24  order snz_cen_extfam_uid18, after(snz_cen_extfam_uid13)
25  order snz_cen_dwell_uid13-ur_snz_cen_dwell_uid18, after(snz_cen_hhld_uid18)
26
27  drop c13
28  order d13c-t13, after(snz_cen_extfam_uid18)
29
30  order pdbirth-pd_age, after(snz_moh_uid)
31
32  * Labelling variables
33  label variable snz_uid "snz_uid"
34  label variable snz_moh_uid "snz_moh_uid"
35  label variable pdbirth "Birthday (Personal Details)"
36  label variable pd_age "Age in 2020 (Personal Details)"
37
38  label variable snz_cen_uid13 "ind_uid Cen13"
39  label variable snz_cen_uid18 "ind_uid Cen18"
40
41  label variable snz_cen_hhld_uid13 "hhld_uid Cen13"
42  label variable snz_cen_hhld_uid18 "hhld_uid Cen18"
43
44  label variable snz_cen_dwell_uid13 "dwell_uid Cen13"
45  label variable ur_snz_cen_dwell_uid18 "dwell_uid Cen18"
46
47  label variable snz_cen_fam_uid13 "fam_uid Cen13"
48  label variable snz_cen_fam_uid18 "fam_uid Cen18"
49
50  label variable snz_cen_extfam_uid13 "extfam_uid Cen13"
51  label variable snz_cen_extfam_uid18 "extfam_uid Cen18"
52
53
54  * d13c variable label, 0 = Non-Maori Descent in Census 2013, 1 = Maori Descent in Census
    ↪  2013
55  label variable d13c "Maori Descent Cen13"
56  label define d13clbl 0 "Non-Maori Desc" 1 "Maori Desc"
57  label values d13c d13clbl
58  note d13c: Maori Descent from Census 2013 - Indicator Variable
59
60
61  * e13c variable label, 0 = Non-Maori Ethnicity 2013, 1 = Maori Ethnicity in Census 2013
62  label variable e13c "Maori Ethnicity Cen13"
63  label define e13clbl 0 "Non-Maori Ethn" 1 "Maori Ethn"
64  label values e13c e13clbl
65  note d13c: Maori Ethnicity from Census 2013 - Indicator Variable
66
67
68  * t13 variable label, 1 = Recorded in TK13
69  label variable t13 "Present in TK13"
70  label define t13lbl 1 "TK13 Record"
71  label values t13 t13lbl
72  note t13: Present in Te Kupenga 2013 - Indicator Variable
73
74  codebook d13c
```

```
codebook e13c
codebook t13


**generating binary var for Maori vs non (Maori can be by descent, ethnicity or both)

gen mbinary=0
replace mbinary=1 if d13c==1 | e13c==1
order mbinary, after(e13c)

* mbinary variable label, 0 = Non-Maori, 1 = Maori
label variable mbinary "Maori Descent/Ethnicity Cen13"
label define mbinarylbl 0 "Non-Maori" 1 "Maori"
label values mbinary mbinarylbl
note mbinary: Maori Descent/Ethnicity from Census 2013 - Indicator Variable

tab mbinary, missing

* Generating binary of those present in C18 also
gen c18_pres=1 if snz_cen_uid18!=.
replace c18_pres=. if snz_cen_uid18==.

order c18_pres, after(snz_cen_extfam_uid18)

* c18_pres variable label, 1 = Present in C18
label variable c18_pres "Present in C18"
label define c18_preslbl 1 "C18 Record"
label values c18_pres c18_preslbl
note c18_pres: Original sample present in Census 2018 - Indicator Variable

codebook c18_pres

* Generating binary of those present in TK13 AND C18
gen TK13_c18=1 if t13==1 & snz_cen_uid18!=.
order TK13_c18, after(t13)

* TK13_c19 variable label, 1 = TK13 AND C18 Record
label variable TK13_c18 "Present in TK13 & C18"
label define TK13_c18lbl 1 "TK13&C18 Record"
label values TK13_c18 c18_preslbl
note TK13_c18: Original TK13 sample present in Census 2018

codebook TK13_c18


**** PREDICTOR CLEANING

** Sex cleaning, 1=male, 2=female

rename cen_ind_sex_code sex

tab sex, missing

* Creating labels for two sex categories
label variable sex "Sex Cen13"
label define sexlbl 1 "Male" 2 "Female"
label values sex sexlbl
note sex: Sex from Census 2013 - Indicator Variable

```

```
tab sex, missing

** Age cleaning, excluding <15

tab cen_ind_age_code13, missing

gen agec13=1 if inrange(cen_ind_age_code13,0,14)
replace agec13=2 if inrange(cen_ind_age_code13,15,24)
replace agec13=3 if inrange(cen_ind_age_code13,25,34)
replace agec13=4 if inrange(cen_ind_age_code13,35,44)
replace agec13=5 if inrange(cen_ind_age_code13,45,54)
replace agec13=6 if inrange(cen_ind_age_code13,55,64)
replace agec13=7 if inrange(cen_ind_age_code13,65,74)
replace agec13=8 if inrange(cen_ind_age_code13,75,113)

rename cen_ind_age_code13 agecont13
order agec13, after(agecont13)

tab agec13, missing

* Creating labels for age-bands
label variable agec13 "Age-Bands Cen13"
label define agelbl 1 "0-14" 2 "15-24" 3 "25-34" 4 "35-44" 5 "45-54" 6 "55-64" 7 "65-74" 8
↪  "75 and above"
label values agec13 agelbl
note agec13: Age-Bands from Census 2013

label variable agecont13 "Age Cen13 (Continuous)"

* 2018 Age

tab cen_ind_age_code18, missing

gen agec18=1 if inrange(cen_ind_age_code18,0,14)
replace agec18=2 if inrange(cen_ind_age_code18,15,24)
replace agec18=3 if inrange(cen_ind_age_code18,25,34)
replace agec18=4 if inrange(cen_ind_age_code18,35,44)
replace agec18=5 if inrange(cen_ind_age_code18,45,54)
replace agec18=6 if inrange(cen_ind_age_code18,55,64)
replace agec18=7 if inrange(cen_ind_age_code18,65,74)
replace agec18=8 if inrange(cen_ind_age_code18,75,113)

rename cen_ind_age_code18 agecont18
order agec18, after(agecont18)
order agecont18-agec18, after(agec13)

tab agec18, missing

* Creating labels for age-bands
label variable agec18 "Age-Bands Cen18"
label define agelbl 1 "0-14" 2 "15-24" 3 "25-34" 4 "35-44" 5 "45-54" 6 "55-64" 7 "65-74" 8
↪  "75 and above"
label values agec18 agelbl
note agec18: Age-Bands from Census 2018

label variable agecont18 "Age Cen18 (Continuous)"


* 2018 Census records - Birth Month & Birth Year
```

```
order cen_ind_birth_month_nbr18-cen_ind_birth_year_nbr18, after(agec18)


* Age in 2020 (beginning of outcome period for vaccinations)

* Age at beginning of outcome period for vaccinations

gen pd_age_cat=1 if inrange(pd_age,0,14)
replace pd_age_cat=2 if inrange(pd_age,15,24)
replace pd_age_cat=3 if inrange(pd_age,25,34)
replace pd_age_cat=4 if inrange(pd_age,35,44)
replace pd_age_cat=5 if inrange(pd_age,45,54)
replace pd_age_cat=6 if inrange(pd_age,55,64)
replace pd_age_cat=7 if inrange(pd_age,65,74)
replace pd_age_cat=8 if inrange(pd_age,75,113)

* Creating labels for age-bands
label variable pd_age_cat "Age-Bands in 2020 - Personal Detail"
label define agelbl 1 "0-14" 2 "15-24" 3 "25-34" 4 "35-44" 5 "45-54" 6 "55-64" 7 "65-74" 8
↪  "75 and above"
label values pd_age_cat agelbl
note pd_age_cat: Age-Bands during 2020 - Personal Detail

tab pd_age_cat, missing

label variable pd_age "Age in 2020 (Continuous) - Personal Detail"

order pdbirth-pd_age pd_age_cat, after(sex)



***** INDIVIDUAL LEVEL *****
** Income cleaning

** INDIVIDUAL Income cleaning (2013)
tab cen_ind_ttl_inc_code13, missing

* Coding to income level, 0=loss/no income, 1=below 15k, 2=below 30k, 3=below 50k, 4=below
↪  100k, 5=above 100k
gen cinc13=0 if inrange(cen_ind_ttl_inc_code13,11,12)
replace cinc13=1 if inrange(cen_ind_ttl_inc_code13,13,15)
replace cinc13=2 if inrange(cen_ind_ttl_inc_code13,16,18)
replace cinc13=3 if inrange(cen_ind_ttl_inc_code13,19,21)
replace cinc13=4 if inrange(cen_ind_ttl_inc_code13,22,24)
replace cinc13=5 if inrange(cen_ind_ttl_inc_code13,25,26)

* Creating labels for income levels
label variable cinc13 "Census 2013 Individual Income Categories (0-5)"
label define inclbl 0 "Loss/No Income" 1 "$1 - $15,000" 2 "$15,001 - $30,000" 3 "$30,001 -
↪  $50,000" 4 "$50,001 - $100,000" 5 "$100,001 and above"
label values cinc13 inclbl
note cinc13: 2013 Census Individual Income from April 2012 - 31st March 2013

codebook cinc13

tab cinc13, missing

** INDIVIDUAL Income cleaning (2018)
tab cen_ind_ttl_inc_code18, missing
```

```
247
248 * Coding to income level, 0=loss/no income, 1=below 15k, 2=below 30k, 3=below 50k, 4=below
    ↪  100k, 5=above 100k
249 gen cinc18=0 if inrange(cen_ind_ttl_inc_code18,11,12)
250 replace cinc18=1 if inrange(cen_ind_ttl_inc_code18,13,15)
251 replace cinc18=2 if inrange(cen_ind_ttl_inc_code18,16,18)
252 replace cinc18=3 if inrange(cen_ind_ttl_inc_code18,19,21)
253 replace cinc18=4 if inrange(cen_ind_ttl_inc_code18,22,24)
254 replace cinc18=5 if inrange(cen_ind_ttl_inc_code18,25,26)
255
256 * Creating labels for income levels
257 label variable cinc18 "Census 2018 Individual Income Categories (0-5)"
258 label define inclbl 0 "Loss/No Income" 1 "$1 - $15,000" 2 "$15,001 - $30,000" 3 "$30,001 -
    ↪  $50,000" 4 "$50,001 - $100,000" 5 "$100,001 and above"
259 label values cinc18 inclbl
260 note cinc18: 2018 Census Individual Income from April 2017 - 31st March 2018
261
262 codebook cinc18
263
264 tab cinc18, missing
265
266 ** Highest Qualification (2013)
267
268 * Coding to education level, 0=no qual, 1=school, 2=post school, 3=uni
269 gen edu13=0 if cen_ind_highest_qual_code13==0
270 replace edu13=1 if inrange(cen_ind_highest_qual_code13,1,4)
271 replace edu13=2 if inrange(cen_ind_highest_qual_code13,5,10)
272 replace edu13=3 if inrange(cen_ind_highest_qual_code13,11,14)
273
274 * Creating labels for edu categories
275 label variable edu13 "Highest Qualification Categories - Census 2013 (0-3)"
276 label define edulbl 0 "No Qualification" 1 "School-level Qualification" 2 "Post-school
    ↪  Qualification" 3 "Degree Qualification"
277 label values edu13 edulbl
278
279 codebook edu13
280
281 tab edu13, missing
282
283 ** Highest Qualification (2018)
284
285 * Coding to education level, 0=no qual, 1=school, 2=post school, 3=uni
286 gen edu18=0 if cen_ind_hst_qual_code18==0
287 replace edu18=1 if inrange(cen_ind_hst_qual_code18,1,4)
288 replace edu18=2 if inrange(cen_ind_hst_qual_code18,5,10)
289 replace edu18=3 if inrange(cen_ind_hst_qual_code18,11,14)
290
291 * Creating labels for edu categories
292 label variable edu18 "Highest Qualification Categories - Census 2018 (0-3)"
293 label define edulbl 0 "No Qualification" 1 "School-level Qualification" 2 "Post-school
    ↪  Qualification" 3 "Degree Qualification"
294 label values edu18 edulbl
295
296 codebook edu18
297
298 tab edu18, missing
299
300 order cinc13-edu18, after(agec18)
```

```
drop cen_ind_ttl_inc_code13 cen_ind_ttl_inc_code18 cen_ind_highest_qual_code13
    cen_ind_hst_qual_code18


** Disability variablecleaning

*2013 Disability
rename cen_ind_dsblty_ind_code13 dsblty13
replace dsblty13=. if dsblty13==7 | dsblty13==9

tab dsblty13, missing

* Creating labels for two disability categories
label variable dsblty13 "Disability Cen13"
label define dsblty13lbl 0 "No Disability" 1 "Have Disability"
label values dsblty13 dsblty13lbl
note dsblty13: Disability Classification from Census 2013

tab dsblty13, missing

* 2018 Disability

rename cen_ind_dsblty_ind_code18 dsblty18
replace dsblty18=. if dsblty18==7 | dsblty18==9
tab dsblty18, missing

* Creating labels for two disability categories
label variable dsblty18 "Disability Cen18"
label define dsblty18lbl 0 "No Disability" 1 "Have Disability"
label values dsblty18 dsblty18lbl
note dsblty18: Disability Classification from Census 2018

tab dsblty18, missing

order dsblty18, after(dsblty13)

tab dsblty13 dsblty18, missing


***** HOUSEHOLD LEVEL *****
* 2013 Household Income
tab cen_hhd_ttl_inc_hhld_code13, missing

* Coding to UPDATED hhld income level, 1 = $20,000 or less, 2 = $20,001 - $30,000, 3 =
    $30,001 - $50,000, 4 = $50,001 - $70,000, 5 = $70,001 - $100,000, 6 = $100,001 -
    $150,000, 7 = $150,000 >, 8 = Not Stated
gen th_inc13=1 if inrange(cen_hhd_ttl_inc_hhld_code13,11,16)
replace th_inc13=2 if inrange(cen_hhd_ttl_inc_hhld_code13,17,18)
replace th_inc13=3 if inrange(cen_hhd_ttl_inc_hhld_code13,19,21)
replace th_inc13=4 if inrange(cen_hhd_ttl_inc_hhld_code13,22,23)
replace th_inc13=5 if cen_hhd_ttl_inc_hhld_code13 == 24
replace th_inc13=6 if cen_hhd_ttl_inc_hhld_code13 == 25
replace th_inc13=7 if cen_hhd_ttl_inc_hhld_code13 == 26


* Creating labels for income levels
label variable th_inc13 "Hhld Income Cen13"
```

```
label define th_inclbl 1 "$20,000 or less" 2 "$20,001 - $30,000" 3 "$30,001 - $50,000" 4 "$50,001 - $70,000" 5 "$70,001 - $100,000" 6 "$100,001 - $150,000" 7 "$150,001 and above"
label values th_inc13 th_inclbl
note th_inc13: Total Household Income from Census 2013

codebook th_inc13
order th_inc13, after(cen_ind_birth_year_nbr18)

tab th_inc13, missing

drop cen_hhd_ttl_inc_hhld_code13

* 2018 Household Income
tab cen_hhd_total_hhld_income_code18, missing

* Coding to UPDATED hhld income level, 1 = $20,000 or less, 2 = $20,001 - $30,000, 3 = $30,001 - $50,000, 4 = $50,001 - $70,000, 5 = $70,001 - $100,000, 6 = $100,001 - $150,000, 7 = $150,000 >, 8 = Not Stated
gen th_inc18=1 if inrange(cen_hhd_total_hhld_income_code18,11,16)
replace th_inc18=2 if inrange(cen_hhd_total_hhld_income_code18,17,18)
replace th_inc18=3 if inrange(cen_hhd_total_hhld_income_code18,19,21)
replace th_inc18=4 if inrange(cen_hhd_total_hhld_income_code18,22,23)
replace th_inc18=5 if cen_hhd_total_hhld_income_code18 == 24
replace th_inc18=6 if cen_hhd_total_hhld_income_code18 == 25
replace th_inc18=7 if cen_hhd_total_hhld_income_code18 == 26

* Creating labels for income levels
label variable th_inc18 "Hhld Income Cen18"
label define th_inclbl 1 "$20,000 or less" 2 "$20,001 - $30,000" 3 "$30,001 - $50,000" 4 "$50,001 - $70,000" 5 "$70,001 - $100,000" 6 "$100,001 - $150,000" 7 "$150,001 and above"
label values th_inc18 th_inclbl
note th_inc18: Total Household Income from Census 2018

codebook th_inc18
order th_inc18, after(th_inc13)

tab th_inc18, missing

drop cen_hhd_total_hhld_income_code18

tab th_inc13 th_inc18, missing


** Household Composition cleaning

* 2013 Household Composition cleaning
*0 = one person hhld, 1 = one couple hhld, 2 = couple with children, 3 = one parent with children, 4 = two family, three or more family hhld, 5 = other multiperson hhld

gen hhldcomp13a=0 if inlist(cen_hhd_hhld_comp_code13,511)
replace hhldcomp13a=1 if inrange(cen_hhd_hhld_comp_code13, 100, 122)
replace hhldcomp13a=2 if inrange(cen_hhd_hhld_comp_code13,131,142)
replace hhldcomp13a=3 if inrange(cen_hhd_hhld_comp_code13,151,162)
replace hhldcomp13a=4 if inrange(cen_hhd_hhld_comp_code13,200,241)
replace hhldcomp13a=4 if inlist(cen_hhd_hhld_comp_code13,311)
replace hhldcomp13a=5 if inrange(cen_hhd_hhld_comp_code13,400,431)

```

```
label variable hhldcomp13a "Hhld Comp Cen13a"
label define hhldcomp13albl 0 "1 person" 1 "1 Couple" 2 "Couple w Children" 3 "1 Parent w Children" 4 "2> Families" 5 "Other Multiperson"
label values hhldcomp13a hhldcomp13albl

codebook hhldcomp13a

tab hhldcomp13a, missing

order hhldcomp13a, after(hhldcomp13)

drop cen_hhd_hhld_comp_code13


*2018 Household Composition cleaning

*0 = one person hhld, 1 = one couple hhld, 2 = couple with children, 3 = one parent with children, 4 = two family, three or more family hhld, 5 = other multiperson hhld

gen hhldcomp18a=0 if inlist(cen_hhd_composn_code18,511)
replace hhldcomp18a=1 if inrange(cen_hhd_composn_code18, 100, 122)
replace hhldcomp18a=2 if inrange(cen_hhd_composn_code18,131,142)
replace hhldcomp18a=3 if inrange(cen_hhd_composn_code18,151,162)
replace hhldcomp18a=4 if inrange(cen_hhd_composn_code18,200,241)
replace hhldcomp18a=4 if inlist(cen_hhd_composn_code18,311)
replace hhldcomp18a=5 if inrange(cen_hhd_composn_code18,400,431)

label variable hhldcomp18a "Hhld Comp Cen18a"
label define hhldcomp18albl 0 "1 person" 1 "1 Couple" 2 "Couple w Children" 3 "1 Parent w Children" 4 "2> Families" 5 "Other Multiperson"
label values hhldcomp18a hhldcomp18albl

codebook hhldcomp18a

tab hhldcomp18a, missing

order hhldcomp18a, after(hhldcomp18)

drop cen_hhd_composn_code18


** Housing Quality (Mould/Dampness) 2018 ONLY

* = 0 if NOT mouldy/damp
gen housingqual18=0 if cen_dwl_damp_code==3 | cen_dwl_mould_code==3
*ALSO =0 if sometimes mouldy/damp
replace housingqual18=0 if cen_dwl_damp_code==2 | cen_dwl_mould_code==2

* = 1 if ALWAYS mould/dampness
replace housingqual18=1 if cen_dwl_damp_code==1 | cen_dwl_mould_code==1

codebook housingqual18

label variable housingqual18 "Damp/Mould Cen18"
label define hqllbl 0 "No/sometimes damp/mould" 1 "Always damp/mould"
label values housingqual18 hqllbl

tab housingqual18, missing

```

```
463 order housingqual18, after(cen_dwl_mould_code)
464 drop cen_dwl_damp_code cen_dwl_mould_code
465
466 ** Household Crowding, collapse into crowded (one or more bedrooms needed) vs. not
467
468 gen crowding18=0 if inlist(cen_hhd_can_crowding_code,3,4,5)
469 replace crowding18=1 if inlist(cen_hhd_can_crowding_code,1,2)
470
471 codebook crowding18
472
473 label variable crowding18 "CAN Crowding Cen18"
474 label define crowding18lbl 0 "Not crowded" 1 "Crowded"
475 label values crowding18 crowding18lbl
476
477 tab crowding18,missing
478
479 order crowding18, after(hhldcomp18a)
480 drop cen_hhd_can_crowding_code
481
482
483 ***** GEOGRAPHIC LEVEL *****
484
485 ** NZDEP 2018
486 codebook cen_ind_nzdep2018
487 rename cen_ind_nzdep2018 nzdep2018
488
489 * Creating labels for nzdep levels
490 label variable nzdep2018 "NZDep Cen18"
491 note nzdep2018: Census 2018 NZDep Index (1=least, 10=most)
492
493 order nzdep2018, after(housingqual18)
494
495 ** NZDEP 2018 QUINTILES
496 codebook nzdep2018
497
498 gen q_nzdep2018=1 if inlist(nzdep2018,1,2)
499 replace q_nzdep2018=2 if inlist(nzdep2018,3,4)
500 replace q_nzdep2018=3 if inlist(nzdep2018,5,6)
501 replace q_nzdep2018=4 if inlist(nzdep2018,7,8)
502 replace q_nzdep2018=5 if inlist(nzdep2018,9,10)
503
504 * Creating labels for nzdep quintile levels
505 label variable q_nzdep2018 "NZDep Cen18 Quintiles"
506 label define q_nzdep2018lbl 1 "Dep 1&2" 2 "Dep 3&4" 3 "Dep 5&6" 4 "Dep 7&8" 5 "Dep 9&10"
507 label values q_nzdep2018 q_nzdep2018lbl
508 note q_nzdep2018: Census 2018 NZDep Quintiles (1=least, 5=most)
509
510 order q_nzdep2018, after(nzdep2018)
511
512
513 ** OUTCOME CLEANING
514 **** Vaccination outcome
515
516 * Labelling vacc variables
517 label variable totdose "Vaccine Total Dose Count - Status"
518 label variable bdose "Booster Dose Count - Status"
519 label variable firv_date "First Vacc Date - Status"
520 label variable fullv_date "Fully Vacc Date - Status"
521 label variable firb_date "First Booster Date - Status"
```

```
522
523 label variable r_actd "Collapsed Vacc Date (Recent) - Activity"
524 label variable r_dhb "DHB of Vacc (Recent) - Activity"
525
526
527 *defining vaccination categories, to reflect increasing vaccination levels
528 * vcat variable: 0=Non-Vacc, 1=Partially, 2= Fully,
529 gen vcat=2 if vstat=="FULLY VACCINATED"
530 replace vcat=0 if vstat=="NULL"
531 replace vcat=0 if vstat==.
532 replace vcat=1 if vstat=="PARTIALLY VACCINATED"
533
534 label variable vcat "Vacc Status Category"
535 label define vcatlbl 0 "No Vaccination" 1 "Partially Vaccinated" 2 "Fully Vaccinated"
536 label values vcat vcatlbl
537
538 tab vcat, missing
539
540 order vcat, before(totdose)
541
542
543 save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR.dta", replace
544
545
546 ***********************************************
```

```
1 /*
2 Data Merging  - Dataset merging to mortality and overseas spell
3 */
4
5
6 * Using saved fully linked dataset
7 use "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR.dta"
8
9 ** Beginning to drop those who left country and/or died before 2020 (Vaccination period)
10
11 tab vcat if year(f_actd) <2020
12
13 * setting start date to beginning of vaccination period (using beginning of 2020)
14 gen stdate=mdy(1,1,2020)
15 format stdate %td
16
17 * left country date
18 gen lcday=substr(pos_applied_date,1,9)
19 gen lcdate=date(lcday,"DMY")
20 format lcdate %td
21 drop lcday
22
23 ****** OVSP BEFORE COVID VACC PERIOD
24 * Table of those who permanently left the country before period start date in 2020
25 codebook snz_uid if lcdate<stdate
26
27 ** Ovsp & Vaccination records present
28 * those who have permanently left the country, but have vaccination records after start date
   ↪  (vaccinated, non and partially)
29 tab vcat if pos_last_departure_ind=="y" & lcdate<stdate
30
31 * those who have permanently left the country, but have vaccination records after start date
   ↪  (only those with vaccinations (full and partially)
```

```
tab vcat if pos_last_departure_ind=="y" & lcdate<stdate & vcat!=0

* amount to be dropped from Ovsp data
sum snz_uid if lcdate<stdate


****** DEATHS BEFORE COVID VACC PERIOD
* creating death date variable
gen dthdate=mdy(dia_dth_death_month_nbr,15,dia_dth_death_year_nbr)
format dthdate %td

* Table of those who died before period start date in 2020
codebook snz_uid if dthdate<stdate
sum snz_uid if dthdate<stdate

** Deaths & Vaccination records present
* those who have vaccination record, but died before 2020. Viewing for linkage bias?

* Vaccinated, but died before 2020 - this gives the value for impossible scenario
tab vcat if dia_dth_death_year_nbr<2020


*** VIEWING TABLE OF THOSE EXCLUDED FROM SAMPLE USING ABOVE CONDITIONS

* using above conditions, showing the number to be dropped
codebook snz_uid if (pos_last_departure_ind=="y" & lcdate<stdate) | dthdate<stdate |
    (dia_dth_death_year_nbr<2020 & dthdate<f_actd)

** VIEWING TABLE OF THOSE INCLUDED IF ABOVE ARE DROPPED
codebook snz_uid if pos_last_departure_ind!="y" | lcdate>stdate | dthdate>stdate |
    dthdate>f_actd

codebook snz_uid if ((pos_last_departure_ind=="y" & lcdate>stdate) | dthdate>stdate |
    (dia_dth_death_year_nbr>2020 & dthdate>f_actd))


** CREATING INDICATOR VARIABLES for left country and deaths

* FOR THOSE WHO LEFT COUNTRY BEFORE COVID VACC PERIOD
gen lc_before =.
replace lc_before=1 if lcdate<stdate

* FOR THOSE WHO DIED BEFORE COVID VACC PERIOD
gen dth_before =.
replace dth_before=1 if dthdate<stdate

save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR.dta", replace

**********************************************
```

```

* Set up libraries;
libname security ODBC dsn=idi_clean_202206_srvprd  schema=security;
libname cen ODBC dsn=idi_clean_202206_srvprd schema=cen_clean;
libname tkp ODBC dsn=idi_adhoc schema=clean_read_TK;
libname Tori "/nas/DataLab/MAA/MAA2021-43/Tori/Thesis";


```

```
*Extracting TK13 from ADHOC, but joining to security concordance table first;

proc sql;
        create table TK13_securitycon as
                select e.*, b.*
                        from (select a.*
                                        from tkp.TK_2013 a
                                       ) b
                        left join (select c.snz_uid, coalesce(d.snz_cen_uid) as
                        ↪  snz_cen_uid_sc
                                        from cen.census_individual_2013 c
                                        left join security.concordance d
                                        on c.snz_uid = d.snz_uid) e
                on e.snz_cen_uid_sc = b.snz_cen_uid;
quit;

proc export data=TK13_securitycon dbms=csv
outfile="/nas/DataLab/MAA/MAA2021-43/Tori/Thesis/TK13_securitycon.csv";
run;
```

```
*/
Data Merging - TK Variables
*/

** USING NEW TK13 DATASET, JOINED TO SECURITY CONCORDANCE TABLE

*merging to full dataset (without deaths and overseas dropped from sample)

import delimited "I:\MAA2021-43\Tori\Thesis\TK13_securitycon.csv"
codebook snz_uid

merge 1:1 snz_uid using "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR.dta"
codebook snz_uid

tab t13 _merge, missing


*checks after merge
tab lc_before if _merge==3, missing
tab lc_before t13, missing

tab dth_before if _merge==3, missing
tab dth_before t13, missing


drop _merge



*dropping those not in TK13

keep if t13==1

codebook snz_uid


save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR_TK13merge.dta", replace
```

```
38
39
40  ***********************************************
41
42  */
43  TK13 - Merging TK18 Variables to dataset \& Cleaning
44  */
45
46  * TK variables
47  use "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR_TK13merge.dta"
48
49  * whanau doing well
50  codebook qwhawhanaudoingwell
51  gen doingwell=qwhawhanaudoingwell
52  label variable doingwell "Whanau doing well? (0 = extremely badly, 10 = extremely well)"
53
54  codebook doingwell
55  tab vcat doingwell, missing
56
57  order doingwell, after(cen_hhd_jen_hhold_income18)
58
59  * whanau getting along
60
61  gen getalong=.
62  replace getalong=11 if qwhawhanaugetalong==11
63  replace getalong=12 if qwhawhanaugetalong==12
64  replace getalong=13 if qwhawhanaugetalong==13
65  replace getalong=14 if qwhawhanaugetalong==14
66  replace getalong=15 if qwhawhanaugetalong==15
67  replace getalong=88 if qwhawhanaugetalong==88
68
69  *cleaning whanau get along variable
70  label variable getalong "Whanau Get Along"
71  label define getalonglbl 11 "Very Well" 12 "Well" 13 "Neither well/bad" 14 "Badly" 15 "Very
    ↪  Badly" 88 "DK"
72  label values getalong getalonglbl
73
74  codebook getalong
75  tab vcat getalong, missing
76
77  * connection to turangawaewae
78  tab ashbinary qtttconnectturangawaewae, missing
79
80  gen conTWW=.
81  replace conTWW=11 if qtttconnectturangawaewae==11
82  replace conTWW=12 if qtttconnectturangawaewae==12
83  replace conTWW=13 if qtttconnectturangawaewae==13
84  replace conTWW=14 if qtttconnectturangawaewae==14
85  replace conTWW=15 if qtttconnectturangawaewae==15
86  replace conTWW=16 if qtttconnectturangawaewae==16
87  replace conTWW=88 if qtttconnectturangawaewae==88
88
89  *cleaning turangawaewae variable
90  label variable conTWW "Connection to Turangawaewae"
91  label define conTWWlbl 11 "Very Strong" 12 "Strong" 13 "Somewhat" 14 "Weak" 15 "Very Weak"
    ↪  16 "None" 88 "DK"
92  label values conTWW conTWWlbl
93
94  codebook conTWW
```

```
95
96  order conTWW, after(getalong)
97
98
99  * importance of culture
100
101 gen cultureimp=.
102 replace cultureimp=11 if qvapimpofculture==11
103 replace cultureimp=12 if qvapimpofculture==12
104 replace cultureimp=13 if qvapimpofculture==13
105 replace cultureimp=14 if qvapimpofculture==14
106 replace cultureimp=15 if qvapimpofculture==15
107 replace cultureimp=88 if qvapimpofculture==88
108
109 *cleaning importance of culture variable
110 label variable cultureimp "Importance of culture"
111 label define cultureimplbl 11 "Very" 12 "Quite" 13 "Somewhat" 14 "A little" 15 "None" 88
    ↪  "DK"
112 label values cultureimp cultureimplbl
113
114 codebook cultureimp
115
116 tab vcat cultureimp, missing
117
118
119 * Trust in healthcare system to treat fairly
120 codebook qcdtinsttrust_health
121 gen thealthfair=qcdtinsttrust_health
122
123 label variable thealthfair "Trust of Fair Healthcare"
124
125 codebook thealthfair
126
127 * Any discrimination while trying to get medical care
128 codebook qcdtanydiscrim_medical
129 gen meddiscr=.
130 replace meddiscr=0 if qcdtanydiscrim_medical==12
131 replace meddiscr=1 if qcdtanydiscrim_medical==11
132
133 label variable meddiscr "Any discrimination during medical care?"
134 label define meddiscrlbl 0 "No" 1 "Yes"
135 label values meddiscr meddiscrlbl
136
137 codebook meddiscr
138
139 tab vcat meddiscr, missing
140
141
142 order getalong-meddiscr, after(doingwell)
143 order pho_enroldate-phodhb_merge, after(meddiscr)
144
145
146 ** Household Crowding, collapse into crowded (one or more bedrooms needed) vs. not
147
148 gen crowding13=0 if inlist(houcrowddv,3,4,5)
149 replace crowding13=1 if inlist(houcrowddv,1,2)
150
151 codebook crowding13
152
```

```
label variable crowding13 "CAN Crowding TK13"
label define crowding13lbl 0 "Not crowded" 1 "Crowded"
label values crowding13 crowding13lbl

tab crowding13, missing

order crowding13, before(crowding18)

** NZDEP 2013
codebook nzdep2013

* Creating labels for nzdep levels
label variable nzdep2013 "NZDep TK13"
note nzdep2013: TK13 NZDep Index (1=least, 10=most)

order nzdep2013, before(nzdep2018)


** NZDEP 2013 QUINTILES
codebook nzdep2013

gen q_nzdep2013=1 if inlist(nzdep2013,1,2)
replace q_nzdep2013=2 if inlist(nzdep2013,3,4)
replace q_nzdep2013=3 if inlist(nzdep2013,5,6)
replace q_nzdep2013=4 if inlist(nzdep2013,7,8)
replace q_nzdep2013=5 if inlist(nzdep2013,9,10)

* Creating labels for nzdep quintile levels
label variable q_nzdep2013 "NZDep TK13 Quintiles"
label define q_nzdep2013lbl 1 "Dep 1&2" 2 "Dep 3&4" 3 "Dep 5&6" 4 "Dep 7&8" 5 "Dep 9&10"
label values q_nzdep2013 q_nzdep2013lbl
note q_nzdep2013: TK13 NZDep Quintiles (1=least, 5=most)

order q_nzdep2013, after(nzdep2013)

codebook q_nzdep2013


save "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR_TK13merge.dta", replace

***********************************************
```

B.6 Te Kupenga 2013 sample (COVID-19 Vaccinations)

```
/*
TK13 Analysis - Frequency Tables \& Regression Analysis
*/

use "I:\MAA2021-43\Tori\Thesis\COVID_Project\Data\ID13_Cen18_CIR_TK13merge.dta"

/*
* Opening dataset

* INCLUDES:
- TK13 only sample
- TK13 variables
```

```
- No individuals dropped (who have died or left country), due to issues with using JK
↪  weights for sample
*/

* Census 2013 to 2018 transition tables (unweighted)

tab edu13 edu18, missing
tab cinc13 cinc18, missing
tab iinc13 iinc18, missing
tab th_inc13 th_inc18, missing
tab dsblty13 dsblty18, missing
tab hhldcomp13a hhldcomp18a, missing
tab crowding13 crowding18, missing
tab q_nzdep2013 q_nzdep2018, missing


**** USING TK13 JK WEIGHTS

** TK13 JK weight set up
svyset _n [pweight=finalwgt], jkrweight(finalwgt_*) vce(jackknife) singleunit(missing)


*sex
        *cell proportions
svy: tab vcat sex, missing cell se format(%7.4f)

        *row proportions
svy: tab vcat sex, missing row se format(%7.4f)

        *table counts
svy: tab vcat sex, missing count se format(%14.3g)


*age
        *cell proportions
svy: tab vcat pd_age_cat, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat pd_age_cat, missing row se format(%7.4f)
        *table counts
svy: tab vcat pd_age_cat, missing count se format(%14.3g)


*education (2018)
        *cell proportions
svy: tab vcat edu18, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat edu18, missing row se format(%7.4f)
        *table counts
svy: tab vcat edu18, missing count se format(%14.3g)


*individual income (census 2018)
        *cell proportions
svy: tab vcat cinc18, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat cinc18, missing row se format(%7.4f)
        *table counts
svy: tab vcat cinc18, missing count se format(%14.3g)

```

```
71
72  *individual income (ird 2018)
73          *cell proportions
74  svy: tab vcat iinc18, missing cell se format(%7.4f)
75          *row proportions
76  svy: tab vcat iinc18, missing row se format(%7.4f)
77          *table counts
78  svy: tab vcat iinc18, missing count se format(%14.3g)
79
80
81  *disability (census 2018)
82          *cell proportions
83  svy: tab vcat dsblty18, missing cell se format(%7.4f)
84          *row proportions
85  svy: tab vcat dsblty18, missing row se format(%7.4f)
86          *table counts
87  svy: tab vcat dsblty18, missing count se format(%14.3g)
88
89  *connection to turangawaewae
90          *cell proportions
91  svy: tab vcat conTWW, missing cell se format(%7.4f)
92          *row proportions
93  svy: tab vcat conTWW, missing row se format(%7.4f)
94          *table counts
95  svy: tab vcat conTWW, missing count se format(%14.3g)
96
97  *culture importance
98          *cell proportions
99  svy: tab vcat cultureimp, missing cell se format(%7.4f)
100         *row proportions
101 svy: tab vcat cultureimp, missing row se format(%7.4f)
102         *table counts
103 svy: tab vcat cultureimp, missing count se format(%14.3g)
104
105
106 * Trust in healthcare system to treat fairly
107         *cell proportions
108 svy: tab vcat thealthfair, missing cell se format(%7.4f)
109         *row proportions
110 svy: tab vcat thealthfair, missing row se format(%7.4f)
111         *table counts
112 svy: tab vcat thealthfair, missing count se format(%14.3g)
113
114
115 * Any discrimination while trying to get medical care
116         *cell proportions
117 svy: tab vcat meddiscr, missing cell se format(%7.4f)
118         *row proportions
119 svy: tab vcat meddiscr, missing row se format(%7.4f)
120         *table counts
121 svy: tab vcat meddiscr, missing count se format(%14.3g)
122
123
124 *household factors
125
126 * How are whanau doing?
127         *cell proportions
128 svy: tab vcat doingwell, missing cell se format(%7.4f)
129         *row proportions
```

```
svy: tab vcat doingwell, missing row se format(%7.4f)
        *table counts
svy: tab vcat doingwell, missing count se format(%14.3g)


* How well are whanau getting along?
        *cell proportions
svy: tab vcat getalong, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat getalong, missing row se format(%7.4f)
        *table counts
svy: tab vcat getalong, missing count se format(%14.3g)


* Household composition
        *cell proportions
svy: tab vcat hhldcomp18a, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat hhldcomp18a, missing row se format(%7.4f)
        *table counts
svy: tab vcat hhldcomp18a, missing count se format(%14.3g)

*Household crowding
        *cell proportions
svy: tab vcat crowding18, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat crowding18, missing row se format(%7.4f)
        *table counts
svy: tab vcat crowding18, missing count se format(%14.3g)

*Household quality
        *cell proportions
svy: tab vcat housingqual18, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat housingqual18, missing row se format(%7.4f)
        *table counts
svy: tab vcat housingqual18, missing count se format(%14.3g)

*Household income
        *cell proportions
svy: tab vcat th_inc18, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat th_inc18, missing row se format(%7.4f)
        *table counts
svy: tab vcat th_inc18, missing count se format(%14.3g)


* Geographic variables

*NZDep 2018 (quintiles)
        *cell proportions
svy: tab vcat q_nzdep2018, missing cell se format(%7.4f)
        *row proportions
svy: tab vcat q_nzdep2018, missing row se format(%7.4f)

        *table counts
svy: tab vcat q_nzdep2018, missing count se format(%14.3g)


```

```
189
190
191 ************* viewing characteristics of those who left country or died********
192
193 codebook lcdth_before
194
195 *unweighted
196 tab pd_age_cat if lcdth_before ==1, missing
197 tab sex if lcdth_before==1, missing
198 tab q_nzdep2018 if lcdth_before==1, missing
199
200 *weighted
201 svy: tab pd_age_cat if lcdth_before ==1, missing
202 svy: tab sex if lcdth_before==1, missing
203 svy: tab q_nzdep2018 if lcdth_before==1, missing
204
205 ************************************* REGRESSIONS
    ↪  ****************************************
206
207 /*
208 * With full TK13 sample dataset
209 - No one dropped
210 */
211
212 ** TK13 JK weight set up
213 svyset _n [pweight=finalwgt], jkrweight(finalwgt_*) vce(jackknife) singleunit(missing)
214
215 ********
216 * ordered logistic regressions
217
218 svy: ologit vcat i.sex ib(5).iinc18 thealthfair doingwell if lcdth_before!=1, base
219 svy: ologit vcat i.sex i.pd_age_cat i.dsblty18 if lcdth_before!=1, base
220 svy: ologit vcat i.q_nzdep2018 if lcdth_before!=1, base
221
222 * interactions
223 svy: ologit vcat i.sex##i.pd_age_cat ib(5).iinc18 thealthfair doingwell if lcdth_before!=1,
    ↪  base
224
225 * TK individual var
226 svy: ologit vcat i.sex i.pd_age_cat ib(3).edu18 ib(5).iinc18 i.dsblty18 i.conTWW doingwell
    ↪  i.conTWW i.cultureimp thealthfair i.meddiscr if lcdth_before!=1, base
227
228 *all household vars (incl TK specific)
229 svy: ologit vcat i.sex i.pd_age_cat ib(3).edu18 ib(5).iinc18 i.dsblty18 doingwell i.getalong
    ↪  i.crowding18 ib(7).th_inc18 i.hhldcomp18a i.housingqual18 if lcdth_before!=1, base
230
231 * all variables
232 svy: ologit vcat i.sex i.pd_age_cat i.dsblty18 ib(3).edu18 ib(5).iinc18 i.conTWW
    ↪  i.cultureimp thealthfair i.meddiscr doingwell i.getalong i.crowding18 i.hhldcomp18a
    ↪  i.housingqual18 ib(7).th_inc18 i.q_nzdep2018 if lcdth_before!=1, base
233
234 *geographic (nzdep)
235 svy: ologit vcat i.sex i.pd_age_cat ib(3).edu18 ib(5).iinc18 i.dsblty18 i.q_nzdep2018 if
    ↪  lcdth_before!=1, base
236
237 * multinomial logistic regression
238
239 *individual
```

```
240 svy: mlogit vcat i.sex i.pd_age_cat ib(3).edu18 ib(5).iinc18  i.dsblty18 if lcdth_before!=1
    ↪  , rrr base
241
242 *nzdep only
243 svy: mlogit vcat i.q_nzdep2018 if lcdth_before!=1, rrr base
244
245 ***********************************************
```