

NZ Health Survey data in the IDI

This guide describes the New Zealand Health Survey (NZHS) datasets in the IDI and the permissions with which they can be used. We summarise some of the key areas of the survey content and the importance of accounting for the sampling strategy in producing survey estimates.

Specific health-related surveys that focus on oral health, nutrition, tobacco, alcohol use or mental health are outside the scope of what is covered here.

Key points

- The NZHS provides access to a range of relevant health variables that are otherwise difficult or impossible to extract from administrative data.
- Currently the NZHS data in the IDI can only be linked to other Ministry of Health (MOH) datasets.
- The NZHS uses a stratified multi-stage cluster sampling design, which needs to be accounted for in the analysis of prevalence estimates and their confidence intervals (and generally for other statistical inferences, e.g. measures of association).

Background

The NZHS is a nationally representative survey, providing information on the health and wellbeing of New Zealanders, both children and adults. The survey has been conducted annually since 2011 (with continuous ongoing sampling, and some content rotated each year) by the Ministry of Health and CBG Health Research (<https://www.cbg.co.nz>).

National health surveys were also conducted in 1992 (adults only), 1995/96, 2001/02 (adults only), and 2006/07. De-identified data from these surveys are available to researchers in confidentialised unit record files (CURFs), but not via the IDI. CURFs are available on application to the Ministry of Health, and include the NZHS datasets in the IDI (see <https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/new-zealand-health-survey>).

What NZHS data can I find in the IDI?

Since the March 2021 refresh, NZHS datasets from 2011/12 to 2018/19 have been made available for IDI users with the appropriate permissions. Data from 2011/12 to 2018/19 child and adult face-to-face surveys can currently be found in the IDI_Adhoc database. These are the IDI table names:

- `<Clean_read_MOH_NZHS.moh_nzhs_adult>` (15+ year olds)
- `<Clean_read_MOH_NZHS.moh_nzhs_child>` (0–14 year olds).

The IDI data includes the core content of the NZHS questionnaire that stays relatively consistent over time, comprising topics such as: long-term health conditions, perspectives and experiences of the health system, health behaviours (e.g. physical activity, fruit and vegetable intake, smoking and alcohol), health status, functional difficulties, socio-demographic information and anthropometric measurements (e.g. body-mass index).

Additional modules are added to the health survey in selected years (e.g. with less than 15,000 people in a given year) covering priority topics in more detail such as: racial discrimination (2016/17, 2020/21), dietary habits, food security and alcohol use (2019/20), Covid-19 (2020/21), mental health and substance use, wellbeing and food security (2021/22). Most of the module data is not currently available in the IDI. Many of these modules have independent weighting schemes to reflect participation in this additional content.

The target population for the survey is the New Zealand usually resident population of all ages including individuals living in non-private accommodation. 99% of this target population is eligible for sampling. The survey is designed to have an annual sample size of approximately 14,000 adults and 5,000 children.

The key challenges for using NZHS data in the IDI are the current limits on linking survey data to non-health data and accounting for the survey sampling design in analyses. Consideration should also be given to other contextual changes over time, such as Covid-19.

Limits on linking

The current permissions for using the NZHS data in the IDI, require that the data is not linked to datasets outside of health. NZHS data in the IDI can only be linked to other health data. IDI projects granted access to the NZHS must only have access to health data.

However, this may change in the future. MoH and Stats NZ are in discussion about whether to allow linking of NZHS to non-health datasets.

Accounting for the sampling strategy

The NZHS has a multi-stage, stratified, probability-proportional-to-size (PPS) sampling design. This sampling strategy must be taken into account to report accurate prevalence estimates from the survey data and their standard errors/confidence intervals which reflect the target population. There are options available in common software packages to account for the survey stratum, clusters and weighting.

Full details on the NZHS sampling strategy are available in Ministry of Health technical reports (see Sample Design from 2015/16 New Zealand Health Survey, linked below).

The NZHS uses a dual sampling frame where participants are selected from an area-based sample or an electoral roll sample (areas with a high number of households with at least one adult has indicated Māori descent). This is to increase the sample size for Māori.

The sampling design is stratified by DHB. Within each DHB, a sample of areas (primary sampling units, PSUs: here Stats NZ meshblocks) is selected, with a sample of households selected from each area, and a sample of one adult (aged 15+) and up to one child (under 15) is selected from each household. Currently there is no variable available for cluster, but this may be added in future iterations of the health surveys in the IDI. There is a DHB variable which can be used to identify these strata.

Each survey dataset also reports annual weights (e.g. linkedwgt), which are designed to make estimates representative of the target population (usual residents of New Zealand). These weights reflect the probability of each respondent being selected into the sample (including consenting to be linked to other health data) and are calibrated to population benchmarks.

In the IDI environment, complex sampling can be accounted for using most major software packages, e.g. the PROC SURVEY family in SAS; survey package in R (<https://cran.r-project.org/web/packages/survey/survey.pdf>); and commands used with the svy: prefix in Stata. There are methods for age-standardising weighted data and comparing weighted estimates between surveys.

The main available method (in the IDI datasets) for accounting for the complex survey design is through the use of jackknife replicate weights, which have been pre-calculated for researchers as a set of 100 replicate weights using the delete-a-group jackknife method (Kott, 2001). Fuller technical details on the statistical theory underlying complex samples and replicate weighting can be found in other sources (e.g. Lumley (2011), Methodology Report (2020) linked below). Users should contact MoH or a statistician if they are not confident in using the weights correctly.

These weights need to be correctly specified to the software package. A number of statistical analysis packages, including SAS, Stata and R, can calculate standard errors using jackknife weights, which are then typically automatically applied to construct confidence intervals for estimates. Note that an extra set of weights is calculated and available for the subset of respondents who have their height and weight measured (e.g. linkedMwgt) and should be used when analysing these measurements.

For further information:

Methodology reports are available for each year of the NZHS (July to June), e.g.

Ministry of Health, New Zealand Health Survey. Methodology Report 2019/20 (2020). Available at:

<https://www.health.govt.nz/system/files/documents/publications/methodology-report-2019-20-new-zealand-health-survey-nov20.pdf>

More information on the sample design is available from: Ministry of Health, New Zealand Health Survey.

Sample Design from 2015/16 New Zealand Health Survey (2016). Available at:

<https://www.health.govt.nz/system/files/documents/publications/sample-design-2015-16-nzhs-dec16.pdf>

Impact of Covid-19

National and regional lockdowns during in 2020 and 2021 in response to Covid-19 outbreaks caused delays in recruitment to the NZHS for face-to-face interviews. This has affected the timing and coverage of survey responses nationally and particularly in the Auckland region. For example, the response rate dropped a few percentage points, there were fewer respondents, and the 2019/20 data were only collected for a 9-month period rather than 12 months (a small number of indicators seem to have a seasonal element where this matters). The Covid-19 pandemic and associated restrictions may have affected trends in some survey responses, for example, Covid-19 economic impacts and any changes in access to healthcare. These effects may vary by age, gender, ethnicity and region; and should be considered in any time series analyses that covers this period.

Conclusion

NZHS data is an excellent source of information on a selection of important health variables that are not collected in administrative data, such as behavioural risk factors. We hope that a stronger understanding about the dataset can help researchers make better use of this data in research to improve health outcomes in Aotearoa New Zealand.

Further information

The New Zealand Health Survey, Ministry of Health website:

<https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/new-zealand-health-survey>

Access to survey microdata at the Ministry of Health:

<https://www.health.govt.nz/nz-health-statistics/national-collections-and-surveys/surveys/access-survey-microdata>

The data dictionary for the IDI health survey is available in the IDI Wiki, under IDI→Metadata→MOH, or can be requested by emailing Access2Microdata-SharedMailbox@stats.govt.nz.

Other references:

Kott PS (2001). The Delete-a-Group Jackknife. *Journal of Official Statistics* 17, 521–526.

Lumley T (2011). *Complex surveys: a guide to analysis using R*. John Wiley & Sons.

Original version published 26 January 2022, written by Andrea Teng, James Stanley, Mel Duncan, Chloe Lynch, Sheree Gibb