Explanatory Notes

Modelled estimates for small areas

based on the

2017-18 National Health Survey

Prepared by the Health Section,
Australian Bureau of Statistics

for the Public Health Information Development Unit, a unit of Torrens University Australia Limited

Release: 21 November 2019

LICENSE CONDITIONS

This customised report carries the following licence:

Creative Commons Attribution 4.0 International

You are free to re-use, build upon, and distribute this material, even commercially. The entire report may be included as an appendix in your work for reference if you wish.

Under the terms of this license, you are required to attribute ABS material in the manner specified (but not in any way that suggests that the ABS endorses you or your use of the work).

ABS material used 'as supplied'
Provided you have not modified or transformed ABS material in any way, for example by:

- changing the ABS text
- calculating percentage changes
- graphing or charting data
- deriving new statistics from unpublished ABS statistics

Material contained in this customised report may be reused provided one of the following attributions is given:

Source: Australian Bureau of Statistics
or
Source: ABS

Derivative material
If you have modified or transformed ABS material, or derived new material from those of the ABS in any way, one of the following attributions must be used:

Based on Australian Bureau of Statistics data
or
Based on ABS data

Citing customised reports
If you are required to cite material from this report please be guided by the examples below.

In-text and reference list/bibliography

- In-text:
  (ABS 2019)
- In reference list/bibliography:
  ABS 2019, Customised report.

In-text only

- (ABS, Customised report, 2019)

CONTENTS

1 INTRODUCTION

The Public Health Information Development Unit (PHIDU) requested the Australian Bureau of Statistics (ABS) provide modelled estimates of characteristics associated with health at a small area level for the Australian population. To meet this request, and by mutual agreement between ABS and PHIDU, the ABS has provided modelled estimates for public release based on the National Health Survey, 2017-18 (NHS). These explanatory notes accompany the modelled estimates for small areas, provided as Excel worksheets, and describe the methodology used to produce them, as well as how to use them.

2 PURPOSE

The 2017-18 NHS is the most recent in a series of Australia-wide health surveys conducted by the ABS. The survey was designed to collect a range of information about the health of Australians to enable reliable estimates at the national and state and territory level. The sample size was too small to produce reliable estimates for areas with small populations such as Population Health Areas (PHAs). To produce reliable and detailed estimates at these geographical levels, models were created using the detailed NHS data (based on selected persons) and applied to data for small areas on ABS Estimated Resident Population (ERP), the 2016 ABS Census of Population and Housing, and administrative sources.

3 METHODOLOGY USED

A modelled estimate can be interpreted as the likely value for an area based on the demographic information we have for that area. The process of producing modelled estimates at the PHA level on indicators measured in the NHS consisted of the following components, described in detail in sections 3.1 to 3.10:

1. Identification of the outcome variables
2. Selection of the predictor variables
3. Identification of the geographical regions
4. Identification of population at risk estimates
5. Scoping the data
6. Creation of binary and proportion variables
7. Aggregating observations and merging datasets
8. Model selection
9. Creation of modelled estimates
10. Assessment of the modelled estimates

3.1 Identification of the outcome variables

The 2017-18 NHS is the most recent in a series of Australia-wide health surveys conducted by the ABS. The survey was conducted in all states and territories and across urban, rural and remote areas of Australia (excluding very remote areas) from July 2017 to June 2018. The survey included around 21,000 people in over 16,000 private dwellings.

The survey was designed to collect a range of information about the health of Australians, including:

- prevalence of long-term health conditions;
- health risk factors such as smoking, overweight and obesity, alcohol consumption and physical activity; and
- demographic and socioeconomic characteristics.

Indicators for modelling are often referred to in literature as outcome variables, dependent variables or response variables. From the 2017-18 NHS, modelled estimates about people from the overall Australian population(a) with the following characteristics were produced at the PHA level:

- self-assessed health as fair or poor, 15 years and over
- diabetes mellitus, long-term, all ages
- mental and behavioural conditions, long-term and current, all ages, by sex
- heart, stroke, and vascular disease, long-term, all ages
- asthma, long-term and current, all ages
- Chronic Obstructive Pulmonary Disease (COPD), long-term and current, all ages
- arthritis, long-term and current, all ages
- osteoporosis, long-term and current, all ages
- high or very high psychological distress, 18 years and over, by sex
- high blood pressure (≥ 140/90 mmHg), 18 years and over
- overweight (but not obese), 18 years and over, by sex
- obese, 18 years and over, by sex
- current smoker(b), 18 years and over, by sex
- alcohol – exceeded 2009 lifetime risk guidelines, 18 years and over, by sex
- low, very low, or no exercise in the last week, 18 years and over
- met guidelines for daily fruit intake, 18 years and over
- diabetes mellitus, long-term, 25 to 64 years, by sex
- overweight (but not obese), 2 to 17 years, by sex
- obese, 2 to 17 years, by sex
- mental and behavioural conditions, long-term and current and currently studying or employed, 16 to 30 years, by sex

- mental and behavioural conditions, long-term and current and employed, 16 to 64 years, by sex

(a) modelled estimates do not include persons usually resident in non-private dwellings, very remote areas of Australia and discrete Aboriginal and Torres Strait Islander communities.

(b) In the case of current smoker, this data item was also collected in the concurrent ABS Survey of Income and Housing. Data for the current smoker model were pooled from both surveys, a sample size of approximately 45,000 people.

For more information about the outcome variables, including definitions, see the worksheet 'Notes' within each of the Excel spreadsheets delivered to PHIDU, the Glossary, or Explanatory Notes on the ABS Website.

Modelled estimates were provided for several of these outcome variables in a previous statistical consultancy for PHIDU using data from the 2014-15 National Health Survey. The definitions of such outcome variables are consistent between the 2017-18 and 2014-15 National Health Surveys. Modelled estimates for these outcome variables are comparable between the two surveys.

More information about the 2017-18 NHS can be found on the ABS Website (ABS catalogue number (Cat. No.) 4364.0.55.001).

3.2 Selection of the predictor variables

In order to predict outcome variables, predictor variables are required on both the NHS dataset and a small area dataset containing population, Census, and administrative data. Predictor variables are also referred to in literature as explanatory variables or independent variables. Predictor variables were created if data were available for small areas for all of urban, rural, and remote Australia and if there was an expectation that they might be good predictors of the outcome variables.

For age and sex predictor variables, data at the small area level were obtained from ABS ERP data from Regional Population Growth, Australia, 2017-18 (Cat. No. 3218.0). This is described below in section 3.4.

For other demographic variables on the NHS, data at the small area level were obtained from the 2016 Census of Population and Housing, as this was the most up-to-date comprehensive source of demographic data due to the depth of information at small geographical levels.

If appropriate, demographic variables on the NHS and Census were adjusted to make them more closely aligned. Variables that were available at the small area but not collected in the NHS were added to the NHS data; these variables included other demographic variables on the Census, geographic variables, and variables from administrative sources.

Predictor variables that relate to the geographical areas where people reside included:

- remoteness area;
- socio-economic indexes for areas (SEIFAs) – population-weighted deciles at the Statistical Areas Level 1 (SA1) level;
- state and territory;
- section of state (major urban/other urban/bounded locality/rural balance);
- Greater Capital City Statistical Area (GCCSA)/balance of state; and;
- design area type (categorises inner city, large and small urban towns, rural towns and remote areas within states and territories for designing the sample of the NHS).

Sources of data included:

- Australian Statistical Geography Standard (ASGS): Volume 5 - Remoteness Structure, July 2016 (Cat. No. 1270.0.55.005);
- Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016 (Cat. No. 2033.0.55.001);
  Australian Statistical Geography Standard (ASGS): Volume 4 - Significant Urban Areas, Urban Centres and Localities, Section of State, July 2016 (Cat. No. 1270.0.55.004); and
- Australian Statistical Geography Standard (ASGS): Volume 1 - Main Structure and Greater Capital City Statistical Areas, July 2016 (Cat. No. 1270.0.55.001).

Predictor variables from administrative sources included:

- births in 2017;
- deaths in 2017;
- immigration within Australia and overseas migration in 2016-17;
- population density in 2017;
- dwelling transfers and median sale prices in 2016;
- personal income tax data for employee earnings, investment, own business or superannuation income in 2012-13;
- recipients of age pensions, disability support pensions, Newstart allowances, carer allowances, health care cards, pensioner cards, Family tax benefits and other benefits in the quarter to March 2016;

- attendance at public hospitals for various conditions and procedures (2016-17), deaths from selected causes (2011 to 2015), Home and Community Care Program (HACC) clients (2012-13), participation in vocational education and training (2015), development of children (2015), immunisations (2011 to 2016), and bowel cancer screening (2012-13).
- Use of services from the Medicare Benefits Schedule (MBS) and transactions from the Pharmaceutical Benefits Scheme (PBS) in 2016.

Data for the above were obtained, in the same numeric order, from:

- Births, Australia, 2017 (Cat. No. 3301.0)
- Deaths, Australia, 2017 (Cat. No. 3302.0)
- Data by Region, 2013-18 (Cat. No. 1410.0)
- Data by Region, 2013-18 (Cat. No. 1410.0)
- Data by Region, 2013-18 (Cat. No. 1410.0)
- Estimates of Personal Income for Small Areas, 2012-13 (Cat. No. 6524.0.55.002)
- Department of Social Services (DSS) Payment Demographic Data, http://www.data.gov.au/dataset/dss-payment-demographic-data
- Public Health Information Development Unit (PHIDU) May 2019 release, http://phidu.torrens.edu.au,
- Multi Agency Data Integration Project (MADIP) (Cat. No. 1700.0)

Within most types of predictor variables (as discussed above), several explanatory variables representing separate categories or data items were tested. The variables tested for significance are listed in Appendix 1 and in the Excel spreadsheets delivered to PHIDU.

3.3 Identification of the geographical regions

The geographical regions required for this request were at a level where direct survey estimates are not available or their sample errors are too high for the direct survey estimates to be useful. Modelled estimates were provided at the PHA level. PHAs were developed by PHIDU in consultation with state and territory health agencies and are comprised of a combination of whole Statistical Areas Level 2 (SA2s) (39.7% of PHAs) and aggregates of SA2s with relatively small populations. For further information, refer to Population Health Areas: Overview on PHIDU's website.

3.4 Identification of population at risk estimates

The base data source used to compile the population at risk estimates for this request is the ABS Estimated Resident Population (ERP) data from Regional Population Growth, Australia, 2017-18 (Cat. No. 3218.0).

This data was then adjusted to match the scope of the NHS and to sum to NHS population state by age by sex estimates (described below in section 3.5).

The adjusted ERP data were also used for denominators in the calculations of proportions of persons at risk i.e. 'population at risk' estimates included in the Excel spreadsheets provided to PHIDU. It is important to note that these population estimates are not official estimates and were created solely for analysis of the NHS modelled estimates and will not match other population data for PHAs.

3.5 Scoping the data

The modelled estimates for small areas are applicable to persons who were usual residents of private dwellings to match the scope of the NHS. They exclude:

- non-private dwellings, for example hospitals and aged care facilities;
- areas classified as very remote
- areas classified as discrete Aboriginal and Torres Strait Islander communities.

Adjustments were made to the ERP data, by using ratios of private to non-private dwellings, calculated from the 2016 Census to approximate exclusion of the persons residing in non-private dwellings, and then summed to the NHS population state by age by sex estimates. These are the 'population at risk' estimates included in the Excel spreadsheets provided to PHIDU.

Adjustments were also made to the Census data, specifically the predictor variables obtained from the Census (described above in section 3.2). Identification of persons' type of dwelling is possible on the Census datasets for respondents at home on Census night so persons residing in non-private dwellings were easily removed from the small area dataset. However, for persons who were not at home on Census night, information is not collected to determine if the dwelling they usually reside in is a private or non-private dwelling; therefore, their records were deleted from the small area dataset. This means that an assumption has been made that the people who were away from home on Census night and live in private dwellings have the same health characteristics as the people who were at home in a private dwelling.

Removal of very remote areas and discrete Aboriginal and Torres Strait Islander communities from the ERP and Census data file was approximately done by deleting persons residing in PHAs that had more than 20% of their population in SA1s classified as very remote or in discrete Aboriginal and Torres Strait Islander communities.

Additional exclusions that were applied included:

- PHAs with an adjusted ERP of fewer than 100 residents aged 18 years and over
- residents of Other Territories
- foreign diplomatic personnel and their families were excluded from the modelled estimates because they are not included in Australia's ERP, the Census or the NHS.

See the worksheet "Excluded PHAs" within each of the Excel spreadsheets delivered to PHIDU for the PHAs excluded.

While out of scope for the NHS, members of non-Australian defence forces (and their dependents) stationed in Australia were unable to be removed from the modelled estimates because they could not be identified in Australia's ERP.

For more information on the survey scope and coverage, see the Explanatory Notes of National Health Survey: First Results, 2017-18 Explanatory Notes (ABS catalogue number 4364.0.55.001).

3.6 Creation of binary and proportion variables

On the NHS dataset outcome variables were created as binary variables to make them suitable for the type of modelling undertaken (logistic regression). On both the NHS and small area datasets, predictor variables that were categorical were also created as binary variables. An observation took the value of 1 if an individual had a characteristic of interest and 0 otherwise. For example:

1. in the case of obesity, the outcome variable for obese took the value of 1 if an individual was obese and 0 if the individual was not obese; and

2. in the case of labour force status, the predictor variable for employed took the value of 1 if an individual was employed and 0 if the individual was unemployed, not in the labour force or aged 0-14 years.

Variables in administrative data were converted to proportions of their areas' population with the characteristic of interest. For example:

- a person can live in an area with a proportion of its population receiving a disability support pension.

In addition, binary variables were created denoting ranges of the administrative data variables. For example:

- for fertility rate, a binary variable was created to denote whether the person lived in an area with a fertility rate between 2 and 2.5.

3.7 Aggregating observations and merging datasets

All the datasets were aggregated by combining observations/respondents with the same SA2, design area type, five year age groups and sex. This decreases the size of the datasets (especially the Census dataset) to increase the efficiency of the modelling process.

The Census, adjusted ERP and administrative datasets were merged into one small area dataset. A number of the SA2 by age by sex groups had a non-zero adjusted ERP with no corresponding combination within the Census dataset. For the most part this was due to population growth between the time the Census was undertaken and when the adjusted ERP was created, or because areas had very small populations; for all but a small number of the affected areas the effect is insignificant. Given that Census data is required in order to derive appropriate modelled estimates, the affected groups have been excluded from estimates at PHA level. It is not expected that these exclusions will have a significant impact on modelled estimates at the PHA level.

3.8 Model selection

Models were created for each outcome variable independently. The model selection method uses the data files to measure the relationship between the outcome variable and possible predictor variables to determine one set of significant predictor variables. This method assumes that the relationships observed in the survey data overall also hold at the small area level.

The models used to determine these relationships were logistic regression models. As part of any model selection process an appropriate significance level must be chosen for determining which explanatory variables to include in the models. The 0.05 (95%) level is most commonly used; however, due to NHS' relatively large sample sizes, the Bayesian Information Criterion (BIC) was used to reduce the risk of over-fitting.

The models were applied to small area data using 2017-18 adjusted ERP for the population counts (described above in section 3.4), summed to create Australia level modelled estimates and compared with reliable direct survey weighted estimates to see if the model adequately predicted the outcome variable. Some models were improved with the addition of less significant predictor variables and interactions of some predictor variables.

3.9 Creation of modelled estimates

The relationships selected (described above) were then fitted using random effects logistic regression models. A mixed estimate comprised of modelled and survey data is then produced for each PHA. A mixed/composite estimate gives results for each small area that reflect the best trade-off between the accuracy of the direct survey weighted estimate and the error associated with the modelled estimate. So, for a small area that happens to have a low sampling error (because of a large sample size within that small area, for example), more weight will be given to the direct estimate when calculating an estimate for that small area. On the other hand, for a small area with high sampling error, more weight will be given to the model based prediction as this will be more reliable in calculating the estimate for that small area. This takes advantage of what is known about PHAs from the survey to improve the modelled estimates.

A pro rata adjustment was then made to the modelled estimates so that they summed to national direct survey estimates. Where modelled estimates were produced broken down by sex, the adjustment was also made broken down by sex. The associated errors resulting from the modelling process, which improve on direct survey estimates' errors, were not adjusted.

The modelled estimates supplied in the MS Excel spreadsheets delivered to the PHIDU are in the form of counts (number of persons) and their relative error for the PHAs. Prevalence proportions (percentage of population at risk in each small area) and their 95% confidence intervals (CIs) have also been calculated. The denominators used in the calculation of proportions at risk were the unofficial population estimates for each PHA (based on adjusted ERP) described above in section 3.4.

3.10 Assessment of the modelled estimates

Various measures were taken to examine the modelled estimates. Modelled estimates were compared with direct survey estimates from the 2017-18 NHS for areas that were sampled. For the survey estimates, 95% CIs were calculated. These were plotted against the modelled estimates to see if the majority of modelled rates fell within the CIs of the NHS estimates.

Relative root mean squared errors (RRMSEs) of the modelled estimates were examined to ensure that the majority were of suitable quality.

Also, the number, range, and applicability of explanatory variables included in the models used to create the small area estimates were considered.

Finally, lists of the 20 PHAs with the highest rates, and the 20 PHAs with the lowest rates; and choropleth maps were produced to assess whether the modelled estimates aligned with expectations. Where available, modelled estimates were compared with modelled estimates produced from the NHS 2014-15.

Please see Appendix 2 for a quality summary for modelled estimates.

4 ACCURACY OF RESULTS

The process undertaken in providing modelled estimates overcomes much of the volatility at the PHA level caused by sampling error. However, it should be remembered that the estimates provided are still subject to errors.

The errors associated with the modelled estimates for small areas fall into four categories, as follows:

1. Sampling error
2. Non-sampling error
3. Modelling error
4. Prediction error

These errors are combined into an overall measure of accuracy, the relative root mean squared error (RRMSE), described in section 4.5.

4.1 Sampling Error

Sampling error is introduced into estimates because NHS data were collected for only a sample of dwellings. Therefore, they are subject to sampling variability; that is, modelled estimates may differ from those that would have been produced if all dwellings had been included in NHS. Furthermore, the smaller the sample obtained within a small area, the greater the sampling error associated with that small area's modelled estimates will be.

4.2 Non-Sampling Error

The imprecision due to sampling error should not be confused with inaccuracies that may occur because of imperfections in reporting by respondents and recording by interviewers, and errors made in coding and processing data. Inaccuracies of this kind are referred to as non-sampling error, and they occur in any enumeration, whether it be a full count (Census) or a sample. Unlike the other sources of error, non-sampling error is not measurable and therefore isn’t accounted for in the measured error (direct or modelled) that accompanies ABS estimates. Every effort is made to reduce non-sampling error to a minimum by careful design of questionnaires, intensive training and supervision of interviewers, and rigorous procedures.

4.3 Modelling Error

Modelling error is introduced by model misspecification. This can occur when the choice of model is incorrect, a key explanatory variable is left out or an inappropriate explanatory variable is included. In practice, it is rarely the case that all determinants of health indicators will be available as good quality small area data to be able to be included as predictor variables in the models. Therefore, the variables chosen in the models may result in incorrect modelled estimates for certain small areas, particularly those unusual small areas that do not follow the typical associations between the available predictor variables and the health indicators. The models that have been chosen have been tested against a range of possible alternative models; however, they are only the most preferred models subject to available predictor variables.

4.4 Prediction Error

A strong model does not guarantee statistically accurate modelled estimates. Prediction error is a measure of the statistical accuracy of the model predictions.

4.5 Relative Root Mean Squared Error (RRMSE)

A measure of the quality of the modelled estimates is the RRMSE. The RRMSE is primarily a measure of prediction error but in its calculation it also inherits some aspects of modelling and sampling error. The RRMSE generally decreases as the population size increases, and is used to assess the reliability of modelled estimates.

As a general rule of thumb, estimates with RRMSEs less than 25% are considered reliable for most purposes, estimates with RRMSEs between 25% and 50% should be used with caution and estimates with RRMSEs greater than 50% are considered too unreliable for general use.

Some areas/groups have high RRMSEs and the accuracy of their modelled estimates can be improved by aggregating them to larger regions/groups. The method and examples for calculating RRMSEs for aggregated areas/groups are provided in an Excel spreadsheet delivered to PHIDU.

A confidence interval (CI) provides a range of values, within which it is estimated that the true population value lies. To assist with the calculation of CIs for aggregated areas, examples of this calculation are provided in the Excel spreadsheets delivered to PHIDU.

5 USING MODELLED ESTIMATES

The small area modelled estimates can be interpreted as the expected prevalence for a typical area in Australia with the same characteristics. For some PHAs, there will be large differences between the modelled estimates and the actual number of people with the characteristic of interest. One explanation for this is that significant local information about particular PHAs exists but has not been collected for all areas and cannot be incorporated into the models. This sort of information is usually not measurable, and relies on local or expert knowledge.

Small area modelled estimates should be viewed as a tool that when used in conjunction with local area knowledge as well as the consideration of the modelled estimates reliability, can provide useful information that can assist with decision making for small geographic areas.

Modelled estimates have been confidentialised to ensure they meet ABS requirements for confidentiality. PHAs with populations or modelled counts that didn't meet the confidentiality rules have modelled estimates comprised solely of the modelled component; rather than the mixed/composite estimator described above in section 3.9. This means that no sampled contribution is included in such modelled estimates, regardless of whether sample exists in these PHAs.

Areas or groups can be aggregated together using examples provided in the MS Excel spreadsheets delivered to PHIDU. Example 1 shows how to aggregate RRMSEs of several small areas. Aggregation of small areas should be done taking into account local knowledge about these areas.

The reliability of the resulting aggregated estimate should be assessed in terms of the error values, CIs, and what is known about the 'new' small area or aggregation group.

Appendix 1 LIST OF PREDICTOR VARIABLES CONSIDERED

This summarises the types of variables that were tested for significance in the models.

Relating to persons:

- sex
- age
- year of arrival in to Australia
- country of birth
- country of birth of parents
- main field of highest non-school qualification
- Indigenous status
- industry of employment
- labour force status
- occupation
- level of highest non-school qualification
- level of highest educational attainment
- registered marital status
- relationship to household reference person
- level of highest year of school completed
- attends university or TAFE
- hours worked in the past week
- personal income
- ancestry
- employee or owner of business
- employment sector
- needs assistance with core activities
- number of children ever born to female
- number of employees of owner businesses
- provided unpaid assistance to a person with a disability
- religion
- social marital status
- unpaid domestic work
- volunteer
- speaks English
- method of travel for work
- Australian citizenship
- whether at the same address five years ago
- whether at the same address one year ago

Relating to the dwelling that persons reside in:

- number of bedrooms
- dwelling structure
- landlord type

- tenure type
- household composition/type
- equivalised gross weekly household income
- gross household income
- household with Indigenous persons
- number of persons usually resident
- number of children usually resident
- family blending
- family composition/type
- household five year mobility indicator
- household has an Internet connection
- household one year mobility indicator
- mortgage amount
- rent amount
- labour force status of family
- number of motor vehicles
- housing suitability – number of bedrooms compared to need

Relating to areas where persons reside in:

- remoteness area
- state and territory
- greater capital city statistical area (GCCSA)
- design area type
- SEIFA Index of Economic Resources (IER) (c)
- SEIFA Index of Education and Occupation (IEO) (c)
- SEIFA Index of Relative Socio-Economic Advantage and Disadvantage (IRSAD) (c)
- SEIFA Index of Relative Socio-Economic Disadvantage (IRSD) (c)
- numbers of births and deaths, and fertility rates
- dwelling, house and unit sales
- median house, unit and dwelling sales prices
- income inequality measures for total income earners
- population density (persons/sq. km)
- internal migration
- overseas migration
- participation in vocational education and training
- children developmentally at risk or on track in selected domains
- children fully immunised
- HPV vaccine coverage
- participation in the National Bowel Cancer Screening Program (NBCSP)
- hospital admissions for selected causes
- average deaths and avoidable deaths from selected causes
- proportion receiving various government benefits
- proportion receiving services / transactions from the Medicare Benefits Schedule (MBS) or Pharmaceutical Benefits Scheme (PBS)

(c) Socio-economic indexes for areas (SEIFAs) – population-weighted deciles at the Statistical Areas Level 1 (SA1) level.

Appendix 2 QUALITY SUMMARY FOR MODELLED ESTIMATES

Measures of prediction accuracy (RRMSEs and CIs) are included in the output provided and can be used to assess the overall reliability for each of the models. The average RRMSE across small areas was calculated for each of the twenty models used to construct the twenty-one tables of output(d).

The distribution of the estimates across the areas within age groups and sexes was as expected for all outcome variables. For example, modelled estimates for overweight adults and obese adults were consistently higher for males than for females.

It is important to consider the number, range, and applicability of explanatory variables included in the models used to create the modelled estimates. Most of the outcome variables had a good range of explanatory variables included in the models. Modelled estimates for outcome variables: Overweight (but not obese) children, and Obese children; are however based on models containing a relatively small number of predictor variables. Reasons for this may include a low sample count for the outcome variable and/or small variation/similar characteristics within the sample for the outcome variable. Caution should be applied when interpreting the modelled estimates for these outcome variables.

When determining the overall reliability of each model, average RRMSEs less than 25% have been summarised as ‘reliable’. Average RRMSEs between 25% and 50% have been summarised as ‘use with caution’. Models with average RRMSEs greater than 50% have been summarised as ‘unreliable for general use’. The number and range of predictor variables is also considered when determining the overall reliability of each model.

(d) The modelled estimates for Diabetes Mellitus, 25 to 64 years used the same model as for Diabetes Mellitus, all ages. Therefore for Diabetes Mellitus, there are two tables of output but only one model used.

Table	Average RRMSE (PHA)	Reliability
Self-assessed health as fair or poor, 15 years and over	10.5%	Reliable
Diabetes Mellitus, long-term, all ages	14.9%	Reliable
Mental and behavioural conditions, long-term and current, all ages	13.8%	Reliable

Mental and behavioural conditions, long-term and current, all ages, by sex	14.0%	Reliable
Heart, stroke, and vascular disease, long-term, all ages	21.9%	Reliable
Asthma, long-term and current, all ages	12.3%	Reliable
Chronic Obstructive Pulmonary Disease (COPD), long-term and current, all ages	26.5%	Use with caution
Arthritis, long-term and current, all ages	14.2%	Reliable
Osteoporosis, long-term and current, all ages	19.9%	Reliable
High or very high psychological distress, 18 years and over	13.1%	Reliable
High or very high psychological distress, 18 years and over, by sex	13.6%	Reliable
High blood pressure (≥ 140/90 mmHg), 18 years and over	6.1%	Reliable
Overweight (but not obese), 18 years and over	6.3%	Reliable
Overweight (but not obese), 18 years and over, by sex	6.6%	Reliable
Obese, 18 years and over	7.8%	Reliable
Obese, 18 years and over, by sex	7.9%	Reliable
Current smoker, 18 years and over	13.7%	Reliable
Current smoker, 18 years and over, by sex	13.4%	Reliable
Alcohol – exceeded 2009 lifetime risk guidelines based on previous 7 days consumption, 18 years and over	14.7%	Reliable
Alcohol – exceeded 2009 lifetime risk guidelines based on previous 7 days consumption, 18 years and over, by sex	15.7%	Reliable
Low, very low, or no exercise, 18 years and over	4.5%	Reliable
Met guidelines for recommended daily fruit intake, 18 years and over	7.6%	Reliable
Diabetes Mellitus, long-term, 25-64 years	17.8%	Reliable

Diabetes Mellitus, long-term, 25-64 years, by sex	18.7%	Reliable
Overweight (but not obese), 2-17 years	7.5%	Use with caution
Overweight (but not obese), 2-17 years, by sex	8.4%	Use with caution
Obese, 2-17 years	14.1%	Use with caution
Obese, 2-17 years, by sex	15.4%	Use with caution
Mental and behavioural conditions, long-term and current and currently studying or employed, 16-30 years	6.9%	Reliable
Mental and behavioural conditions, long-term and current and currently studying or employed, 16-30 years, by sex	8.6%	Reliable
Mental and behavioural conditions, long-term and current, and employed, 16-64 years	14.0%	Reliable
Mental and behavioural conditions, long-term and current, and employed, 16-64 years, sex	14.4%	Reliable