

# Queensland preventive health survey methods

Adult and child surveys



# Contents

---

Supporting resources	2
Background	3
Survey data collection methods summary	3
Regional results	3
Available regions	3
Statistical methods	4
Prevalence	4
Calculation—direct method	4
Calculation—synthetic estimation for small areas	5
Synthetic estimation—interpretations and implications	5
Similarities and differences— QPHS and PHIDU/ABS	6
Trend results	7
Dissemination	7
State and regional reporting	7
Dissemination mechanism	8
Results available in QSAS	8
Quality assurance	8
Queensland Health disclaimer	8
Relative standard errors (RSE) and releasability	9
Confidence intervals (CI)	9
Contact information	9
Appendix	10
Data collection methods summary	10
Sample size information	12
Glossary	13
Version control	14

## Supporting resources

Available from [www.health.qld.gov.au/phsurvey](http://www.health.qld.gov.au/phsurvey)

1. PDF—[Conditions of use](#)
2. PDF—[Survey methods \(adult and child\)](#)
3. Excel—[Adult survey sample size information](#)
4. Excel—[Child survey sample size information](#)
5. Excel— Data download—[State adult/child results](#)
6. Excel—Data download—[Regional HHS & PHN adult/child results](#)
7. Excel—Data download—[Regional LGA adult results](#)
8. Visualisation—[QSAS survey measures](#)

# Background

Health data can be broadly classified as administrative or survey data. Administrative data sets are typically collected as part of providing a health service and includes all the individuals who have received that service. It is typically direct counts of service episodes or consumers and basic statistics can be directly calculated (for example sums). Survey data is typically collected to answer a specific question and is based on a subset (sample) of the population. Administrative data typically include significantly higher numbers of health consumers than can be surveyed (for example, over 2.5 million hospitalisations per year compared to surveys of 2,000 to 15,000 participants). These differences mean that the data sets have different strengths and limitations, use different statistical methods, and results may need to be used for different purposes or interpreted differently.

This document summarises the methodology for the Queensland preventive health telephone survey (QPHS; [www.health.qld.gov.au/phsurvey](http://www.health.qld.gov.au/phsurvey)). The Preventive Health Branch (PHB) is the data custodian for the QPHS and can be contacted using [Population\\_Epidemiology@health.qld.gov.au](mailto:Population_Epidemiology@health.qld.gov.au).

## Survey data collection methods summary

**Data source:** the Queensland preventive health telephone survey series (QPHS).

**Mode:** surveys are conducted by computer assisted telephone interview (CATI) using an external provider. All interviews are conducted in English.

### Sampling frame

- Adult: 2002 to 2014—random digit dialling (RDD) of landline telephone numbers.
- Child: 2011, 2013, 2014—random digit dialling (RDD) of landline telephone numbers.
- Adult and child: 2015 to current—a list-based frame that provides near-complete coverage of the Queensland residential population. It is maintained by the Queensland Government for official statistical reporting purposes under the *Statistical Returns Act 1896* and includes both landline and mobile numbers.

### Eligibility

- Adult survey: persons 18 years and older that reside in a residential household.
- Child survey: primary parent or caregiver reports on the health and lifestyle of a child aged 5–17 years in the household who is randomly selected using the ‘next birthday’ rule.

**Survey weights:** all data are population weighted.

**Data collection dates:** Data collection dates exclude school holidays.

See [Data collection methods summary](#) for more information by survey and year.

## Regional results

### Available regions

- Hospital and Health Service areas (adult and child)

- Primary Health Networks (adult and child)
- Local government areas (adult only)
- Statistical areas 3 (by request or in the internal Queensland Health Planning Portal)
- Statistical areas 2 (synthetic estimates in the internal Queensland Health Planning Portal)

Results for LGAs with small population size may not meet reliability criteria or be unavailable. In some areas, contiguous LGAs were combined to achieve sufficient numbers of participants for reporting (note that limited information for the unaggregated LGA may also be in QSAS). Aggregated LGAs are:

- Rest of Central West HHS: Barcoo (S), Blackall Tambo (R), Boulia (S), Diamantina (S) and Winton (S) LGAs
- Rest of North West HHS: Burke (S), Carpentaria (S), Cloncurry (S), Doomadgee (S), McKinlay (S), Mornington (S) LGAs
- Rest of South West HHS: Bulloo (S), Paroo (S), Quilpie (S) LGAs.

This method could not be applied to all small areas and consequently not all LGAs are represented.

Regional boundaries change over time—for example the Queensland Government de-amalgamated four LGAs in 2014. QSAS Users are responsible for defining their regions of interest and knowing whether comparability may be affected by historical boundary changes. To assist, LGAs are identified with their respective ABS Australian Statistical Geography Standard (ASGS) code in the sample size tables. Of note, there are instances where boundaries have changed but the region name has remained identical.

See [Sample size information](#) for detailed information by region.

## Statistical methods

### Prevalence

Results are expressed as a population weighted prevalence or mean.

**Prevalence** is the number of persons with a health condition or behaviour at a point in time divided by the population at risk and is reported as a percentage.

**Mean** is the average of a result across the population or subpopulation.

### Calculation—direct method

When there are a sufficient number of participants in an area or with particular characteristics of interest (for example, by socioeconomic status), survey prevalence results can be calculated directly. Methods appropriate for survey data (for example, use of survey weights) are used to ensure that results can be generalised to the population of interest. Statistical criteria are used to ensure the results are reliable and of sufficient quality to be reported. Unfortunately, this method cannot typically be applied for small areas, such as Statistical Area 2 (SA2) mainly due to the smaller sample that will render most estimate unreliable.

## Calculation—synthetic estimation for small areas

For planning purposes, focus has increased on providing health data by small areas such as SA2s or SA3s. For survey data with smaller sample sizes, directly calculated prevalence estimates for these small areas is typically not feasible or reliable and other techniques must be used. For example, the Public Health Information Development Unit (PHIDU) at Torrens University reports modelled estimates for small areas for prevalence estimates from the National Health Survey (NHS). The Australian Bureau of Statistics (ABS) also releases data from the NHS using the same method as outlined in National Health Survey: First Results methodology.<sup>1</sup>

PHB has developed a similar method to PHIDU/ABS to generate small area risk factor prevalence information. Briefly, the process involves generating direct estimates from the small area and then creating a modelled estimate based on the population characteristics of the small area. The direct and modelled estimates are combined based on sampling variation to obtain the QPHS final synthetic estimate (referred to as the modelled estimate by PHIDU/ABS). For areas without sufficient sample for directly calculated results, this estimate will provide a reasonable approximation to the true prevalence of the health characteristic for the area.

Currently the QPHS releases synthetic estimates for SA2 adult persons only.

## Synthetic estimation—interpretations and implications

Appropriate use and interpretation of any synthetic estimate (QPHS or PHIDU/ABS) requires an understanding of the different statistical techniques described above. Specifically, because a component of the synthetic estimates is based on a state-wide modelled estimate, synthetic estimates are less sensitive in identifying local level changes, especially over the short term or in specifically targeted subgroups. This particularly impacts their use for the following purposes:

1. **To assess local interventions:** Synthetic estimates may not be useful in situations where localised interventions are being evaluated.
2. **Trend analysis:** Synthetic estimates are limited for monitoring trends, especially in the short term. Trend analysis using synthetic estimates will simply reflect trends occurring across the state as a whole.
3. **Subgroups:** Subgroups of interest often differ from the state-wide sociodemographic profile meaning that synthetic estimates will be less accurate for some specific groups. For example, for small areas with high Aboriginal and Torres Strait Islander populations, the modelled results component would be influenced by the state-wide results which could reduce the reliability for localised Aboriginal and Torres Strait Islander populations.

---

<sup>1</sup> <https://www.abs.gov.au/methodologies/national-health-survey-first-results-methodology/2017-18#appendix-4-modelled-estimates-for-small-areas>

A related issue involves *scaling*. Unlike administrative count data, synthetic estimates are challenging to scale so that smaller areas will sum to results for larger areas. In practice, this is similar to the issue with survey weighted data overall—weighted subpopulation estimates cannot be used to generate totals (for example, LGA results cannot be summed to produce accurate HHS results). This also affects the ability to generate results by sociodemographic characteristics (so the current preferred release is ‘person’ level data) or for some indicators with more than two outcomes. For indicators and sociodemographic characteristics, dichotomisation can be used to generate consistent results. A dichotomised indicator for physical activity would be:

**Table 1: Example of dichotomised physical activity indicators**

Reported prevalence in 3 categories	Dichotomised to 2 categories
Category 1 = Not active on any day	Category 1 = Insufficient activity (Not active on any day + Less than 5 sessions or 150 minutes)
Category 2 = Less than 5 sessions or 150 minutes	Category 2 = Sufficient activity (At least 5 sessions and 150 minutes)
Category 3 = At least 5 sessions and 150 minutes	NA

## Similarities and differences— QPHS and PHIDU/ABS

The QPHS method is based on the method above with the following similarities and differences:

- The PHIDU/ABS method utilises data from other ABS surveys such as the Census and the Household Income survey as well as estimated residential population (ERP) data. Some of this information isn’t currently available for Queensland, therefore, the proposed method will be based on QPHS data and ERP provided by the Statistical Services Branch (SSB).
- Synthetic estimates for the QPHS will use negative binomial/Poisson regression techniques for prevalence measures rather than logistic regression as this approach is more appropriate for modelling rates and counts. Linear regression is used to model means in the QPHS.
- Both the PHIDU/ABS and proposed QPHS methods use a combination of direct estimates and modelled estimates depending on its sampling variation.
- Because the PHIDU/ABS are reported as counts, results can be scaled so that smaller levels (for example SA2s) can be summed accurately to larger areas (for example SA3s). Because the QPHS reports as prevalence (rates), it is not feasible to scale the results so that smaller areas sum to larger areas.
  - This adds complexity to data management and, in some cases, interpretation.

- Mixing direct and synthetic estimates within geographic levels will add additional interpretation complexity. To reduce this complexity, currently only direct estimates are provided for SA3 and only synthetic for SA2.

The above similarities and differences mean that SA2 estimates in QSAS

- Are available only for **adult** QPHS data and at person-level only (not by sociodemographic characteristics)
- Do not include a confidence interval (CI) as the modelled methodology does not enable straight forward interpretation of CIs
- Do not have a comparable statewide modelled result so should not be compared to Queensland direct estimates; comparisons between SA2s is recommended by ranking
- **Cannot** be combined or 'rolled up' to generate higher level results (for example by SA3, LGA, HHS or PHN).

## Trend results

Trend analysis is the most appropriate method to assess changes over time because the effects of sampling variability are minimised as the number of data points increase. When data have been consistently collected over a period of time, trend analysis is more reliable than comparing two data points because it minimises the effects of sampling variability.

Trend results are expressed as annual percentage change (APC). APC measures the average percentage increase or decrease per year. For both state and regional trends, annual data are used. This contrasts with prevalence where two years of data are pooled.

Trends that are significantly increasing or decreasing are reported, specifically:

- Adults—analysis is undertaken for sociodemographic subgroups with significant differences in the rate of change over time reported
- Children—only person-level trends are available due to sample size limitations. Regions without sufficient sample size for reporting are omitted.

PHB trend analysis methods are described in [Trends in preventive health risk factors 2002–2013](#).

## Dissemination

### State and regional reporting

Official government statistics for Queensland are based on **annual survey data**. Regional results require combining (pooling) **two years of data** for adequate sample size.

Pooled Queensland results are provided for comparative purposes only; official Queensland statistics must always reference the annual results.

## Dissemination mechanism

The Queensland survey analytic system (QSAS; [www.health.qld.gov.au/phsurvey](http://www.health.qld.gov.au/phsurvey)) is the primary dissemination mechanism for the QPHS. QSAS provides QPHS results in various formats, for example interactive graphs/dashboards, downloadable machine-readable data, or custom requests).

QSAS replaced hard copy reports in 2015. Older data were reviewed and, in some cases, reanalysed to ensure results are comparable over time. Prior reports have been superseded by QSAS results.

QSAS contains **aggregate** results. The underlying unit record file (URF) data is based on surveys with complex sampling designs requiring analysis by specialist statistical software. Using QSAS results to calculate additional statistics (for example results for combined regions, indicators, or sociodemographic subgroups) is statistically inappropriate and will not produce valid results. Such approaches are strongly discouraged.

## Results available in QSAS

Routine QSAS products include:

- Headline key health indicators for adults and children
  - Queensland—persons, males, females
  - Regions—persons
- Adult risk factor trends
  - Queensland—persons, males, females, selected additional subgroups
  - Regions—persons
- Child risk factor trends—Queensland persons
- Adult and child detailed results by subgroups
  - Queensland—age, sex, socioeconomic status, remoteness
  - Regions—age, sex, socioeconomic status.
- Machine readable downloadable data are available for most of these products and often includes additional subgroup results that are not displayed on dashboards.

For the most up-to-date information on indicators, sociodemographic subgroups and regional data available in QSAS see the visualisation—[QSAS survey measures](#)

## Quality assurance

### Queensland Health disclaimer

The information in QSAS was prepared by the Queensland Department of Health, Preventive Health Branch. The Department has taken great care to ensure the information is correct and accurate. The Department does not guarantee, and accepts no legal liability whatsoever arising from, or connected to, the use of this information.



## Relative standard errors (RSE) and releasability

RSE are used to exclude results that do not meet statistical reliability criteria.

Releasability based on RSE is defined such that estimates with an RSE

- less than 25% are considered reliable (and are displayed in graphs/tables)
- between 25–50% should be interpreted with caution
- greater than 50% are not considered reliable (excluded from graphs/tables).

In any reproductions of these data, Users are strongly encouraged to identify results with RSE limitations. Results are also suppressed to protect participants' privacy, specifically where there are less than 50 total participants or less than 10 with the characteristic of interest. QSAS conditions of use specify that Users must not generate or disseminate suppressed results.

## Confidence intervals (CI)

Results will be presented with 95% CIs for statistical methods where confidence intervals can be calculated and interpreted. CIs are a function of sample size and the prevalence of the health indicator being investigated and allows Users to:

- assess the precision and reliability of results
- apply a conservative method to determine statistical differences—if CIs do not overlap the estimates are significantly different.

CIs are similar to margins of error in that they provide the range of values that would likely contain the result if the entire population was included in the survey.

## Contact information

[Population\\_Epidemiology@health.qld.gov.au](mailto:Population_Epidemiology@health.qld.gov.au)

Preventive Health Branch

Prevention Division

Department of Health | Queensland Government

15 Butterfield Street Herston QLD 4006

# Appendix

## Data collection methods summary

**Table 2: Adult survey methodology by year**

Survey year	Survey start	Survey end	Sample size	Survey time (minutes)	Response rate	Weighted to ERP year
2002 <sup>1</sup>	8/4/2002	13/6/2002	2,481		75%	2001
2004 <sup>1</sup>	27/4/2004	28/6/2004	2,231	15:00	71%	
2006 <sup>1</sup>	13/10/2006	26/11/2006	1,521		66%	2004
2008 <sup>1</sup>	10/6/2008	4/7/2008	2,002	20:42	47%	
2009 <sup>2</sup>	27/01/2009	25/03/2009	7,571	15:30	56%	2008
2010 <sup>3</sup>	29/10/2009	22/02/2010	8,959	12:07	65%	2009
2011 <sup>3</sup>	11/03/2011	6/06/2011	12,164	15:44	44%	2009
2012 <sup>3</sup>	3/10/2011	28/03/2012	19,398	13:26	81%	2011
2013 <sup>3</sup>	14/02/2013	22/05/2013	7,791	16:05	77%	2011
2014 <sup>3</sup>	2/12/2013	1/07/2014	14,787	16:00	68%	2012
2015 <sup>4</sup>	19/03/2015	29/06/2015	12,568	17:04	65%	2013
2016 <sup>4</sup>	4/11/2015	14/03/2016	11,948	14:18	64%	2014
2017 <sup>4</sup>	10/10/2016	26/03/2017	12,694	12:48	70%	2016
2018 <sup>4</sup>	18/10/2017	29/03/2018	12,386	14:18	68%	2016
2019 <sup>4</sup>	22/10/2018	28/03/2019	12,562	14:24	63%	2017
2020 <sup>4</sup>	21/10/2019	11/03/2020	12,536	13:50	61%	2018

1 Omnibus survey (limited indicators available)

2 Self reported adult health survey (SRAHS)

3 Self reported health survey (SRHS)

4 Adult preventive health telephone survey

**Table 3: Child survey methodology by year**

Survey year	Survey start	Survey end	Sample size	Survey time (minutes)	Response rate	Weighted to ERP year
2011 <sup>1</sup>	08/06/2011	28/07/2011	2,484	12:05	86%	2009
2013 <sup>1</sup>	14/02/2013	22/05/2013	2,467	14:09	88%	2011
2014 <sup>1</sup>	6/12/2013	9/06/2014	2,986	11:06	94%	2013
2015 <sup>2</sup>	03/06/2015	26/06/2015	2,521	14:51	80%	2013
2016 <sup>2</sup>	10/02/2016	15/03/2016	2,504	11:42	84%	2014
2017 <sup>2</sup>	6/03/2017	29/03/2017	2,393	17.54	77%	2016
2018 <sup>2</sup>	9/10/2017	7/11/2017	2,633	13.30	79%	2016
2019 <sup>2</sup>	17/10/2018	16/11/2018	2,518	16:06	74%	2017
2020 <sup>2</sup>	14/10/2019	14/11/2019	2,505	13:36	73%	2018

1 Child health status (CHS)

2 Child preventive health telephone survey

## Sample size information

**Table 4: Adult pooled datasets**

Pooled dataset	Sample size	Weighted to ERP year
2009–10	16,530	2008
2011–12	31,562	2011
2013–14	22,578	2013
2015–16	24,516	2014
2017–18	25,080	2016
2018–19	24,948	2018
2019–20	25,098	2018

**Table 5: Child pooled datasets**

Pooled dataset	Sample size	Weighted to ERP year
2013–14	5,453	2013
2015–16	5,025	2014
2017–18	5,026	2016
2018–19	5,151	2018
2019–20	5,023	2018

# Glossary

Term	Definition
<b>Confidence intervals</b>	In general, a range of values expected to contain the true value 95% of the time (95%CI).
<b>Estimate–direct</b>	Is the estimate of the health characteristic derived from a sampling process and may include population weighting to weight observations so that they reflect population characteristics.
<b>Estimate–mixed</b>	Is the combination of the direct and modelled estimates based on sampling variability. The method for synthetic estimates outlined in this report is a type of mixed method as it is derived using both the direct and modelled estimates. This is also the method used by PHIDU/ABS, however, additional sociodemographic variables from other ABS survey are include in the modelling process.
<b>Estimate–modelled</b>	A modelled estimate can be interpreted as the expected prevalence of a health characteristic for an area based on the demographic information available for that area. Note that PHIDU uses this term to mean the final estimate whereas QPHS uses it to refer to the modelled component of the final synthetic estimate.
<b>Estimate–synthetic</b>	Any method that attempts to infer any health characteristic of a small area based on a sample from a larger area. This is usually done because of limitations of sample size for smaller areas. The techniques used can include simulation and/or regression methods.
<b>Relative standard error</b>	Standard error measures how much a survey estimate is likely to deviate from the actual population. It is expressed as a number. By contrast, relative standard error (RSE) is the standard error expressed as a fraction of the estimate and is usually a percentage $RSE\% (x) = (\text{standard error}(x) / x) * 100$ .
<b>Sample variability (between-sample)</b>	Because samples are a subset of the population, variation across samples is expected. Between sample variability measures how different results are between samples. This is a function of sample size with larger samples having less variability.

## Version control

Version	Date	Comments
Version 1	4 November 2020	Initial release
Version 2	Xx August 2021	Added statistical methods for small areas
Review date	2 years from latest release	