NSW Health Survey: Review of the Weighting Procedures

Professor David Steel
Head, School of Mathematics and Applied Statistics,
University of Wollongong

Introduction

The NSW Health Survey is a telephone survey of the population of usual residents of NSW in private households. A sample of approximately 17,000 people is selected throughout the calendar year. The population is stratified into 17 strata corresponding to the 17 area health services. A sample of households is selected using Random Digit Dialling (RDD). One person is selected at random from the usual residents in the selected household. Children are included in the survey with the information provided by a person aged over 16.

The sampling frame was created by using 4 digit prefixes that contain working residential numbers (WRNs) in the Electronic White Pages (EWPs). Banks of 10 contiguous numbers with no numbers in the EWP were removed from the sampling frame. To improve the efficiency of the data collection the frame was matched with the Yellow Pages to remove business numbers.

The stratification of the RDD sampling frame according to area health services was achieved by geo-coding all the numbers present in the EWP using the listed address. Numbers not listed in the EWP with the same prefix are allocated to the area with the highest proportion of numbers listed in the EWP.

The resulting eligible numbers are then randomly ordered and divided into regions. The survey is spread throughout the year, by enumerating approximately a quarter of the sample each quarter. Numbers are called with an aim of achieved a prescribed sample of 250 respondents in each area. However, the responding sample in each quarter will not be exactly 25% of the annual sample.

Weighting Procedures

We will use the following subscripts:

- *i* denotes a person
- *j* denotes a household
- *h* denotes a stratum, i.e. area health service
- *a* denotes a age–sex cell

Also

- *N* denotes population size
- *n* denotes sample size
- *T* denotes number of telephone lines in the population
- *t* denotes number of telephone lines in the sample
- *s* denotes the sample

The probability of selection of person *ijh* is

$$\pi_{ijh} = t_h \frac{T_{jh}}{T_h} \frac{1}{N_{jh}}$$

The associated raw person weight that accounts for the different selection probabilities is then

$$w_{ijh} = \pi_{ijh}^{-1} = \frac{T_h}{t_h} \frac{N_{jh}}{T_{jh}}$$

These weights can be adjusted to agree with externally derived population benchmarks, N_{ah}. Let $\hat{N}_{ah} = \sum_{ijh \in s_{ah}} w_{ijh}$ be the survey based estimate of N_{ah}. The resulting post-stratified weight for $ijh \in a$ is then

$$W_{ijh} = \frac{N_{ah}}{\hat{N}_{ah}} w_{ijh}$$

The factor $\frac{T_h}{t_h}$ cancels in the calculation of W_{ijh}, so that if we let $z_{jh} = \frac{N_{jh}}{T_{jh}}$, then $W_{ijh} = \frac{N_{ah}}{\sum_{ijh \in s_{ah}} z_{jh}} z_{jh}$.

These weights can be summed to produce estimates of totals for any category and will agree with the external age–sex benchmarks. That is $\sum_{ijh \in s_{ah}} W_{ijh} = N_{ah}$, $\sum_{ijh \in s_h} W_{ijh} = N_h$ and $\sum_{ijh \in s} W_{ijh} = N$.

The weights can be scaled to sum to the total sample size, resulting in the scaled weights

$$SW_{ijh} = W_{ijh} \frac{n}{N}$$

In general there is no reason to scale the weights except perhaps if chi-squared tests are to be carried out.

Appendix 1 describes the weighting process currently in use in the notation used here. The result is a post-stratification weight which is equivalent to W_{ijh}. The calculation of W_{ijh} takes 9 steps, but could be achieved in 3. Another weight is also calculated called the sampling weight, which calibrates to the stratum benchmark population, ignoring age and sex, and is then scaled to the total sample size. It is unclear what the sampling weight is used for.

Adequacy of the current weight components adjustments.

In general sample weighting procedures serve three purposes. They are designed to:

- account for different chances of selection,
- adjust for variation of the sample from the population due to random selection,
- adjust for potential non-response bias.

The different chances of selection are accounted for by using the inverse of the probability of selection. In this survey the selection probabilities vary because

- different sampling fractions are used to select telephone numbers in each region,
- the probability of selection of a household is proportional to the number of telephone lines connected,
- selecting one person per household means that a person's selection probability is inversely proportional to the number of usual residents in the selected household.

The weighting procedure used appropriately accounts for the variation in selection probabilities. (See page1, weighting procedures)

These factors all lead to an increase in the variation of the weights. Equal size samples for each area are used because the purpose of the survey is to produce area estimates.

Cases of a large number of usual residents in the household can lead to very large weights. Such cases can arise due to families with a large number of children. In some cases a large number of adults can arise because of the special nature of the dwelling in which the selected telephone line has be selected. In the 2003 survey there were 15 households with greater than 8 persons in them. It is suggested that such cases be double checked, at the time of interview if possible, to determine what is the reason for such a large number.

Cases of several telephone lines into a household can lead to relatively low weights. In the 2003 survey there were 13 households with greater than 3 telephone lines. It is suggested that such cases be double checked, at the time of interview if possible, to determine what is the reason for such a large number.

A useful check of the weights is to examine their distribution. Examination of the coefficient of variation (CV) of the weights and the upper and lower tails of the distribution can be useful in identifying problems with the weighting. Because a major cause of variation in the weights is use of different sampling fraction between areas, these analyses should be done for each area as well as for the entire sample. In some case trimming of weights may have to be considered (see page 4, weight trimming procedures). In the 2003 survey the CV of the weights within a stratum varied from 0.58 to 0.71 and the ratio of the maximum to median weight varied from 4.0 to 8.4 across the health areas.

The post-stratified weights attempt to account for the probability weighted sample profile possibly differing from the population due to random selection and non-response. This is achieved by calibrating the survey estimated after weighting for unequal selection probabilities to age–sex population benchmarks for each area. This is a standard approach. Issues arise as to the precise level of detail of the age categories that should be used. Again there is also the issue of the adjustment of the ABS population estimates to the population of usual residents in private households.

Calibration of the estimates to age–sex population benchmarks is a common approach. One reason is that such benchmarks are available annually from the ABS for sub-state areas. Another reason is that many variables, including health variables, are related to age and sex. Using post-stratification by age–sex effectively corrects for the weighted sample being different from the population due to non-response and random variation. It eliminates for non-response bias if non-respondents and respondents have the same mean within age–sex cells in a region.

There is a limit to the fineness of the post-stratification. If the number of respondents in a post-stratification cell is small, there can be large variation in the associated weights. There is also an non-negligible chance of zero respondents in a cell. For these reasons it is advisable to limit the number of post-strata. A rule of thumb is that a cell should have an expected number of respondents of at least 6 (see page 4, weight trimming procedures).

In the past adult and child data were collected and weighted separately. Five year age groups were used for adults and single year age groups for children.

Calibration of weighting procedure to socioeconomic status

Weighting procedures can be developed that calibrate the survey estimates to external estimates. For this to be worthwhile there must be such external estimates available reliably and regularly at an appropriate geographic level. Estimates from major ABS surveys could be considered. To use these figures at the health area level the ABS figures would have to be obtained at this geographic level. An investigation on what figures are available at a useful frequency, timeliness and geographic level could be undertaken. Use of ABS survey data has its own problems. Regional estimates from such surveys can be volatile. Comparability of data collected in the health survey and ABS survey is also an issue. While there may be minor inaccuracies in the reporting of age and sex in the Health Survey, the potential for other variables, for example employment status, of major inconsistencies is probably much greater. Census variables or variables for which information is available less than annually can be considered, However, assumptions would have to be made concerning how quickly the distribution of these variables changes over time.

Calibration to state level auxiliary variables can also be considered, but given the focus on area level estimates the benefits are likely to be small.

For calibration to socioeconomic variables to be worthwhile, the values of the key variables of interest should be related to the socioeconomic variables. This can be examined through an analysis of the current survey data. In practice there would probably have to be a trade-off in terms of the number of t age categories that could be used in conjunction with any socioeconomic variables. So an analysis would examine this trade-off.

Non-response rates should also have to vary according to the socioeconomic variables for their use in estimation to be of value. To determine if this condition applies, the age–sex weighted survey estimates of the socioeconomic variables should be compared with external information on them.

Although use of auxiliary population information on socioeconomic variables is unlikely to be feasible, comparison of weighted survey estimates obtained from the NSW Health Survey with other published estimates would be useful in assessing the overall quality of the survey. In performing these comparisons the possible sampling errors on both sets of estimates need to be taken into account as well as any differences in definitions, coverage and questions used.

Household weights

Characteristics of the household in which a responding person resides can also be considered for use in the estimation and weighting. This is a possibility only for household characteristics for which information is collected in the survey. The issues associated with using household characteristics are the same as using person level socioeconomic characteristics. That is availability of reliable external estimates with the required frequency, timeliness, definitions, geographic level, that are sufficiently related to the key variables of interest and variation in non-response.

Implications of combined sampling of adults and children

In the current survey a single person is selected from a selected household completely at random irrespective of age. This means that there is no control on the age–sex profile of the sample. It also means that adults and children can be treated in the same way in the estimation and weighting. The weighting procedure appropriately accounts for the probability of selection of a person being inversely proportional to the total number of usual residents in the selected household.

There is no reason due to the specified selection method that the weighted data should under-represent adults in households without children. The selection probabilities of adults for the current design are different from those in a design in which the sample was selected from adults in the household. This is reflected in the weighting for the current survey using the total number of people in the household of all ages rather than the number of adults. However, the current survey should use the selection probabilities used in the design used not an alternative that was not used.

The selection of a person within the selected household is a potential source of bias if the selection process is not followed. In as much as any selection bias within households is related to age and sex, it is accounted for by the age–sex component of the weighting.

Weighted data from child and adult dataset

It is planned for separate data sets will be made available for adults and children. Adult data sets will be available annually whereas the children data set will be available biennially, that is based on two years sample.

If the data were made available for each year the weights can be used directly. Because of the selection of one person from each selected household the adult and children sample for a particular year must be weighted together.

The only complication of the dissemination strategy is how to handle the child data being released for 2 years. A straightforward approach is to use the original weight divided by 2. This means that any estimates produced refer to the average of the 2 years in question, which is presumably the intention.

Appropriate use of population benchmarks using quarterly weighting

The discussion so far has assumed that the weighting is performed for the annual survey. The survey is enumerated using quarterly non-overlapping samples. If only annual estimates are required then the weighting can be done using the entire annual survey. An alternative is to implement the weighting on a quarterly basis. Quarterly weighting would adjust for any differences of the age–sex profile of the sample over the year. This may be important for variables with a strong seasonal effect. However, any benefits would have to be balanced against the need to reduce the number of age categories used in the weighting process.

Quarterly weighting would be beneficial if quarterly estimates are required. However, the weighting cells should be reviewed to ensure a reasonable number of respondents in each one. Given a quarterly sample of approximately 125 in each sex, then categories should be chosen so that they contain at least 6/125=0.04, that is 4.8% of the male or female population in order to ensure that the expected sample in each category of at least 6 respondents in each quarter. This requirement implies no more than about 20 age categories. Analysis of the population distribution across ages should be undertaken to decide the categories to be used. Any merging of age categories should be based on the similarity of the means of important variables in the categories to be merged.

Production of quarterly estimates would mean that figures can be produced in a more timely manner. However, analysis of changes between quarters could be unreliable as the sample size is small and the estimates are based on non-overlapping samples. Rolling estimates for 4 consecutive quarters can be produced.

If quarterly weighting is used and estimates for several quarters are required then this can be done by dividing each weight by the number of quarters involved. This procedure includes the case of producing annual estimates by averaging the estimates for each quarter.

Weighting of sub-samples

Some questions are only asked in less than all quarters. Quarterly weights can be applied. Estimates for more than one quarter can then be produced by simple averaging.

In some quarters some questions may be asked for a randomly selected sub-sample., which is often one third of the main sample. Denote the sub-sample by *ss.* There are several options available:

1. Scale the weights W_{ijh} of those in the sub-sample to the stratum population N_h by applying the factor $\frac{N_h}{\sum_{i \in ss_h} W_{ijh}}$
2. Scale the weights W_{ijh} of those in the sub-sample to the stratum age–sex population N_{ah} by applying the factor $\frac{N_{ah}}{\sum_{i \in ss_{ah}} W_{ijh}}$
3. Re-weight the sub-sample separately, which could lead to problems with small sample sizes
4. Weight the sub-sample to the main sample.

Option 2 is suggested.

Weight trimming procedures

The design of the sample means that weights can vary considerably. Some extreme weights may be traced to large numbers of usual residents and/or telephone lines. Such cases should be investigated as they may reflect special cases that deserve investigation and be related to the precise definition of the population in the scope of the survey. The scope of the survey excludes people resident in institutions. The population benchmarks are adjusted to allow for people resident in institutions. Extreme weights may be due to selection of a telephone number in a business or institution of some other sort of special dwelling.

The choice of weighting categories in the post-stratification should try to reduce the variation of the weights. However, given the large variation in area sampling fractions, number of people within households and telephone lines a considerable degree of variation should be expected. Additional variation due to non-response can also be expected. A hopefully small number of extreme weights may be experienced. The weight associated with a respondent is the number of people that they effectively represent and will be there contribution to any cell in a table that they belong to. Extreme weights increase the variance and instability of estimates.

Automatic weight trimming can be considered. Trimming of an extreme weight can reduce the variance an instability of estimates, but may also introduce bias. Procedures can be introduced into the steps used to calculate the weights to reduce the possibility of extreme weights. For example truncation of the number of people or telephone lines in the calculation of the weight that account for the variation in the selection probabilities. Even if such procedures are used, there will usually be a need to check the final weights. Trimming of a weights will involve adjustment of the weights for other respondents in order to preserve the calibration to the population benchmarks.

Various methods have been suggested to implement weight trimming, although no simple rules of thumb have been proposed. Recommendation of a particular methods or a simple rule would require some analysis of the survey weights and data.

Based on the 2003 weights there does not seem to be a need to trim the weights. Particularly high or low weights are likely to correspond to a household containing a large number of people or telephone lines respectively. Such households may be unusual or possibly due to some sort or reporting or processing error. It is suggested than for weighting purposes the number of people be truncated at 10 and the number of telephone lines truncated at 5. This will affect only a very small number of households (6 and 3 respectively in 2003), but should reduce the most extreme weights.

If we assume that the number of respondents in post-stratification cell has a Poisson distribution then for a cell with an expected sample take of 5 the probability of a zero respondents is 0.0067 or 1 in 148. If the expected sample take is 6 the probability becomes 0.0025 or 1 in 403. The use of single year age groups up to age 18 and 5 year age groups above 18 implies 66 (?) age–sex groups in each are, that is 1122 in all. So even if with a threshold of 6 we should expect to encounter the issue of zero responding sample in a weighting cell. To reduce this probability to a rare event would need us to use say a value of 10 resulting in a probability of 1 in 22,026. Even so the weighting system needs to be able to cope with the occasional case of zero sample, which can be handled by merging age categories.

Use of weights in trend analysis

Weighted estimates can be examined over time. The weights used will be based on the ABS population estimates for the year in question. Changes observed will be the net result in changes in changes in the composition of the population and changes in the means within subpopulations. The sub-populations can be defined according to any variables, but using age and sex is fairly common. Consider the estimate of the population total of a variable Y for the period *t*,

$$\hat{Y}_t = \sum_{ah} N_{aht} \hat{\bar{Y}}_{aht}$$

The changes can be examined controlling for the change in composition of the population by age and sex, by the common process of age–sex standardization.

Then there are at least two ways of standardizing the estimate for age–sex.

(1) Use the age–sex structure for some base period, $\tilde{N}_{ah}$, but the current period's estimate of the mean within age–sex cell, resulting in

$$\tilde{Y}_t = \sum_{ah} \tilde{N}_{ah} \hat{\bar{Y}}_{aht}$$

(2) Use the age–sex structure for the current period, $\tilde{N}_{ah}$, but the base period's estimate of the mean within age–sex cell, $\hat{\bar{Y}}_{ah0}$, resulting in

$$\breve{Y}_t = \sum_{ah} N_{aht} \hat{\bar{Y}}_{ah0}$$

Other forms of standardization are also possible.

The standardized estimate $\tilde{Y}_t$ can be implemented for all variables by included a single additional weight that calibrates the current survey to the age–sex benchmark for the base period.

In analysing these changes sampling errors should also be taken into account.

Production of an age–sex standardized series should probably be done on an experimental basis at first. The original series should still be produced even if a standardized one is compiled.

Weighting for new health areas

Changes may occur to the geographic definition of health areas. In such circumstances it is necessary to produce estimates from pervious surveys for the new health areas. This will involve coding the respondents to the old survey to the new areas. Estimates can then be produced from the old survey according the new areas by using the original weights and treating the new areas as a cross-classifications.

If population benchmarks are available for the time period of the older survey according the new heath areas then the original weights can be adjusted to agree with these benchmark. Denote the new health areas by the subscript *k*. The adjustment correspond to multiplying the original weights, W_{ijh}, by the factor $\frac{N_{ak}}{\sum_{i \in s_{ak}} W_{ijh}}$, where s_{ak} is the sample of units in the new health areas in age–sex cell *a*.

Appendix 1: Current Weighting Process

1. Calculate pselect = z_{jh}.

2. Calculate apop= N_{ah}, which is an estimate of the residential population. It is calculated using apop=pop(1-nrespop), where pop comes from the ABS mid-year estimated resident population. The factor nrespop is an estimate of the proportion of the usual resident population that is non-residential, that is in institutions.
(What is the source of the date used to calculated nrespop. What does it attempt to exclude. Are retired or aged care dwelling excluded or included in the apop figures)

3. Calculate ahspop= N_h, NSWpop= N.

4. Calculate sumpsel= $\sum_{ijh \in s_h} z_{ijh}$, ahssamp= n_h, NSWsamp= n.

5. Calculate hhwgt=pselect*(ahssamp/sumpsel)= $z_{jh} \frac{n_h}{\sum_{ijh \in s_h} z_{jh}}$.

6. Calculate spersons= $\sum_{ijh \in s_{ah}} z_{jh} \frac{n_h}{\sum_{ijh \in s_h} z_{jh}}$, then

sprop=spersons/ahssamp= $\sum_{ijh \in s_{ah}} z_{jh} \frac{1}{\sum_{ijh \in s_h} z_{jh}}$

7. Calculate pprop=apop/ahspop= $\frac{N_{ah}}{N_h}$.

8. Calculate agesxwgt=pprop/sprop= $\frac{N_{ah}}{N_h} \frac{\sum_{ijh \in s_h} z_{jh}}{\sum_{ijh \in s_{ah}} z_{jh}}$,

9. (a) Calculate Post-stratification weights
wgt=(ahspop/ahssamp)*agesxwgt*hhwgt

$$= \frac{N_h}{n_h} \frac{N_{ah}}{N_h} \frac{\sum_{ijh \in s_h} z_{jh}}{\sum_{ijh \in s_{ah}} z_{jh}} z_{jh} \frac{n_h}{\sum_{ijh \in s_h} z_{jh}}$$

$$= N_{ah} \frac{z_{jh}}{\sum_{ijh \in s_{ah}} z_{jh}} = W_{ijh}$$

(b) Calculate sampling weights

wgtsamp=hhwgt*(ahspop/NSWpop)/(ahssamp/nswsamp)

$$= z_{jh} \frac{n_h}{\sum_{ijh \in s_h} z_{jh}} \frac{N_h}{N} \frac{n}{n_h}$$

$$= z_{jh} \frac{N_h}{\sum_{ijh \in s_h} z_{jh}} \frac{n}{N}$$

The sampling weight calibrates to the stratum benchmark population and is then scaled to the total sample size.