Weighting matters: a practical application

D Hindmarsh [12], C L Birrell[2], D Steel[2], L Corscadden[1]

[1]Bureau of Health Information, Level 11 67 Albert Avenue Chatswood, 2067, NSW Australia
[2] National Institute for Applied Statistics Research Australia (NIASRA),
School of Mathematics and Applied Statistics
University of Wollongong
NSW 2522 Australia

Abstract
Introduction: The calculation and application of weights is an important step in producing estimates from a sample survey. Practical application of survey weighting requires a number of decisions on how to handle issues, such as extreme weights and small cell sample sizes. Generally, discussion centres around the extreme weights themselves, rather than the impact of the extreme weights on results. This paper aims to identify diagnostics that may assist in making decisions about aggregation of small cells and trimming of extreme weights, partly by assessing impacts on results. It uses a case study where there is high variability in the survey weights due to the survey design.
The 2017-18 New South Wales (NSW) Emergency Department survey covers 82 hospitals, ranging from tertiary teaching hospitals to small local hospitals. Unbiased estimates are required for each hospital, as well as for 16 administrative areas (Local Health Districts (LHD)) and for NSW. Quarterly data are used to report key performance indicators at the LHD level.

Method: Several options for modifying weights were applied to data from the New South Wales (NSW) Emergency Department (ED) survey. We introduce a novel diagnostic we call the deviation sum of squares - the sum of squares of deviations of the sum of weights vs the population. As well as looking at the deviation sum of squares and other information regarding the distribution of the weights, we considered the impact of the different weights on the estimates of the proportion in the most positive response option of six questions, and on the score-based performance indices reported quarterly to the NSW Health system.

Results: Although weights need to be applied at the level of the stratification, the actual mechanism used to create the weight had a minimal effect on results for key variables at the three reporting levels (hospital, LHD, NSW) and performance indices at LHD level. The deviation sum of squares, together with the maximum weight can assist in informing decisions about aggregation of cells and trimming of weights.

Conclusion: Simple diagnostics of the weight distribution, maximum weight and how closely the weight totals agree with the population by stratum can be used to assist in the decision-making about how to modify weights. Results for the NSW ED survey appear reasonably robust to the method used to create the weights, provided the weights are created at the stratum level, and have a low deviation sum of squares.

Key Words: Survey analysis, Weighting, Trimming

1. Introduction

Weighting is an important step in preparing survey data for analysis. Most papers that discuss weighting focus on rules concerning the number of observations in a stratum and when and how to aggregate weighting cells (Potter, 1988, Botman et al., 2000, Karlton and Flores-Cervantes, 2003), and how to deal with extreme weights (Chowdhury et al., 2007, Battaglia et al., 2006, Battaglia et al., 2009), usually by trimming, due to the effect of extreme weights on the mean square error (MSE) of estimates (Potter, 1990). Few papers discuss how decisions on weighting affect the survey estimates and associated MSEs.

Generalised regression weighting methods allow weights to be created for margins rather than for each individual stratum cell. Iterative processes can be used until the sum of the weights meet the appropriate marginal population totals – called benchmarks (Bell 2000). The Australian Bureau of Statistics (ABS) devised the GREGWT macro in SAS (Bell 2000) to undertake generalised regression methods for use in weighting its surveys (for instance, Watson, 2004). It has also been used for surveys in New Zealand (Clark et al, 2008).

This paper applies a range of methods to aggregate small cells with large weights and/or trim large weights to see how weighting decisions affect results in an existing survey. The ultimate aim is to see if an automated method of dealing with small cells and large weights can provide similar results to the current manual method and particularly if diagnostics regarding the weights can help with decision-making. Of particular interest is the influence of cells with high weights on estimates of hospital, local health district (LHD) and NSW-level results and whether the influence was greater for larger hospitals with a small sampling fraction. Given that the aim of trimming extreme weights is to reduce the relative MSE (RMSE) then that should be the basis of any diagnostics. This paper focuses on the effect of different weighting methods on the distribution of the weights, RMSEs of the weights and the concomitant effects on quarterly estimates for six questions, and key performance indicators (KPIs) used for quarterly performance reporting.

The remainder of the paper is as follows. Section 2 provides a brief outline of the NSW Emergency Department Patient Survey 2017-18, a brief description of the six questions and the two KPIs used to test the weights. Section 3 describes the various methods of calculating weights and associated diagnostics. The deviation of the sum of the weights from the population is introduced as a useful diagnostic. Section 4 follows with the analysis of the six questions and the KPIs using the various weights, followed by discussion and conclusions.

2. The NSW Emergency Department Patient Survey, 2017–18

Responses for a questionnaire about quality of care while at the Emergency Department (ED) were obtained from 15,995 patients visiting one of 82 EDs in New South Wales (NSW) between 1 July 2017 and 30 June 2018. The weighted response rate was 24% (accounting for differential response rates applied to the age strata). The eligible patient populations range from 2,000 to 64,000 in EDs and sampling fractions range from 2% to 22%.

This survey is one of a suite of surveys run through the NSW Patient Survey program administered by the Bureau of Health Information. The NSW Patient Survey Program website (http://www.bhi.nsw.gov.au/nsw_patient_survey_program) provides extensive information about sample selection and weighting in the technical supplements available for each survey. The following information is provided as background for this paper.

2.1 Survey design

Sampling is designed to provide sufficient respondents at the hospital level for reporting. Smaller district hospitals are sampled for reporting annually; larger metropolitan and tertiary hospitals are sampled for reporting quarterly. Sample sizes are decided at the hospital level, and then sampled proportional to size within each hospital by age (0-17, 18-49, 50+) and stay type (admitted, discharged) strata. Differential response rates are applied to take into account the fact that patients aged 50 years and older are more likely to respond than any patients or parents of patients aged less than 50.

2.1 Weighting

Quarterly weights are created to obtain quarterly results by LHD and for NSW as a whole. In order to ensure that quarterly results include all hospitals, those hospitals that are sampled for annual reporting are collapsed into a single group by LHD for quarterly weighting. The current weighting protocol is presented in Appendix 1.

When all four quarters of data are received, the weights for hospitals sampled for annual reporting are adjusted based on the full year of data. This paper focuses on the effects of decisions for weighting at the quarterly level.

2.1 Weighting

We selected six questions and two score-based KPIs created on a quarterly basis for internal reporting at the LHD level. The six questions and the KPIs and improvement measures created using the score-based measures are shown in Table 1. The questions have between three and five response options.

Table 1: Questions used for testing the different weighting methods and the associated score-based measures

KPI or improvement measure in which the question is included	*Question number*	*Question text*
Patient-centred care improvement measure	Q16	Did the ED health professionals explain things in a way you could understand?
Patient engagement (discharged patients) KPI	Q19	Were you involved, as much as you wanted to be, in decisions about your care and treatment?
Patient-centred care improvement measure	Q23	Health professionals worked together very well
Patient-centred care improvement measure	Q29	Did you feel you were treated with respect and dignity while you were in the ED?
Patient engagement (discharged patients) KPI	Q53	Thinking about your illness or treatment, did an ED health professional tell you about what signs or symptoms to watch out for after you went home?
Overall experience KPI	Q63	Overall, how would you rate the care you received while in the ED?

Assessment of differences is based on a test of non-overlapping confidence intervals between the LHD or hospital and the NSW level result, using 12 months of data.
The score-based KPI measures for overall care and patient engagement were also calculated using the six different weight methods. Scores for each question are allocated in the following manner: the most positive category is given a score of 10, the most negative category given a score of 0 and remaining categories scored linearly. Missing responses, and responses such as 'Don't know/can't remember' are excluded from the scores. These two KPIs are made up of four to seven questions (NSW Health, 2019).

3. Weighting methods, decisions and diagnostics

Six different weighting methods were created as shown in Table 2. They are presented in ascending order of complexity. The current manual method is version 3.

Table 2: Summary of Weighting Methods

Weight method	*Question number*
0. **Same** weight (identical to no weight)	Total eligible pop/total number of respondents
1. **Hospital** level weights	Hospital eligible pop/total number of respondents per hospital
2. **Stratum level** weights	Zero cells merged with the adjacent age-stratum cell, no other changes
3. **Current method**	Stratum level weights, zero cells merged, manually check cells with less than 6 responses and cell wgt > median wt, GREGWT macro used to set max to 400 **(highly manual).**
4. **GREGWT* with added margins by volume**	As per current, then benchmark margins to 3 volume based hospital groupings, max 400
5. **GREGWT macro with multiple margins**	No manual aggregation apart from merging zero cells, 6 marginal benchmarks: Benchmark 1 = quarter lhd agest3 Benchmark 2 = quarter lhd agest2 Benchmark 3 = quarter lhd_hos servicecategory Benchmark 4 = quarter peergrp4rpt lhd Benchmark 5 = quarter lhd age_strata Benchmark 6 = quarter peergrp4rpt agest2 = ages 0-18 and 18-49 combined agest3 = ages 18-49 and 50+ combined lhd_hos = LHD combined with modified hospital code Peergrp4rpt = Proxy of size and complexity of cases seen at hospital. Annually reported peer groups combined to a single group

*The GREGWT macro (Bell 2000) is used by permission of ABS.

Three diagnostics of the weights are calculated: the maximum; the coefficient of variation (CV) of the weights (i.e. standard deviation divided by mean); and a measure called the deviation sum of squares. The CV can be used in the calculation of the design effect (DEFF) due to weighting. The deviation sum of squares (SSDev) is calculated as the sum of the squared deviations of the sums of the weights to the population summed across the six strata, all hospital-LHD groupings and four quarters (Equation 1).

.

$$SSDev = \sum_{k=1}^{4}\left[\frac{1}{\Sigma\Sigma n_{ijk}}\sum_{i}^{QH}\ \sum_{j}^{6}\frac{n_{ijk}}{N_{ijk}}\left(n_{ijk}w_{ijk} - N_{ijk}\right)^2\right] \quad (1)$$

where

n_{ijk} = number of respondents for i^{th} hospital group ($i = 1 \ldots QH$,where QH is the total number of hospital groups used for quarterly reporting, j^{th} stratum ($j = 1 \ldots 6$), and k^{th} quarter $k = 1 \ldots 4$);

w_{ijk} = weight for i^{th} hospital group, j^{th} stratum, k^{th} quarter and

N_{ijk} = eligible population for i^{th} hospital group, j^{th} stratum, k^{th} quarter.

Table 3 provides the summary of these measures for the six weight methods used in this assessment.

Table 3: Maximum weight, Coefficient of variation of weights and sums of squares of deviations (SSDev) of 6 different weighting methods, Emergency Department Patient Survey 2017-18

Weight method	*Maximum weight*	*CV weights*	*SSDev*
0. **Same** weight (identical to no weight)	115	0	9380
1. **Hospital** level weights	257	39.5%	4581
2. **Stratum level** weights	1059	62.0%	0
3. **Current method**	400	59.7%	31
4. **GREGWT* with added margins by volume**	400	57.8%	346
5. **GREGWT macro with multiple margins**	400	59.8%	33

Of these methods, the most promising option is using weight method 5 (GREGWT with six marginal benchmarks and trimmed to a maximum of 400). It had among the lower SSDev and a similar CV of the weights at NSW level to that for the current method, without the need to manually assess small cells.

4. Analysis of the six questions

There is no 'gold standard' for any of these questions, as the only data comes from surveys. The sample sizes for each hospital are designed to give estimates with a margin of error of ±7% for each reporting period. If the more automated method of weighting creates results that are within, say, 1% of the current method, and does not adversely affect KPIs being reported, then the weighting method can be updated. As with all reports on results for the NSW Patient Survey Program, the weighting method is presented in the technical supplement published at the same time as the report. The range in results for the six questions for each of the weight methods is shown in Table 4.

Weighting has clearly had an impact on the results as shown by the minimum and maximum values. For most questions, when all respondents are weighted equally (weight method 0) is quite different to the range for those weighted as per normal protocol (weight method 3). Results for weight method 1 are consistently closer to results for weight method 0 than those using weight method 3, suggesting that it is insufficient to weight at the

hospital level. The under 50 population is oversampled to compensate for the lower response rate of this group, so weighting at the hospital level does not adjust for any such oversampling. Results for weight method 2, which includes a weight of over 1000, are very similar to the current method for most questions. Weight method 4 also has similar ranges to the current method but the minimums for question 16 and 53 are 0.7 and 2 percentage points different to the current method respectively. In contrast the range in results for the weight method 5 is very similar to weight method 3, and any difference is only 0.1 or 0.2 percentage points respectively.

Table 4: Minimum and maximum percent, by hospital, in most positive response category for six questions, by weight method

	Question	Weight method					
		0^	1	2	3*	4	5
Minimum	Q16 They *always* explain things well	67.0	67.4	66.9	66.9	67.6	66.9
	Q19 Definitely involved in decisions	52.3	52.7	51.6	51.6	51.6	51.8
	Q23 Health professionals worked together *very well*	39.8	39.9	40.0	39.9	39.9	40.1
	Q33 Definitely treated with respect and dignity	72.6	72.6	70.7	70.7	70.7	70.6
	Q53 They told me *completely* what to watch for after departing ED	45.2	45.8	42.5	42.5	44.5	42.4
	Q63 Overall *very good* rating of care	37.6	37.8	37.5	37.9	37.9	37.9
Maximum	Q16 They *always* explain things well	91.5	91.4	96.3	96.3	96.3	96.3
	Q19 Definitely involved in decisions	83.7	83.3	84.9	84.9	84.9	84.9
	Q23 Health professionals worked together *very well*	80.2	80.3	79.1	79.1	78.5	78.9
	Q33 Definitely treated with respect and dignity	94.3	94.2	95.8	95.8	95.8	95.8
	Q53 They told me *completely* what to watch for after departing ED	74.6	74.8	73.5	73.5	73.5	73.4
	Q63 Overall *very good* rating of care	82.8	82.9	80.3	80.3	80.6	80.3

^ - equal weight per respondent * - current weighting method

The estimated standard errors of results were also similar between the current weights weight method 3 and weight methods 2, 4 and 5 (Table 5). In contrast, weight method 0 and weight method 1 have lower estimated standard errors. This result is expected, as both of these methods effectively treat the survey as a simple random sample, with the weights for the latter just adjusting for patient volume at the hospital level.

In terms of performance measurement using the current method based on overlapping confidence intervals, there are only minor differences in the number of results that would be flagged using weight method 2 to weight method 5 (Figure 1). All four methods would give the same number of LHDs that are worse than the NSW result, and there are only minor differences in numbers of LHDs that would be considered ‘better’ than NSW. There would be slightly fewer results that are flagged at hospital level, both as being ‘worse’ than NSW using weight method 4 or 5 than weight methods 2 or 3. Weight method 0 would

have around 8 more flagged as worse. Weight method 1 would flag over 10 more hospitals as being better than NSW than any other weight method – including for weight method 0. There are small differences in the number results that would be flagged as 'better' among all the other weight methods.

Table 5: Minimum and maximum standard error of percent in most positive response category for six questions, by weight method

	Question	Weight method					
		0^	1	2	3*	4	5
Minimum	Q16 They *always* explain things well	0.31	0.35	0.38	0.37	0.37	0.37
	Q19 Definitely involved in decisions	0.39	0.43	0.46	0.46	0.46	0.46
	Q23 Health professionals worked together *very well*	0.39	0.42	0.47	0.46	0.46	0.46
	Q33 Definitely treated with respect and dignity	0.28	0.31	0.34	0.34	0.33	0.34
	Q53 They told me *completely* what to watch for after departing ED	0.49	0.53	0.56	0.55	0.55	0.55
	Q63 Overall *very good* rating of care	0.37	0.41	0.45	0.45	0.44	0.45
Maximum	Q16 They *always* explain things well	5.02	5.27	6.56	6.56	6.56	6.57
	Q19 Definitely involved in decisions	6.01	6.00	8.01	7.99	7.99	8.03
	Q23 Health professionals worked together *very well*	6.07	6.07	7.91	7.92	7.92	7.92
	Q33 Definitely treated with respect and dignity	4.91	5.08	6.70	6.70	6.70	6.71
	Q53 They told me *completely* what to watch for after departing ED	9.34	9.35	10.09	10.29	10.29	10.24
	Q63 Overall *very good* rating of care	5.61	5.82	7.21	7.21	7.21	7.28

^ - equal weight per respondent * - current weighting method

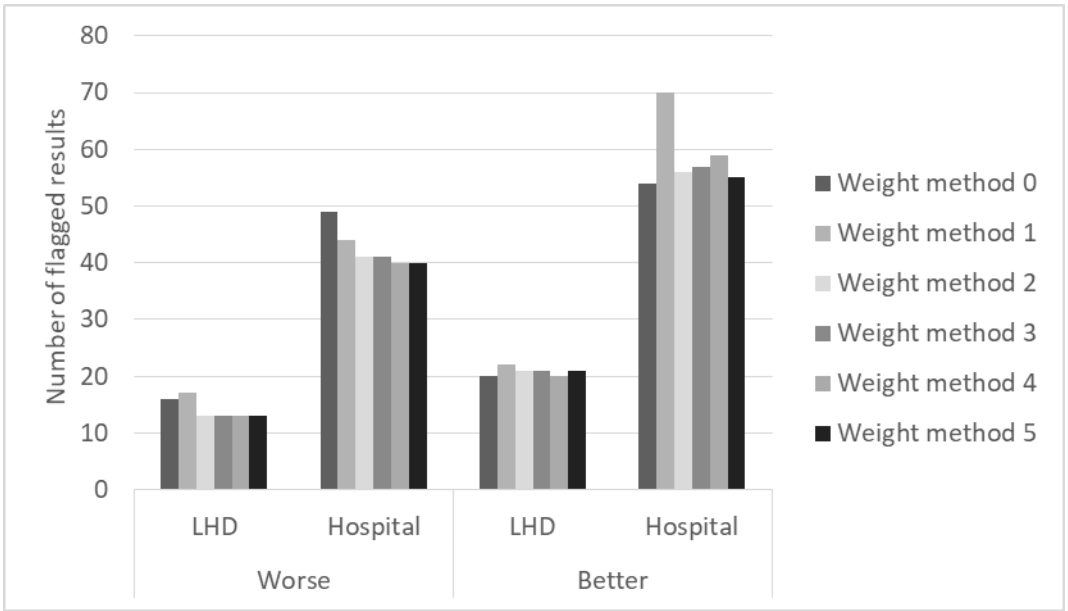

Figure 1: Significance summary at hospital and LHD level summarised over the 6 questions, ED 2017-18

Based on benchmark cut-offs for quarterly performance measured at the LHD level, the effect of the weighting on the patient engagement (discharged patients) KPI is negligible for flagged results both below and above benchmarks (Figure 2). For the overall experience KPI, the same number of LHDs are below the benchmark for weight method 5 and weight method 2 compared with weight method 3 (the current weight system), but weight method 2 and weight method 5 result in one more LHD result above the benchmark than for weight method 3. Weight method 4 differs by one in the number below and above the benchmarks. In contrast use of weight method 0 and weight method 1 for the overall experience KPI have a different pattern, with a lower number below benchmark and a higher number above benchmark.

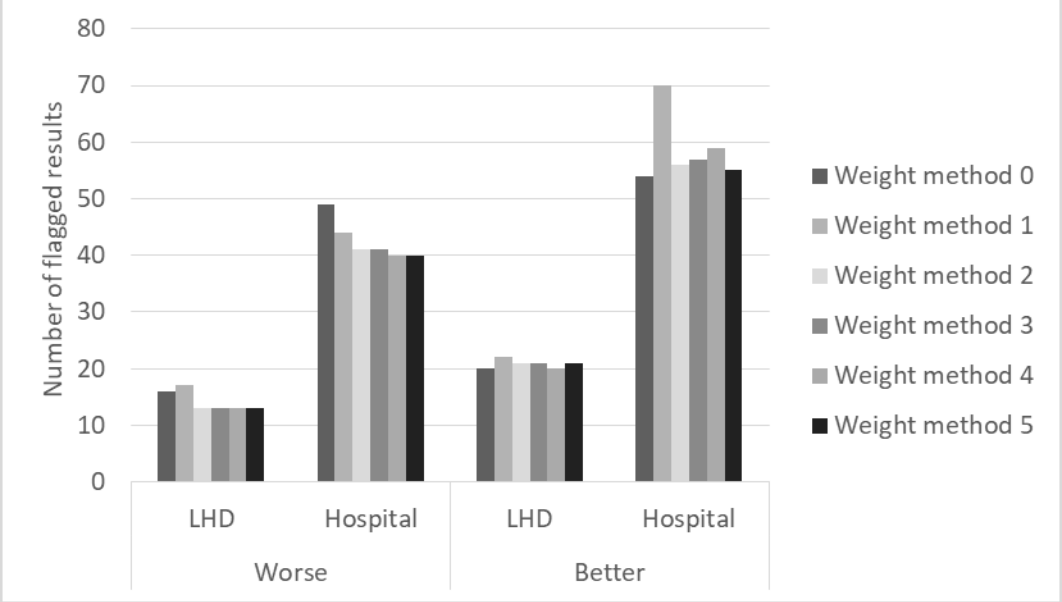

Figure 1: Number of LHDs below or exceeding KPI benchmarks, summarized over four quarters and 17 LHDs, ED 2017-18

5. Conclusions

It is clear that weighting at the stratum level has made a difference compared to when all respondents are giving equal weighting or when weighting is undertaken at the hospital level and ignores the effect of age and stay type. Results for the six questions when weights were created at the hospital level were similar to those when all respondents were given equal weighting.

The differences between results based on the current weighting method and any of the four methods that weighted at the stratum level were very minor in this example and definitely within the expected margin of error (±7%). The effect of the large maximum weight in weight method 2 would depend on the number of responses in the cell, how close the responses are to the observed pattern and the numbers in the other cells for that hospital or LHD. Weight method 4 requires inclusion of a metric based on volume at hospital level. We consider that weight method 5 – the method with six benchmarks is easier to implement in practice than weight method 4. The advantage of the proposed method is that it uses general regression, can be coded quickly and avoids the manual method that is currently utilised.

The diagnostics that were useful in determining the best method were the maximum weights and the deviation sum of squares. This process will be implemented in parallel to the current weighting method for two other BHI surveys to check whether the results are consistent prior to being used to replace the current method.

The use of results to selected questions to gauge the impact of the different weights was very informative. It can quantify the effect of several weighting methods in a transparent manner, which is very useful when creating results that will be used for performance measurement.

Acknowledgements

The authors acknowledge the support of the Bureau of Health Information in allowing access to the NSW Patient Survey Program data for this study.

References

Battaglia, M. P., Frankel, M. R. and Link, M. (2006) An examination of poststratification techniques for the Behavioral Risk Factor Surveillance System. In Online Proceedings of the Survey Research Methods Section, Joint Statistical Meeting. American Statistical Association. http://www.amstat.org/sections/srms/Proceedings/y2006 /Files/JSM2006-000144.pdf. Accessed 29th August 2011

Battaglia, M. P., Izrael, D., Hoaglin, D. C. and Frankel, M. R. (2009) Practical considerations in raking survey data. Survey Practice. http://surveypractice.org/ 2009/06/29/raking-survey-data. Accessed 9th April 2011

Botman, S., Moore, T., Moriarity, C. and Parsons, V. (2000) Design and estimation for the National Health Interview Survey, 1995-2004. Vital Health Stat, 2, 130

Clark, R., Devlin, M., Gerritsen, S., Templeton, R. and Stefanogiannis, N. (2008) Methodology Report for the 2006/07 New Zealand Health Survey. Report, NZ Ministry of Health, Wellington, New Zealand.http://www.health.govt.nz/publication/ methodology-report-2006-07-new-zealand-health-survey. Accessed 5th August 2011.

Chowdhury, S., Khare, M. and Wolter, K. (2007) Weight trimming in the National Immunization Survey. In Online Proceedings of the Survey Research Methods Section, Joint Statistical Meeting. American Statistical Association. http://www.amstat.org /sections/SRMS/proceedings/y2007/Files/JSM2007- 000077.pdf. Accessed 29th August 2011.

Henstridge, J. (2001) The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Weighting and imputation. Discussion Paper Series 3/01, Melbourne Institute of Applied and Social Research. www.melbourninstitute.com/downloads/hilda/ Discussion_Papers/hdps301.pdf. Accessed 25th February 2011.

Hindmarsh (2013) Small area estimation for health surveys. PhD. University of Wollongong Thesis Collection 1954-2016. https://ro.uow.edu.au/theses/3746/. Accessed 26th September 2019.

Karlton, G. and Flores-Cervantes, I. (2003) Weighting methods. Journal of Official Statistics, 19, 81–97.

NSW Health (2018) Information Bulletin 2018_048. 2018-19 KPI and Improvement Measure Data Supplement. System Information and Analytics, NSW Health. https://www1.health.nsw.gov.au/pds/ActivePDSDocuments/IB2018_048.pdf (accessed 25 September 2019)

Potter, F. J. (1988) A survey of procedures to control extreme sampling weights. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 453–458. American Statistical Assocation. URL: http://www.amstat. org/Sections/Srms/Proceedings/papers/1988_083.pdf. Accessed 29th August 2011.

Potter, F. J. (1990) A study of procedures to identify and trim extreme sampling weights. In Proceedings of the Section on Survey Research Methods, American Statistical Association, 225–230. American Statistical Assocation. URL http://www.amstat.org/Sections/Srms/Proceedings/papers/1990_034.pdf. Accessed 29th August 2011.

Watson, N. (2004) The Household, Income, and Labour Dynamics in Australia (HILDA Survey): Wave 2 Weighting. Technical Paper Series No. 4/04, Melbourne Institute of Applied Economic and Social Research. URL: www.melbourneinstitute.com /Bibliography357downloads/hilda/Technical_Papers/htec404.pdf . Accessed 25th February 2011.

Watson, N. and Fry, T. R. L. (2002) The Household, Income, and Labour Dynamics in Australia (HILDA Survey): Wave 1 Weighting. Technical Paper Series No. 3/02, Melbourne Institute of Applied Economic and Social Research. www.melbourneinstitute.com/downloads/hilda/Technical_Papers/htec302.pdf. Accessed 25th February 2011.

Appendix 1 Current method of weighting at the quarterly level

In order to report quarterly results for NSW and each LHD as well as for hospitals specifically sampled for quarterly reporting, it is necessary to weight the entire dataset, not just hospitals that are sampled for quarterly reporting.

Hospitals that are sampled for annual reporting are coded as CHSP (combined hospital); hospitals sampled for quarterly reporting maintain their current code. This modified hospital code is combined with the LHD for quarterly weighting, so the number of LHD_hos categories per LHD is the number of hospitals sampled for quarterly reporting plus one.

If there are any age-stay type cells with eligible population but no respondents they are combined with the adjoining age group within the stay type.

Initial quarterly weights are obtained for each LHD_hos by calculating the design weight at the level of LHD_hos.

$w_i = N_i/n_i$

where

w_i is the initial (design) weight of the i^{th} LHD_hos, Ni is the eligible population of the i^{th} LHD_hos and n_i is the sample size for the i^{th} LHD_hos.

Note that all strata within the same LHD_hos have the same weight at this point.

The GREGWT macro is used to provide initial weights at the stratum level, using w_i as the initial weight, with a benchmark for each quarter of age stratum x stay type x LHD_hos. This gives identical results to those obtained by stratifying using the 6 strata within each LHD_hos; the advantage of using the GREGWT macro is that it produces detailed diagnostics.

We find the median weight across all respondents and identify any cell above the median with less than 6 respondents. The weight of small cells with weights less than median are ignored because these are not influential. Population, sample size and weight of all six strata within affected LHD-hos combinations are extracted. In a very manual process, two analysts independently decide whether to combine the small cell for each LHD_hos and if so, how they should be combined. These decisions are based on considerations such as the

weight of the stratum with small sample size relative to the weight of other strata within the LHD_hos, and whether the weight of the small cell will be increased by aggregating. An example of this is shown in Table A1. The median weight was 65.2. The decision was to merge with the 18-49 year age group. Note the extreme difference in population between the Admitted and non-Admitted strata, so weights for non-admitted are much larger than for admitted, despite the fact that we have stratified the sample proportionately across the age-group-stay type strata.

Table A1: Example of weighting decisions for one LHD_hos grouping

Stay type	*Age stratum*	*Number of responses*	*Eligible population*	*Weight following phase 1*	*Flagged cell*	*Decision*	*Final cell weight*
Admitted	Under 18	9	194	21.6			21.6
Emergency	18-49	10	526	52.6			52.6
	50+	19	1223	64.4			64.4
Non-admitted	Under 18	4	2276	569.0	****	Merge with 18-49	365.3
Emergency	18-49	12	3568	297.3			365.3
	50+	18	2556	142.0			142.0

The stay type or age stratum of cells that are to be aggregated are updated in both the population and response dataset and the GREGWT macro is run again.

A similar check is undertaken after the second round. Sometimes a small cell still has a weight greater than the median but it is deliberately kept because the weights are acceptable relative to other cells for that hospital.

Once satisfied with the small cells, a summary of the weights is produced, including a histogram of the weight distribution at the NSW level, a brief summary (min, max etc) of weights, design effect and the ratio of max to median weight for all respondents and also by LHD and hospital. The sum of the weights is compared to the population at the hospital level and within age strata, stay type and then all 6 strata. Sometimes this summary identifies further outliers or hospitals with high ratios of max:median in which case additional decisions are sometimes made to combine other cells and/or put a cap on the max weight. This requires going back and rerunning the GREGWT process, adding the maximum weight criterion and creating an updated summary of weights file.