2014

Inclusion of mobile telephone numbers into an ongoing population health survey in New South Wales, Australia using an overlapping dual -frame design: final weighting strategy

Margo L. Barr
University of Wollongong, margo.barr@doh.health.nsw.gov.au

Raymond A. Ferguson
NSW Ministry of Health, rferg@doh.health.nsw.gov.au

Phil J. Hughes
University of Wollongong, phil.hughes@uow.edu.au

David G. Steel
University of Wollongong, dsteel@uow.edu.au

Inclusion of mobile telephone numbers into an ongoing population health survey in New South Wales, Australia using an overlapping dual -frame design: final weighting strategy

Abstract

Background In 2012 mobile telephone numbers were included into the ongoing NSW Population Health Survey (NSWPHS) using an overlapping dual-frame design. Previously in the NSWPHS the sample was selected using random digit dialling (RDD) of landline telephone numbers. The survey was undertaken using computer assisted telephone interviewing (CATI). Details about the methodology, call outcomes and representativeness of the sample in the first quarter of 2012 are published elsewhere. The weighting strategy also needed to be revised to manage the differing probabilities of selection by frame and to adjust for the increased chance of selection of dual-phone users. This paper describes and details the final weighting strategy adopted to properly combine the data from the two overlapping sample frames in the NSWPHS and the benchmark populations used, based on the limited information available in Australia.

Methods/Design Estimates of the number of telephone numbers for the landline and mobile phone frames, used to calculate the differing probabilities of selection by frame, for NSW and by stratum, were obtained by apportioning Australian estimates as none were available for NSW. The weighting strategy was then developed by calculating person selection probabilities, selection weights, applying a constant composite factor to the dual-phone users sample weights, and benchmarking to the latest NSW population by age group, sex and stratum.

Conclusions The inclusion of mobile telephone numbers, through an overlapping dual-frame design, improved the coverage of the survey and an appropriate weighing procedure is feasible, although it added substantially to the complexity of the weighting strategy. Access to accurate Australian, State and Territory estimates of the number of landline and mobile telephone numbers and type of phone use by at least age group and sex would greatly assist in the weighting of dual-frame surveys in Australia.

NIASRA

NATIONAL INSTITUTE FOR APPLIED STATISTICS RESEARCH AUSTRALIA

National Institute for Applied Statistics Research Australia

The University of Wollongong

Working Paper

01-14

Inclusion of mobile telephone numbers into an ongoing population health survey in New South Wales, Australia using an overlapping dual-frame design: final weighting strategy.

Margo L Barr, Raymond A Ferguson, Phil J Hughes and David G Steel

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522. Phone +61 2 4221 5435, Fax +61 2 4221 4845. Email: anica@uow.edu.au

Inclusion of mobile telephone numbers into an ongoing population health survey in New South Wales, Australia using an overlapping dual-frame design: final weighting strategy.

Margo L Barr[1,2§], Raymond A Ferguson[1], Phil J Hughes[2] and David G Steel[2]

[1] Centre for Epidemiology and Evidence, NSW Ministry of Health, 73 Miller Street, North Sydney, Australia

[2] National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong, Australia

[§]Corresponding author

Email addresses:

MLB: margo.barr@doh.health.nsw.gov.au

DGS: david_steel@uow.edu.au

RF: rferg@doh.health.nsw.gov.au

PH: phil.hughes@uow.edu.au

Abstract

Background

In 2012 mobile telephone numbers were included into the ongoing NSW Population Health Survey (NSWPHS) using an overlapping dual-frame design. Previously in the NSWPHS the sample was selected using random digit dialling (RDD) of landline telephone numbers. The survey was undertaken using computer assisted telephone interviewing (CATI). Details about the methodology, call outcomes and representativeness of the sample in the first quarter of 2012 are published elsewhere. The weighting strategy also needed to be revised to manage the differing probabilities of selection by frame and to adjust for the increased chance of selection of dual-phone users. This paper describes and details the final weighting strategy adopted to properly combine the data from the two overlapping sample frames in the NSWPHS and the benchmark populations used, based on the limited information available in Australia.

Methods/Design

Estimates of the number of telephone numbers for the landline and mobile phone frames, used to calculate the differing probabilities of selection by frame, for NSW and by stratum, were obtained by apportioning Australian estimates as none were available for NSW. The weighting strategy was then developed by calculating person selection probabilities, selection weights, applying a constant composite factor to the dual-phone users sample weights, and benchmarking to the latest NSW population by age group, sex and stratum.

Conclusions

The inclusion of mobile telephone numbers, through an overlapping dual-frame design, improved the coverage of the survey and an appropriate weighing procedure is feasible, although it added substantially to the complexity of the weighting strategy. Access to accurate Australian, State and Territory estimates of the number of landline and mobile telephone numbers and type of phone use by at least age group and sex would greatly assist in the weighting of dual-frame surveys in Australia.

Background

Since 2002 information about the health of the New South Wales (NSW) population has been obtained using the NSW Population Health Survey (NSWPHS) [1]. This survey is a continuous sample survey of approximately 15,000 persons each year. The survey is stratified by health administration area and equal numbers are selected from each of the strata, using random digit dialling (RDD) of landline phone numbers and computer assisted telephone interviewing (CATI). Because of the growing number of mobile-only phone users in the population, estimated to be 19% in Australia in 2011[2], mobile telephone numbers were included in 2012 using an overlapping dual-frame design; with type of phone use defined as mobile only, landline only and dual-phone users; thus people with a mobile phone and living in a household with a landline could now be selected though either the landline or mobile phone sampling frames. Details about the methodology, call outcomes and representation of the sample in the first quarter of 2012 are provided in Barr et al [3], and the questions in the questionnaire are available from the survey website [1].

In the previous landline based samples for the NSWPHS, equal sample sizes were used in each stratum, and therefore the probability of selection varied by stratum. Moreover, as one person was randomly selected from each selected household, the probability of selection also varied by household size. Weights were calculated for use in survey estimation to account for the differences in probabilities of selection and then benchmarked to the latest NSW population by age group, sex and stratum as shown in Steel 2004 [4] and summarised in appendix A. The use of equal probabilities to select landline phones in each stratum meant that the factor $\frac{T_h}{t_h}$, which is the ratio of telephone numbers T_h in stratum *h* to the number of telephone numbers in the sample t_h, cancelled in the previous calculation of the weights, and so the actual number of landline telephone numbers in each of the strata did not need to be known. However, with the inclusion of the mobile phone frame this is not the case and the number of landlines and mobile telephone numbers in the population for each stratum needed to be estimated. In 2011 the Australian Communication and Media Authority (ACMA) estimated that there were 29.28 million mobile telephone numbers and 10.54 million landline telephone numbers in Australia [2]. Estimates, however, are not routinely provided by State, let alone by health administration area.

As the previous NSWPHS samples came from a single frame the weighting did not need to account for the differing chances of selection by type of phone use. However, with the inclusion of the mobile telephone numbers, using an overlapping dual-frame design, dual-phone users now have an increased chance of selection because they could be selected from either frame.

There is currently a growing body of knowledge on issues and methods to deal with overlapping frames as summarised in the American Association for Public Opinion Research (AAPOR): Cell Phone Task Force Report [5], and in particular the use of composite weights to adjust for the increased chance of selection of dual-phone users. Hartley 1962 and 1974 [6,7] first described the calculation of these composite weights in overlapping frames. We use the notation of *A* for landline frame, *B* for the mobile frame, *Y* for the population total of interest, *y* for the estimator, *a* for landline only component, *b* for mobile only component and *ab* for dual phone users component. In this case the composite estimator is defined as $y_{comp} = y_a + y_b + y_\lambda$ where the estimate for the overlap population is $y_\lambda = y_{ab}^A + (1-\lambda) y_{ab}^B$ with y_{ab}^A and y_{ab}^B being the estimators for persons with both mobile and landlines from frame A and B respectively and the composite factor being between 0 and 1 ($0< \lambda <1$). Most overlapping dual frame surveys conducted to date have used a constant composite factor λ and the most common value is 0.5 [8,9,10].

Calculation of weights, in an overlapping dual-frame design, ideally requires type of phone use benchmarks as well as population benchmarks [5]. In the USA type of phone use benchmarks, at the national level, are collected using the National Health Interview Survey (NHIS) [11], where questions on residential phone use have been included since 1963 and mobile phone use since 2003. The most recent report for the USA is January to June 2012 where they estimated that 56.1% of adults lived in a household with a landline and a mobile phone, 7.8% lived in a household with a landline but no mobile phone, 34.0% lived in a household with only a mobile phone, 1.9% lived in a household without a mobile phone or a landline phone, and 0.2% of adults lived in a household where the phone status was unknown [12]. The highest mobile-only phone rates were in un-related adults with no children (75.9%), young adults (60.1% in 25 to 29 year olds, 49.5% in 18 to 24 year olds, and 55.1% in 30-34 year olds), house renters (58.2%), and people within poor households (51.8%) [12].

Currently there is no equivalent source of information on type of phone use in Australia. The first estimates of landline phone use from an equivalent national survey, the Australian Health Survey (AHS) conducted by the Australian Bureau of Statistics (ABS), are expected to be available in 2014 [13]. There are currently no plans to collect mobile phone use in this national survey. However, landline and mobile phone use questions have been included in the Roy Morgan Single Source Survey (RMSSS) since 2005 [14]. The most recent published estimates from this survey are for June 2011when they estimated that 74% of adults in Australia lived in a household with a landline and a mobile phone, 5% lived in a household with a landline but no mobile phone, and 19% lived in a household with only a mobile phone; with the highest mobile-only phone rates being in young adults (37% in 18 to 24 year olds) [2].

This paper describes and details the final weighting strategy adopted to properly combine the data from the two overlapping sample frames in the NSWPHS and the benchmark populations used, based on the limited information available in Australia.

Methods/Design

Within a stratum the landline sample was selected using equal probability of selection of landline telephone numbers and then random selection of one person from the selected household. In the mobile phone sample an equal probability sample of mobile telephone numbers in Australia was selected and screened for adult residents in NSW. If the respondent has one or more children one child was selected at random.

Final weighting strategy

For the sampling design used person selection probabilities for the landline frame and mobile frame were derived as follows:

person *ijh* from the landline frame	***adult*** *i* from the mobile frame	***child*** *c* from parent *p* from the mobile frame
$\pi_{ijh}^{A} = \frac{t_h^A}{T_h^A}\frac{T_{jh}^A}{N_{jh}}$	$\pi_i^B = \frac{t^B}{T^B}\frac{T_i^B}{N_i}$	$\pi_{cj}^B = \pi_p^B \frac{N_{cp}}{N_{cj}}$

Where: *i* denotes an eligible person; *c* denotes a child of an eligible person; *p* denotes a parent; *h* denotes the stratum; *j* denotes a household; *N* denotes population size; *T* denotes number of telephone numbers in the population; *t* denotes number of telephone numbers in the sample; *A* denotes landline frame; *B* denotes mobile frame. For the design used $N_i = 1$ and N_{cp} is the number of parents that a child selected through a parent in the mobile phone frame has and N_{cj} is the number of children in the household of the parent. The weights were then the inverse $w = \pi^{-1}$ in each situation.

The sample weights of the dual phone-users were then adjusted using the composite factor λ set at 0.5. So for those dual phone-users selected from:

- the landline frame the composite weights were $w_{ijh}^{\lambda} = \lambda w_{ijh}^{A}$
- the mobile frame the composite weights were $w_i^{\lambda} = (1-\lambda) w_i^{B}$

Benchmarking to the reference population was then performed as per previous years by adjusting the weights for differences between weighted estimates of the age and sex structure obtained from the combined landline and mobile phone sample and ABS mid-year population estimates for each stratum, N_{dh} [15]. This was achieved by summing the weights for the age and sex cell *d* in stratum *h*, to produce a survey estimate of the population in that cell, $\hat{N}_{dh}$ and then multiplying the weights by $\frac{N_{dh}}{\hat{N}_{dh}}$. If these population estimates also included type of phone use, then these could be used to further improve the estimation. However, this information is not available in Australia.

Estimation of number of telephone numbers in NSW

The weights described above require the number of landline telephones in stratum *h*, T_h^A, and the number of mobile telephone numbers in NSW, T_{NSW}^B. As there was no specific NSW residential landline telephone data T_h^A available we divided the number of residential landline telephone numbers in Australia, using the ACMA estimate [2], by the proportion of the population in that stratum, using the ABS estimates [15], after having first adjusted it by the percentage of the population who had landline phones in that stratum, using the RMSSS estimates [13]. As there was no specific NSW mobile telephone data T_{NSW}^B available we divided the number of

mobile telephone numbers in Australia, using the ACMA estimate [2], by the proportion of the population in NSW, using the ABS estimates [15], having first adjusted it by the percentage of the population in NSW who had mobile phones, using the RMSSS estimates [13].

These procedures produce estimates as follows:

$$T_h^A = \frac{N_h^A P_h^A}{N_{Aust}^A P_{Aust}^A} T_{Aust}^A \text{ and } T_{NSW}^B = \frac{N_{NSW}^B P_{NSW}^B}{N_{Aust}^B P_{Aust}^B} T_{Aust}^B,$$

where P_h^A denotes the proportion of people living in a household with a landline phone in stratum *h* and P_{NSW}^B is the proportion of people in NSW with a mobile phone.

Table 1 shows the estimated number of telephone numbers by frame for NSW. We estimated that there were 3.5 million residential landline telephone numbers and 9.8 million mobile telephone numbers in NSW and landline numbers in the strata ranged from 23,764 in Far West health administration area to 443,603 in Hunter New England health administration area.

[INSERT TABLE 1 NEAR HERE]

Calculation of the weights

Data from the NSWPHS for the first quarter of 2012 was used to test the weighting strategy. This consisted of data on 3395 respondents with 2171 (64%) from the landline frame, with 17.6% being landline-only, and 1224 (36%) from the mobile frame, with 25.8% being mobile-only. Data needed to be available for all core weighting variables including age, sex, stratum, number of landline phones, number of mobile phones they personally have, and eligible persons in the household. If the respondent refused to provide their age or sex the interview was terminated. If values could not be imputed for missing and/or erroneous core weighting variables then the record was removed from the dataset.

Table 2 shows a summary of the data management required. Data needed to be imputed for 29 respondents for number of landline phones in the household and 26 respondents for number of mobile phones personally have. Table 3 shows a summary of the sampled and reported strata. The majority of respondents

recruited through the landline frame were, using postcode/suburb and/or local government area provided by the respondent during the interview, in the same stratum as initially allocated, with the majority of the mismatches being within the metropolitan health administration areas where telephone numbers are more transportable. Table 2 also shows that all of the respondents recruited through the mobile frame, except for 17, could be allocated to a stratum using postcode/suburb and/or local government area provided by the respondent during the interview. This resulted in 3378 respondents, 2933 adults and 445 children, for which weights could be calculated.

[INSERT TABLE 2 AND TABLE 3 NEAR HERE]

Table 4 shows the summary statistics by frame for the sample divided by number of telephone lines in the population, telephone lines in the household divided by eligible persons in household, person selection probabilities, person weights, and the composite weights for dual phone-users. Average person weights were 3.3 times higher for the mobile-only respondents, 1.3 times higher for the landline-only respondents and 1.7 times higher for dual-phone users in the mobile frame compared to the dual-phone users in the landline frame.

[INSET TABLE 4 NEAR HERE]

Table 4 also shows the summary statistics for the person weights, composite for dual-phone users, scaled back to the number of respondents in the sample and for the weights for the dual-frame when benchmarked to the NSW population by age group, sex and stratum. The mean final weight was 2,152, ranging from 14 for a 76 year old female dual-phone user in Far West Health administration area recruited through the landline frame to 21,807 for a 76 year old male landline-only phone user in South East Sydney health administration area recruited through the landline frame. The distributions of the final weights are shown in Figure 1. Figure 1 also shows the distributions of the final weights by frame and type of phone use for comparison.

[INSET FIGURE 1 NEAR HERE]

Weights are used to eliminate bias that would arise from ignoring the differences in selection probabilities and also improve estimates by adjusting to known population benchmarks. The increase in sampling variance

due to weighting is reflected in the Weight effects, which were also calculated using $weff = n\frac{\sum w_i^2}{(\sum w_i)^2}$ where: n denotes sample size and w denotes weights. This is the component of the design effect due to weighting.

Table 5 shows the weight effects for each of the weighting parameters. The overall weight effect was 1.93. Weight effects varied by: age group, from 1.55 in 25-34 years to 2.24 in 65 plus years; sex, from 1.83 in males to 1.97 in females; and stratum, from 1.41 in North Sydney health administration area, to 3.24 in Mid North Coast health administration area. These effects are similar to, and in many cases less than, the effects found in the corresponding quarter of the 2011 NSWPHS when only a landline based sample was used.

[INSET TABLE 5 NEAR HERE]

Discussion

The development of the weighting strategy, weighted for the person selection probabilities by frame, composite weights applied to dual-phone users, and benchmarked to the NSW population, was more complex than it had been for the previous landline frame. It was however encouraging that the weighting effects were similar to those found in previous years.

The need to estimate the number of telephone numbers for NSW and by stratum from the Australia figures, used to calculate the differing probabilities of selection, highlighted the desirability to be able to access accurate information at least at the State and Territory level. Access to more accurate type of phone use benchmarks would have also allowed weighting by type of phone use. We considered using the type of phone use totals collected by RMSSS [13] to generate benchmark populations by age group, sex, stratum and type of phone use. However, after conducting a sensitivity analysis we concluded that potential errors in the type of phone use estimates provided by age group, sex and stratum, which were well below the design level of the survey, were likely to impact on the NSWPHS health indicator estimates.

The compositing factor λ used for the composite weights was set at 0.5. However the use of 0.5 as the composite factor assumes that all sampled units respond. Skinner (1991) and Skinner and Rao (1996) have explored ways to reduce non-response bias by raking the estimates to type of phone use totals from an

independent source [16,17]. However, when Brick (2006) applied these to the Current Population Survey (CPS) he found that none of the suggested estimation schemes substantially reduced the non-response bias of the estimate [18]. So with overlapping dual-frames design surveys being relatively new in Australia the use of $\lambda = 0.5$ as the compositing factor seemed appropriate. It is possible to determine a value of this factor that minimises the sampling variance of the estimator, but this value will be variable specific. Moreover, it is likely that for various reasons, the estimates obtained for the overlapping component of the population, obtained from the two sampling frames do not have the same expectation, and using $\lambda = 0.5$ ensures that the two frames are given equal prominence in the estimation. Although further research needs to be undertaken to explore other estimation schemes using Australian data.

Conclusions

The inclusion of the mobile telephone numbers through an overlapping dual-frame design, improved the coverage of the survey and an appropriate weighing procedure is feasible, although it added substantially to the complexity of the weighting strategy. Access to accurate Australian, State and Territory estimates of the number of landline and mobile telephone numbers and type of phone use by at least age group and sex would greatly assist in the weighting of dual-frame surveys in Australia.

.List of abbreviations

AAPOR	American Association for Public Opinion Researchers
ABS	Australian Bureau of Statistics
ACMA	Australian Communication and Media Authority
AHS	Australian Health Survey
CATI	Computer assisted telephone interviewing
NHIS	National Health Interview Survey
NSW	New South Wales
NSWPHS	NSW Population Health Survey
RDD	Random digit dialling

RMSSS Roy Morgan Single Source Survey

Competing interests

Nil

Authors' information

MLB is a PhD student with the National Institute for Applied Statistics Research, University of Wollongong, Wollongong, Australia

Authors' contributions

MLB developed the overall concepts and planned the study; analysis the data, wrote the methods and results, wrote the introduction and discussion and finalised the manuscript. RAF managed the data, checked the analysis programs and commented on drafts of the manuscript. PH provided development and operational advice, checked the underlying logic of the analysis and commented on drafts of the manuscript and DGS provided development and analysis advice, checked the underlying logic of the weighting and commented on drafts of the manuscript. All authors read and approved the final manuscript.

Acknowledgments

We acknowledge the interviewing staff and supervisors at the Centre for Epidemiology and Evidence, NSW Ministry of Health for collecting the data and providing their comments. We also acknowledge the respondents for participating in the survey.

References

1. NSW Ministry of Health, NSW Population Health Surveys http://www.health.nsw.gov.au/surveys/Pages/default.aspx

2. Australian Communications and Media Authority (ACMA): Communications report 2010–11. ACMA; 2011

3. Barr ML, van Ritten JJ, Steel DG and Thackway SV. Inclusion of mobile phone numbers into an ongoing population health survey in New South Wales, Australia: design, methods, call outcomes, costs and sample representativeness. BMC Medical Research Methodology 2012, 12:177

4. Steel D: New South Wales Population Health Survey: Review of the Weighting Procedure. In *Commissioned Report to the Centre of Epidemiology and Research* NSW Department of Health of Australia; 2004.

5. The American Association for Public Opinion Research (AAPOR): Cell Phone Task Force Report: New considerations for survey researchers when planning and conducting RDD telephone surveys in the US with respondents reached via cell phone numbers. 2010.

6. Hartley HO. Multiple Frame Surveys. Proceedings of the Social Statistics Section, American Statistical Association 1962, 203-6.

7. Hartley HO. Multiple Frame Methodology and Selected Application. Sankhyā, Ser. C, Part 3, 36, 99-118.

8. Brick JM, Cervantes IF, Lee S, Norman G: Non-sampling errors in dual frame telephone surveys. Survey Methodology 2011, 37(1):1–12. OR Benford, R., Tompson, T., Fleury, C., Feinberg, G., Feinberg, B., Speulda, N., and Weber, A. 2009. **"**Cell Phone and Landline – Considerations for Sample Design, Estimates, Weighting, and Costs. Paper presented at the 64th annual conference of the American Association for Public Opinion Research; Hollywood, FL.

9. Lohr SL. Dual frame surveys: Recent developments and challenges Proceedings of the 45th Meeting of the Italian Statistical Society 2010. (Sharon Lohr, Dual Frame Surveys: Recent Developments and Challenges, David Haziza, Resampling methods for variance estimation in the presence of missing survey data, Emilia Rocco, Using auxiliary information and non parametric methods in weighting adjustments

10. Wolter, K. M., Smith, P., & Blumberg, S. J. (2010). Statistical foundations of cell-phone surveys. Survey Methodology 2010. 36(2); 203-215.

11. National Health Interview Survey. http://www.cdc.gov/nchs/nhis.htm

12. Blumberg SJ, Luke JV: Wireless substitution: Estimates from the National Health Interview Survey. January - June 2012, National Centre for Health Statistics. 2012. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201212.PDF

13. Australian Bureau of Statistics, Australian Health Survey 2011-2013 (AHS) http://www.abs.gov.au/australianhealthsurvey

14. Roy Morgan Single Source Survey. http://www.roymorgan.com/products/single-source/

15. Australian Bureau of Statistics: Census quickstats. New South Wales: ABS; 2011. http://www.censusdata.abs.gov.au/census_services/getproduct/census/2011/quickstat/1.

16. Skinner CJ. On the efficiency of raking ratio estimation for multiple frame surveys. Journal of the American Statistical Association 1991, 86:779-84.

17. Skinner CJ and Rao NK, Estimation in dual frame surveys with complex designs. Journal of the American Statistical Association 1996. 91:349-56

18. Brick, J.M., Dipko, S., Presser, S., Tucker, C. and Yuan, Y. (2006). ―Nonresponse Bias in a Dual Frame Sample of Cell and Landline Numbers, *Public Opinion Quarterly*, 70(5), 780-793.

Table 1: Number of telephone numbers by frame for NSW

Health administration area (stratum for landline frame)	Landline frame		Mobile frame	
	% stratum with landline	Estimated number of lines	% stratum with landline	Estimated number of lines
Sydney	74.0%	254015		
South Western Sydney	79.0%	406768		
South Eastern Sydney	76.0%	381287		
Illawarra Shoalhaven	82.0%	194868		
Western Sydney	79.0%	385908		
Nepean Blue Mountains	84.0%	177441		
Northern Sydney	86.0%	431456		
Central Coast	82.0%	162390		
Hunter New England	84.0%	443603		
Northern NSW	85.0%	157109		
Mid North Coast	81.0%	106940		
Southern NSW	82.0%	97434		
Murrumbidgee (inc Albury LGA)	82.8%	153043		
Western NSW	80.0%	137306		
Far West	90.0%	23764		
TOTAL	80.8%	3,513,333	85.8%	9,385,073

Table 2: Management of missing and erroneous data.

Variable	n	Raw data					Imputed data				
		Mis	**Mean**	**Med**	**Min**	**Max**	**Management of missing values and excessive values**	**Mean**	**Med**	**Min**	**Max**
Persons in household (landline frame)	2171	0	2.51	2	1	10	Set to 1 if missing and to 10 if greater than 10	2.51	2	1	10
Children in household (mobile frame)	139	0	1.73	2	1	5	Set to 1 if missing and to 6 if greater than 6	1.73	2	1	5
Landline lines in household (landline frame)	2171	**10**	1.03	1	0	5	Substitute with 1 if 0 or missing and to 5 if greater than 5	1.03	1	1	5
Landline lines in household (mobile frame)	1224	**19**	0.77	1	0	3	Substitute with 0 if missing and to 5 if greater than 5	0.76	1	0	3
Mobile phone numbers (landline frame)	2171	**15**	0.91	1	0	**6**	Substitute with 0 if missing and to 5 if greater than 5	0.90	1	0	5
Mobile phone numbers (mobile frame)	1224	**11**	1.10	1	1	5	Substitute with 1 if 0 or missing and to 5 if greater than 5	1.10	1	1	5

Table 3: Management of missing and inconsistent data on health administration areas in NSWPHS

Health administration area	Landline frame					Mobile frame	
	As per stratum	Derived from survey questions	Diff	Change to	Changes from	As per stratum	Derived from survey questions
Mobile	0	0	0	None	None	1224	17
Sydney (Syd)	170	141	-29	SES (8), NBM (1) to Syd	Syd to: SWS (7), SES (13), WS (7), NBM (1), NS (8), CC (1), HNE (1)	0	162
South Western Sydney (SWS)	146	153	7	Syd (7), WS (1), NBM (2) to SWS	SWS to: WS (2), NBM (1)	0	161
South Eastern Sydney (SES)	65	73	8	Syd (13), WS (2), MNC (1), FW (2) to SES	SES to: Syd (8), IS (1), WS (1)	0	140
Illawarra Shoalhaven (IS)	113	114	1	SES (1) to IS	None	0	59
Western Sydney (WS)	123	133	10	Syd (7), SWS (2), SES (1), NBM (1), NS (7) to WS	WS to: SWS (1), SES (2), NBM (2), NS (3)	0	153
Nepean Blue Mountains (NBM)	143	142	-1	Syd (1), SWS (1), WS (2) to NBM	NBM to: Syd (1), SWS (2), WS (1), WNSW (1)	0	58
Northern Sydney (NS)	133	137	4	Syd (8), WS (3) to NS	NS to: WS (7)	0	166
Central Coast (CC)	165	164	-1	Syd (1) to CC	CC to: HNE (2)	0	46
Hunter New England (HNE)	204	208	4	Syd (1), CC (2), MNC (1) to HNE	None	0	106
Northern NSW (NNSW)	108	107	-1	None	NNSW to: MNC (1)	0	33
Mid North Coast (MNC)	316	315	-1	NNSW (1) to MNC	MNC to: SES (1), HNE (1)	0	21
Southern NSW (SNSW)	206	206	0	None	None	0	34
Murrumbidgee (M) including Albury LGA	84	85	1	FW (1) to M	None		44
Western NSW (WNSW)	97	98	1	NBM (1) to WNSW	None	0	22
Far West (FW)	98	95	-3	None	FW to: SES (2), M (1)	0	2
TOTAL	2171	2171				1224	1207

Table 4: Summary of the person selection probability, composite and benchmark weight statistics for each of the frames.

Group	Phone type	Description	Formula	Sum	Ave	Median	Min	Max
Landline Frame (n=2171)								
Adult and children (n=2171)	All types (n=2171)	Interviews divided by universe of telephone numbers	$\frac{t_h^A}{T_h^A}$	2.68	0.0012	0.0007	0.00017	0.0041
		Lines in household divided by eligible persons in household	$\frac{T_{jh}^A}{N_{jh}}$	1216.69	0.5699	0.50000	0.11111	3.0000
		Person selection probability (π_{ijh}^A)	$\frac{t_h^A}{T_h^A}\frac{T_{jh}^A}{N_{jh}}$	1.59	0.0007	0.0003	0.00003	0.0082
		Selection weight (w_{ijh}^A)	$\frac{1}{\pi_{ijh}^A}$	8939582	4113.94	2864.6	121.31	35214.76
	Landline only (n=383)	Selection weight (w_{ijh}^A)	$\frac{1}{\pi_{ijh}^A}$	1074321	2805.02	1725.43	121.31	29345.64
	Both (n=1788)	Selection weight (w_{ijh}^A)	$\frac{1}{\pi_{ijh}^A}$	78765261	4394.00	2911.00	169.30	35214.76
		Composite weight $\left(w_{ijh}^{\lambda}\right)$ (where $\lambda = 0.5$)	λw_{ijh}^A	3932630	2197.00	1455.50	84.65	17607.38
Mobile Frame (n=1207)								
Adults (n=1069)	All types (n=1069)	Interviews divided by universe of telephone numbers	$\frac{t^B}{T_B}$	0.14	0.0001	0.0001	0.00013	0.0001
		Mobile phones for person divided by eligible persons (where N_i=1)	$\frac{T_i^B}{N_i}$	1168.00	1.0947	1.00000	1.00000	5.0000

Group	Phone type	Description	Formula	Sum	Ave	Median	Min	Max
		Person selection probability (π_i^B)	$\frac{t^B}{T^B}\frac{T_i^B}{N_i}$	0.15	0.0001	0.00013	0.00013	0.0007
		Selection weight (w_i^B)	$\frac{1}{\pi_i^B}$	7819874	7328.84	7655.04	1531.01	7655.04
	Mobile only (n=284)	Selection weight (w_i^B)	$\frac{1}{\pi_i^B}$	2071325	7319.17	7655.04	1913.76	7655.04
	Both (n=785)	Selection weight (w_{i^B})	$\frac{1}{\pi_i^B}$	5748549	7332.33	7655.04	1531.01	7655.04
		Composite weight $\left(w_i^\lambda\right)$	$(1-\lambda)w_i^B$	2874274	3666.17	3827.52	765.50	3827.52
Children (n=138)	All types (n=138)	Parents probability of selection	π_p^B	0.02	0.0001	0.0001	0.00013	0.0003
		Number of parents divided by eligible children in household	$\frac{N_{cp}}{N_{cj}}$	177.57	1.2867	1.00000	0.33333	2.0000
		Person selection probability (π_{cp}^B)	$\pi_p^B\frac{N_{cp}}{N_{cj}}$	0.03	0.0002	0.0001	0.00004	0.0005
		Selection weight (w_{cp}^B)	$\frac{1}{\pi_{cp}^B}$	964534	6989.38	7655.04	1913.76	22965.11
	Mobile only (n=26)	Selection weight (w_{cp}^B)	$\frac{1}{\pi_{cp}^B}$	158842	6109.31	3827.52	1913.76	15310.07
	Both (n=112)	Selection weight (w_{cp}^B)	$\frac{1}{\pi_{cp}^B}$	805692	7193.68	7655.04	1913.76	22965.11
		Composite weight (w_{cp}^λ)	$(1-\lambda)w_{cp}^B$	402846	3596.84	3827.52	956.88	11482.55

Group	Phone type	Description	Formula	Sum	Ave	Median	Min	Max
Both frames (n=3378)								
Adults and children (n=3378)	All types (n=3378)	Selection weight (composite for both users) -see note (a)	w_i^U	10514239	3112.56	2934.56	84.65	29345.64
		Selection weight (composite for both users) scaled back to the number of respondents	w_i^{U*}	3378	1.00000	0.8698	0.04779	10.999
		Post stratification weight (benchmarked to the population by age x sex x health admin) (W_i^U)	$\frac{N_{dh}}{\hat{N}_{dh}} w_i^{U^*}$	7272086	2152.78	1634.97	13.54	21807

(a) The weight w_i^U is the selection weight relevant to the segment of the overall sample from which the respondent was selected. For those respondents accessible through both the landline frame and the mobile phone frame it is the composite weight.

Table 5: Weight effects by weighting parameters for quarter 1 of the 2012 and 2011 NSWPHS

Category		n	SUM(WGT)2	(SUMWGT)	(SUMWGT)2	2012 *weff*	2011 *weff* (n=3377)
Age Group	0-13 years	368	7297166859	1244521	1548832668784	1.73	1.58
	14-24 years	317	5728404905	1066508	1137439271404	1.60	1.71
	25-34 years	397	4372748462	1057202	1117675032746	1.55	1.73
	35-44 years	346	4278905532	974108	948886376182	1.56	1.76
	45-54 years	489	3262991785	995006	990036601734	1.61	1.91
	55-64 years	624	2097445465	852381	726553045256	1.80	1.93
	65 plus	837	3136171943	1082361	1171505485852	2.24	1.63
Sex	Males	1429	16560322718	3600556	12964003293103	1.83	2.13
	Females	1949	13613512232	3671530	13480134523526	1.97	2.54
Health administration area	Syd	303	1698048663	585360	342646633987	1.50	1.80
	SWS	314	4303110764	892880	797234926549	1.69	1.62
	SES	213	5079590457	843566	711603697584	1.52	1.81
	IS	173	1303216701	391278	153098535888	1.47	1.82
	WS	286	3618759102	846389	716374051549	1.44	1.65
	NBM	200	1062941408	347524	120772881923	1.76	1.86
	NS	303	3343021760	846173	716008052067	1.41	1.80
	CC	210	1022421509	320135	102486405420	2.09	2.16
	HNE	314	4347558425	885170	783525875790	1.74	1.74
	NNSW	140	1082404196	300456	90273555553	1.68	1.68
	MNC	336	451722818	216328	46797881462	3.24	1.93
	SNSW	240	462055826	205377	42179613548	2.63	2.31
	M	129	885322373	241598	58369453477	1.84	1.89
	WNSW	120	1025192088	268286	71977640717	1.71	2.29
	FW	97	18833284	30750	945569265	1.93	1.80
Overall		3378	30173834950	7272086	52883238281997	1.93	2.37

Appendix A: Previous landline weighting strategy

1. Calculation of the raw person weight that accounts for the different selection probabilities.
   The probability of selection of a household is proportional to the number of telephone landline and is given by $\frac{T_{jh}}{T_h} t_h$. Given a household is selected the probability a person is selected is $\frac{1}{N_{jh}}$. The probability of selection of the *i* th person in the *j* th household is the product of these two probabilities and so the corresponding weight is:

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

2. Adjust the weights to agree with externally derived population benchmarks, N_{dh}.

With $\hat{N}_{dh} = \sum_{ijh \in s_{dh}} w_{ijh}$ being the survey based estimate of N_{dh}. The resulting post-stratified weight for $ijh \in d$ is then

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

This allowed the factor $\frac{T_h}{t_h}$ to cancel in the calculation of W_{ijh}, so that if $z_{jh} = \frac{N_{jh}}{T_{jh}}$, then $W_{ijh} = \frac{N_{ah}}{\sum_{ijh \in s_{dh}} z_{jh}} z_{jh}$.

3. The weights are then summed to produce estimates of totals for any category and will agree with the external age-sex benchmarks. That is $\sum_{ijh \in s_{dh}} W_{ijh} = N_{dh}$, $\sum_{ijh \in s_h} W_{ijh} = N_h$ and $\sum_{ijh \in s} W_{ijh} = N$.

where

- *i* denotes an eligible person
- *h* denotes a strata *j* denotes eligible the household
- *d* denotes an age-sex cell
- *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

Figure 1: Percentage of final weights, overall, by type of phone use and by frame, quarter 1 2012 NSWPHS.

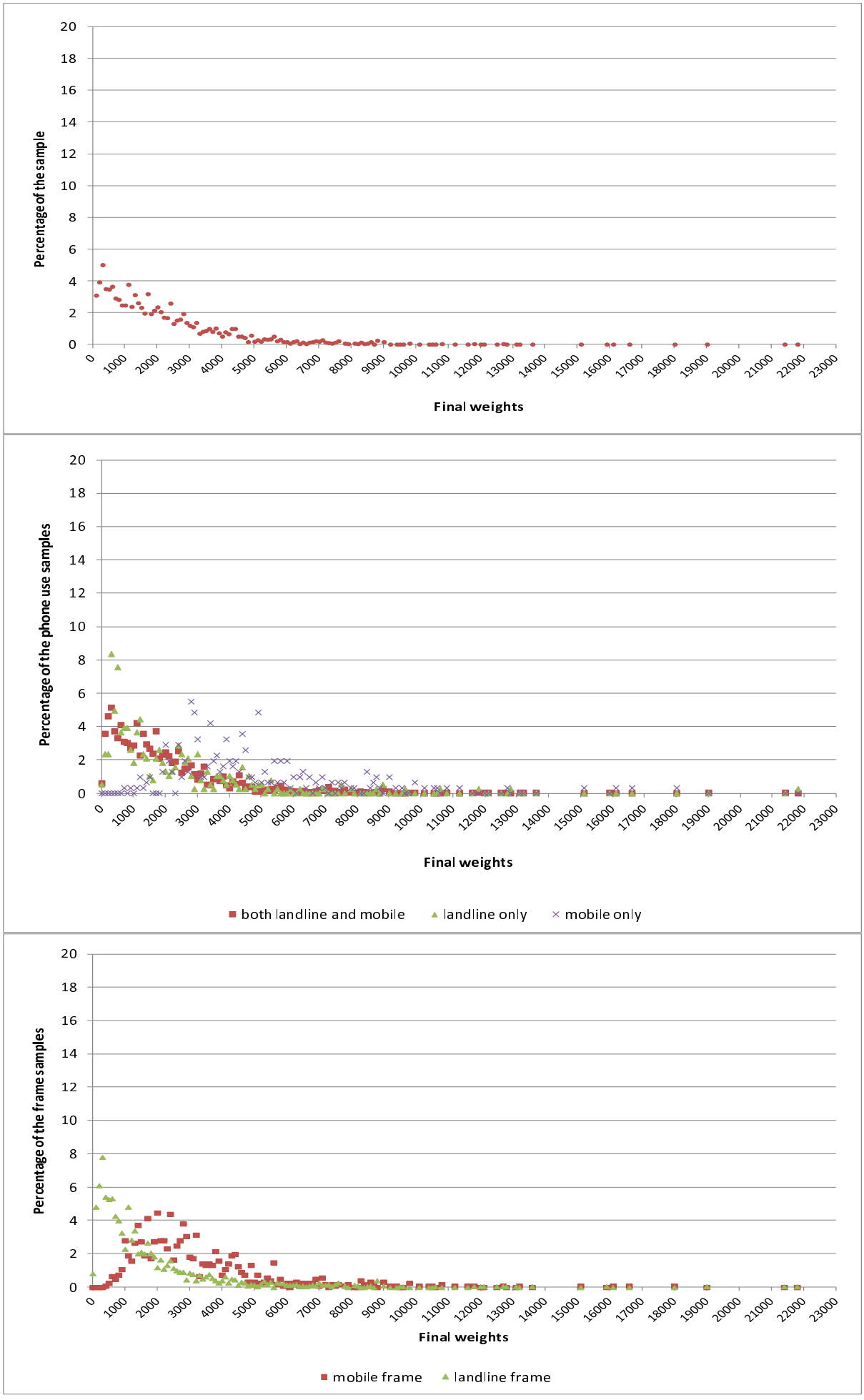

Figure 1