Author’s response to reviews

Title: Developing a weighting strategy to include mobile phone numbers into an ongoing population health survey using an overlapping dual-frame design with limited benchmark information.

Authors:

Margo L Barr (meyes@doh.health.nsw.gov.au)
Raymond A Ferguson (rferg@doh.health.nsw.gov.au)
Phil J Hughes (phil.hughes@uow.edu.au)
David G Steel (david_steel@uow.edu.au)

Version: 3 Date: 20 August 2014

Author's response to reviews: see over
To the Editor

Thank you for your comments about our paper. We are very pleased that our paper is of interest to your journal and seen to be an important topic and timely, considering the current prevalence of mobile phone only use.

We have examined the reviewer’s major, minor essential revisions and discretionally revision and have addressed each of them as documented below.

Regards

Margo

Reviewer 1: Regina Bernal

Title: 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

Overview: The international scientific production exposes three main issues: the validity of estimates obtained in telephone survey, because of the exclusion of households without landline; increasing nonresponse; methodological procedures to obtain valid estimates. This study deals with current and important issues in the telephone survey. Due to advancement in technology, other mode data collection, such as the Internet or cell phone, can be used with landline interviews to increase the coverage of the target population. The "Dual Frame" method and "Mixed-Mode Design" (Groves and Lepkowski, 1985; SIRKEN and Casady, 1988) use the sampling frame for cell phones and households for the sample the landline. However, there are issues to be resolved, such as the effect on estimates of the variables of interest, due to the different mode and the use of cell phone, which implies a change of the element in the sample.

Major revisions:

Issue 1: Background. The authors don’t explain why is important to include the cell phone in the health survey. The problem is related with the noncoverage bias. I suggest for the authors that include the paragraph that explain about noncoverage bias. (KISH. L. Survey Sampling, 1965; KALTON. G. Compensating for Missing Survey Data, 1983)

Response 1: The second sentence has been changed in order to explain why it is important to include mobile phones in the health survey to the following:

‘Because of the potential for coverage bias from the growing number of mobile-only phone users in the population, estimated to be 19% in Australia in 2011 [2], mobile phone numbers were included in 2012 using an overlapping dual-frame design. Coverage bias is a product of the proportion of the population not covered and the difference in the mean of the variable of interest between the covered group and the non-covered group’ [3]. Evidence from the National Health Interview Survey (NHIS) in the US has shown the mobile-only phone users substantial different for the health indicators: five or more drinks in one day at least once in
the past year (17.5% v 30.5% - 74% higher), current smokers (14.5% v 24.3% - 68% higher), and ever diagnosed with diabetes (10.8% v 6.2% - 43% lower) [4].

**Issue 2: Methods; Page 4.** The estimators are: 
\[ \hat{\lambda}_{comp} = \hat{\lambda}_a + \hat{\lambda}_b + \hat{\lambda}_A \] 
and 
\[ \hat{\lambda}_A = \lambda \hat{\lambda}_{ab} + (1 - \lambda)\hat{\lambda}_{ab} \]

**Question:** The problem with using dual frame is to define the estimator of variance. What is the statistical software used for analysis? What is estimator of variance used for statistical inference?

**Response 2:** We agree that variance is an important technical issue but a detailed discussion is beyond the scope of this paper. We have included a comment and some relevant references.

Specifically we have added the following to the discussion text:

‘The AAOPR Cell Phone Task Force Report [7], acknowledges that variance estimation for dual frame sample designs is somewhat more complex than for single frame designs. This issue is considered by Lohr and Rao (2000) and summarised in Lohr (2009)’

**Issue 3: Methods; Page 6.** Doubt: In the third paragraph, the text is not clear about the adjustment weight. The authors use the reference population in the computational process for calculate weight post stratification by sex and age?

**Response 3:** Sentence was changed to:

‘Benchmarking to the reference population was then performed, as per previous years, by adjusting the weights obtained from the combined landline and mobile phone sample, by age and sex to the ABS mid-year population estimates for each stratum, [19].’

**Issue 4: Method; Page 7.** If \( P_h^A \) is an estimate then replace \( P_h^A \) by \( \hat{P}_h^A \)

**Response 4:** Yes these proportions were estimated from the Roy Morgan Single Source Survey and so we have inserted the word "estimated’ and modified the formulas to \( \hat{P}_h^A \) and \( \hat{P}_{NSW}^B \) accordingly.

The sentence now reads:

‘These procedures produce estimates as follows:

\[ T_h^A = \frac{N_h \hat{P}_h^A}{N_{Aust}^A} T_{Aust}^A \] and \[ T_{NSW}^B = \frac{N_{NSW} \hat{P}_{NSW}^B}{N_{Aust}^A} T_{Aust}^A \]

Where \( \hat{P}_h^A \) denotes the estimated proportion of people living in a household with a landline phone in stratum \( h \) and \( \hat{P}_{NSW}^B \) is the estimated proportion of people in NSW with a mobile phone.’

**Issue 5: Method; Page 9.** Add bibliographic references of the measure of Weight effects (wEff). It is important to calculate the coefficient of variation (CV) of the weights because they have high variability, as shown by the Figure 1. The CV greater than 20% there is a loss of accuracy in the estimate (see Kish).
Response 5: The following references were added for the weight effects:


The coefficient of variation (CV) of the weights were added into Table 3

We have also expanded the discussion of weff and their connection with design effects. As you have noted the close relationship between weff and the CV, readers can now see the effect of the CV of the weights directly from table 3. Specifically we have modified the sentence to read:

‘The weight effects were calculated using \[ \text{weff} = n \frac{\sum w_i^2}{(\sum w_i)^2} \] where: \( n \) denotes sample size and \( w \) denotes weights [20-22]. The weight effect is the design effect due to weighting and is equal to \( 1 + C_w^2 \), where \( C_w \) is the coefficient of variation of the weights (i.e. the standard deviation of the weights divide by the mean of the weights) and is a standardised measure of the variation of the weights.”

Table 3 shows the weight effects and coefficient of variation of the weights for each of the weighting parameters for the first quarters of 2012. As shown in Table 3 the overall weight effect for the first quarter of 2012 was 1.93 and the coefficient of variation of the weights was 0.96. 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. The highest coefficient of variation of the weights was 1.5 for Mid North Coast health administration area. In both the previous landline only survey and the dual frame approach weights vary because of use of difference selection probabilities between strata, the sampling of one person per household and the calibration to age-sex benchmarks. Also as shown in Table 3 the weight effects for 2012 were 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.’

Issue 6: Result; Page 20. Is the overall SUMWGT (7.272.06) igual the target population?

Response 6: Yes 7.272 million is the target population.
Reviewer's report 2: Reviewer: Carol Pierannunzi

Title: 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

Overview: I read with interest the manuscript entitled “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.” This paper is on a timely topic that is being faced by many surveys which rely on telephone interviews. Most government surveys which rely on telephone interviews are conducted on both landline and cell phones, some with dual frame samples and some with overlapping samples. Australia seems to be at a disadvantage in constructing design weights, due to the lack of information on phone use and phone ownership at sub-national geographic levels. This is less of a problem in the US, but is likely to be a situation faced by a number of researchers who have geographic areas of interest where phone use information is unavailable. Overall the paper is a description of the issues that the researchers faced and their actions taken to deal with weighting without control totals. The methods are appropriate and draw from current established practices for dealing with probability of selection. I think that many readers would benefit from the discussions of the steps taken by these researchers. I think that the discussion would appeal to more readers if it was less specific and focused on the issue of weighting effects (cutting the imputation and data management sections)

Major revisions:

Issue 1: I think that the paper would be better cast as a discussion of establishment of design weights when control totals are not known. It reads more like a story of what happened rather than an academic paper on the appropriate use of weights. It is also cast as a description for just one case, but in fact it would be more useful to readers as a description of a method that was used in NSW but which can be used in any scenario where there are no benchmarks.

Response 1: We agree that a weighting strategy was developed for the case when population benchmarks by type of phone user are not available. We are making this point earlier and more explicitly and so have modified the abstract to read:

‘The weighting strategy needed to be significantly expanded to manage the differing probabilities of selection by frame, including that of children of mobile-only phone users, and to adjust for the increased chance of selection of dual-phone users. This paper describes the development of the final weighting strategy to properly combine the data from two overlapping sample frames accounting for the fact that population benchmarks for the different sampling frames were not available. The number of phones available by frame were only available at the national level but not for the state or regional level for which estimates are required’.

Issue 2: Parts of the discussion section should be moved to methods section of the paper. For example, the discussion on how they acknowledge that the composite factor of .5 is somewhat arbitrary includes the information that they had considered other control totals. It would help the reader understand the methods if this had been presented earlier in the paper. I was wondering about
this decision until late in the paper. In addition the information that the sample was disproportionate within each of the health administration area needs to come earlier in the paper (it is not clean until after the reader is internally asking questions about probability of selection).

Response 2: We have moved information about why the composite factor of 0.5 was chosen to the background, so the following sentence is now included:

‘So with overlapping dual-frames design surveys being relatively new in Australia [5,8,14,15] the use of \( \lambda = 0.5 \) as the compositing factor was considered appropriate.’

We have also provided information that the sample was disproportionate within each of the health administration area in the background so the following sentence is now included:

‘In the previous landline based samples for the NSWPHS, equal sample sizes were used in each stratum, even though the populations differed substantially and therefore the probability of selection varied by stratum.’

Issue 3a: The tables are cumbersome. I do not think that most readers will need the information in Table 2. The explanation in the text of imputation is fine. This is not the purpose of the paper and is distracting. Likewise the information in table 3 is not helpful for the reader to understand the design weighting process. A brief discussion in the text would be better.

Response 3a: We have removed Tables 2 and 3 and just included the important information in the text as follows:

‘For the landline frame imputation was used for number of persons in household (1 if missing and 10 if greater than 10), number of landlines phones in household (1 if 0 or missing and 5 if greater than 5), number of personal mobile phones (substitute with 0 if missing and to 5 if greater than 5). For the mobile frame imputation was used for number of children in household (1 if missing and 6 if greater than 6), number of landlines in household (substitute with 0 if missing and to 5 if greater than 5) and number of personal mobile phones (substitute with 1 if 0 or missing and to 5 if greater than 5)’

‘Data needed to be imputed, using these rules for 29 respondents for number of landline phones in the household (10 from landline frame and 19 from the mobile frame) and 26 respondents for number of personal mobile phones (15 from the landline frame and 11 from the mobile frame). The majority of respondents (97%) 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 (55/72; 76%) where phone numbers are more transportable.’

Issue 3b: A table comparing the old weighting process (using just the landline respondents) and the new process with examples of similarities/ differences in the weights is needed and essential to the conclusion that the weighting processes are “similar to, and in many cases, less than, the effects found” in earlier iterations of the survey. The authors tell us that the weights are similar, but it needs to be illustrated in tabular format with the actual weights/ weighting effects.
**Response 3b:** We do not think we need to also include a table comparing the old weighting process (using just the landline respondents) and the new process with examples of similarities/differences in the weights. We are not concluding that the weighting processes are "similar to, and in many cases, less than, the effects found" in earlier iterations of the survey we are concluding that the weight effects are "similar to, and in many cases, less than, the effects found" in earlier iterations of the survey. The weight effects for the previous year (2011) when only the landline frame was only used is already included in Table 5 (now 3) and we have references and provided an Appendix that contains the previous weighting procedure for the landline only frame.

**Minor Essential Revisions:**

**Issue 4:** I would consider taking the children out of the weighting discussion in the paper. I understand that the weights for the children were needed in the real world, but this is a discussion of a method, and the fact that there are only 138 (with only 26 in the cell phone sample), is problematic.

**Response 4:** As the study being described is a whole of population study I think that it is important to include the children from the mobile frame in the weighting strategy otherwise there would be a sector of the population who have a zero probability of selection, ie children of people who do not have a landline phone. We have thus expanded the background section of the paper to better describe the survey procedures and explain the inclusion of the children is important.

So the sentence:

‘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]’

has been expanded to read:

‘The landline phone sample procedures were the same as in previous years. The mobile phone sample procedures were as follows; NSW residents were selected using RDD of mobile phone numbers using CATI and the mobile phone owner was selected. If the respondent had one or more children one child was also selected at random in order to ensure that children of people who did not have a landline were also included. Further details about the methodology, call outcomes and representation of the sample in the first quarter of 2012 are provided in Barr et al [5], and the questions in the questionnaire are available from the survey website [1].’

We have also expanded the abstract to include mention of children of mobile-only phone users:

‘The weighting strategy needed to be significantly expended to manage the differing probabilities of selection by frame, including that of children of mobile-only phone users, and to adjust for the increased chance of selection of dual-phone users.’

**Issue 5:** I had a couple of questions on the Figures which were not addressed in the discussion. Tell the readers why there are so many low weights for the mobile frame (even through the means of the weights were higher than for the weights of the landline). I think that I understand why this could be true, but the authors do not address it. Overall I thought the Figures were helpful and appropriate to the discussion. I would do something to label them more effectively.
Response 5: The following has been added to the text to explain why there are few low weights for the mobile frame:

‘Most of the variability in the weights is due to the stratification by health administration area with equal number of respondents being selected for each health administration area which is disproportionate to the populations. Because there is no geography on mobile numbers no stratification can occur and very few of the mobile sample comes from rural areas and so there are few low weights. So once benchmarked to the populations the urban areas get quite high weights and rural areas quite low weights. For example in Far West 82% of the weights are less than 500, whereas in South East Sydney only 2% of the weights are less than 500 and 43% are 4000 or greater.

The graphs have been redrawn with the Y axis being the number of respondents so that it is easy to see how many respondents have large or small weights overall and by the frame and by phone use, rather than how they were previously represented (ie percentage of group).

Issue 6: The title of the paper correctly makes the reader assume that the paper will focus on a strategy. If the focus of the paper is changed to be more of a discussion of appropriate methods for developing design weights where there are unknown controls, I would also adjust the title. It is not just about NSW—it is a method for creation of weights in any context where there are no benchmarks.

Response 6: We would like to keep the paper with the existing focus ie that it focuses on a weighting strategy used for the NSWPHS. When we were preparing this paper, although there are several review articles on weighting sample surveys, they rarely go into enough detail so that the authors could understand the issues and or duplicate the processes. We thus think it is important to focus on one strategy in this paper rather than changing the focus to a paper that discusses appropriate methods for developing design weights as suggested.

However to better highlight the limited benchmark information in Australia we are happy to modify the title to:

‘Developing a weighting strategy to include mobile phone numbers into an ongoing population health survey using an overlapping dual-frame design with limited benchmark information’

Issue 7: There are a number of typos (p. 7 first line under Calculation of the weights “was” should be changed to “were”; p. 7 “9.8 million mobile telephone numbers” appears as “9.38” in the corresponding table—-one of those is off).

Response 7: Have changed p. 7 first line under Calculation of the weights to “were” and p. 7 have changed the text to “9.38 telephone numbers”

Discretionary revisions
**Issue 8:** There needs to be some discussion on why the dual frame sample was not used or considered and the effect that a dual frame would have had on the design weighting steps. This can go in the background or methods section.

**Response 8:** This remark is a little confusing and as it is a discretionary revision we have not made any modifications. If it is about why we chose an overlapping dual-frame design, rather than a screening dual-frame design, this was discussed in Barr et al 2012.

**Issue 9:** There needs to be some explanation of what other data might have been used to estimate phone use. There is no mention of whether there are data on the number of households and number of phones (landline and/or mobile) within any geographic area. I did not see any mention of the number of households. Perhaps the authors did look or have these data and they were not useful, but it is not clear.

**Response 9:** Such data are not available in Australia and comments on what data are available have been included in the paper.

**Quality of written English:** Needs some language corrections before being published

**Response:** We have engaged a language expert who has provided additional editorial comments.

To improve the readability the phrase:

‘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’

It has been moved down the background section and rewritten as:

‘In the overlapping dual-frame design there are three types of phone use; mobile-only, landline-only and dual-phone users—people with a mobile phone and living in a household with a landline phone—who could now be selected though either the landline or mobile phone number sampling frames’.
Reviewer's report 3: Eleonora Dal Grande

Title: 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

Overview: The authors have conducted a statistical study on how to weight survey data using data from two overlapping sampling frames for Australia.

Major Compulsory Revisions

**Issue 1:** Abstract – Background. The sentence “Details about the methodology….” should not be in the abstract.

*Response 1:* Have removed the sentence “Details about the methodology….” as suggested.

**Issue 2:** Background; Paragraph 1. Can you provide a statement on why including mobile telephone numbers is important and is a problem?

*Response 2:* Have added the following statement to describe why including mobile telephone numbers is important and is a problem.

‘Because of the potential for coverage bias from the growing number of mobile-only phone users in the population, estimated to be 19% in Australia in 2011 [2], mobile phone numbers were included in 2012 using an overlapping dual-frame design. Coverage bias is the product of the proportion of the population not covered and the difference in the mean of the variable of interest between the covered group and the non-covered group [3]. Evidence from the National Health Interview Survey (NHIS) in the US has shown the mobile-only phone users substantial different for the health indicators: five or more drinks in one day at least once in the past year (17.5% v 30.5% - 74% higher), current smokers (14.5% v 24.3% - 68% higher), and ever diagnosed with diabetes (10.8% v 6.2% - 43% lower) [4].’

**Issue 3:** Background. Literature from Australia in this field needs to be cited or discussed: Pennay D, Vickers N. Dual-frame Omnibus Survey. Technical and methodological summary report. The Social Research Centre, 2012;

*Response 3:* The report has been acknowledged in the background section and added it to the reference list.

**Issue 4:** Background. I would like to know how this paper is novel and how it is different from the current literature.

*Response 4:* The weighting methods described in the report Dual Frame Omnibus Survey were for Australia overall and utilised rim weighting rather than including an overlap adjustment. This paper describes a weighting strategy which needs to deal with disproportionate stratification of the landline frame, when data needs to be collected about children as well as adults, and how to apply an overlap adjustment.
The following sentence has been added to the introduction to highlight what this paper adds to the current literature:

‘However the most recent detailed description of dual frame weighting available in Australia from the Dual-frame Omnibus Survey conducted in 2012 did not need to deal with disproportionate stratification of the landline frame, data needing to be collected about children as well as adults, and how to apply an overlap adjustment. [8]’.

**Issue 5: Methods/design, calculation of the weights; Paragraph 2, 2nd sentence.** What method was used for imputation?

**Response 5:** I have removed Tables 2 and 3 and included the imputation information in the text as follows:

‘For the landline frame imputation was used for number of persons in household (1 if missing and 10 if greater than 10), number of landlines phones in household (1 if 0 or missing and 5 if greater than 5), number of personal mobile phones (substitute with 0 if missing and to 5 if greater than 5). For the mobile frame imputation was used for number of children in household (1 if missing and 6 if greater than 6), number of landlines in household (substitute with 0 if missing and to 5 if greater than 5) and number of personal mobile phones (substitute with 1 if 0 or missing and to 5 if greater than 5).’

**Minor essential revisions**

**Issue 6: Document overall.** The authors swap between “phone” and “telephone”, can they change it so it is consistent throughout the document.

**Response 6:** We had tried to use the term “phone” when referring to the handset type and had tried to use “telephone” when referring to numbers. We agree that that was not consistent and so have changed them all to phone.

**Issue 7: Abstract-Background.** Since this is journal is for international audience, please spell out NSW.

**Response 7:** We have changed “NSW” to “New South Wales”

**Issue 8: Abstract: Background; Paragraph 6.** Space between “June 2011” and “when”.

**Response 8:** We modified the sentence to read:

‘It was estimated in the 2010-11 report that as at June 2011, 74% of adults in Australia lived in a household with a landline and a mobile phone…’

**Issue 9: Methods/design, calculation of the weights; Paragraph 2, 2nd sentence.** Sentence a bit clumsy, suggest “… and 26 respondents for number of mobile phones personally have” to “… and 26 respondents for number of personal mobile phones”

**Response 9:** We have changed the sentence “… and 26 respondents for number of mobile phones personally have” to “… and 26 respondents for number of personal mobile phones”