

Paragraph numbers added by Dr Don Harding. The document has not been changed in any other way.

Commentary on a survey conducted in relation to the Safety Net case, 2004

Ian Gordon PhD, AStat
Director
Statistical Consulting Centre
University of Melbourne

7 April, 2004

1 Introduction

1. The Statistical Consulting Centre was commissioned by the Australian Council of Trade Unions (ACTU) to provide expert statistical commentary on some matters arising in the Safety Net Review 2004, being heard in the Australian Industrial Relations Commission.

2. A survey has been carried out, sponsored by the Commonwealth Government, and a report of the survey has been provided to me. Some supplementary tables relating to the survey have also been provided. In addition, I have been provided with the survey form of the Australian Bureau of Statistics (ABS) in relation to employee earnings and hours, and with transcripts from the current Safety Net hearing, dated 22 March 2004 and 23 March 2004.

3. Hence the documents with which I have been provided and upon which I rely for this commentary are:

1. "Minimum wages in Australia: an analysis of the impact on small and medium businesses", by Don Harding and Glenys Harding, dated March 2004;
2. "Minimum wages in Australia: an analysis of the impact on small and medium businesses: Supplementary tables to ACTU", by Don Harding and Glenys Harding, dated 19 March 2004;
3. A survey form from the Australian Bureau of Statistics (ABS) entitled "Survey of employee earnings and hours: Employee form" (4 pages);
4. Transcripts from the current hearing, which I have accessed via the internet, at <http://www.airc.gov.au/documents/Transcripts/220304c20032508.htm> for the hearing of 22 March 2004 and at <http://www.airc.gov.au/documents/Transcripts/230304c20032508.htm> for the hearing of 23 March 2004.

4. I have been asked in specific terms to comment on three issues arising out of the survey carried out on behalf of the Commonwealth and described in the Hardings' report: response rate, standard errors and the location of information about the Safety Net concept in the Commonwealth survey, and on any other matters I consider important.

2 Response rate and non-response bias

5. The response rate in the survey in question was between 20% and 22%.

6. In my opinion, the size of this response rate is low enough to have serious concerns about the reliability of the results of the survey.

7. It is important that the response rate be high, to reduce the possibility of response bias. Response bias occurs when the respondents to a survey have different opinions and characteristics to those who do not respond. By the nature of the case, it is generally not possible to ascertain the relevant survey information from non-respondents. Accordingly, the response bias in a particular instance cannot be measured.

8. A high response rate is vital: the higher the response rate, the less the scope for response bias. Conversely, in a survey with a low response rate, the scope for response bias is large.

9. If the problem of non-response is ignored, this amounts to the assumption that all characteristics in the non-respondents are present to the same degree and proportion as in the respondents. This is an assumption that response bias is absent.

10. However, there are many examples that illustrate that assuming the absence of response bias is dangerous. In the history of sampling there are numerous documented cases where information obtained (to varying degrees) on non-respondents has shown response bias. I myself was involved in a survey of young men who had been convicted of a drink-driving offence 10 years prior to the survey. Of the total group we attempted to survey, approximately two thirds could be traced and interviewed. Independently of the survey, we were able to obtain data on alcohol-related traffic offences that had occurred in the total group in the intervening 10 years since the index offence. Among those interviewed, the percentage with 2 or more such offences was 15%; among the non-respondents, the percentage of 2 or more such offences was 30%. This clearly indicated that the non-respondents, as a group, had tangibly different histories and characteristics than the respondents. In this case this was sufficient for us to decide not to publish the results of the survey, on the grounds that the unknown response biases were likely to produce misleading results.

11. It is sometimes assumed, possibly implicitly, that there is no obvious reason why the non-respondents in a given survey would be different from the respondents, and therefore that the results from the responding fraction of the sample can be validly extrapolated to the whole. This argument is problematic for two reasons. Firstly,

it relies on the imagination of the researcher; there could be non-obvious reasons that apply. Secondly, non-respondents are intrinsically different from respondents in at least one respect: their response status. Experienced survey scientists understand that a behavioural characteristic such as this—the decision to respond to the survey—is most unlikely to be unrelated to all other characteristics.

12. It is for this reason that very low response rates are taken to be evidence of an inadequate survey, no matter what other features the survey has. Of course, the threshold of “very low” is ultimately arbitrary.

13. It is difficult to be definitive about what constitutes an acceptable response rate under any circumstances. I am prepared to say that in almost all situations, a response rate of over 90%, for example, is sufficient to render the results reliable for quantitative conclusions.

14. Equally, it is difficult to say what constitutes an unacceptably low response rate under any circumstances. However, statisticians generally agree that there is a threshold below which response rates are unacceptably low. They may disagree about the actual value of the threshold.

15. Many writers decline to construct specific guidelines, for the above reasons. However, Babbie writes: “A response rate of at least 50 percent is generally considered adequate for analysis and reporting. A response rate of at least 60 percent is considered good, and a response rate of 70 percent or more is very good”.¹

17. In commenting on these sorts of guidelines, Lohr writes: “I believe that giving such absolute guidelines for acceptable response rates is dangerous and has led many survey investigators to unfounded complacency; many examples exist of surveys with a 70% response rate whose results are flawed.”²

18. That is, Lohr’s view is that a response rate even as high as 70% may be too low.

19. In my opinion, a response rate of lower than 30% means that the results of a survey must be regarded as unreliable, due to possible response bias. This survey has a rate substantially below that threshold. Hence I have serious concerns about its reliability.

20. I am surprised that Dr Harding did not accept (PN 801) that a response rate of 20% to 22% is low. In my opinion, this view is contrary to the consensus among statisticians.

21. Further, in the Hardings’ reports there are a number of points at which the response rate of the survey is conceded to be “low”: paragraph 71, page 20; paragraph 73, page 21 (“the low response rate”); paragraph 10, page 5.

22. In the Commonwealth survey there are certainly possibilities for response bias.

23. For example, it may be that employers who are supportive of the Government’s

¹Babbie E.R. (1973). *Survey Research Methods*. Belmont, California: Wadsworth, page 182.

²Lohr S.L. (1999). *Sampling: Design and Analysis*. California: Duxbury, page 281.

industrial relations policy were more likely to respond, because they saw it as an opportunity to express their view and possibly to influence policy.

24.

It may be that employers who were impossible to contact had characteristics that tended to be different from those who could be contacted. We don't know this, of course. But it is a realistic possibility, and the low response rate means that such a possibility may have affected the available results in a way that we cannot apprehend.

25.

The Hardings quoted some work by Professor Jon Krosnick which examined the demographic representativeness of 20 or so (the number varies depending on what was recorded) media and government surveys in the USA, and assessed whether the response rate was related to the representativeness. They quoted a conclusion of Professor Krosnick's. I think it is useful to consider the detail of Krosnick's report, which I accessed via the internet at <http://www.psy.ohio-state.edu/social/krosnick.htm>.

26.

It is worth noting, in passing, that the seven Government sponsored surveys in Krosnick's data have response rates in the range 19% to 87%, with the average response rate being in the range 36% to 62% (depending on the definition used); hence the current survey is at the very low end of these figures.

27.

Krosnick found that:

- The 18–25 year age group was under-represented in 86% of the surveys;
- The correlation between response rate and age was -0.80 , which was very statistically significant. (The correlation is a measure of association which ranges between -1 and $+1$, hence values close to these show very strong correlation.) This means that the higher the response rate, the lower the average age;
- African-Americans were under-represented in 95% of the surveys;
- Subjects with less than high school education were under-represented in 100% of the surveys;
- Post-graduates were over-represented in 100% of the surveys;
- The correlation between response rate and education was -0.74 , which was very statistically significant. This means that the higher the response rate, the lower the average level of education;
- People with household income of over \$75,000 were under-represented in 94% of surveys;
- There was “slightly larger error for race and income than would be expected based on sampling error only”;
- There was “more error for education than for sampling error alone.”

28.

Krosnick formed the view that the above evidence was consistent with the proposition that “even surveys with relatively low response rates have excellent demographic representativeness”.

29. I disagree. In my opinion, his research demonstrates exactly the concern we should have about low response rates: in his data, low response rates were associated with biases in the estimation of relevant characteristics of the target population; in his case, demographic characteristics.

30. However, this is only a starting point. Demographic representativeness does not, of itself, guarantee an unbiased result for the question of interest, unless demographics are the only factors influencing the response.

31. What we really want in a survey is representativeness with respect to the question of interest, and representativeness with respect to other variables—such as demographic features—can only ever be a proxy for that.

32. In any case, the Krosnick research relates to surveys of the general population in the USA. As Dr Harding observed in cross-examination, “unfortunately we don’t have comparable work for business surveys”. There are important differences between the natures and contexts of surveys of business and surveys of the general population, and I do not agree with Dr Harding that you can, with any security, “extrapolate” from the Krosnick research to a surveys of small businesses in Australia.

33. Dr Harding quoted Australian Bureau of Statistics research in which its Statistical Clearing House looked at simulated data, and (separate) real data relating to a survey of businesses, to see whether non-response bias is a problem. Not surprisingly, the paper Dr Harding cited concluded:

“Overall, we can conclude that non-response bias can have significant detrimental effects on the accuracy of survey estimates. These effects can be reduced through higher response rates. From this, it is recommended that steps should be taken to ensure that the response rates achieved in any particular survey are as high as possible.”

34. I agree with this conclusion, which does not support the contention that low response rates may be adequate.

35. There are examples in the classic sample survey literature that demonstrate the way in which the bias can successively reduce with increasing response rate. For example, a sample of fruit tree growers in North Carolina³ used three successive mailings. The first mailing yielded 10% of the target population, and had an average of 456 fruit trees per grower. The second mailing yielded 17% of the target population, and had an average of 382 fruit trees per grower, and the third mailing yielded 14% of the target population, and had an average of 340 fruit trees per grower. In this particular case, the population average was known and was equal to 329 fruit trees per grower, so the three successive mailings had averages which became closer and closer to the true value, although all of them had a positive bias.

36. The Hardings made a claim about the likely direction of a non-response bias across the strata used to construct the weighted estimates. They wrote (7.114, page 114)

³Cochran WG. (1977). *Sampling Techniques (3rd edn.)*. New York: Wiley, p 360.

that “it is unlikely that the bias in all of the stratum [sic] will be of the same sign”. This assertion was tested in cross-examination at PN823–PN831. The discussion there is not clear to me, but in any case I do not believe a case has been made for the assertion. To the extent that I understand Dr Harding’s argument, it seems to amount to the proposition that any response bias will be random: sometimes positive, sometimes negative, and hence, perhaps, tending to cancel out overall. However, the nature of a non-response bias means that it can have a substantial systematic, non-random component. When the same bias operates in all strata, it is likely to do so in the same manner.

37.

Hence I consider it possible, and likely in some circumstances, that a non-response bias will operate across all strata, and therefore have the same sign in all strata. It will do so in circumstances where the bias is related to characteristics not captured by the stratification. The mere construction of strata does not in itself solve the problem, because there can be features of businesses that are not the basis of the strata, but that are nevertheless related to the questions of interest, and the inclination to respond. I have made observations above about the way such a bias might operate, but there are a number of possible such characteristics. Again, a low response rate gives much scope for such biases to have an effect.

38.

In the overview of the report, the Hardings offer three main reasons to support the contention that the low response rates do not create a significant bias.

1. The report argues (paragraph 75, page 21) that the decision not to respond was made before the business was told the survey contained some questions about the Safety Net, and that therefore non-response is unrelated to Safety Net issues.

Firstly, this raises the question of what potential respondents were actually told, prior to the decision to respond. As far as I am aware, there is no information about this in the report. However, it is relevant to the validity of the survey. What the interviewer said, and what answers may have been given to respondents’ questions at this point, are important. If a potential respondent asked, for example: “Who’s behind this survey? Is it a government thing? Why do they want to know about this from me?” . . . what kind of answer from the interviewer was standard protocol? The content of any preamble, and of any interviewer comment, could influence the propensity to respond to the survey. Without knowledge of this I cannot assess whether this may have been the case.

Secondly, and rather obviously, whether or not someone responds may be associated with Safety Net issues, without the decision being directly provoked by knowledge that the survey is about those issues. It could be, for example, that respondents who are harder to contact are inclined to have particular Safety Net views. Or business respondents who “hate surveys” and therefore refuse to participate may tend to have particular Safety Net views.

2. A second reason offered by the Hardings (paragraph 76) is that “non response was attributable to factors that are best described as random and would not, therefore, be expected to over (or under) select respondents [in the categories of interest]”.

In my opinion, this amounts to a bald assertion. I am not aware of the Hardings' justification for this assertion.

3. The third reason given is that the weights used in the survey are constructed by post-stratification (paragraph 77, page 21).

Paragraph 7.115, states that the variables industry, size and region are the main variables that explain variation *in the sample* (my emphasis), and “thus, by extension, these variables are likely to capture the bulk of any variation in the probability of non-response”. But this argument is a non-sequitur. It is easy to see that there could be variables not measured which are related to the outcomes of interest, and they could be more important than the measured variables, in which case the measured variables might not capture non-response. The Hardings essentially concede the failure of this argument in stating in the text that follows this argument: “...although we cannot be certain of this without conducting and analysing a survey of non-respondents”.

I agree with this latter statement. Since a survey of non-respondents has not, in fact, been conducted, we cannot be certain that non-response is mainly explained by the stratification variables. In my opinion, we cannot even be confident that this is the case.

39.

Finally, in the detail of the response rate calculations, there are some matters that are not clear to me.

- In Table 7.5:
 - Apparently there were 651 calls to businesses with under 20 full-time employees that were considered ineligible. I do not understand this.
 - In two other categories, namely “mobile phone” and “number not working” there must be some debate about whether the appropriate category is “not eligible”.
- At 7.92, page 116, it is noted that “9887 [calls] is a large number relative to the 1800 completed interviews”. (Table 7.5 indicates the figure is 1804 interviews, actually.) Since the figure of 9887 refers to the previous wave of the survey, the point is presumably even more forceful in reference to the survey under discussion, for which 14,439 calls were made to achieve 1800 completed interviews.

40.

In my opinion, the response rate in the Commonwealth survey means that the results cannot be regarded as reliable.

3 Precision of estimates

41.

In paragraph 12 of the report (page 5) the Hardings state that “Inference requires standard errors” and that “standard errors are desirable”. They also state that good

practice involves the reporting of measures of precision, such as confidence intervals (paragraph 7.125).

42. The report contains no standard errors or confidence intervals. According to the authors, they were unable to calculate them (paragraph 7.126, page 116).

43. The best way to reflect the precision of an estimate is by a confidence interval, for example, a 95% confidence interval. This is a range of values within which we are confident—95% confident, for example—that the true value lies. Since most confidence intervals use standard errors directly in their calculation, obtaining standard errors is particularly important, as a vital input to the confidence interval.

44. For example, it might be estimated on the basis of a survey that 13% of school children aged 15 have tried cigarette smoking. This is a point estimate, but even assuming that it has been obtained from a properly designed study with high response rate, we need to supplement the point estimate with a confidence interval for meaningful inference. In that case, we might obtain a 95% confidence interval of (10% to 16%).

45. Consider, however, another example. A random sample of AFL supporters is taken, which shows that 50% of them are St. Kilda supporters. The survey was appropriately random, with 100% response rate. What meaningful inference can be drawn? It is not reasonable to draw a conclusion of substance without a 95% confidence interval for the true percentage of St. Kilda supporters, on the basis of the sample.

46. Suppose that when this calculation is done, the 95% confidence interval is found to be (1% to 99%). That is, the data are telling us that we can be very confident that the true percentage of St. Kilda supporters is somewhere in the range 1% to 99%.

47. Since the percentage of interest must be in the range 0% to 100%, we have hardly made an advance by finding out that we are very confident that the true value is in the range 1% to 99%. The data giving rise to such a point estimate and confidence interval are a random sample of two supporters in which one is a St. Kilda supporter.

48. Now consider the estimate of a reduction of about 14,000 jobs over a three month period attributable to the 2003 SNA (Table 3.3, page 49). We do not know what the confidence interval for this figure is. Is it $14,000 \pm 5,000$? Could it be $14,000 \pm 20,000$? On the evidence of Dr Harding under cross-examination, we do not know. This latter example would be a confidence interval of (-6,000 to 34,000). We would be claiming that we are 95% confident that the true reduction in jobs lies between -6,000 jobs (which is actually an *increase* of 6,000 jobs) and 34,000 jobs.

49. In those circumstances, we cannot rule out that the data are consistent with no reduction at all in the actual population of interest, or even an increase.

50. The detailed technical reason for the lack of standard errors and confidence intervals is unclear to me. It appears to arise from a concern that the sampling was done using what is known as a “negative binomial” approach: sampling until a given number of subjects was obtained. This approach does not, in itself, cause any material difference to the way in which standard errors are calculated from the binomial

case. However, as I say, the concern the Hardings raise is not clear to me.

51. In any case, the need for some measure of precision is fundamental.

52. I should also note that, given the difficulty that the Hardings have flagged, any subsequent provision of confidence intervals or standard errors would need careful scrutiny. Indeed, I make the observation that the point estimates they have calculated are not really checkable on the basis of the material I have seen. It should be possible to provide stratum weights and estimates, so that such checking, at least in some key cases, could be done.

53. In my opinion, the absence of confidence intervals for key estimates of interest means that we cannot be confident about the corresponding population figures in quantitative terms.

4 Knowledge of the Safety Net concept

54. The ACTU sought my views on the importance of respondents' understanding of the Safety Net adjustment. I have perused the cross-examination of Dr Harding on that point, at PN574 to PN587. I am not personally aware of the details of the ABS survey of employee earnings and hours in 2002 (EEH survey), and concerns that the ABS had about publishing the results of the survey on the basis that respondents in that survey had not understood the concept of the safety net, other than via information supplied to me by the ACTU.

55. The ACTU has provided me with a copy of an email from the ABS which states that in the EEH survey, "With respect to the safety net adjustment, it became apparent that this term and the associated concept were not widely understood by respondents. In addition to underreporting of entitlement to the safety net adjustment among award-only employees, there was frequent incorrect reporting of safety net entitlements among employees who had no link between their remuneration and an underlying award." As a result of these concerns on the part of the ABS, these survey results have not been released.

56. At the very least, this finding from the nation's premier survey organisation must give concern about this issue in any survey which attempts to obtain information about the safety net adjustments.

57. In a fundamental sense, it is important to be confident that a survey is eliciting responses about the matter of interest, and when the topic has some complexity, it is useful to test that understanding in some way.

58. In the case of the Commonwealth survey, it appears that the understanding of the concept was assumed to be satisfactory, given the preamble at question 11, without explicit testing of this assumption. It therefore comes down to a matter of judgment as to whether this assumption is reasonable.

59. I do not consider myself to have the relevant expertise to make such a judgment.

5 Other matters

60.

There are some important anomalies, in my opinion, in the form of the questions on the actual effect of the 2003 Safety Net adjustment, and the question about future prospects.

1. Consider question 21a. This question is only asked of firms that said at question 20 that they increased the size of their total workforce over the past three months. Why is there no option allowing for the position that the firm would have decreased its workforce in the absence of a Safety Net wage increase in 2003? The absence of such an option means that the question is “leading”, in the sense of having a restricted range of possible responses.
2. I think the term “lost” in question 21c is inappropriate and may have biased responses. This is because this question is only asked of firms that have already indicated that they have increased their workforce, and then indicated that they would have increased the workforce by less without a 2003 Safety Net wage increase. Jobs are not being “lost” in such cases; rather, jobs are being created, but fewer than would otherwise be the case. In my view, this wording has the potential to confuse, in the context in which it is used.
3. Consider question 23a, which is only asked of firms that said at question 20 that they decreased the size of their total workforce over the past three months. The structure of this question should mirror question 21a: but it does not. In my opinion, it is seriously flawed, for several reasons.
 - (a) As is the case for question 21a, the responses should allow for more options. How does someone respond if they would have reduced the workforce by less if there had been no Safety Net wage increase? Or if they would have increased the workforce? The absence of these options is worrying, and could have consequential effects.
 - (b) It is not clear to me why, given that only one possible response was offered for “If there had been no Safety Net ...”, it was the option that the reductions would have been greater. As I have just argued, I believe that all of the logically possible options should have been offered, but the one that has been offered seems to be one which assumes that the Safety Net will support the retention of employment. I note that none of the 80 firms that reported a decrease used this option (Table 7, Supplementary Tables).
 - (c) There is the logical difficulty that a firm may hold views 2 and 3 concurrently. In a forced choice question it is strongly desirable to arrange mutually exclusive options. Again, there is no option allowing for the position that the firm would have decreased its workforce by less in the absence of a Safety Net wage increase in 2003.
4. Consider question 24a. In my opinion, it is hard enough to obtain answers to counter-factual questions about the past (“If Peter Everitt had played in the 1997 AFL Grand Final, would St. Kilda have won?”). This is for the obvious reason that it is difficult to isolate the possible effects of a single change in past

circumstances, holding all other things equal. However, it is even more difficult, in my opinion, to get reliable information about people's future intentions given hypothetical scenarios. This is what attempted at question 24a. Again, the options are confusing. How does someone answer, if they think they would reduce the number of employees, but not by as much as they otherwise would have done? Such a respondent has a choice of responses 2 and 4; it is not clear to me which one they might choose. However, they lead to very different consequential questions: option 2 leads to question 24c, which is about the number of jobs **saved**, while option 4 leads to question 24d, which is about **losing** jobs. Further, if option 2 is given, the following options should also be allowed:

- “My business would increase its workforce by more than it would otherwise have done.”
- “My business would increase its workforce by less than it would otherwise have done.”
- “My business would reduce its workforce by more than it would otherwise have done.”

61.

These structural issues in the detail of questions are not trivial. They had the potential to lead to confusion, in my opinion, and hence unreliable results.

6 Summary

62.

In my opinion:

- The response rate in the Commonwealth survey was low, and this means that the results cannot be regarded as reliable.
- The absence of confidence intervals for key estimates of interest means that we cannot be confident about the corresponding population figures in quantitative terms.
- There was confusing wording in several key questions which had the potential to cause unreliable results.

63.

In my opinion, the low response rate alone means that it is not possible to extrapolate reliably from the Commonwealth survey to the target population as a whole, with any confidence. The other factors accentuate the difficulty.