

QUALITY INDICATORS FOR HOUSEHOLD SURVEYS

Methodology Advisory Committee meeting

November 2000

AUTHORS

David Finlay
Population Survey Development
Email: david.finlay@abs.gov.au
Phone: (02) 6252 7296
Fax: (02) 6252 6530

Karen Quine
Population Survey Development
Email: karen.quine@abs.gov.au
Phone: (02) 6252 7624
Fax: (02) 6252 6530

Matthew Cronin
Household Collection Support
Email: matthew.cronin@abs.gov.au
Phone: (02) 6252 5594
Fax: (02) 6252 8015

Quality Indicators for Household Surveys

Introduction

1 At the end of last year, the Australian Bureau of Statistics (ABS) Population Survey Practices Steering Committee (PSPSC) indicated support for developing a standard set of quality indicators relevant to most household surveys run by the ABS. Since then, the Statistical Services Branch, Population Surveys Branch and Social and Labour Division of the ABS have been working jointly on the development of a "core" set of standardised quality indicators. These quality indicators cover all components of the survey cycle, and together they provide a consolidated framework of indicators from which users can assess the quality of survey data and ABS can monitor quality issues.

2 As many of the quality indicators are still being drafted, this is more a discussion paper than a final piece of research. We are endeavouring to compile a core set of indicators, however it is expected that many of these indicators will evolve and improve over time. The purpose of this paper is to provide the MAC with an overview of the quality indicators that ABS plans to implement for household surveys, and to propose a set of indicators that could be used externally. Feedback is sought from the MAC on the proposed indicators and associated methodologies, as well as confirmation that our directions are sound.

3 It should be noted that this project is one of several ABS projects that are currently investigating the issue of quality. Other research currently being conducted by the ABS on data quality issues are: National Accounts Quality Measures (John Zarb), Qualifying Quality (Bill Allen), Quality Measures for Time Series (Gemma van Halderen, Nick von Sanden) and Response Rates as Quality Measures (Graeme Brown).

Executive Summary

4 The ABS is currently developing a standard set of quality indicators relevant to household surveys run by the ABS. They will provide a consolidated framework of indicators from which users can assess the quality of survey data and ABS can monitor quality issues. Data quality encompasses many aspects, especially accuracy, timeliness, relevance and cost of production.

5 The quality indicators will be implemented in three phases, based on the importance of the indicator and the extent of work required to implement it. Phase 1 indicators (which this paper focuses on) involves those indicators that are relatively easy to implement in terms of data availability and cost. Phases 2 and 3 indicators require greater effort in terms of developing methodology, setting up procedures and resolving cost implications.

6 Many facets of the survey cycle are covered in the core set of quality indicators, from aspects such as frame quality and sample design, through to the questionnaire and timeliness. However, accurately measuring the quality of some aspects, especially the questionnaire, is difficult to do quantitatively and hence these

quality indicators should be used with caution. Although many of the quality indicators will only be used internally within the ABS, some of them may be useful to external clients.

7 Feedback is sought from the MAC on all aspects of the quality measures project, but in particular the comprehensiveness of the indicators and the set of quality indicators for external use. It is hoped that input from the MAC will feed into the design and development of the quality indicators.

Purpose of Quality Indicators

8 Quality indicators for household surveys are useful for:

- Assisting survey managers and data users to assess the quality of survey data and determine whether that quality is sufficiently good for the intended use of survey outputs.
- Providing an objective basis for continuous improvement of the survey process.
- Enabling managers to optimise the resources spent on different parts of the survey process.

9 Data quality can involve a broad range of issues and several researchers have documented their findings in this regard. For example, G. Brackstone identifies the following six dimensions of data quality: relevance, accuracy, timeliness, accessibility, interpretability and coherence (BRACKSTONE, G. (Dec 1999) *Managing Data Quality in a Statistical Agency, Survey Methodology Vol 25 No. 2* pp. 139-150). For the purposes of this paper, quality of data encompasses accuracy, timeliness, relevance and cost of production. Both internal and external users of ABS household survey data should be able to use, and should actually be using, information about the quality of our data. This information should be used to help make appropriate judgements on the quality of the data when making informed decisions.

10 Monitoring of appropriate quality indicators provides a mechanism for ensuring that survey outputs are of an appropriate quality for their intended use. In addition to routine monitoring of quality indicators, survey managers should conduct special investigations aimed at identifying and reducing/removing the factors which would (or would likely) cause a reduction of quality. For continuous improvement of the survey operation, it is important that quality indicators on components of a survey operation be evaluated on a regular basis.

11 Although many of the quality indicators covered here will only be used internally within the ABS, some of them may be used externally (e.g. the undercoverage rate may be useful for users interested in how well the survey represents the population, or the amount of testing undertaken may be of interest to external clients funding a particular survey). Hence we need to ensure that the quality indicators are relevant to users of ABS data. This would include the training of users, to help them become aware of what quality indicators exist, the benefits they provide, and how to understand and use them.

12 It is envisaged that in future a standard set of quality indicators will be produced for each ABS household survey. By standardising the measures to be

produced, their definitions and the infrastructure for their production and dissemination, both internal and external users will be able to compare the quality of data between different surveys, as well as monitor, for repeated surveys, the change in quality over time. In addition, survey areas will be able to adopt best practices in quality management and help raise ABS's capability in the measurement and dissemination of quality information.

13 We acknowledge that in addition to the core set of standard quality indicators that encompass all surveys, there may be other important quality indicators which are specific to the content or methodology of each survey, but are not included in the standard set because of their special nature. For example, the quality indicators outlined in this paper are weak in the area of relevance (i.e. to what extent the survey meets the needs of users), due to the difficulty in defining indicators that are applicable across all household surveys. Therefore, relevance indicators should be included when specifying indicators specific to each survey.

Summary of Quality Indicators

14 The quality indicators will be implemented in three phases, based on the importance of the indicator and the extent of work required to implement it:

- Phase 1 - these are the indicators that should be relatively easy to implement as the relevant data are easily available and costs are not prohibitive. It is planned to have facilities available for most indicators set up by early 2001.
- Phase 2 - these are the indicators that should be feasible to implement, data could be made available and costs are unlikely to be prohibitive, but further work is required to develop the methodology for measurement, or to set up the procedures and systems for collecting the data.
- Phase 3 - these are the indicators currently not able to be implemented. Considerable work is needed to resolve the measurement standards and methodology. These indicators are likely to require information to be specially collected, and the cost implications would need to be considered.

15 The proposed "core" set of quality indicators and their associated implementation priorities have been summarised in the table below. More detailed information on those indicators that form Phase 1 of the project can be found in the next section.

| Category | Quality Indicator | Phase of Implementation |
|--------------------|--|--------------------------------|
| Data specification | Extent of meeting key objectives | 2 |
| | Data items specified v. implemented | 2 |
| | Percentage of data items deviating from ABS concept and classification standards | 2 |
| Frame quality | Sample loss rates | 1 |
| | Time between blocking and enumeration (months) | 2 |
| | Time since last growth revision (i.e. how recently the frame has been revised to allow for recent population growth) | 1 |

| | | |
|-----------------------------|--|--------------------------------|
| Coverage | Undercoverage rate | 1 |
| Sample design | Standard errors for key data items | 1 |
| | Design effects for key data items | 1 |
| | Relative standard error (RSE) realisation ratio for key data items | 1 |
| | Sample size realisation ratio | 1 |
| Questionnaire | Percentage of modules (topics) cognitive tested | 1 |
| | Number of rounds and extent of field testing | 1 |
| | Percentage of modified (repaired) modules | 1 |
| | Testing strategies employed | 1 |
| | Number of queries raised during field operations/input processing | 1 |
| | Sequencing "complexity of pathways" | 3 |
| | Questionnaire module ratings "confidence in meeting data specifications" | 3 |
| Interviewing process | Average number of visits/telephone calls made to attempt contact | 2 |
| | Percentage of households requiring follow-up | 2 |
| | Other Responsible Adult (ORA) rate (i.e. the proportion of interviews where data is gathered from someone other than the person to whom the questions refer) | 1 |
| Household response rates | Full response rate | 1 |
| | Partial response rate | 1 |
| | Total response rate | 1 |
| | Full non-response rate | 1 |
| | Refusal rate | 1 |
| | Non-contact rate | 1 |
| Category | Quality Indicator | Phase of Implementation |
| | Non-response due to technical difficulties rate | 1 |
| | Other non-response rate | 1 |
| | Processing reject rate | 1 |
| | Item response rates for key data items | 2 |
| Provider (respondent) load | Average household interview time (mins) | 1 |
| | Average time required by respondents outside of interview by household (mins) | 1 |
| | Percentage of respondents who needed to source records "other than those expected" | 2 |
| Coding | Coding error "discrepancy" rate | 3 |
| Editing and data adjustment | Edit failure rates | 2 |
| | Over editing rate | 2 |
| | Imputation rate for key data items | 2 |
| | Data adjustment measure for key data items | 3 |

| | | |
|--------------------------------------|--|---|
| Data revision | Changes to derivation specifications | 1 |
| | Versions of publication Main Unit Record Files (MURFs) | 1 |
| | Versions of public release Confidentialised Unit Record Files (CURFs) | 1 |
| | Number/percentage of data items containing errors identified in the publication MURF | 1 |
| | Number/percentage of data items containing errors identified in the public release CURF | 1 |
| | Number of errors identified in a publication | 1 |
| Timeliness | Timeliness of field enumeration | 1 |
| | Timeliness of the publication MURF | 1 |
| | Timeliness of the first (initial) publication | 1 |
| | Timeliness of the public release CURF | 1 |
| Output usage and client satisfaction | Rating of how well the data is used | 2 |
| | Client satisfaction rating(s) | 3 |
| Costs summary | Enumeration cost per fully responding household | 2 |
| | Overall costs: 1) Development 2) Field enumeration 3) Processing 4) Analysis and Dissemination 5) Other Processes and planning/management 6) Overall | 2 |

Phase 1 Quality Indicators

Frame Quality

Sample loss rates

16 Sample loss rates are calculated in terms of dwellings rather than persons, as it is only dwellings that are listed before enumeration. There are two components to sample loss: unoccupied dwellings, and all persons (in a dwelling) out on scope and coverage. The sample loss rate for unoccupied dwellings (e.g. dwellings vacant, demolished, under construction, listed in error) is the percentage of all dwellings in the sample that are unoccupied. The sample loss rate due to all persons out on scope and coverage is the percentage of dwellings in the sample where all persons in that dwelling are outside the scope and coverage of the survey.

17 Sample loss rates are necessary both as measures of frame quality/efficiency and for the determination of the sample required to meet a determined responding sample size. A high sample loss rate indicates a high proportion of "dud" units on the frame, and results in inefficiencies in collection activities as well as in uncertainty

of final sample size. Low sample loss rates are more likely to occur closer to reselection when the frame is reasonably new.

Time since last growth revision

18 Time since last growth revision indicates how recently the frame has been revised to allow for recent population growth. The aim of growth revision is to update the list of dwellings in the strata and then ensure an efficient sample of the stratum takes place in terms of the size and clustering of the sample. Growth revisions are carried out for stratum that have experienced 'large' dwelling growth since the last growth revision.

19 This quality indicator measures the timeliness of the sample frame. A long time since the last growth revision means that the representation of new dwellings will be quite variable and hence increase the variability of the resulting estimates. An out-of-date frame can also cause some workload problems due to growth occurring in selected blocks.

Coverage

Undercoverage rate

20 The undercoverage rate provides an indication of the undercoverage of the sample. Under-representation of the target population would lead to potential bias of the results. The non-covered population is usually estimated for through benchmarking the sample to the target population. When the non-covered population differs from the covered population in terms of the variable of interest, the estimates would be subject to "undercoverage bias". The proportion of the target population not covered by the sample, i.e. the survey undercoverage rate, would provide an indication of the magnitude of this bias.

21 An undercoverage rate equal to zero indicates that there is no undercoverage in terms of the sample size, i.e. the population estimates based on the sample represent the target population. An undercoverage rate greater than zero indicates that there is undercoverage of the sample. The higher the value of the undercoverage rate the greater the amount of undercoverage, therefore the greater the likelihood of having biased results (even though adjustment for undercoverage is made in benchmarking).

Sample Design

Standard errors for key data items

22 Relative standard errors (RSEs) provide a useful measure of the statistical accuracy of the complex sample design, and indicate the magnitude of error due to sampling. The RSEs, for the purposes of these quality indicators, are to be calculated from the survey data for key data items only and are not to be estimated using a model. A higher RSE indicates a poorer quality estimate, whereas a lower RSE indicates an estimate of higher quality.

Design effects for key data items

23 This quality indicator is the ratio of the achieved RSE for a complex survey design to the RSE for a simple random sample (SRS), for the same sample size. This provides a useful measure of the statistical accuracy of a complex sample design relative to a SRS of the same sample size.

24 The lower the design effect the higher the efficiency of the sample. A stratified sample is generally expected to have a design effect less than one, as stratification increases the efficiency of the sample, compared to a SRS. A cluster sample is expected to have a design effect greater than one, as clustering reduces the efficiency of the sample.

RSE realisation ratio for key data items

25 This is the ratio of the achieved RSE to the design RSE for a key data item. The RSE realisation ratio provides a useful measure of the statistical accuracy of the complex sample design relative to that expected. A RSE realisation value of less than one indicates an estimate of higher quality than planned, but a less efficient sample design; a RSE realisation value equal to one indicates that the efficiency of the sample is equal to that expected; and a RSE realisation ratio greater than one indicates an estimate of lower quality than expected and an inadequate sample size. Realisation ratios can also be used to quantify the extent of control over the sample design. This indicator would feed into any review of the methodology used to estimate achieved sample sizes.

Sample size realisation ratio

26 This is the ratio of the achieved sample size (in terms of fully responding households) to the expected sample size, i.e. the proportion of the expected final sample size that was actually achieved through enumeration and then used to produce estimates.

27 This indicator can be used to monitor the efficiency of the sample design and its deterioration over time, as well as to quantify the extent of control over the sample design. It will feed into the development/monitoring/review of operational budgets, and any review of the methodology used to estimate achieved sample sizes.

28 Ideally the value of the realisation ratio should equal one. A realisation ratio equal to one indicates that the achieved final sample size is equal to that expected, and the quality of the data and the survey costs should meet expectations. A realisation ratio less than one indicates a smaller achieved final sample size than expected, and a realisation ratio greater than one indicates that the achieved final sample size is larger than expected.

Questionnaire

29 Compiling quantitative measures of the quality of a questionnaire is very difficult to do, so the following indicators are at best proxies. All quality indicators relating to a questionnaire must be considered in context for their meaning to be

useful. This may include knowing whether the survey is new or a repeat, the length of development time and the overall testing strategy.

Percentage of modules (topics) cognitive tested

30 A module, or topic, can be thought of as a set of contiguous questions concerned with the collection of a measure, or measures, for a single domain or sub-domain. In practice a module can be one question (e.g. union membership) or much longer (e.g. 36 questions for a self assessment of health).

31 A module is deemed to be cognitive tested if 50% or more of the questions within the module have been subjected to at least one round of cognitive interviews. The cognitive interviews may have been conducted in the current round of development for a particular survey or any other development conducted by the ABS (e.g. another survey, the rolling test program or a previous iteration of the current survey if it is a repeat). For each module there needs to be a history kept of whether that module has been cognitive tested.

32 The main use of this indicator is to complement the "Testing strategies employed" indicator by indicating the extent of cognitive testing the final collection vehicle has been subjected to. If used loosely this indicator could be used as a proxy for respondent comprehension of the questionnaire. This indicator does not reflect the total cognitive testing effort of the survey developers, as the indicator will exclude modules dropped from the final survey.

33 It is important to realise that this indicator does not show whether the problems identified in cognitive interviewing have been remedied, or whether all problems have been identified (though it may be assumed that gross problems have been identified). It provides a broad indicator of the extent of cognitive testing conducted on the final questionnaire. A higher value for this indicator would mean that more modules in the final questionnaire have been subjected to cognitive testing.

Number of rounds and extent of field testing

34 Field tests include any test where Population Survey Operations (PSO) interviewers or ABS permanent officers conduct interviews based on household samples maintained by PSO. This includes skirmishes, pilot tests and dress rehearsals. A round of field testing is any temporally discrete period where PSO interviewers or ABS permanent officers conduct interviews based on PSO maintained household samples.

35 The extent of the field test will be calculated using two characteristics:

- the proportion of the final questionnaire tested; and
- the number of fully responding households involved in the field test.

36 These two characteristics, when multiplied for a given test, will give an indication of how many fully responding households have been through the final questionnaire as a result of the test. For example, a field test of one quarter of the final questionnaire involving 300 fully responding households means that the extent

of the field test was 75 households. The reason the proportion of the final questionnaire is required in this indicator is to ensure that large tests of small components of the questionnaire do not give the illusion of an extensively tested questionnaire.

37 This quality indicator is best used as an explanatory variable to the "Testing strategies employed" indicator to describe the extent of field testing conducted. It can also loosely be used as a proxy for the mechanical accuracy and procedural refinement of the questionnaire. This quality indicator could also be used as an indicator of respondent load during survey development. It should be noted that this indicator describes the amount of testing undertaken but does not provide an indicator of the quality of that testing.

38 The more rounds of field testing and the higher the sample size, the more field testing has been conducted and potentially the more accurate the questionnaire and better refined procedures than when field testing was initiated.

Percentage of modified (repaired) modules

39 A modified or repaired module comes in two forms. The first is where a pre-existing module is modified in order to reduce non-sampling error (e.g. better question wording, better layout, open v. closed response questions) or to meet changed data specifications (e.g. new or altered data item definition, new population flows). "Standard modules" modified for a given collection can also be included in this category.

40 The second form of repaired or modified module are "new" modules being developed. New modules are composed predominantly of questions that have not been asked in previous survey modules together. A module is considered new if 20% or more of questions in the module have not previously been asked together in a survey module with a similar population flow. These modules will very often be refined, repaired or modified during development because of their unique nature.

41 This indicator will show the extent to which the questionnaire was re-engineered, or re-created, during development and thus indicate the impact of development on the final questionnaire. However the interpretation of this indicator is difficult as the final quality indicator can be influenced by: the amount of new modules in the survey; the amount of data item re-specification; whether tested modules make it into the final collection vehicle; and the breadth of testing. As such, this quality indicator would be best seen in the context of other questionnaire quality indicators and not interpreted as a lone figure.

Testing strategies employed

42 This quality indicator will be represented by a simple table showing a list of possible testing strategies and a Yes/No indicator of whether the strategy was employed in the development of the questionnaire. Each testing strategy only needs to be applied to any part of a questionnaire for this quality indicator to record the strategy as being used in the development of the questionnaire. The testing strategies to be included are:

- Field testing
- Split sample comparisons
- Cognitive interviewing
- Respondent debriefing
- Interviewer debriefing
- Focus groups
- Expert panels
- Observational coding
- Follow-up questions
- Record referral
- Other (e.g. useability tests and the use of trace files)

43 This quality indicator will show in broad terms the types of testing conducted on a given questionnaire and therefore act as a proxy for the types of problems testing was most likely to uncover and rectify. The extent to which these testing strategies were employed is not stated in this indicator but can be shown by other quality indicators concerning the questionnaire development.

Number of queries raised during field operations/input processing

44 The purpose of this indicator is to determine the number and type of problems experienced in the implementation of a survey collection strategy. Most problems raised during the field phase of a collection are currently logged on a database by broad area of concern (General, Processing, Coding and the Questionnaire). These query database categories are multi-purpose and are used for queries as well as dissemination of other important information.

45 The simplest interpretation would be that a decrease in this quality indicator would mean an increase in the quality of the questionnaire, field procedures and/or processing. However a query raised does not give any indication of the extent of the problem, and known problems may be dealt with in training and therefore never result in a query though the problem exists.

Interviewing process

Other Responsible Adult (ORA) rate

46 The ORA rate indicates the percentage of the total sample that is made up of responses sourced solely from an "other responsible adult" under the Any Responsible Adult (ARA) methodology. The ARA methodology allows for any responsible adult in the household to answer questions on behalf of other household members. Where responses are partially or completely obtained from the person to whom the questions refer, interviews are recorded as "Person him/herself".

47 This indicator is used to show what proportion of responses are gathered from someone other than the person to whom the questions refer. Measures of the ORA rate reflect the potential size of errors due to proxy reporting. Although cheaper and likely to lead to higher response rates, the quality of data is usually

lower, the higher the ORA rate. However this is not always the case as it depends on the reliability of the proxy responses and the type of information being collected.

Household response rates

48 Households in the active survey sample (i.e. households sampled less sample loss) are each classified into one (and only one) of the following categories:

1) Fully responding household

49 Fully responding households are those households where a 'useable' questionnaire is available for all in-scope members of that household.

2) Partially responding household: refused

3) Partially responding household: non-contact

4) Partially responding household: non-response due to technical difficulties

5) Partially responding household: other non-response

50 Partially responding households are those households where a 'useable' questionnaire is available for one or more in-scope members of that household but is not available for all in-scope members of that household.

6) Fully non-responding household: refused

7) Fully non-responding household: non-contact

8) Fully non-responding household: non-response due to technical difficulties

9) Fully non-responding household: other non-response

51 Fully non-responding households are those households where a 'useable' questionnaire is not available for any member of that household.

52 Rules will be defined to assist in classifying households to the above categories.

53 A 'usable' questionnaire is a questionnaire whose data appears in the publication Main Unit Record File (MURF) and for which no key data item has been imputed or has been derived using imputed data. A key data item is a data item (a particular characteristic of units in a population which is measured or observed) that is essential for the survey to meet its objectives. Refer to paragraph 74 for a definition of the publication MURF.

54 Following the above categorisation of households the following household response rates can be calculated:

Full response rate

55 This is the percentage of households in the active survey sample (i.e. the number of households sampled, less sample loss) that fully responded.

Partial response rate

56 This is the percentage of households in the active survey sample that partially responded.

Total response rate

57 This is the percentage of households in the active survey sample that either fully or partially responded. This is the sum of the full response rate and the partial response rate.

Full non-response rate

58 This is the percentage of households in the active survey sample that fully non-responded. The sum of the total response rate and the full non-response rate add to 100%.

Refusal rate

59 This is the percentage of households in the active survey sample where one or more members of that household refused to provide all the required information.

Non-contact rate

60 This is the percentage of households in the active survey sample who could not provide information because one or more members of that household could not be contacted.

Non-response due to technical difficulties rate

61 This is the percentage of households in the active survey sample for which no responses were received due to technical difficulties, for one or more members of that household, e.g. questionnaires lost/destroyed in transit.

Other non-response rate

62 This is the percentage of households in the active survey sample who could not provide information because of language, sickness, disability, death (within the reference period), etc. for one or more members of that household.

63 Note that the above four response rates (refusal rate, non-contact rate, non-response due to technical difficulties rate, and other non-response rates) sum to the sum of the partial response rate and the full non-response rate.

Processing reject rate

64 This is the percentage of households in the active survey sample for which responses were excluded as a result of missing or inaccurate information identified during post-survey processing, for one or more members of that household.

65 Household response rates can be used by:

- survey designers, to allocate sample to strata having regard to sample loss and non-response, and to evaluate the design of questions.
- field operation managers, to monitor the performance of surveys when they are in the field or after they have been completed, and to evaluate the collection methodology.
- survey processing staff, to assist them when adjusting, weighting or imputing for missing data.
- data users, who may require information to detect trends in response or analyse non-response bias
- users of microdata, who may require information on the survey response for particular analyses, for reweighting, or comparison and combination with other data sources, or assessment of the accuracy of the data.
- data users, to assess the quality of specific variables.

66 Lower response rate values indicate a greater probability for poorer quality estimates, and vice versa. However it should be noted that the quality of the estimates will be dependent on the methodology in place to account for non-response.

Provider (respondent) load

Average household interview time (mins)

67 Household interview time is the time taken to completely enumerate a survey at a household. Average household interview time can be calculated for all interviews, first time interviews, or second time or more interviews. It could also be broken down by personal visits and telephone interviews.

68 This indicator can be used to measure the burden on respondents by measuring the amount of their time required to provide the survey data. Higher values indicate higher loads being placed on households, and may have a detrimental effect on the quality of estimates and the ABS's reputation. This indicator could also be used to produce a measure of how much each response costs to collect in the field.

Average time required by respondents outside of interview by household (mins)

69 This indicator will assess the time taken by respondents to complete self administered questionnaires (SAQs) outside of the time taken for the administration of any interviewer-mediated components of the survey, e.g. the completion of diaries. The information would be collected via a self-report in response to a question in the SAQ. This indicator should only be calculated for fully responding households in order to be on an equivalent basis to the "normal" ABS population survey respondent load estimate of interview time per fully responding household. This would allow the direct addition of average interview time and time taken outside of interview to calculate a measure of respondent load.

70 An increase in this value represents an increase in the respondent load outside of interview for fully responding households. However this may not mean

overall respondent load has increased if interview time has changed. As such this measure should always be interpreted in the context of the average interview time for fully responding households.

Data revision

Changes to derivation specifications

71 This indicator is simply the ratio of the number of changes to derived item specifications (i.e. a detailed list specifying derivations of groups of response items) to the number of derived items initially specified. It should count all:

- changes arising from incorrectly specified derivations.
- additional derivations.
- discarded derivations.

72 This measure will provide an indication of:

- the suitability of the initial set of derivation specifications (thus reflecting the effectiveness of the negotiation and consultation between clients and the ABS).
- the accuracy of the logic used to specify the derivations.

73 A high percentage indicates a large number of changes to specifications relative to the set of derived items initially agreed.

Versions of publication MURFs

74 A publication Main Unit Record File (MURF) is defined as that version of the unit record which all parties mutually agree is suitable for the preparation of the initial publication manuscript.

75 This indicator records the number of times that a publication MURF is derived in order to change the data file. The changes might be required to correct errors found in the data, to add or modify derivations of data items or to repair programming errors. In counting each iteration of the publication MURF, the type and source of the changes must also be recorded in order to make judgements about the quality of the first version of the publication MURF.

76 The number of versions of the publication MURF, coupled with the reasons for the changes, is an indicator of the quality of the initial version of the publication MURF. The count on its own indicates the efficiency of the error detection and correction process.

Versions of public release CURFs

77 A Confidentialised Unit Record File (CURF) is available for public release once it has received approval from the Statistician. This indicator records the number of times that a CURF released to clients is re-derived, and a new version subsequently published. The changes might be required to correct errors found in the data, to add or modify derivations of data items, to repair programming errors or to rectify confidentiality problems. In counting each version of a CURF, the type and

source of the changes applied must also be recorded, in order to make judgements about the quality of the initial file released.

78 The fewer versions of the CURF that are produced the better. High numbers of re-derives indicates that the initial specification of the CURF was inadequate and/or the data it contains was of poor quality.

Number/percentage of data items containing errors identified in the publication MURF

79 Errors in the MURF should be identified and fixed during the input and output processing stages of the survey. Errors identified after the publication MURF has been handed over can cause unnecessary rework of analysis and dissemination activities if the errors need to be fixed and result in a revised publication MURF, or in lesser quality data if they are deemed tolerable and the MURF is not subsequently amended.

80 Lower numbers and percentages of errors are obviously desirable. Identified errors which are not deemed significant enough to amend the publication MURF are of lesser consequence in terms of delays to the publication, but have a greater impact on the quality of the data as they remain in the final data.

Number/percentage of data items containing errors identified in the public release CURF

81 This item is an indicator of the quality of the public release CURF and the quality of the data processing survey process. Where errors are considered important enough to revise and re-release the file, it is also a measure of relative efficiency. Errors in the CURF should be identified and fixed during the processing stage of the survey.

82 Lower numbers and percentages of errors are obviously desirable. Identified errors which are not deemed significant enough to amend and re-release the public release CURF are of lesser consequence in terms of frustration and rework of analysis by clients, but have a greater impact on the quality of the data as they remain in the final data.

Number of errors identified in a publication

83 This quality indicator is the number of errors identified in a publication after release. An error is any identified mistake which has some impact on the accuracy or understanding of the final publication estimates. For each error, some information about the error needs to be collected and stored.

84 This item is an indicator of the quality of the final publication. Errors in the publication can be caused from errors in producing the publication from a correct MURF, or can come from errors in the MURF. Errors in the MURF should be identified and fixed during the input and output processing stages of the survey. Errors in the preparation of the publication should be identified and fixed before release of the publication.

85 Lower numbers of errors are obviously desirable. Significant errors, which warrant the issue of a corrigenda or reissue of the publication are the most serious errors, and can jeopardise the ABS's reputation of producing high quality statistics. Less significant errors may have a lesser impact on the ABS's reputation, but reduce the quality of the ABS published data.

86 The above Data Revision indicators may be used to monitor the quality of ABS's products. A large number of revisions may lead to poor client satisfaction which is likely to have a detrimental effect on the ABS's reputation. These indicators may also indicate poor procedures.

Timeliness

Timeliness of field enumeration

87 This is measured (in three parts) as the number of months to commence field enumeration from:

- 1) when the survey is proposed to the Population Surveys Planning Committee (PSPC)
- 2) survey approval
- 3) provision of final data item specifications

88 This indicator can provide an overall measure of the timeliness of the survey, and can be used to focus efforts towards improving timeliness. It should be noted that this is an indicator of the lead time to get a survey in the field, rather than an indicator of the efficiency of the field operation itself.

Timeliness of the publication MURF

89 This is measured (in three parts) as the number of months to produce a publication MURF from:

- 1) when the survey is proposed to the PSPC
- 2) survey approval
- 3) completion of final enumeration

90 The production of the initial publication MURF is a significant milestone in the survey processing cycle, and enables work to commence on producing tables for the publication. Currently the ABS sets a target of 7 months following enumeration for production of the MURF. A shorter timeframe means that data is likely to be of more relevance to users for decision making purposes.

Timeliness of the first (initial) publication

91 This is measured (in three parts) as the number of months to produce the first publication from:

- 1) when the survey is proposed to the PSPC

- 2) survey approval
- 3) completion of field enumeration

92 In order to meet legislative requirements, and to satisfy the public interest, the ABS publishes results from all its collections as soon as possible after data processing has been completed. Initial publications usually contain descriptive information on methodology, some commentary on the data, tables and information about sampling errors. The target date for the release of Special Supplementary Survey initial publications is 9 months after enumeration. Publications released after this time would be viewed as late, impact on users and should be accompanied by an explanation.

Timeliness of the public release CURF

93 This is measured (in three parts) as the number of months to the date a CURF can be released from:

- 1) when the survey is proposed to the PSPC
- 2) survey approval
- 3) completion of field enumeration

94 The production of the final CURF is the last significant milestone in the survey processing cycle, and is now a high priority for the ABS. Currently an 11 month target date following enumeration has been established for the production of the CURF. A shorter timeframe means that data is likely to be of more relevance to users for decision making purposes.

Quality indicators for external use

95 As mentioned earlier, many of the quality indicators will be used internally by the ABS to monitor data quality and establish areas for improvement within surveys. However some indicators would be useful for consumption by external users in order to assess the quality of ABS survey data. The method of disseminating this information is still being explored, but feedback on the type of indicators useful to external users would be appreciated.

96 The following sub-set of Phase 1 quality indicators is being proposed for external use:

- Undercoverage rate
- Standard errors for key data items
- Percentage of modules (topics) cognitive tested
- Number of rounds and extent of field testing
- Percentage of modified (repaired) modules
- Testing strategies employed
- ORA rate
- Household response rates (i.e. fully responded and partially responded rates)
- Timeliness of the first (initial) publication
- Timeliness of the public release CURF

Discussion points for MAC

97 The authors would like feedback and comments from the MAC on the following aspects:

- The comprehensiveness and relevance of the quality indicators.
- Whether the set of quality indicators for external use is adequate.
- The overall scope and usefulness of the quality indicators project.
- The methodologies for the Phase 1 quality indicators.