



DEPARTMENT OF  
**FAMILY AND  
COMMUNITY  
SERVICES**



THE UNIVERSITY OF  
**MELBOURNE**

## **HILDA PROJECT TECHNICAL PAPER SERIES**

**No. 4/02, October 2002**

---

### **Assessing the Quality of the HILDA Survey Wave 1 Data**

*Nicole Watson and Mark Wooden*

---

**The HILDA Project was initiated, and is funded, by the Commonwealth  
Department of Family and Community Services**



## Contents

<b>INTRODUCTION.....</b>	<b>2</b>
<b>SAMPLE REPRESENTATIVENESS .....</b>	<b>3</b>
<b>MISSING DATA.....</b>	<b>9</b>
ITEM NON-RESPONSE .....	9
<i>Missing Data on Income</i> .....	9
<i>Questionnaire Design Flaw: Persons Living with Both Parents</i> .....	12
<i>Skip Sequence Error: Permanently Unable to Work</i> .....	13
INCOMPLETE HOUSEHOLDS .....	14
<b>ACCURACY OF THE DATA .....</b>	<b>15</b>
QUESTIONNAIRE DESIGN ISSUES .....	15
<i>Child Care Costs</i> .....	15
<i>Current Wage and Salary Income Reported for People Without Jobs</i> .....	15
<i>Calendar</i> .....	16
<i>Marital Status</i> .....	16
<i>Time Use</i> .....	17
<i>Leave Entitlements</i> .....	17
DATA COLLECTION ISSUES .....	18
<i>Sex</i> .....	18
<i>Date of Birth</i> .....	18
<i>Working Hours</i> .....	19
<i>Income From Incorporated and Unincorporated Businesses</i> .....	20
<i>Interviewer Observations</i> .....	20
<i>Mode Effects and Social Desirability / Acquiescence Bias</i> .....	21
CROSS-FORM COMPARISONS: CHRONIC HEALTH CONDITION, LABOUR FORCE STATUS AND MARITAL STATUS .....	22
DATA CODING AND ENTRY .....	23
DATA MATCHING.....	23
COMPARISONS WITH ABS DATA .....	24
<b>SUMMARY .....</b>	<b>27</b>
<b>REFERENCES.....</b>	<b>28</b>
<b>APPENDIX 1: INCIDENCE OF MISSING CASES.....</b>	<b>29</b>

## **Introduction**

The quality of data generated by any survey is a function of three main elements. These are:

- (i) the representativeness of the sample;
- (ii) the completeness of responses (i.e., the extent to which there is missing data, or item-non-response); and
- (iii) the accuracy of the data (which in turn is a function of the questionnaire design, the accuracy of the responses provided, and the accuracy with which responses are recorded, coded and entered into the database).

In addition, for the HILDA Survey, added concerns arise as a result of the need to match data, both across individuals within the same households and within individuals. This is a consequence of both the presence of multiple instruments and the desire to construct household-level variables.

This paper provides an assessment of the data obtained for the first wave of the HILDA Survey (conducted in 2001) against each of these criteria.

## Sample Representativeness

As discussed in Watson and Wooden (2002), the sample selected for the first wave of the HILDA Survey was intended, with some exceptions (most notably persons living in remote parts of Australia), to be representative of all usual residents of private dwellings in Australia. However, like all voluntary surveys, the fact that not all selected sample members will agree to participate raises the possibility that the achieved sample will have characteristics that diverge from the population it purports to represent.<sup>1</sup> The survey estimates may not, therefore, provide an unbiased description of the population from which the sample is drawn. Furthermore, attempts to use these data to estimate behavioural responses may need to take into account the self-selected nature of the sample in order to ensure parameter estimates are not biased.

It is thus important to establish in what ways the responding sample differs from the population. We can assess this by first comparing the responding sample with the selected sample. This is possible since all selected dwellings were fully enumerated prior to interviews commencing. The range of information we have about the full sample, however, is extremely narrow, limited to just location, type of dwelling structure, the security features of the dwelling, an interviewer assessment of the external condition of the dwelling, and the visibility of high-rise buildings in the immediate vicinity. A comparison between the selected and responding samples on each of these five criteria is thus provided in Table 1. Note, however, that comparisons between the selected and responding samples are not strictly valid given the former relates to dwellings whereas the latter relates to households. For most purposes, however, this distinction can be safely ignored, with dwellings with more than one household representing less than three per cent of the sample of in-scope dwellings containing respondents.

Table 1 indicates that non-response is not random. Most obviously, there is a clear difference in the geographic distribution of the responding sample compared with the selected sample, with residents in Sydney under-responding compared with residents elsewhere in Australia. This is not particularly surprising, and is a common feature of other voluntary surveys in Australia (Bednall, Cavenett and Shaw 2000). The usual explanation provided is that people in Sydney are under more time pressures and hence are far more reluctant to spend part of their time participating in surveys. In addition, we also suspect that part of the explanation lies in greater difficulties making contact with the occupants of selected dwellings in Sydney because of a relatively high incidence of dwellings with locked gates, gatekeepers and other devices intended to screen access. Table 1 provides some support for this hypothesis, with response rates lower in dwellings with locked gates, though this lower rate of response is only significant if such dwellings also have intercom access.

With respect to the other dwelling characteristics considered – dwelling structure, condition of dwelling and the visibility of high-rise buildings – there are no significant differences between the selected and responding sample. In summary, Table 1 suggests that the only issue of concern is the geographic distribution of the responding sample – analyses of the HILDA Survey will thus need to take into account the under-representation of households in Sydney.

---

<sup>1</sup> Interviews were completed with at least one household member at 66% of all selected households in wave 1 of the HILDA Survey (Watson and Wooden 2002).

**Table 1: Characteristics of Wave 1 HILDA Survey Samples<sup>a</sup>**

	<i>HILDA selected sample (dwellings)</i>	<i>HILDA responding sample (households)</i>	<i>Prob. of zero diff. b/w selected &amp; responding samples</i>	<i>2001 Census (dwellings)</i>	<i>Prob. of zero diff. b/w Census &amp; selected sample</i>
Area of usual residence					
Sydney	21.0	16.9	<0.01	20.3	ns
Rest of NSW	13.4	14.6	<0.05	12.8	ns
Melbourne	17.6	16.7	ns	17.6	ns
Rest of Victoria	6.7	7.5	<0.05	6.9	ns
Brisbane	8.8	8.8	ns	8.5	ns
Rest of Queensland	10.4	11.8	ns	10.6	ns
Adelaide	6.1	6.1	ns	6.1	ns
Rest of South Australia	1.9	2.3	<0.05	2.2	<0.05
Perth	7.3	7.7	ns	7.2	ns
Rest of Western Australia	2.4	2.8	ns	2.6	ns
Tasmania	2.6	2.8	ns	2.6	ns
Northern Territory	0.6	0.5	ns	0.9	ns
ACT	1.4	1.5	ns	1.6	<0.01
Dwelling structure <sup>b</sup>					
Separate house	75.9	77.3	ns	75.3	ns
Semi-detached	9.7	10.0	ns	8.9	ns
Flat	13.3	11.7	ns	13.1	ns
Other	0.9	0.9	ns	1.9	<0.01
Condition of dwelling					
Very good / excellent	33.7	33.8	ns		
Good	35.6	36.2	ns		
Average	25.2	25.0	ns		
Poor	4.9	4.6	ns		
Very poor / almost derelict	0.5	0.4	ns		
Security features of dwelling					
Locked gate (w/o intercom)	3.3	2.8	ns		
Locked door / gate (w i/c)	5.6	4.1	<0.05		
Security guard, doorman, etc.	1.3	1.1	ns		
Bars on windows	4.9	4.9	ns		
Security door	37.0	37.4	ns		
No trespassing sign	0.5	0.5	ns		
Beware of dog sign	2.2	2.2	ns		
Evidence of a dangerous dog	2.5	2.3	ns		
No junk mail sign	3.5	3.3	ns		
Neighbourhood watch sign	3.7	3.6	ns		
High-rise buildings in area					
A lot – more than 50%	1.7	1.0	ns		
A fair bit – 20-50%	1.0	0.8	ns		
One or two	2.4	2.1	ns		
None at all	94.9	96.1	ns		

Notes: a HILDA Survey responses have been adjusted to account for variation in the probability of selection.  
b Excludes a small proportion of cases where dwelling structure type not able to be classified.  
ns Denotes difference not significantly different at the 95% confidence level.

As a check on the quality of the selection process we also report, in Table 1, figures from the 2001 Census (which was conducted just prior to commencement of wave 1 of the HILDA Survey) on the distribution of private dwellings in Australia by both location and structure type. As can be seen, the selected sample for the HILDA Survey appears to conform closely to the Census data and thus suggests that, as was intended, the initial sample was not biased in any significant way.

What about the characteristics of the individual members of our responding sample? Obviously we cannot know what the characteristics of the members of non-responding households are. Any assessment of the representativeness of the sample of individual household members will thus have to be based on a comparison with some benchmark sample. The most obvious choice here is the Monthly Population Survey (MPS), conducted by the Australian Bureau of Statistics (ABS). The sample for the MPS is large (with the sampled dwellings covering about 0.5 per cent of the Australian population), the scope of coverage is broad, and response rates are acknowledged to be high.

In Table 2, therefore, we compare the distribution of selected characteristics of the sample of individual respondents to wave 1 of the HILDA Survey with those from a comparable month of the MPS – typically October 2001.<sup>2</sup> Before proceeding, it should be noted that the comparisons are not completely straightforward. Most importantly, the coverage of the MPS is broader than that of the HILDA Survey and includes persons living in both institutions and in remote areas.

Table 2 indicates that the HILDA sample is noticeably different from the broader population in a number of ways. First, and as expected given the distribution of responding households, Sydney residents are under-represented in the HILDA individual sample. The first column in Table 2 indicates that according to the MPS for October 2001, persons living in Sydney, comprised 21.5 per cent of the Australian population aged 15 years and over. In contrast Sydney residents make up only 16.9 per cent of the sample of persons completing a HILDA Person Questionnaire (PQ). Note that almost one quarter of the lower response rate in Sydney is due to a relatively high incidence of non-response *within* households. This can be seen by comparing the figures reported in column 2 with those reported in column 3. Sydney residents actually comprise 18 per cent of all HILDA sample members (as determined from the Household Form), which is more than a full percentage point higher than the proportion that completed a PQ.

Second, the HILDA sample has an under-representation of men and an over-representation of women, which is not uncommon in voluntary surveys. A large proportion of this differential (over 60 per cent), however, is explained by male household members not participating in the PQ interview.

Third, married persons are over-represented (and unmarried persons under-represented). In part, this was to be expected given the population for HILDA excludes persons living in institutions. It was also expected that it would be more difficult to make contact with persons living alone. Nevertheless, the size of the differential is much larger than expected (though possibly might reflect differences in the way *de facto* partnerships are defined and treated in the HILDA Survey compared with the MPS).

---

<sup>2</sup> Fifty per cent of all of the personal (PQ) interviews in wave 1 were completed by early October 2001.

**Table 2: Selected Wave 1 Individual Sample Characteristics Compared with Population Estimates from the ABS Monthly Population Survey (persons aged 15 years or over)**

	<i>ABS<sup>a</sup></i>	<i>HILDA PQ respondents<sup>b</sup></i>	<i>HILDA: All hh members<sup>b</sup></i>
Area of usual residence			
Sydney	21.5	16.9**	18.0**
Rest of NSW	12.2	14.5**	14.1**
Melbourne	18.4	17.3*	17.5
Rest of Victoria	6.7	7.5	7.3
Brisbane	8.6	8.8	8.8
Rest of Queensland	10.0	11.5*	11.4
Adelaide	5.8	6.1	5.8
Rest of South Australia	2.0	2.4*	2.3
Perth	7.3	7.5	7.4
Rest of Western Australia	2.5	2.8	2.7
Tasmania	2.4	2.8	2.7
Northern Territory	0.9	0.5**	0.5**
ACT	1.6	1.6	1.6
Sex			
Male	49.3	47.4**	48.6*
Female	50.7	52.6**	51.4*
Age (years) at 30 Sept 2001			
15-19	8.8	8.7	9.4*
20-24	8.9	7.4**	7.8**
25-34	18.7	18.7	18.7
35-44	19.0	21.7**	21.4**
45-54	17.1	17.1	17.1
55-64	11.8	12.0	11.7
65 or over	15.6	14.4*	13.9**
Marital status			
Married (including de facto)	58.7	63.4*	62.7**
Not married	41.3	36.6**	37.3**
Relationship in household			
Family member			
Husband, wife or partner			
with child <15	23.1	26.6**	26.2**
without child <15	36.7	36.3	36.0
with dependants	28.0	30.3**	29.9**
without dependants	32.0	32.7	32.3
Lone parent			
with child <15	3.2	3.6*	3.4
w dep students but w/o children <15	0.6	0.6	0.6
without dependants	1.6	1.4	1.5
Dependent student	6.8	6.2*	5.7**
Non-dependent child	8.4	6.2**	8.1
Other family member	2.3	2.1	2.4
Non-family member			
Lone person	11.9	14.0**	12.9*
Not living alone	5.3	3.0**	3.3**

**Table 2 (cont'd)**

	<i>ABS<sup>a</sup></i>	<i>HILDA: PQ respondents<sup>b</sup></i>	<i>HILDA: All household members<sup>b</sup></i>
Indigenous status			
Indigenous	1.7	1.8	
Non-indigenous	98.3	98.2	
Birthplace			
Born in Australia	72.4	74.4**	
Born outside Australia			
Main English-speaking country	10.2	10.9	
Other country	17.5	14.7**	
Labour force status <sup>c</sup>			
Employed			
Full-time	42.1	41.7	
Part-time	17.4	19.5**	
Unemployed	4.3	4.4	
Not in the labour force	36.3	34.5*	
Employment status in main job (employed persons only)			
Employee	86.0	87.0	
Employer	3.6	3.9	
Own account worker	10.0	8.4**	
Contributing family worker	0.4	0.8**	

Notes: \*\* and \* denotes significantly different from the ABS population estimate at the 99% and 95% confidence levels respectively.

a With the exception of indigenous status and employment status, the ABS estimates come from the Monthly Population Survey for October 2001. In the case of the two exceptions, data for August 2001 are used. With the exception of country of birth and relationship in household, the population that these estimates apply to is all civilians aged 15 years and over. The figures for country of birth and relationship in household exclude persons living in an institution.

b The HILDA estimates are also for people aged 15 years and over, but include defence force personnel and exclude people living in remote areas of Australia and those living in special dwellings. They have also been adjusted to account for variation in the probability of selection.

c We vary from the usual ABS definition in defining full-time work solely on the basis of usual hours worked (rather than on a combination of usual hours and actual hours worked).

Source: ABS data are from *The Labour Force, Australia* (cat. no. 6203.0), August 2001 and October 2001 issues.

Fourth, dependent students, non-dependent children and non-family members not living alone appear to be under-represented (and couples with children under 15 and lone persons are over-represented). That said, users of these data need to be aware that the questions used in the HILDA Survey to determine the household relationships are quite different to the ABS Census questions. One strength of the HILDA Survey was that the relationship of every person in the household to every other person was collected, whereas the ABS Census collects the relationships between a reference person and every other person. In cases where a person is related to a member of the household but not to the reference person, the approach taken in the HILDA Survey will lead to more people being identified as family



members and fewer being identified as non-family members. However, it is unlikely that these differences in the questions could explain the 3.1 percentage point gap between the ABS Census and the HILDA Survey for couples with children under 15 years of age. At least part of the observed difference in the proportion of dependent students and non-dependent children can be explained by the amount of information available to classify these adult children. The dependent student status for persons that did not provide an individual interview had to be approximated from the employment question on the HF. This may have caused some people to be classified as a non-dependent child when they were actually a dependent student. However, it is clear that both dependent students and non-dependent children were less likely to respond to the HILDA Survey. Finally, adults living alone were more likely to respond, possibly because the modest cash incentive, which was household-based rather than individual-based, was more likely to have a positive effect on small households.

Fifth, immigrants from a non-English-speaking background comprise only 14.7 per cent of the HILDA sample, which compares with a population estimate of 17.5 per cent. We suspect this difference reflects the greater difficulties communicating with persons for whom English was not their first language. In addition, there may be a greater suspicion of government-backed surveys among this sub-population.

Of the other characteristics considered, the differences from ABS benchmarks are generally small and often insignificant. The age composition of the HILDA individual sample, for example, is quite close to the MPS, even though the HILDA sample excludes persons living in institutions, which will tend to mean a lesser representation of older people. That said, it is true that people in their earlier 20s appear to be under-represented. The breakdown by labour force status is also similar, though persons working part-time are over-represented while persons outside the labour force status are under-represented. We suggest that these differences might be explained by both the over-representation of married women and by the exclusion of persons living in institutions. Finally, when comparing our sample of employed persons with that reported in the August 2001 MPS, we again find only small differences. Specifically, there does appear to be an under-representation of own account workers (that is, the self-employed) in the HILDA sample.

Overall, while there are clear sources of bias in the HILDA data, we do not believe that the size of these discrepancies is so large as to discredit the data. Furthermore, the sources of bias which appear to be of greatest importance – differences in rates of response across both sex and location – are relatively easily corrected through the application of population weights.

## Missing Data

### Item Non-response

Another potential source of response bias is item non-response. That is, while a member of a selected household may agree to an interview, they may then subsequently either refuse, or be unable, to answer some of the questions asked. Frequency counts from the interview data (that is, data collected as part of the Household Form, Household Questionnaire or Person Questionnaire), however, indicate that missing data is generally not a large problem.<sup>3</sup> Indeed, for the majority of variables there were no missing cases at all, and where there were missing cases, the incidence was generally relatively low – less than 2 per cent.

This was not true of the Self-completion Questionnaire (SCQ), with item non-response rates averaging 2.4 per cent per item in this instrument. This higher rate of non-response is to be expected given the self-completion nature of this instrument.<sup>4</sup>

A summary of the questions where missing cases may be a problem is provided in Appendix 1. As suggested by the comments provided in that Appendix, there are often good reasons why a relatively high rate of missing data on some specific questions is expected (e.g., the question requires extensive recall or knowledge the respondent may not have). Indeed, in some cases explicit provision is made for ‘don’t know’ responses. Further, for a small minority of questions we are unable to clearly separate respondents who are unable to answer from those for whom the question does not apply (though more detailed analysis of respondent characteristics may assist here).

Appendix 1, however, does highlight at least two potentially serious problems. First, many of the questions requiring respondents to provide a monetary value are associated with relatively high rates of non-response. Second, a questionnaire design flaw has resulted in considerable missing data in the section on family background.

### *Missing Data on Income*

The high incidence of missing data for monetary items suggests either a strong aversion to income and wealth-related questions on the part of sample members or a low level of knowledge about such matters. Such findings are not unexpected, with recent analyses of item non-response data from the German Socio-Economic Panel (GSOEP) almost exclusively focusing on the impact on income and wealth information (Ripahn and Serfing 2002, Schrapler 2002). Indeed, item non-response in the HILDA Survey compares favourably with that reported in the first wave of the GSOEP. According to Schrapler (2002, p. 8), almost 13 per cent of all employed persons failed to report a figure for current (last month’s) gross earnings in the first wave of the GSOEP.<sup>5</sup> By comparison, the most comparable figure in the

---

<sup>3</sup> Missing data is defined here to include both refusals to provide an answer and the inability to provide an answer (i.e., a response of ‘don’t know’). In the database, however, these two different types of response are clearly distinguished.

<sup>4</sup> Note that many of the questions in the SCQ are intended to be combined to form scales. In such cases, the impact of missing data is typically negated by using the average score on completed items in the same scale. In the scoring system used for the SF-36, a general health diagnostic measure administered as part of the HILDA Survey, for example, the respondents average score on completed items within a sub-scale is used provided the respondent answered at least 50 per cent of the items in the scale (Ware et al. 1993, p. 6.16).

<sup>5</sup> Item non-response does fall with time, both because persons who do not answer all questions are more likely to attrit in subsequent waves, and because sensitivity declines with time in the panel.

HILDA Survey is just 7.3 per cent. Moreover, in the HILDA Survey persons responding that they did not know their gross earnings were asked if they knew their net earnings, from which gross earnings could then be imputed with a reasonably high degree of accuracy. This reduces the effective item non-response rate on current wage and salary income to just 3.3 per cent.

Nevertheless, it is still true that the income section is affected by relatively high rates of item non-response. This is particularly true of financial year income data, and especially business income. As reported in Table 3, we are unable to calculate a financial year business income total for 23.5 per cent of persons earning such income. Overall, there are just over 2100 cases (almost 15 per cent of the individual sample) where calculating total income is either not possible or will require having to impute at least one of the income components.

**Table 3: Sources of Missing Financial Year Income Data**

<i>Income source</i>	<i>Refused (no.)</i>	<i>Don't know (no.)</i>	<i>Valid cases</i>	<i>% missing</i>
Wages and salaries	21	568	8179	7.2
Government benefits and pensions	12	53	4621	1.4
Business	13	499	2182	23.5
Investments	25	1113	13969	8.1
Other	4	106	13969	0.8
TOTAL INCOME			13969	14.7

None of this matters, however, if item non-response is random or at least uncorrelated with the variables of interest for analysis. As a crude test of this, we estimated a logit model of the likelihood of individual respondents not providing all the necessary information needed to construct a measure of financial year income. The results are presented in Table 4, and three conclusions are immediately apparent. First, the overall model estimated is significant, indicating that response is not random. Second, while there are systematic associations, the overall predictive power of the model is still quite poor, suggesting that the problem of bias generated by non-response on income questions is not likely to be serious. Third, relatively few variables exhibit statistically significant associations with response / non-response. That said, a number of these variables are likely to be correlated with income and hence a correlation with non-response is problematic for analyses of income. For example, we find evidence that response to the income questions is correlated with sex, with females being less likely to provide complete income data. We also find a strong correlation with labour force status. In particular, persons outside the labour force, who can be expected to have relatively low incomes, were more likely to provide complete income information, possibly because they have less information to report on. Also statistically significant were:

- location – with the likelihood of income data being provided greatest for persons residing in Tasmania and Perth and relatively low for people living in Sydney and in non-metropolitan Western Australia;

**Table 4: Predicting Presence of Financial Year Income Data – Logit Results**

<i>Variables</i>	<i>Coeff.</i>	<i>Standard error</i>	<i>Odds ratio</i>	<i>P&gt;  Z </i>
Constant	3.861	0.483		<0.001
Sex (female=1)	-0.199	0.058	0.820	0.001
Age	-0.003	0.009	0.997	0.708
Age squared (x 100)	-0.013	0.010	1.000	0.166
Marital status [base = Married]				
De facto	0.160	0.095	1.173	0.090
Separated	0.192	0.147	1.212	0.192
Divorced	0.479	0.122	1.614	<0.001
Widowed	0.236	0.131	1.267	0.072
Never married	0.041	0.086	1.042	0.634
Country of birth [base = Australia]				
Overseas: Main English-speaking country	0.063	0.081	1.065	0.436
Overseas: Other country	0.192	0.105	1.212	0.066
Speaks language other than English	-0.159	0.111	0.853	0.151
Aboriginal / Torres Strait Islander	0.431	0.226	1.540	0.056
Place of residence [base = Melbourne]				
Sydney	-0.223	0.081	0.800	0.006
Rest NSW	-0.126	0.089	0.881	0.156
Rest Victoria	-0.056	0.107	0.945	0.599
Brisbane	-0.147	0.101	0.863	0.149
Rest Qld	-0.104	0.097	0.902	0.283
Adelaide	0.124	0.120	1.132	0.302
Rest SA	-0.146	0.149	0.864	0.326
Perth	0.456	0.127	1.578	<0.001
Rest WA	-0.379	0.148	0.684	0.010
Tasmania	0.632	0.199	1.881	0.001
Northern Territory	-0.044	0.351	0.957	0.900
ACT	-0.138	0.194	0.871	0.477
Education attainment [base = degree]				
Postgraduate degree	-0.006	0.170	0.994	0.972
Graduate diploma	-0.061	0.135	0.941	0.653
Diploma	-0.135	0.110	0.874	0.223
Certificate – level III / IV	-0.216	0.100	0.805	0.032
Certificate – level I / II	0.050	0.145	1.052	0.731
Certificate – level not defined	-0.191	0.140	0.826	0.171
Completed Year 12	-0.069	0.113	0.933	0.542
Completed Year 11 or less	-0.092	0.101	0.912	0.360
Education level unknown	-0.377	0.164	0.686	0.022
Labour force status [base = employed FT]				
Employed part-time	-0.140	0.069	0.869	0.040
Unemployed	-0.209	0.126	0.812	0.099
Not in the labour force	0.684	0.082	1.982	<0.001
Occupational status	-0.002	0.001	0.998	0.121
No occupation status	0.348	0.184	1.417	0.059
% employed since left education	0.027	0.103	1.028	0.789
% unemployed since left education	-0.034	0.268	0.967	0.899

**Table 4 (cont'd)**

<i>Variables</i>	<i>Coeff.</i>	<i>Standard error</i>	<i>Odds ratio</i>	<i>P&gt;  Z </i>
Importance of:				
Home	-0.023	0.015	0.977	0.126
Financial situation	-0.049	0.016	0.953	0.002
Involvement in community	0.001	0.012	1.001	0.906
Health	-0.012	0.020	0.988	0.570
Family	0.001	0.024	1.001	0.952
Leisure time	0.031	0.014	1.032	0.026
Religion	-0.013	0.008	0.987	0.086
Interview situation				
Understanding	-0.221	0.056	0.802	<0.001
Poor eyesight	-0.350	0.259	0.705	0.176
Hearing problems	-0.087	0.196	0.912	0.659
Reading difficulties	0.128	0.243	1.137	0.598
English as second language	0.069	0.161	1.071	0.671
Other language problems	0.313	0.444	1.368	0.480
Telephone interview	0.439	0.383	1.550	0.253
Assisted interview	0.720	0.183	2.054	<0.001
Others present, not assisted	0.176	0.055	1.192	0.001
Likelihood ratio – chi-squared (56)	387.6			
Pseudo R-squared	0.034			
N	13876			

- whether an individual is divorced or not – divorced people appear more likely to provide income data, perhaps because the divorce process increases awareness of income streams;
- the importance attached to one's financial situation – with those rating finances as important far less likely to respond, possibly because they are more likely to regard income questions as sensitive; and
- the importance attached to leisure time – with the direction of association the reverse of that for the financial situation (that is, those rating leisure time as important are more likely to respond).

Finally, we also included controls for the interview situation. Not surprisingly, respondents assessed as having a better understanding of the questions were more responsive (though the causality here could easily run in the reverse direction). We also found evidence that the presence of other household members during the interview tended to enhance respondent cooperation.

#### *Questionnaire Design Flaw: Persons Living with Both Parents*

Appendix 1 also indicates rates of missing data of between 8 and 13 per cent for questions about parents and family background. During testing such questions were not found to work well for young respondents who were still living at home with their parents. The decision was thus taken to sequence respondents in this category past many of the questions in this section

on the assumption that the appropriate responses can be derived directly from the parents. This, however, proved to be a mistake for at least two reasons. First, derivation of responses requires parents in the household to have responded, and of course this was not always the case. Second, the questions on parental employment status were based on what the parents were doing when the respondent was aged 14, which in many cases cannot be derived from the responses of the respondents' parents.

In total, 1231 respondents skipped this sequence of questions because they were still living at home with both their parents (which represents almost 9 per cent of the total individual sample). However, as documented in Table 5, for most of these cases we were still able to derive the missing information on the basis of the parents' responses. Consider, for example country of birth of the father. By using the information provided by the parent in the household we were able to reduce the number of missing cases from 1314 to just 220.

**Table 5: Response Summary, Questions B7 to B16, PQ**

	<i>Answer provided by respondent</i>	<i>Don't know / refusal by respondent</i>	<i>Answer derived from parents' responses</i>	<i>Not derivable from parents</i>	<i>Valid N</i>
Presence of siblings	12722	16	1166	65	13969
No. of siblings	12110	8	1128	103	13349
Whether oldest when growing up	12115	3	1128	103	13349
Country of birth of father	12655	83	1094	137	13969
Country of birth of mother	12697	41	1163	68	13969
Father's employment status when R = 14 years	12631	244	868	226	13969
Father's occupation	12029	215	816	363	13417
When growing up, whether father unemployed for 6 months + Mother's employment status when R = 14 years	11833	681	902	192	13608
Mother's occupation	12629	177	764	399	13969
	8858	472	517	467	10314

#### *Skip Sequence Error: Permanently Unable to Work*

Another source of missing data arose from the way many interviewers treated respondents identified as permanently unable to work. Following the practice adopted by the ABS in the Labour Force Survey, in determining whether a respondent was working in a paid job, we provided a separate code for persons indicating that they were permanently unable to work. Such respondents would then be sequenced past all the questions concerning both current employment (Section E) and job search activity and work preferences (Section F). It was not intended that respondents would be directly asked whether they were unable to work. This option was only to be coded if the information was volunteered. There were a total of 307 persons (2.2 per cent of the sample) who were coded as permanently unable to work here.

While these persons were not to be asked any questions about current employment, they were asked questions about previous employment. Following this series of questions, the

interviewer was then required to check the answer to the question on current work status before deciding whether to sequence the respondent to the section for persons not in the labour force. Unfortunately the question immediately preceding this check concerned the reason respondents ceased their last job and included the category sickness, disability or injury. It appears some interviewers took such a response to imply a permanent inability to work, and hence a great many more persons were sequenced past the section for persons not in the labour force than should have been. Indeed, the total number of persons sequenced past Section F because they were permanently unable to work was 759, more than double the number that should have been sequenced in this fashion. None of these individuals were thus asked any questions about why they were not working, whether they were looking for work and whether they still wanted to work. In constructing labour force status, we have thus been forced to assume that all of these individuals are not in the labour force.

Overall, it is our view that the effect of this mistake on the part of some interviewers has had little impact on the data. We strongly suspect that the large majority of these individuals were no longer looking for work nor interested in work and hence most of the responses in Section F can be easily imputed. However, this is not true of the questions on retirement age. As a result these questions are being asked again in Wave 2.

### **Incomplete Households**

A further source of missing data is incomplete households. As noted earlier, and as discussed at greater length in Watson and Wooden (2002), not all eligible adult members of cooperating households agreed, or were able, to be interviewed. There are 810 incomplete households, which represent 10.5 per cent of the household sample. This means the problem of missing data will be magnified when dealing with variables constructed by combining the different responses of household members. The most obvious example here is household income. While we are able to derive a gross financial year income estimate for 85 per cent of all individual sample members without the need for any imputation, an estimate of total household gross income can only be derived in 71 per cent of cases.<sup>6</sup> As reported in Table 6, the main source of missing household income data is incomplete responses by respondents within complete households. This accounts for around 64 per cent of the missing household income information. The remaining 36 per cent is mainly due to incomplete households.

**Table 6: The Sources of Missing Household Income Data**  
(figures in brackets are percentages)

<i>Did all household members respond?</i>	<i>Did all individual respondents provide complete income data?</i>	
	<i>No</i>	<i>Yes</i>
<i>No</i>	207 (2.7)	603 (7.8)
<i>Yes</i>	1430 (18.6)	5442 (70.8)

<sup>6</sup> This figure allows for respondents to report either gross or net wage and salary income.

## Accuracy of the Data

As discussed in Watson and Wooden (2002, p. 14), considerable effort was devoted to monitoring the quality of data being collected during the fieldwork phase. This included requiring interviewers to check their interview forms for consistency and completeness. In addition, supervisors regularly scrutinised the completed workloads submitted by interviewers, and where problems were identified, the work was returned to the interviewer for checking and, where necessary, verification with the household.

Of course, even the most rigorous monitoring exercise will not preclude errors being made, both by the interviewer and by the respondent. Ideally, identification of such errors would involve either some form of post-survey validation study or the matching of the data to records (e.g., annual tax returns). Only a limited validation study was undertaken, involving just over 10 per cent of all questionnaires and only a limited selection of questions. Indeed, the main purpose of this validation exercise was not to check data quality but to verify that interviewers were attending households. Ultimately, a detailed validation study was deemed not possible given privacy concerns, the added burden on respondents and the additional cost involved. Similarly, privacy concerns, together with the likely impact on rates of response, ruled out any attempt to seek respondent consent to match their responses to other administrative data sets.<sup>7</sup> The best we can do in assessing data accuracy, therefore, is to report on some crude and partial indicators of data quality for specific data items of concern, as well as rely on comparisons with estimates from other well established cross-section surveys.

## Questionnaire Design Issues

Several of the data problems identified in the data emanated from the questionnaire design phase. While there was a reasonable amount of testing undertaken prior to Wave 1, some problems with the questionnaire were not identified and corrected at this stage. Nevertheless, the interviewers rated the respondents understanding of the questions mostly being either excellent or good. In less than 6 per cent of cases was the assessment less favourable (usually only 'fair'). The most problematic questions are discussed below.

### *Child Care Costs*

The child care grids in the HQ are very complex. The respondent were required to separate out the hours and child care costs by type of child and by type of care. It is suspected that in some cases, they were not able to do this very well. There was evidence in at least one case where the total cost of a particular type of child care was reported for both the pre-school age children and the school aged children, and it was apparent that the total cost should have been split between these children.

### *Current Wage and Salary Income Reported for People Without Jobs*

There are 13 respondents that reported having current wages and salaries, but were not classified as being currently employed. It is suspected that the income reported is more likely to be business income (for example, spouses who have income from a family business, but do not actually work in the business). It should also be noted that there are 16 respondents recorded as being an employee but reported not receiving wages and salaries. A further 126

---

<sup>7</sup> Indeed, in some cases, respondent consent would not be sufficient. Obtaining access to tax records, for example, would also require legislative amendments.



respondents are recorded as an employee of their own business but reported not receiving wages and salaries, and 414 respondents are recorded as being an employer while still receiving wages and salaries and having only one job. It is possible that these respondents are reporting business income as wages and salary income. Users should treat these cases carefully if their primary focus is on the components of income rather than total income.

### *Calendar*

Inspection of the information collected through the calendar in Section FG reveals several problems in recording the spells of study, employment, and unemployment over the 14 to 18 month calendar period. These recording problems are in addition to the recall problems that some respondents would have undoubtedly had given the length of the period, especially when many spells are involved.

Two problems with the calendar information have been identified:

- Missing job numbers. The interviewers were meant to record the job number for each job held during the calendar period. This would have enabled the identification of jobs that change between being part-time and full-time, as these jobs should have been recorded on two lines with the same job number. However, the interviewers often did not record the job number and, even when it was recorded, this information was (mistakenly) not entered into the processing system. There are 387 cases where a switch between part-time and full-time work occurred, and only a small proportion of these are expected to involve changes in the same job. In light of these difficulties, the distinction between full-time and part-time employment has been dropped from the calendar for Wave 2.
- Inconsistent recording of jobs. The interviewers did not have sufficient instructions on how to treat breaks in employment (such as long-term leave, or infrequent hours). There were 258 cases where jobs were recorded as a broken line, and it is believed that in most cases, these apparently different spells of employment were with the same employer, and typically were the same job.

We also suspect that it may prove difficult to accurately match jobs between waves as there are very few characteristics about the job that can be used to identify it. A seam of two to five months worth of data was collected in wave 1 to assist the matching process (this seam will be collected in both waves). However, it is likely that respondent recall of events at the early part of the calendar for wave 2 will be less accurate than what they were for the latter part of the calendar for wave 1, thus giving rise to difficulties in matching jobs, especially where multiple jobs are held.

### *Marital Status*

The HF and the PQ both referred to the concept of being ‘legally’ married. While the intent of these questions was to ask about ‘registered’ marriages, it is expected that some respondents with de facto relationships considered them to be ‘legal’ because they have certain rights that a married person would have under the Australian legal system. It is unclear how much of a problem this actually is in the wave 1 data. However, in wave 2 we have opted to use the concept of ‘registered’ rather than ‘legal’ marriages. This may mean that there will be some inconsistencies in marital status between wave 1 and wave 2, simply because of the change in wording.

## *Time Use*

The responses to the time use questions in the SCQ have undergone a large amount of checking and editing, but it is likely that some problems still remain. It is clear that a significant portion of the respondents did not understand the question or how to answer it. Three broad problem areas have been identified. First, excessive hours were sometimes reported, suggesting that it may have been difficult for some respondents to think in terms of ‘hours per week’. Second, it was apparent that respondents recorded hours against more than one category if they were doing more than one thing at a time (such as looking after the children and doing the housework) leading to the sum of the hours exceeding what was possible or likely. Third, it was clear that some respondents were confused by the layout of the boxes used to record the number of hours for each activity – hours and minutes were sometimes recorded and it was not always clear when the respondent was answering in this fashion. Users should thus exercise caution in using the time use data, and recognise that these data is not as clean as other data items.

The time use question for wave 2 has been redesigned to address the problems identified in wave 1.

## *Leave Entitlements*

Respondents were asked in the SCQ whether they had access to paid and unpaid maternity leave in their current job. To avoid the use of a skip for males, a ‘not applicable’ option was provided. However, as can be seen in Table 7, 1543 males provided a response to paid maternity leave and 1535 provided a response to unpaid maternity leave. Presumably these males were answering with respect to whether paid or unpaid maternity leave were available to other employees in the workplace. Further, there were 1161 females reporting that paid maternity leave was not applicable for them (the similar figure for unpaid maternity leave being 1151). These females were typically older and it is expected that they believed the question was not applicable to them because they were not planning on having any or any more children.

To overcome these difficulties, the question wording in wave 2 has been changed. In wave 1 the question asked respondents to consider “whether you would be able to use it [maternity leave] if you needed to in your current job”. In wave 2 the question asks respondents to indicate whether maternity leave is available to “you or other employees working at a similar level to you at your workplace”.

**Table 7: Responses to Whether had Access to Paid and Unpaid Maternity Leave by Sex**

<i>Type of leave</i>	<i>Not stated, multi-response (no.)</i>	<i>Not applicable (no.)</i>	<i>Yes, no or don't know (no.)</i>	<i>Valid cases</i>
Paid maternity leave				
Males	83	2709	1543	4335
Females	87	1161	2679	3927
Unpaid maternity leave				
Males	78	2722	1535	4335
Females	100	1151	2676	3927

## Data Collection Issues

### *Sex*

One problem with survey data, especially when using pen-and-paper techniques, is that incorrect codes will be checked (or circled, as is the case with the HILDA Survey interview instruments). Some guide to the extent of this problem is provided by how well interviewers performed the relatively simple task of identifying the sex of each household member. In particular, we can compare the recorded sex with the first name of respondents. While not all first names are gender-specific, we nevertheless have identified a number of cases where the entered responses look suspicious. Specifically, we have identified 25 cases where we strongly suspect that either the respondent's name has been recorded incorrectly (which has no implications for the data) or the incorrect code has been circled (with obvious adverse consequences). There are a further 21 cases where errors seem possible, but we are less certain.

Errors in the coding of sex, of course, can be easily corrected in wave 2 (assuming the sample member responds).<sup>8</sup> Errors on many other variables, however, are not so easily checked. Indeed, in wave 2 only name, sex and date of birth are being verified. Miscoding will therefore affect most questions, though if the coding of sex is a good guide, the incidence of miscodes can be expected to be relatively low.

### *Date of Birth*

Date of birth information is potentially more problematic, with errors arising from incorrect recording, inaccurate reporting and inadequate knowledge (bearing in mind that date of birth information is provided by one household member on behalf of all others in the household).<sup>9</sup> Like sex, date of birth details will be checked in wave 2 and hence it can be expected that changes will be made.

Some indication of the accuracy of the date of birth information, however, can be gleaned by comparing the recorded date of birth with the recorded age at last birthday. Looking at all household members (and not just PQ respondents) there are 919 cases where the reported age in years does not match with the reported date of birth, which represents 4.6 per cent of all cases. This may seem high, but in all but one of these cases, the discrepancy is just one year.<sup>10</sup>

In summary, there may be a sizeable error rate when recording date of birth, which we suspect stems mostly from the fact that individual respondents were not directly asked for their date of birth. The size of the error, however, is small and it can be expected that most errors will be eliminated after wave 2 data have been processed.

---

<sup>8</sup> This, of course, means that any corrections to the wave 1 data will not be made until after the wave 2 data have been collected and received.

<sup>9</sup> There are 86 individuals on the HF where the full date of birth is not recorded. For only six of these cases, however, was a year not provided. For respondents on the PQ, incomplete date of birth information was restricted to just 15 respondents, and there was no case where year of birth could not be provided.

<sup>10</sup> Which in turn reflects in-field monitoring procedure, with supervisors required to return work where the discrepancy exceeded the one-year margin.

## *Working Hours*

One section of the questionnaire with which we were particularly concerned was working hours.

### Working hours compared with one year ago

Respondents were asked to report their current usual hours (E1) and then compare that with the hours usually worked one year ago (E3). As part of the question sequence respondents were asked whether they were currently working more or less than one year ago (E2). Clearly answers at E2 should be consistent with the differences stated at E1 and E3. Despite interviewers being reminded in the questionnaire to ensure answers were consistent, 26 cases were found where responses were not. We strongly suspect, however that the confusion in these few cases arose from the wording of the responses. For example, when asked whether this was more or less than last year, a respondent may have indicated that they worked “less a year ago” which was then incorrectly classified to the precoded category “less than a year ago”. We have thus overwritten answers at E2 to be consistent with the number of hours stated.

### Preferred working hours versus actual working hours

The scope for inconsistency also arose when measuring preferred hours, with questions on both whether more or less hours were preferred and how many hours preferred. Again a few cases (n=7) produced inconsistencies. The data have been altered so that responses are consistent, mainly though altering responses at E6 (the question on whether more or fewer hours are preferred).

### Hours in main job versus hours in all jobs

Almost 9 per cent of the respondents in paid employment indicated that they currently held more than one job. In such instances respondents were asked to indicate how many hours they worked in their main job. Clearly hours worked in the main job cannot exceed hours worked in all jobs. Unfortunately, there were 13 cases where responses suggested that this was so. Further, there were another 73 cases where hours worked in main job equalled hours in all jobs, which while technically possible seems unlikely. We have no way of determining with any confidence what the correct response should be and hence the data have not been altered. It is thus up to users to determine how to handle these cases. A variable flagging the problem cases has been added to the dataset to assist users in their identification.

### Hours worked at home

Just over 28 per cent of respondents reported that they worked some of their usual hours for their main job at home. These respondents were then asked to indicate how many hours they worked at home. While not re-emphasising that the hours worked at home should only be for their main job, this was the intention of the question. There were 33 cases that reported more hours worked at home than what they work in their main job. Of the 8 cases that had more than one job, there were 4 cases where all hours worked were worked at home, so the hours worked at home in the main job could be corrected. This leaves inconsistencies in 29 cases and it is left to the users to determine how to treat these cases.

### *Income From Incorporated and Unincorporated Businesses*

Respondents were asked whether they worked in their own business, were a silent partner in a partnership, or were a beneficiary from a trust. If they did, then they were asked about their income from these businesses, separating the income from incorporated or unincorporated businesses. There were 72 cases that reported having income from their own business, partnership or trust, but then said they did not have either incorporated or unincorporated businesses. It is expected that these people had a trust but did not know whether it was incorporated or unincorporated. The last question in the section on income asks about any other income not already reported, so we expect that if the individuals received an income from these trusts, the amounts would be recorded there.

### *Interviewer Observations*

The interviewers were required to complete a section on the HF about their observations of the dwelling and the surrounding area as well as a section on the PQ about the interview situation. Unfortunately, not all interviewers completed these sections prior to sending the forms back to the office and when later re-contacted for this information, they could not recall all of the details of the dwelling or interview. As a result, some of this interviewer observation information is missing. The following table shows that the extent of this missing information is very small.

**Table 8: Missing Interviewer Observations for the Dwelling and Interview Situation**

<i>Interviewer observation</i>	<i>Not stated</i>	<i>Don't know</i>	<i>Valid cases</i>	<i>% missing</i>
Dwelling characteristics				
Type of dwelling	9	0	11693	0.1
Condition of dwelling	19	0	11693	0.2
Security features	17	0	11693	0.1
High-rise buildings in area	11	0	11693	0.1
Interview characteristics				
Presence of other adults	17	5	13969	0.2
Influence of others	21	4	5561	0.4
Understanding of questions	12	1	13969	0.1
Suspicion of study	14	1	13969	0.1
Reference to documents	13	1	13969	0.1
Cooperation	11	1	13969	0.1
Problems affecting interview				
Poor eyesight	11	1	13969	0.1
Hearing problems	11	1	13969	0.1
Reading difficulties	12	1	13969	0.1
English as second language	10	1	13969	0.1
Other language problems	10	1	13969	0.1
Assisted interview	13	1	13969	0.1
Reason for assistance	3	0	510	0.6

### *Mode Effects and Social Desirability / Acquiescence Bias*

It is often thought that the mode used to collect survey data can influence responses, especially with respect to subjective questions (see Paulhus 1991 for a review). The major concern here is what is known as social desirability bias, where responses are biased towards some socially acceptable norm (Edwards 1957). More importantly, such bias is expected to be greater when questionnaires are presented in the form of personal interviews rather than as self-administered and anonymous questionnaires. Thus it may be that part of the explanation for the relatively high score on overall life satisfaction in the HILDA sample is social desirability.<sup>11</sup> That is, some of the less happy members of the sample may have been reluctant to reveal their true feelings to a total stranger in a personal interview situation. On the other hand, if these same persons had completed the survey on their own in a situation where anonymity were much more assured, they may have felt less of a need to overstate how satisfied they felt.

Testing for the presence of such bias, of course, requires the presence of multiple methods of survey delivery, and while the HILDA Survey did complement the personal interviews with a self-administered questionnaire (the SCQ), the SCQ was not used as alternative to the personal interview. Instead, the SCQ was used as a vehicle for collecting additional information. Nevertheless, there was one item that was repeated in both instruments. This item provided a measure of self-assessed health, and it seems reasonable to expect that responses on this question would be equally affected by social desirability. A comparison of responses on these two items is consistent with the social desirability hypothesis. Scored on a 5-point scale with 1 indicating excellent health and 5 poor health, the mean score in the PQ was 2.46 and in the SCQ it was 2.52 (after restricting the sample to respondents who answered both questions). In other words, and consistent with the social desirability hypothesis, respondents were inclined to respond less favourably when completing the questionnaire on their own. Nevertheless, this difference, while highly significant in statistical terms, is still quite small in absolute terms (just 0.08).

The responses may also be tempered by the presence of other household members during the interview. Take, for example, life satisfaction. Table 9 shows that respondents that were interviewed with other adults present (besides the interviewer) had a higher mean life satisfaction score than those that were interviewed alone (8.13 compared to 7.85). While approximately half of this difference can be explained by other factors (such as, age, sex, marital status, etc), the remaining half is unexplained, suggesting that the respondents overstate their satisfaction with their life when other adults are present. The size of the differential is again quite small in absolute terms.

When asked about their satisfaction with specific aspects of their lives, the respondents interviewed with others present were again more likely to give higher responses, particularly with aspects of their lives that might be seen as commenting on their life with the other person (such as their satisfaction with their home, financial situation, safety and free time). Nevertheless, the differences are generally small, and in all cases less than was the case with overall life satisfaction.

---

<sup>11</sup> Mean life satisfaction in the HILDA Survey, scored on a 0-10 scale, was 7.96. Moreover 92 per cent of the sample chose a response in the upper half of the scale. Comparison with data from the World Values Survey (see Inglehart and Klingemann 2000) suggests that this makes Australians the most satisfied population on Earth (along with the Dutch).

**Table 9: Average Life Satisfaction score by Whether Other Adults Present During the Interview (0-10 scale)**

<i>Aspect of life</i>	<i>Other adults present during interview</i>	<i>No other adults present during interview</i>
Home	8.16	7.93
Employment opportunities	6.65	6.65
Financial situation	6.17	6.08
Safety	7.98	7.84
Part of local community	6.66	6.70
Health	7.38	7.41
Neighbourhood	8.03	7.98
Free time	6.76	6.60
Overall life satisfaction	8.13	7.85

### **Cross-form Comparisons: Chronic Health Condition, Labour Force Status and Marital Status**

One method for establishing reliability of responses is to ask the question more than once. This, of course, comes at a cost – time constraints mean questions on other topics have to be foregone. As a consequence, there was very little repetition of questions in the HILDA Survey. Nevertheless, there were three questions included in the Household Form (HF) that were either repeated in the PQ, or where other questions were included in the PQ to provide a measure of essentially the same concept. These questions related to the presence of a chronic health or disability condition, labour force status and marital status. Note at the outset, however, that some variation is to be expected given: (i) the dates of completion of the HQ and PQ were not always the same; (ii) the HF was completed by one household member on behalf of all household members; and (iii) questions were not identically worded.

A summary of the incidence of cases where answers differ across the two survey instrument follows:

- Long-term health condition (yes/no) 10.0%
- Labour force status (3 categories) 6.1%
- Marital status (5 categories) 0.4%

Somewhat surprisingly, the item where there was the greatest mismatch was the presence of a long-term chronic health condition. We might have expected less variation given the question was posed in an almost identical fashion in both instruments, and chronic conditions are, by definition, relatively stable. The results, however, may be indicative of the difficulties respondents have dealing with showcards containing long lists.

There is also a considerable divergence with respect to labour force status. This, however, is understandable given both labour force status can change quite quickly and employment status in the HF, unlike the PQ, is self-determined.<sup>12</sup>

Finally, with respect to marital status there is a relatively high concordance across instruments suggesting relatively little measurement error.

### **Data Coding and Entry**

The quality of the data is dependent not only on the questionnaire design, the understanding of the respondents, the accuracy of the interviewers recording the information, but also on the procedures undertaken in the office to code open-ended questions and to enter the data.

Considerable effort was spent in validating the coding and data entry undertaken, as discussed in Watson and Wooden (2002, p17-18). Notably, the data from the HF, HQ and PQ was double entered and subsequent investigations of data problems have revealed the issue rarely lies at the data entry stage of the process and is more likely to have occurred during the questionnaire design or interview.

The quality of the data from the SCQ is lower than the quality from the interviewer-administered questionnaires for two reasons; respondents rather than interviewers completed the forms, and the forms were scanned rather than having manually double entry. Nevertheless, we believe the data from the SCQ to be of reasonably good quality.

The interviewers were instructed not to look at the SCQs after picking them up from the household, due the more sensitive nature of some of the questions asked in the SCQ. The data was accepted 'as is', and the forms were not checked to make sure the respondent had correctly followed the skips or that all of the questions were appropriately filled out. The respondents were not re-contacted when unlikely answers were identified.

The SCQs were scanned and the data was read from these images using a mark-sense recognition system. A sample of 100 SCQs were checked in their entirety against the data scanned in the SCQ database. The average error rate was 0.3 per cent per item – this translates to an average of 1.7 errors in every fourth form. By undertaking checking of the multiple responses and forms with a high proportion of missing data, this error rate fell to approximately 0.13 per cent per item – which translates to an average of 1.5 errors in every eighth form. These errors are typically due to the respondent marking mostly outside of the box, making a light or heavy mark, or due to written numbers being misinterpreted.

### **Data Matching**

The interviewer-administered forms for each household were typically kept together, and no problems have been identified in matching the PQs and HW with the appropriate HF.

The SCQ forms were, however, dealt with separately. The interviewer usually picked up the SCQ forms from the respondent, though the respondent had the option of returning the SCQ forms by mail. However, as the data entry for the SCQ occurred on a different system from the interviewer-administered forms, the SCQs were separated from the rest of the forms for the household for processing.

---

<sup>12</sup> The PQ follows the ABS in using a much longer battery of questions concerning employment and job search activity in order to classify respondents into different labour force categories.



The SCQ data had to be then matched back to the person level data. To assist this process, the person identifier was written on the SCQ form, and the serial number of the SCQ was written on the PQ. As shown in Table 10, 104 of the 13,159 SCQs that were completed and returned could not be matched to a PQ and 413 SCQs could only be matched on one of the identifiers.

**Table 10: Type of SCQ to PQ Match for Returned SCQs**

<i>Type of match</i>	<i>Number</i>
Full match	12642
Person identifier only	243
Serial number only	170
Not matched	104
Total	13159

### **Comparisons with ABS Data**

The suite of household surveys conducted regularly by the ABS is widely accepted as providing the most accurate estimates of population characteristics, and hence it makes sense to compare key estimates from the HILDA Survey with comparable ABS estimates. Indeed, we have already reported such comparisons when assessing the representativeness of the HILDA Survey sample. In Table 11, however, we report further comparisons with ABS data, but with an emphasis on mean values for a range of key variables. Note once again that the comparisons are not always strictly valid. Thus we must take into account the fact that the HILDA estimates exclude residents of non-private dwellings whereas ABS estimates from the labour force include this sub-population (and hence explaining why we restrict most of the comparison of labour market indicators to the population aged 15 to 64 years).

For the most part, the figures presented in Table 11 again suggest again that the HILDA Survey data are generating estimates in line with ABS sources. Indeed, estimates of household size and home size are virtually identical to ABS estimates, while estimates of median home value and average rent are extremely close after adjusting the ABS estimates for inflation. Similarly, estimates of key labour market indicators, such as the employment-population ratio and the unemployment rate, derived from the HILDA Survey data are quite close to ABS estimates from October 2001. Perhaps the most noticeable difference concerns unemployment duration, with the HILDA Survey generating a mean estimate that is around two weeks longer than the ABS estimate from the MPS.

Estimated mean current gross wage and salary income for employees (\$37,190 per year) is also quite close to the estimate derived from the MPS conducted in August. The HILDA estimate is 2.6 per cent higher. Even the difference in mean household income, which is quite large – \$3840 (or 7%) after adjusting ABS estimates from the 196-97 Survey of Income and Housing Costs for price inflation – can be explained. The HILDA estimate, which appears to be biased downwards, excludes households where total household income cannot be computed due to missing data. This biases the HILDA estimate downwards since larger households are far more likely to have missing data and also have, on average, higher total incomes.

**Table 11: HILDA Sample Means Compared with ABS Estimates**

	<i>HILDA</i>		<i>ABS</i>	
	<i>Un-weighted</i>	<i>Weighted</i>	<i>Estimate</i>	<i>Reference period</i>
<i>Labour market</i>				
Employment population ratio (population aged 15-64 yrs)	0.701	0.701	0.692	Oct 2001
Labour force participation rate (%) (population aged 15-64 yrs)	75.1	75.1	74.3	Oct 2001
Unemployment rate (%) (population aged 15-64 yrs)	6.8	6.7	6.8	Oct 2001
Part-time employment share <sup>a</sup> (population aged 15-64 yrs)	0.314	0.293	0.288	Oct 2001
Average usual weekly hours worked (population aged 15-64 yrs)				Oct 2001
Part-time employed	18.6	18.5	18.1	
Full-time employed	46.6	46.3	44.8	
All employed	37.8	38.2	37.2	
Mean duration of unemployment (weeks) <sup>b</sup>	50.0	50.7	48.6	Oct 2001
Duration of current job (employed persons aged 15-69 yrs)				Feb 2000
% < 1 year	20.7	21.3	23.6	
% 1 to <5 years	35.6	36.5	35.6	
% 5 to < 10 years	17.6	17.2	16.4	
% 10 years or more	26.1	25.0	24.4	
<i>Income</i>				
Mean financial year household income (\$) <sup>c</sup>	49010	50390	45180 (54230)	1996-97
Average current annual earnings (\$ / year) – all jobs, employees <sup>d</sup>				Aug 2001
Men	43840	44030	42810	
Women	28500	28980	28680	
All persons	36460	37190	36240	
<i>Housing</i>				
Average no. of bedrooms <sup>e</sup>	3.0	3.1	3.0	Sep-Dec 99
Median house value: owners (\$000) <sup>f</sup>	200.0	200.0	145.2 (200.4)	1997-98
Mean weekly rent: renters (\$) <sup>f</sup>	149	157	140 (156)	Sep-Dec 99
<i>Demographics</i>				
Average household size (no. of persons)	2.6	2.6	2.6	Aug 2001

**Table 11 (cont'd)**

	<i>HILDA</i>		<i>ABS</i>	
	<i>Un-weighted</i>	<i>Weighted</i>	<i>Un-weighted</i>	<i>Weighted</i>
<i>Health (SF-36 scores): Persons aged 18+ yrs</i>				
Physical functioning	82.1	82.1	82.6	1995
Role limit physical	78.2	78.4	79.8	1995
Bodily pain	73.5	73.5	76.8	1995
General health	69.4	69.3	71.6	1995
Vitality	60.4	60.5	64.5	1995
Social functioning	81.4	81.2	84.9	1995
Role limit emotional	81.8	81.8	82.8	1995
Mental health	73.8	73.7	75.9	1995

- Notes:
- a Part-time employment is defined here as usually working less than 35 hours per week. This is slightly different than the definition used by the ABS. The ABS data reported here, however, are based on the same definition as used in the HILDA Survey data.
  - b Some respondents to the HILDA Survey could not provide the date they commenced job search, but were able to indicate a broad range for unemployment duration. For these cases unemployment duration has been (crudely) imputed.
  - c HILDA data relate to the 2000/01 financial year. The figure in parentheses is the ABS estimate inflated by growth in national gross household income per capita over the period 1996-97 to 2000-01.
  - d ABS data have been annualised (by multiplying weekly earnings by 52.14). Employees who worked solely for payment in kind in their main job have been excluded.
  - e The ABS definition relates to bedrooms per dwelling, whereas in the HILDA Survey the data should relate to bedrooms in the home in which the household lives.
  - f Figures in parentheses are the estimated values after adjusting for price inflation (calculated in September 2001 prices).

ABS Sources:

Labour Force, Selected Summary Tables, Australia, Monthly (6291.0.40.001), companion data (available in SuperTable format).

Labour Mobility, Australia, February 2000 (cat. no. 6209.0).

Australian Housing Survey: Housing Characteristics, Costs and Conditions, 1999 (cat. no. 4182.0).

Employee Earnings, Benefits and Trade Union Membership, Australia, August 2001 (cat. no. 6310.0).

Housing Occupancy and Costs, Australia, 1997-98 (cat. no. 4130.0).

1997-98 Survey of Income and Housing Costs, Confidentialised Unit Record File.

1995 National Health Survey: SF-36 Population Norms (cat. no. 4399.0).

## Summary

It can be expected that a survey of this size and scope will have its share of problems. Nevertheless, our overall assessment is that the data generated in wave 1 compare favourably with other surveys of this type around the world.

The key reasons for this assessment are fourfold.

- i. The sample appears to be reasonably representative of the wider population, with the main sources of bias – differences in rates of response across both sex and location – relatively easily corrected through the application of population weights.
- ii. Estimates of mean values on key variables such as the unemployment rate and average annual earnings are in line with population estimates from ABS sources.
- iii. Relatively low rates of non-response on most data items.
- iv. Evidence of a high level of internal consistency in the data.

This is not to say that the data set is not without problems. Incomplete households means that deriving estimates for household-level variables will, in many cases, mean imputing the responses of non-respondents. There are also obvious problems with some specific questions and sections of the survey instruments – the calendar in the PQ and the time use question in the SCQ being the most obvious examples here.

## References

- Bednall, D., Cavenett, P. and Shaw, M. (2000), 'Response Rates in Australian Market Research', Paper presented at the Market Research Society of Society Annual Conference, Sydney, October.
- Edwards, A.L. (1957), *The Social Desirability Variable in Personality Assessment and Research*, Holt, Rinehart & Winston, Inc, New York.
- Inglehart, R. and Klingemann, H. (2000), 'Genes, Culture, Democracy and Happiness', in E. Diener and E.M. Suh (eds), *Subjective Well-being Across Cultures*, MIT Press, Cambridge.
- Paulhus, D.L. (1991), 'Measurement and Control of Response Bias', in J.P. Robinson, P.R. Shaver and L.S. Wrightsman (eds.), *Measures of Personality and Social Psychological Attitudes*, Harcourt Brace Jovanovich, San Diego.
- Ripahn, R.T. and Serfing, O. (2002), 'Item Non-Response on Income and Wealth Questions', Paper presented at the 5th International German Socio-Economic Panel Users Conference, Berlin, 3-4 July.
- Schrapler, P. (2002), 'Respondent Behavior in Panel Studies – A Case Study for Income-Nonresponse by Means of the German Socio-Economic Panel', Paper presented at the 5th International German Socio-Economic Panel Users Conference, Berlin, 3-4 July.
- Ware, J.E., Snow, K.K., Kosinski, M. and Gandek, B. (1993), *SF-36 Health Survey: Manual and Interpretation Guide*, QualityMetric Inc., Lincoln, RI.
- Watson, N. and Wooden, M. (2002), 'The Household, Income and Labour Dynamics in Australia (HILDA) Survey: Wave 1 Survey Methodology', HILDA Project Technical Paper Series No. 1/02, Melbourne Institute of Applied Economic and Social Research, University of Melbourne.

**Appendix 1: Incidence of Missing Cases**  
**(questions where incidence of missing cases exceeds 2% of expected N)**

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
<i>Household Questionnaire</i>				
Q4a	Difficulty finding good quality childcare	7.1	1149	Showcard explicitly provided a don't know / not applicable option.
Q4b	Difficulty finding right person to take care of child	6.0	1149	
Q4c	Difficulty getting care for hours needed	5.0	1149	
Q4d	Difficulty finding care for a sick child	12.2	1149	
Q4e	Difficulty finding care during school holidays	17.8	1149	
Q4f	Difficulty with the cost of child care	8.3	1149	
Q4g	Difficulty juggling multiple childcare arrangements	24.0	1149	
Q4h	Difficulty finding care for a difficult or special needs child	53.6	1149	
Q4j	Difficulty finding a place at the childcare centre of choice	19.5	1149	
Q4k	Difficulty finding a childcare centre in the right location	17.7	1149	
Q4m	Difficulty finding care child is happy with	8.5	1149	
Q8	Child care costs – school age children during school holidays	2.3	1095	
Q12	How Child Care Benefit payment made	6.5	657	Instrument only provided for two options – fortnightly reductions in fees and annual lump sum payments. Other options, however, exist.
R10	Value of home	5.9	5268	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
R13	Value of outstanding home loan	5.5	2249	
R14	Usual home loan repayment	5.8	2249	
R16	Year home loan expected to be paid off	11.3	2249	
R18	Amount owing on home loan from friend, relative, etc.	4.7	342	
R20	Amount owing on any other home loan	10.3	379	
R21	Usual repayment on other home loan	11.9	379	
<i>Person Questionnaire</i>				
B4	Age when parents first separated	3.7	2675	Affected by recall problems.
B7	Presence of siblings	8.9 (0.6)	13969	Sequencing error caused many persons not to be asked this question. The figures in brackets indicate the proportion missing after information provided by parents is used.
B8	Number of siblings	9.3 (0.8)	13349	
B9	Whether oldest when growing up	9.2 (0.8)	13349	
B10a	Father's country of birth	9.4 (1.6)	13969	
B10b	Mother's country of birth	9.1 (1.6)	13969	
B12	Whether father employed at age 14.	13.0 (3.4)	13969	
B13	Father's occupation when growing up	10.3 (4.3)	13417	
B14	Whether father unemployed for 6 months + while growing up	13.0 (6.4)	13608	
B15	Whether mother employed at age 14.	9.6 (4.1)	13969	
B16	Mother's occupation when growing up	14.1 (9.1)	10314	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
C11a	Qualification studying for	3.0	1527	Mainly as a result of insufficient information to enable a classification to be made.
D16	Pay in previous job	8.0	2195	Affected by recall problems.
E5	Reason for working part-time	6.1	2727	Mainly as a result of insufficient information to enable a classification to be made.
E26	Expectation that contract will be renewed	13.1	628	A don't know answer is to be expected in many cases.
E35	Firm size (if more than 1 workplace)	13.3	4570	All of the "don't knows" able to report whether their firm had more or fewer than 100 employees.
F3	Date began looking for work	12.2	728	Affected by recall problems. All respondents could provide an estimate within a range.
F4	Whether receiving Intensive Assistance	2.3	728	
F5a	Whether could start work last week (if actively looking)	2.1	728	
F17	Whether could start work in next 4 weeks (if not in labour force but want to work)	2.8	1248	
F18	Reservation wage	9.0	1976	Difficult concept.
F20	Probability of finding a job (unemployed)	4.2	1976	Difficult concept.
G3	Gross wage income, main job – current	7.3	7723	Only 3.3% could not provide either their gross or net wage.
G7	Gross pay, other jobs – current	11.5	722	8.6% could not provide either their gross or net wage.
G18	Disability Pension, current income	2.4	125	Small sample.
G20	Gross wage income – last financial year (LFY)	8.2	8179	Declines slightly, to 7.1%, if we allow for either gross or net income to be reported.



<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
G24	Gross wage income from incorporated businesses – LFY	15.9	654	
G25b	Dividend income from incorporated businesses – LFY	24.2	124	
G26b	Profit from non-incorporated businesses – LFY	31.1	1585	
G27b	Interest earned (if >\$100) – LFY	19.2	3380	
G28b	Dividend income – LFY	13.7	3901	
G28b	Royalties income – LFY	39.4	71	
G29b	Profit / loss from rental income	21.3	1306	
G31	Newstart income, LFY	2.4	167	
G31	Service pension income, LFY	5.4	56	
G31	Disability support pension income, LFY	3.4	89	
G31	Disability pension income, LFY	9.1	22	Very small sample.
G31	Carer payment income, LFY	9.1	22	Very small sample.
G31	Sickness allowance income, LFY	9.1	11	Very small sample.
G31	Parenting payment income, LFY	4.2	189	
G33	Annuity income, FY	5.7	756	
G33	Child support income, FY	3.3	369	
G33	Workers comp etc. income, FY	9.5	190	
G33	Income from parents, FY	3.8	633	
G33	Income from other persons, FY	6.0	84	
G33	Income from other sources, FY	3.2	250	
H5	Regular child support payments	3.9	456	
H7	Other child-related expenses	12.2	490	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
H15d	Distance of non-resident parent from child [first in grid]	7.2	891	
H18	Regular child support received	3.5	339	
H20	Receipt of income for other child-related expenses	15.9	145	
H22	Number of nights youngest child stays with other parent	2.3	822	
H24	Amount of contact youngest child has with other parent	2.8	822	
J2a	Month of most recent marriage	2.2	9827	
J2a	Year of second marriage (if married more than twice)	4.5	133	
J2c	Years lived together before marriage (2nd marriage)	3.6	55	Small sample.
J2e	Year widowed, first marriage (if married more than once)	4.7	171	
J2e	Year widowed, second marriage (if married more than twice)	12.0	25	Very small sample.
J2f	Year separated / divorced, most recent marriage	2.8	1579	
J2f	Year separated / divorced, first marriage (if married more than once)	4.4	1255	
J2f	Year separated / divorced, second marriage (if married more than twice)	4.7	106	
J2g	Year divorce finalised, second marriage (if married more than twice)	4.7	106	
J4	Month began living with current de facto partner	14.3	1353	
J8	Month began living together in first relationship	52.0	2220	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
J8	Year began living together in first relationship	2.4	2220	
K10	Month moved to current address	9.8	13969	
<i>Self-completion Questionnaire (matched sample only)</i>				
A3a	Physical functioning: Whether health limits vigorous activities	2.9	13055	
A3b	Physical functioning: Whether health limits moderate activities	2.2	13055	
A3c	Physical functioning: Whether health limits lifting or carrying groceries	2.3	13055	
A3d	Physical functioning: Whether health limits climbing several flights stairs	2.7	13055	
A3e	Physical functioning: Whether health limits climbing one flight stairs	3.1	13055	
A3f	Physical functioning: Whether health limits bending, kneeling or stooping	2.5	13055	
A3g	Physical functioning: Whether health limits walking one kilometre	2.5	13055	
A3h	Physical functioning: Whether health limits walking half a kilometre	2.9	13055	
A3i	Physical functioning: Whether health limits walking 100 metres	3.2	13055	
A3j	Physical functioning: Whether health limits bathing or dressing	2.3	13055	
A4a	Role-physical: Cut down amount of time spent on work or other activities	2.2	13055	
A4b	Role-physical: Accomplished less than would like	2.5	13055	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
A4c	Role-physical: Limited in the kind of work or activities	2.7	13055	
A4d	Role-physical: Had difficulty performing work or other activities	2.6	13055	
A5a	Role-emotional: Cut down amount of time spent on work or other activities	2.3	13055	
A5b	Role-emotional: Accomplished less than would like	2.2	13055	
A5c	Role-emotional: Didn't do work as carefully as usual	2.7	13055	
A11a	General health: Seem to get sick a little easier than other people	2.5	13055	
A11c	General health: Expect health to get worse	2.6	13055	
A11d	General health: Health is excellent	2.5	13055	
B5	Daily consumption of alcohol when drinking	3.0	11068	
B11a	Satisfaction with the relationship with their partner	2.3	9568	Not applicable column provided, but seems likely that this was not always used. That is, some of those who should have circled the not applicable option did not provide an answer.
B11b	Satisfaction with their relationship with children	2.5	8981	
B11c	Satisfaction with partner's relationship with their children	3.9	7698	
B11d	Satisfaction with their relationship with stepchildren	26.8	1940	
B11e	Satisfaction with how well the children in the household get along with each other	6.3	5946	
B11f	Satisfaction with their relationship with respondent's parents	3.6	9444	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
B11g	Satisfaction with their relationship with step parents	23.7	2276	
B11h	Satisfaction with their relationship with most recent former spouse or partner	10.1	4383	
B16a	Hours per week spent on household errands	3.2	13055	While respondents were instructed to write in “0” if the activity was not undertaken, we suspect some respondents did not comply with this instruction. Consequently some of the missing cases should have been recorded as zeros.
B16b	Hours per week spent on housework	3.2	13055	
B16c	Hours per week spent on outdoor tasks	3.5	13055	
B16d	Hours per week interacting with children	5.1	13055	
B16e	Hours per week doing volunteer or charity work	4.5	13055	
B16f	Hours per week caring for a disabled spouse or adult relative or elderly parents	5.0	13055	
B16g	Hours per week travelling to work	4.9	13055	
C2b	Could not pay mortgage or rent on time	2.6	13055	
C2c	Had to pawn or sell something	2.2	13055	
C2d	Went without meals	2.1	13055	
C2e	Was unable to heat home	2.3	13055	
C2g	Asked for help from welfare or community organisations	2.1	13055	
C3b	Method of raising \$2000 in an emergency	3.4	10912	
C5	Savings time horizon	2.2	13055	
C7a	Attitudes to borrowing: For a holiday	4.8	13055	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
C7b	Attitudes to borrowing: If income falls unexpectedly	3.6	13055	
C7c	Attitudes to borrowing: For clothes / jewellery	5.4	13055	
C7d	Attitudes to borrowing: For a car	4.2	13055	
C7e	Attitudes to borrowing: For education expenses	5.2	13055	
D1a	It is important to have a paying job in order to be happy	2.5	13055	
D1b	I would enjoy having a job even if didn't need the money	2.8	13055	
D1c	Many working mothers seem to care more about being successful at work than meeting the needs of their children	3.4	13055	
D1d	Many working fathers seem to care more about being successful at work than meeting the needs of their children	3.1	13055	
D1e	If both partners in a couple work, they should share equal in the housework and care of children	2.8	13055	
D1f	Whatever career a woman may have, her most important role in life is still that of being a mother	2.8	13055	
D1g	Whatever career a man may have, his most important role in life is still that of being a father	2.9	13055	
D1h	Mothers who don't really need the money shouldn't work	2.9	13055	
D1i	Children do just as well if the mother earns the money and the father cares for the home and children	3.1	13055	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
D1j	It is much better for everyone involved if the man earns the money and the woman takes care of the home and children	2.8	13055	
D1k	As long as the care is good, it is fine for children under 3 years of age to be placed in child care all day for 5 days a week	2.8	13055	
D1l	A working mother can establish just as good a relationship with her children as a mother who does not work for pay	2.9	13055	
D1m	A working father can establish just as good a relationship with his children as a father who does not work for pay	2.9	13055	
D1n	A father should be as heavily involved in the care of his children as the mother	2.7	13055	
E3a	Paid maternity leave	3.9	4392	A 'don't know' option explicitly provided.
E3b	Unpaid maternity leave	4.1	4389	
E3c	Parental leave	7.2	8262	
E3d	Special leave for caring for family members	5.6	8262	
E3e	Permanent part-time work	5.5	8262	
E3f	Home-based work	5.9	8262	
E3g	Flexible start and finish times	5.2	8262	
F1	Parenting responsibilities	4.6	13055	
F2a	Being a parent is harder than I thought it would be	3.1	4522	A 'does not apply' option explicitly provided.
F2b	I often feel tired, worn out, or exhausted from meeting the needs of my children	3.2	4507	
F2c	I feel trapped by my responsibilities as a parent	3.3	4513	

<i>Form / Qstn #</i>	<i>Variable</i>	<i>Missing cases (%)</i>	<i>Expected N</i>	<i>Comments</i>
F2d	I find that taking care of my child/children is much more work than pleasure	3.3	4510	
F3	Perception of whether doing fair share of the child care	3.7	4573	