

Income non-reporting: implications for health inequalities research

Gavin Turrell

Abstract

Objectives—To determine whether, in the context of a face to face interview, socioeconomic groups differ in their propensity to provide details about the amount of their personal income, and to discuss the likely consequences of any differences for studies that use income based measures of socioeconomic position.

Design and setting—The study used data from the 1995 Australian Health Survey. The sample was selected using a stratified multi-stage area design that covered urban and rural areas across all States and Territories and included non-institutionalised residents of private and non-private dwellings. The response rate was 91.5% for selected dwellings and 97.0% for persons within dwellings. Data were collected using face to face interviews. Income response, the dependent measure, was binary coded (0 if income was reported and 1 for refusals, “don’t know” and insufficient information). Socioeconomic position was measured using employment status, occupation, education and main income source. The socioeconomic characteristics of income non-reporters were initially examined using sex specific age adjusted proportions with 95% confidence intervals. Multivariate analysis was performed using logistic regression.

Participants—Persons aged 15–64 (n=33 434) who were reportedly in receipt of an income from one or more sources during the data collection reference period.

Results—The overall rate of income non-response was 9.8%. Propensity to not report income increased with age (15–29 years 5.8%, 30–49 10.6%, 50–64 13.8%). No gender differences were found (men 10.2%, women 9.3%). Income non-response was not strongly nor consistently related to education or occupation for men, although there was a suggested association among these variables for women, with highly educated women and those in professional occupations being less likely to report their income. Strong associations were evident between income non-response, labour force status and main income source. Rates were highest among the employed and those in receipt of an income from their own business or partnership, and lowest among the unemployed and those in receipt of a govern-

ment pension or benefit (which excluded the unemployed).

Conclusion—Given that differences in income non-reporting were small to moderate across levels of the education and occupation variables, and that propensity to not report income was greater among higher socioeconomic groups, estimates of the relation between income and health are unlikely to be affected by socioeconomic variability in income non-response. Probability estimates from a logistic regression suggested that higher rates of income non-reporting among employed persons who received their income from a business or partnership were not attributable to socioeconomic factors. Rather, it is proposed that these higher rates were attributable to recall effects, or concerns about having one’s income information disclosed to taxation authorities. Future studies need to replicate this analysis to determine whether the results can be inferred to other survey and data collection contexts. The analysis should also be extended to include an examination of the relation between socioeconomic position and accuracy of income reporting. Little is known about this issue, yet it represents a potential source of bias that may have important implications for studies that investigate the association between income and health.

(J Epidemiol Community Health 2000;54:207–214)

A large number of studies have investigated the relation between socioeconomic position and health.^{1–3} Some of this research has been conducted using measures of socioeconomic position that are based on income.^{4,5} With few exceptions, these studies show that persons classified as “low income” typically die at the highest rates for most major causes of death^{6–8} and they experience more ill health and disability.^{9–11}

Despite the popularity and predictive capacity of income based measures, the collection of income data is problematic. Specifically, income is a sensitive and private topic,¹² and as a consequence, income questions asked in the context of survey research are susceptible to high rates of non-response. A search of the published literature in the health and social sciences located nine studies that identified the extent of income non-reporting in their samples (table 1). Based on this evidence, income questions seemingly elicit non-response rates that range between 10 and 25 per cent. This interval is broadly consistent

School of Public Health, Centre for Public Health Research, Queensland University of Technology, Brisbane, Australia

Correspondence to: Dr G Turrell, The University of Michigan School of Public Health, Department of Epidemiology, 109 Observatory Street, Ann Arbor, Michigan 48109-2029, USA

Accepted for publication 5 June 1999

Table 1 Prevalence for the non-reporting of income: published studies in the health and social sciences

Author(s)	Country	Sample size	Data collection method	Non-response prevalence (%)
Wright 1975 ¹³	Holland	1687	Face to face	17.3
Locander and Burton 1976 ¹⁴	US	500	Telephone (men)	18.9–24.6*
Siemiatycki 1979 ¹⁵	Canada	500	Telephone (women)	23.7–35.9
		768	Mail	8.8
		791	Face to face	14.1
Jordan <i>et al</i> 1980 ¹⁶	US	1056	Telephone	18.9
		1210	Face to face	12.0
		303	Telephone	21.0
Hippler and Hippler 1986 ¹⁷	Germany	2057	Face to face	24.2
		1618	Face to face	18.7
		1421	Face to face	7.2–25.1*
Kormendi 1988 ¹⁸	Denmark	651	Telephone	12.8–33.2
Martin-Matthews <i>et al</i> 1991 ¹⁹	Canada	224	Face to face	8.5–38.0*
	Guadagnoli and Cleary 1992 ²⁰	US	292	Mail
de Leeuw 1992 ²¹	Holland	267		13.9
		283		22.3
		254	Mail	14.0
		243	Face to face	17.0
		343	Telephone	18.0

*Response rate varied by type of income question. †Response rate varied by patient sample.

with the observations of other researchers.²² Importantly, the rate varies considerably depending on the data collection method, and the structure, content and wording of the question. All other things being equal, mail surveys tend to elicit the lowest rate of income non-response (compared with telephone surveys) and questions that ask the respondent to provide an exact income amount (rather than nominate an income category) elicit the highest rates.

Rates of item non-response that range between 10 and 25 per cent are clearly a cause for concern. They pose problems when undertaking multivariate analysis, such as the necessary exclusion of all data for those cases not having a value on the missed item (effectively reducing the sample size by 10–25 per cent). If systematically associated with a particular type of income earner, these exclusions may compromise data quality (that is, introduce bias), thus adversely affecting the results and limiting generalisability. A range of statistical procedures has been proposed to deal with missing data in survey research.^{23–24} Inherent in these techniques however, is the assumption that data are missing at random. Studies investigating the characteristics of item non-responders raise doubts about the veracity of these

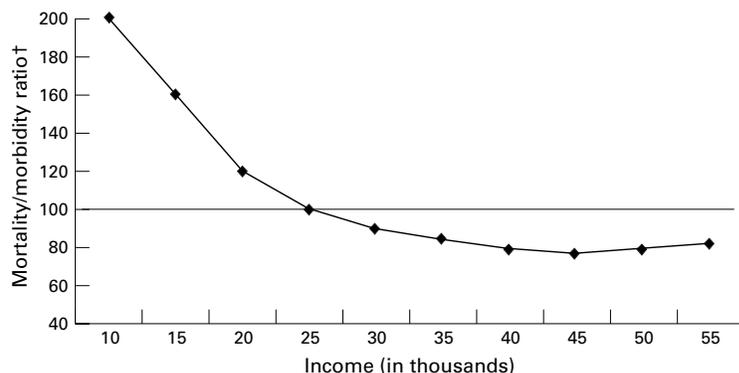


Figure 1 Generalised form of health plotted against income: hypothetical example. *This is a very general representation of the relation between income and health, and its exact form varies by gender and age. The nature and extent of the curvilinear shape was modelled on the findings of Backlund, Sorlie and Johnson³⁷ and others.^{9,38} †A value of 100 represents the mean.

assumptions. Although inconsistencies characterise some of this research, rates of item non-response have been shown to be significantly higher among the elderly,^{25–30} women,^{25–30} the unmarried,²⁹ persons of low socioeconomic position,^{25–30} the physically, cognitively and psychologically impaired,²⁷ and those reporting poorer health.^{29–30} Rather than being an exclusively random phenomenon therefore, item non-response is often systematically related to a respondent's sociodemographic and health characteristics. Of particular importance for this study is the finding that item non-response is higher among low socioeconomic groups. This finding takes on even greater significance when research relating to survey participation is examined. There is now a large literature on this topic. While the evidence presents a mixed and somewhat inconsistent picture overall, one finding in particular is highly reliable: persons of low socioeconomic position are least likely to respond to, or participate in, survey research.^{31–36}

Thus we are confronted with two independent but mutually supporting bodies of literature: one shows that low socioeconomic groups are likely to be under-represented in survey research, and the other indicates that when they do participate, they are more likely to have higher rates of item non-response. This would seem to have serious implications for studies that use income based measures of socioeconomic position, for it suggests that data quality problems that are introduced by survey non-response among low socioeconomic groups are likely to be compounded by problems that are introduced by their higher rates of item non-response. Whether or not this proves to be the case however, depends primarily on how socioeconomic groups differ in terms of their rate of income non-reporting. For example, if low socioeconomic groups are less likely to participate and also less likely to report their income, then income based estimates of socioeconomic differences in health are likely to be considerably under-estimated. If however, income non-reporting is more prevalent among high socioeconomic groups then this will have minimal negative impacts on estimates of the relation between income and health. The shape of this relation is very often curvilinear, with the gradient being steep and near linear at low income levels tapering quickly to a gentler slope at moderate to high income levels.^{9,37,38} Increments of income on the flatter section of the gradient often results in little, if any, additional health benefit. Figure 1 presents a non-specific representation of this relation using hypothetical data to simulate the nature of the curvilinear association between income and health.

This study investigates the issue of income non-response and its likely consequences for health inequalities research by comparing income reporters and non-reporters on the basis of four measures of socioeconomic position, namely, labour force status, occupation, education and main income source.

Method

The data used in this study were collected by the Australian Bureau of Statistics (ABS) as part of the 1995 National Health Survey (NHS). Full details of the Survey's scope and coverage, its research design, sampling procedures and data collection methods have been reported elsewhere.³⁹ Only a brief overview is provided here.

SAMPLE DESIGN

The NHS covered urban and rural areas across the six States and two Territories of Australia, and included non-institutionalised residents of both private (houses, flats, caravans, etc) and non-private (hotels, hostels, boarding houses) dwellings. A total of 23 817 dwellings were included in the original sample, representing approximately 1 in 310 of the Australian population. The dwellings were randomly selected using a stratified multi-stage area design that ensured that all persons in each State and Territory had an approximately equal probability of being chosen. After excluding refusals, non-contacts and other non-responding groups the final sample consisted of 21 787 dwellings (91.5% response rate). Information was obtained on 53 828 children and adults resident in the selected dwellings (97.0% response rate).

DATA COLLECTION AND SURVEY INSTRUMENT

Data were collected by face to face interviews using trained interviewers. Persons aged 18 years or more were personally interviewed. Children aged 15–17 were interviewed with the consent and knowledge of a responsible adult, usually a parent, close relative or guardian. Where consent was not given, and for children aged less than 15 years, a responsible adult was asked to answer on their behalf. A range of questionnaires was administered as part of the interview. Information was sought about the structure of the household and the socio-demographic characteristics of its members, their self reported health, health related actions and health risk factors.

MEASURES

Income response (dependent measure)

Income data were collected on the basis of a two part question from persons aged 15 years or more. The income question related primarily to regular/recurring cash income and excluded non-cash receipts such as income in kind, capital transfers and capital gains and losses. The first part of the question asked respondents to indicate whether in the past financial year they had received income as a result of profit or loss from their own business, rental investment properties, dividends or interest. Respondents who reportedly received income from one or more of these sources were subsequently asked to indicate the amount received (in exact dollars) before tax being taken out, but after the deduction of business expenses. Part two of the income question asked respondents to indicate whether they were currently receiving income from a wage or salary, a family payment, a government pension

or cash benefit, maintenance/child support, superannuation/annuity, workers compensation/accident or sickness insurance, or any other income. Again, if one or more source was nominated, the respondent was asked to indicate the amount received (in exact dollars) before tax being deducted. A respondent's gross personal annual income was then ascertained by summing the values reported in part's one and/or two. Before releasing the NHS data file, the ABS grouped the individual income values into 17 income categories. Two additional categories were also generated: a "not applicable" category containing persons who were not receiving an income from any source when the survey was conducted (includes dependent children, full time students, persons engaged in home duties) and a "don't know/not stated" category. This category contained respondents who reportedly did not know their income, or refused to divulge the amount they earned, or supplied insufficient information for their income to be accurately determined (it was not possible to disaggregate this category and examine each type of response separately). For the purposes of this study, income response was operationalised as a binary variable. Respondents who were categorised in the "don't know/not stated" group were coded 1, and those who were assigned to one of the 17 income categories were coded 0. Respondents assigned to the "not applicable" category were excluded from the study.

Socioeconomic position

Labour force status Questions relating to labour force status were asked of respondents aged between 15 and 64 years. Those aged 15 years or more who were still attending school, and those aged 65 years or over (that is, the retired) were excluded. The labour force measure supplied by the ABS comprised three categories: (1) employed persons, defined as those who reported that they had worked in the week preceding the interview, (2) unemployed persons, defined as those who were not employed during the week preceding the interview but who had actively sought work at some point during the previous month, and (3) persons not in the labour force, defined as those who were neither employed nor unemployed. This last group included persons engaged in home duties on a full time basis, those who were financially independent and did not need to work, and those who were not employed when the survey was conducted and were not actively seeking work.

Occupation Respondents who were employed during the week preceding the interview were asked to indicate their job title and then describe the main tasks or duties that they performed. This information was subsequently coded to the ABS's Australian Standard Classification of Occupations (ASCO).⁴⁰ ASCO is a skill based measure that groups together occupations requiring similar levels of education, knowledge, responsibility, and on the job training and experience. The occupational groupings are hierarchically ordered

based on their relative skill levels, with those occupations having the most extensive skill requirements located at the top of the hierarchy. The NHS data file was publicly released with occupations coded to eight ASCO categories as follows: (1) Managers and Administrators (includes judges, parliamentarians, general and specialist managers, farmers and farm managers and managing supervisors), (2) Professionals (includes architects, surveyors, cartographers and engineers, medical practitioners, dentists, veterinarians, school teachers and university lecturers, lawyers and journalists), (3) Para-Professionals (includes technical officers and technicians, aircraft pilots, registered nurses, police and ambulance officers, inspectors and regulatory officers), (4) Tradespersons (includes toolmakers, fitters, electricians, carpenters, plumbers, vehicle mechanics, cooks and cabinet-makers), (5) Clerks (includes stenographers and typists, business machine operators, receptionists, general and specialist clerks), (6) Salespersons and Personal Service Workers (includes sales representatives and assistants, real estate agents, tellers, bar attendants, travel agents, child care workers and dental nurses), (7) Plant and Machine Operators and Drivers (includes road and rail transport drivers, forklift drivers, fire fighters, crane and machine operators) and (8) Labourers and Related Workers (includes farm hands, cleaners, construction and mining labourers, ushers and door attendants, kitchen hands, storemen/women, caretakers). When undertaking multivariate analyses, this eight level occupation measure resulted in an over-stratified model, with consequent cell size problems. The original variable, therefore, was collapsed into three categories: professionals (groups 1–3), white collar (groups 5 and 6) and blue collar (groups 4, 7 and 8). This three level measure has been used by other researchers who have demonstrated that it clearly discriminates between occupation groups in terms of a range of health indicators.⁴¹

Education As part of the NHS's data collection protocol, approximately half of the adult respondents were asked to provide information about their post-school educational qualifications (the other half were asked to complete the SF36 health and well being questionnaire. This strategy was adopted in an attempt to minimise respondent burden). Based on responses to a number of related questions, respondents were classified by the ABS to one of the following education categories: (1) higher degree, (2) post-graduate diploma, (3) bachelor degree, (4) undergraduate diploma, (5) associate diploma, (6) skilled vocational, (7) basic vocational, and (8) no higher qualifications. A subsequent analysis based on this "half-sample" variable produced small cell sizes after stratifying by sex and age, thus it was necessary to collapse the measure into four categories: bachelor degree or higher (levels 1–3 on the original classification), diploma (levels 4 and 5), vocational (levels 6 and 7), and no higher educational qualification (level 8).

Main income source Respondents who reported receiving an income during the 12 months before the interview (irrespective of whether or not they provided details about the amount received) were asked to identify the main source of their income. Six income source options were made available on the NHS data file: (1) wages or salary (from an employer or from own limited liability company), (2) interest or dividends, (3) profit or loss from one's own business or share in a partnership, (4) government pension or benefit (5) superannuation or annuities, and (6) "other". For the purposes of this paper, only groups 1–4 were included in the analyses. Group 5 was excluded because there were very few working aged persons living off superannuation and annuity payments. Group 6 consisted of persons in receipt of an income from a diverse range of sources, including profit or loss from rental investment properties, family payment, maintenance/child support, workers compensation/accident or sickness insurance, and any other source of income. This group was excluded because of its heterogeneity and the associated interpretive difficulties.

ANALYSIS

Of the 53 828 persons in the original sample, 14 840 (27.6%) were not receiving income from any source at the time the data were collected and were excluded from the present study. Those aged 65 years or older ($n=5554$, 10.3%) were also excluded. This group was not asked to provide information about their labour force status or occupation, and the main income source for the vast majority was a government pension or benefit. The final data set used in the analysis, therefore, comprised 33 434 persons aged 15–64 who reported receiving an income from one or more sources.

The relation between socioeconomic position and income response was initially examined by comparing each socioeconomic group in terms of the proportion who did not report their income. Separate analyses are presented for men and women adjusted for age (15–29 years, 30–49, and 50–64) using direct standardisation (the total population was used as the standard). Prevalence rates are presented with 95% confidence intervals. Given the large sample size, the results of significance tests are not presented because even negligible differences would attain significance. Thus a descriptive approach is applied to the presentation of percentages, with socioeconomic differences of 4% or greater being suggestive of an effect. This is admittedly a somewhat arbitrary cut off. However, power calculations performed at the 95% significance level (two tailed) with 90% power indicated that a 2–3% difference in proportions between socioeconomic groups would be identified as "statistically significant". In the context of this study these differences were not considered important or meaningful, and hence a more conservative difference of 4% was used.

The multivariate relation between socioeconomic position and income non-response was examined using logistic regression, performed

Table 2 Socioeconomic characteristics of income non-reporters: per cent (age adjusted) who did not report their income, by gender

SES indicator	Men			Women		
	Number	%	95% CI	Number	%	95% CI
<i>Labour force status</i>						
Unemployed	972	4.1	1.0, 7.2	621	3.7	0.0, 7.6
Not in labour force	2 162	9.0	6.9, 11.1	5 119	7.0	5.7, 8.3
Employed	13 505	10.9	10.0, 11.7	11 051	10.8	9.8, 11.7
<i>Occupation</i>						
Blue collar	6 485	10.6	9.4, 11.8	1 985	7.3	5.2, 9.4
White collar	2 174	8.8	6.7, 10.9	5 839	10.7	9.5, 11.9
Professionals	4 835	12.3	10.9, 13.7	3 240	12.0	10.3, 13.7
<i>Highest education qualification</i>						
No higher qualifications	3 917	10.0	8.5, 11.5	4 782	8.8	7.4, 10.2
Vocational	2 331	10.9	8.9, 12.9	1 487	11.0	8.5, 13.5
Diploma	732	10.2	6.6, 13.8	789	10.9	7.5, 14.3
Bachelor degree or higher	1 188	11.3	8.5, 14.1	1 189	12.0	9.2, 14.8
<i>Main income source</i>						
Government pension or benefit*	1 706	6.5	5.6, 7.3	3 411	4.6	3.7, 5.4
Wages or salary	10 899	7.4	6.5, 8.3	9 243	8.4	7.4, 9.4
Interest/dividends	309	14.2	8.7, 19.7	500	13.0	8.7, 17.3
Own business or partnership	2 000	23.0	20.9, 25.1	1 053	26.2	23.2, 29.2

*Includes pensions and allowances for the sick and disabled, pensions for lone parents and widowed persons, family payments, special benefits plus other pensions, benefits and allowances. The unemployed were excluded from this group as they were previously identified under the Labour force status variable.

with the Proc Logistic procedure in SAS.⁴² Before undertaking this analysis, the relation was examined on the basis of multi-way cross tabulations (controlling for gender and age). These revealed that the distribution of sample responses across the levels of a number of the measures was insufficient to allow for the testing of models that included two or more socioeconomic indicators. To do so, would have violated a number of the assumptions that underpin the goodness of fit statistics (that is, at least 10 respondents in each group, 80% of expected counts at least 5, all other expected counts greater than 2, and no zero counts). The tabular analysis did reveal however, that it was possible to undertake a more detailed examination of the relation between socioeconomic position and income non-response by confining the analysis to employed respondents whose main income source was either wages or salary, or own business or partnership. Using this sub-set of socioeconomic variables (and categories within variables) the following model was examined:

Table 3 Modelling the association between occupation, income source, age and income non-reporting: men and women

	B	SE	β	Odds ratio	95% CI
<i>Men*</i>					
Intercept	-2.31	0.07	—	—	—
Professionals†	0.165	0.06	0.043	1.18	1.03, 1.34
White collar	-0.072	0.09	-0.014	0.930	0.76, 1.12
Own business or partnership‡	1.23	0.06	0.247	3.44	3.02, 3.93
30–49 years§	-0.211	0.07	-0.058	0.810	0.69, 0.93
15–29 years	-0.552	0.09	-0.137	0.575	0.47, 0.69
<i>Women*</i>					
Intercept	-2.30	0.12	—	—	—
Professionals†	0.562	0.11	0.142	1.75	1.39, 2.20
White collar	0.484	0.11	0.133	1.62	1.30, 2.01
Own business or partnership‡	1.28	0.08	0.204	3.61	3.04, 4.29
30–49 years§	-0.530	0.08	-0.145	0.588	0.49, 0.70
15–29 years	-0.992	0.11	-0.255	0.371	0.29, 0.46

*Non-reporting (men = 1233, women = 950), reporting (men = 11 336, women = 8843). †Reference category = blue collar. ‡Reference category = wages and salary. §Reference category = 50–64 years old.

$\Pr(\text{non-response}) = \alpha + \beta_1 \text{Professionals} + \beta_2 \text{White Collar} + \beta_3 \text{Own Business or Partnership} + \beta_4 \text{30–49 years} + \beta_5 \text{15–29 years}$.

The reference categories were blue collar workers, wage and salary earners and persons aged 50–64 years respectively. Separate models were specified for men and women, and the maximum likelihood estimates for each indicator variable are presented, including odds ratios and 95% confidence intervals. Goodness of fit of the model was assessed by Q_L , the residual deviance, which is distributed with a χ^2 distribution. In addition, the same model was used to generate a probability matrix that estimated the likelihood of someone not reporting their income given a particular combination of age adjusted socioeconomic characteristics. Other researchers examining the relation between income and health have also used this approach.⁴³

Results

Of the 33 434 respondents who were receiving an income from one or more sources, 9.8% (n=3276) did not report the amount of their income. There was no evidence that income non-reporting was related with gender: men 10.2% (1707 of 16 640) and women 9.3% (1569 of 16 794). Age was strongly associated with income non-reporting: 15–29 years (5.8%, 611 of 10 397), 30–49 years (10.6%, 1723 of 16 222) and 50–54 years (13.8%, 942 of 6815). This pattern was similar for both men and women.

Table 2 presents data comparing socioeconomic groups in terms of the proportion who did not report their income. If we accept that a 4% difference is indicative of a meaningful effect, then there was evidence of an association between labour force status and income non-response: rates were lowest for the unemployed and highest among those in paid work. Rates of income non-reporting were not obviously associated with occupation for men, although there was a suggestion of a relation for women, with respondents in professional occupations being more likely than their counterparts in blue collar jobs to not report their income. Among men there was little evidence that education level was related to income non-response, however, for women, there was a suggestion that these variables were associated. A strong pattern of association was evident between main income source and income non-response for both men and women. Rates were highest among those who received income from their own business or partnership and lowest among those who obtained their income from government pensions or benefits (which excluded the unemployed).

Table 3 presents the maximum likelihood estimates from a logistic regression analysis that examined the relation between socioeconomic position and income non-response on the basis of a model that included occupation, income source and age. The deviance statistics indicate that each model fitted the data adequately (for men, $Q_L = 17.8$, df 12, $\Pr > \chi^2 = 0.120$; women $Q_L = 15.5$, df 12, $\Pr > \chi^2 = 0.214$). For men, those in receipt of income from their

Table 4 Estimated probabilities of not reporting income based on a logistic model that included age, occupation and main income source as explanatory variables*

Occupation	Main income source	Age group		
		15–29	30–49	50–64
Men				
Professionals	Own business or partnership	0.18	0.24	0.28
White collar	Own business or partnership	0.15	0.20	0.24
Blue collar	Own business or partnership	0.16	0.21	0.25
Professionals	Wages or salary	0.06	0.08	0.10
White collar	Wages or salary	0.05	0.06	0.08
Blue collar	Wages or salary	0.05	0.07	0.08
Women				
Professionals	Own business or partnership	0.19	0.27	0.38
White collar	Own business or partnership	0.17	0.25	0.36
Blue collar	Own business or partnership	0.11	0.17	0.26
Professionals	Wages or salary	0.06	0.09	0.14
White collar	Wages or salary	0.05	0.08	0.13
Blue collar	Wages or salary	0.03	0.05	0.09

*Separate models were specified for men and women.

own business or partnership were more than three times as likely as wage and salary earners to not report their income (OR 3.44, 95% CI 3.02, 3.93). Professional men were also relatively more likely to not report income than were male blue collar workers (OR 1.18, 95% CI 1.03, 1.34). Male white collar workers were no different than their counterparts in blue collar occupations in terms of their propensity to not report their income. For women, those in professional (OR 1.75, 95% CI 1.39, 2.20) and white collar (OR 1.62, 95% CI 1.30, 2.01) occupations were more likely to not report their income than women in blue collar jobs. The odds of income non-reporting were also higher for women who received income from their own business or partnership: this group was over three times as likely as wage and salary earners to not report their income (OR 3.61, 95% CI 3.04, 4.29). For both men and women, a strong association between age and income non-reporting was evident.

Table 4 presents the results of a logistic regression that estimated the probability of income non-reporting given different combinations of age adjusted socioeconomic characteristics. The pattern of estimated probabilities highlights three important trends. Firstly, the probabilities are noticeably differentiated on the basis of income source. Irrespective of gender, occupation and age, respondents who receive income from their own business or partnership are much more likely to not report the amount they earn compared with their counterparts whose main income source is wages and salary. Secondly, the fact that the estimated probability of income non-reporting varied little by occupation within the same income source group suggests that this latter factor is a far more important determinant of non-response than the former. This is made particularly evident by examining some of the estimated probabilities. Male professionals who earned their income from their own business or partnership were much more likely to not report their income compared with their professional counterparts on a wage and salary. The same pattern is evident when different occupational groups are compared. Rates of income non-reporting are higher among female blue collar workers in their own business or

partnership than among female professionals on a wage or salary. Thirdly, independent of gender, occupation and income source, age has a marked effect on the probability of income non-reporting. For example, among female professionals whose main income source was their own business or partnership, the estimated probability of not reporting income is 38% if aged 50–64 and 19% if aged between 15 and 29.

Discussion

Item non-response associated with income reporting is determined in part by the data collection method and the type of question used to elicit income information. The Australian Health Survey used in this study collected data using face to face interviews and respondents were asked to report their income by nominating an exact amount. The overall rate of income non-response was 9.8%. Compared with the findings of previous studies,^{13–21} this rate is at the lower bound of the range of non-response values.

Rates of income non-reporting did not differ by gender. Propensity to report income was, however, strongly related to age, with older respondents being more likely to not report their income. It remains to be determined why older persons are less likely to report their income in the context of survey research. The only known study to have examined this issue focused on an elderly sample (mean age 73 years).¹⁹ Factors influencing income non-response among elderly retired persons are likely, in the main, to be qualitatively different from those that shape and circumscribe the reporting decisions of “older” persons of working age. Increasing conservatism, however, may be one factor that is common to both age groups. As people age they possibly become more concerned about issues of privacy and confidentiality, and feel less comfortable discussing personally and socially sensitive topics such as income with strangers.¹⁹

The central question examined in this study was whether rates of income non-response varied by socioeconomic position, and the possible implications of any variation in terms of health inequalities research more generally. We know from previous studies that low socioeconomic groups are least likely to participate in survey research.^{31–35} If it was also the case that rates of income non-response were high among those low socioeconomic groups who do participate, then income based estimates of socioeconomic differences in health were likely to be considerably under-estimated. Four measures of socioeconomic position were used in this study: labour force status, occupation, highest educational qualification and main income source. Given that this study was based on a large national sample that achieved a very high response rate across dwellings (91.5%) and people within dwellings (97.0%), we can be confident that each of these four socioeconomic groups were adequately represented.

The results pertaining to occupation and education indicated that the patterning, direction and strength of association with income

non-response was neither sufficiently consistent nor strong to warrant the conclusion that these variables were related in any important or meaningful way. The age adjusted descriptive analysis provided little evidence of a relation between occupation, education and income non-response for men. For women, however, there was a suggested association among these variables, with women in professional occupations and those with a bachelor degree or higher being less likely to report their income. A subsequent logistic regression analysis detected significant differences between some of the occupation groups, although the effect size for men was negligible, and for women, moderate. In both cases it was respondents in higher status occupations who were least likely to report the amount of their personal income. The probability analysis provided further confirmation of these trends. If we interpret the forgoing evidence in light of the general shape of the relation between income and health (fig 1) it seems that minimal (if any) income related bias will result from sample losses that are generated as a consequence of excluding respondents who did not report their income. At best, sample losses are likely to be similar for all points on the socioeconomic distribution, and at worst, the losses will be slightly greater among the higher socioeconomic groups. Given the curvilinear association between income and health however, either scenario will not affect the interpretation and generalisability of this relation in any important way.

Stronger associations were detected between income non-response, labour force status and main income source. The descriptive analysis showed that rates were high among the employed and those who received an income via their own business or partnership, and low among the unemployed, those on a wage or salary and those in receipt of income via a government pension or benefit. A number of interpretations of these findings are possible. Firstly, they may be viewed as evidence of the impact of socioeconomic factors, for they imply that higher socioeconomic position is associated with greater propensity to not report income, and this is consistent with earlier results relating to occupation and education. The probability analysis, however, raises doubts about the veracity of this interpretation. For both men and women, and each age group, the between-occupation variability in propensity to not report income was considerably smaller than variability between the two income source categories. If socioeconomic position were an important factor determining income non-response, we would have needed to have observed the reverse of this (that is, large occupational variability within each income source category).

A second possible interpretation is that the strong association between income source and income non-response is attributable to a “recall” effect. Low rates of income non-reporting, for example, may be a consequence of being in receipt of a regular income from a single, clearly identifiable source, where the amount is constant across extended periods of

KEY POINTS

- Statistical techniques that adjust for missing data assume item non-response is distributed randomly: the results of this study challenge this assumption.
- Older persons and those in higher socioeconomic groups are less likely to report their income in a face to face interview.
- Estimates of the relation between income and health are unlikely to be adversely affected by socioeconomic variability in income non-reporting.

time. Those on a wage or salary, the unemployed, and those in receipt of other government pension or benefits are likely to know and remember how much they receive, and therefore, are more likely to be in a position to report their income in the context of a face to face interview. This is less likely to be the case for those persons in receipt of incomes from diffuse and less regular sources such as profit or loss from their own business or partnership. Previous studies examining general response error¹² and more specific research focusing on income non-response^{17,18} provide support for this interpretation. This work shows that both knowledge and memory underpin a person’s confidence and ability to provide an accurate and an immediate response to a survey question about income.

A third possible interpretation is that rates of income non-response are higher among business people because of a concern about the subsequent reporting of their income information to taxation authorities. The Australian Health Survey is conducted by the Australian Bureau of Statistics, both of which are very “official” sounding entities. Although as part of its ethical protocol, the ABS provides assurances of privacy and confidentiality, these may not be sufficient to assuage the concerns of those in receipt of incomes from their own business or partnership.

In light of all the foregoing evidence, it is concluded that estimates of the relation between income and health are unlikely to be significantly affected by any non-random error associated with socioeconomic variability in income non-reporting. The proportion of respondents who failed to report their income was not substantially different across the occupation and education groups, and the observed differences by labour force status and income source were unlikely to be attributable to socioeconomic influences. We need however, to remain circumspect about the extent to which these results can be generalised. As was noted earlier, income non-response is in part a function of the data collection methods and question format and wording. Had these been different then it would seem reasonable to presume that the overall level of income non-response would also have been different and that a different patterning of association would be evident with socioeconomic position. Clearly, further research is needed to ascertain the reliability of this study’s findings and to

determine the extent to which they can be inferred to other survey and data collection contexts. Finally, the results of this study tell us nothing about whether socioeconomic position is related to the accuracy of income reporting. At present, very little is known about this issue, yet it represents a potential source of bias that may affect estimates of the association between income and health.

Funding: none.

Conflicts of interest: none.

- 1 Kaplan GA, Lynch JW. Whither studies on the socioeconomic foundations of population health? *Am J Public Health* 1997;87:1409–11.
- 2 Fein O. The influence of social class on health status: American and British research on health inequalities. *J Intern Med* 1995;10:577–86.
- 3 Feinstein JS. The relationship between socioeconomic status and health: a review of the literature. *Milbank Q* 1993;71:279–322.
- 4 Berkman LF, Macintyre S. The measurement of social class in health studies: old measures and new formulations. *International Agency for Research on Cancer* 1997;138:51–64.
- 5 Liberatos P, Link BG, Kelsey JL. The measurement of social class in epidemiology. *Epidemiol Rev* 1988;10:87–121.
- 6 Pappas G, Queen S, Hadden W, Fisher G. The increasing disparity in mortality between socioeconomic groups in the United States, 1960 and 1986. *N Engl J Med* 1993;329:103–9.
- 7 McDonough P, Duncan GJ, Williams D, House J. Income dynamics and adult mortality in the United States, 1972 through 1989. *Am J Public Health* 1997;87:1476–83.
- 8 Martelin T. Mortality by indicators of socioeconomic status among the Finnish elderly. *Soc Sci Med* 1994;38:1257–78.
- 9 Blaxter M. *Health and Lifestyles*. London: Tavistock/Routledge, 1990.
- 10 Lynch JW, Kaplan GA, Salonen JT. Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic life-course. *Soc Sci Med* 1997;44:809–19.
- 11 Lynch JW, Kaplan GA, Salonen R, et al. Socioeconomic status and progression of carotid atherosclerosis: prospective evidence from the Kuopio Ischemic Heart Disease Risk Factor Study. *Arterioscler Thromb Vasc Biol* 1997;17:513–19.
- 12 Sudman S, Bradburn NM. *Asking questions: A critical guide to questionnaire design*. California: Jossey-Bass Inc, 1982:195.
- 13 Wright JD. A research note of the non-reporting of income: The Dutch Case. *Sociologica Neerlandica* 1975;2:86–93.
- 14 Locander WB, Burton JP. The effect of question form on gathering income data by telephone. *Journal of Marketing Research* 1976;13:189–92.
- 15 Siemiatycki J. A comparison of mail, telephone and home interview strategies for household health surveys. *Am J Public Health* 1979;69:238–45.
- 16 Jordan LA, Marcus AC, Reeder LG. Response styles in telephone and household interviewing: a field experiment. *Public Opinion Quarterly* 1980;44:210–22.
- 17 Hippler HJ, Hippler G. Reducing refusal rates in the case of threatening questions: the 'door in the face' technique. *Journal of Official Statistics* 1986;2:25–33.
- 18 Kormendi E. The quality of income information in telephone and face to face surveys. In: Groves RM, Biemer PP, Lyberg LE, et al. eds. *Telephone survey methodology*. Canada: John Wiley, 1988:341–56.
- 19 Martin-Matthews A, Ross LR, Cousineau AEG. Obtaining income information from elderly respondents: an assessment of non-response and reliability. *Canadian Journal on Aging* 1991;10:177–98.
- 20 Guadagnoli E, Cleary PD. Age-related item nonresponse in surveys of recently discharged patients. *J Gerontol* 1992;47:206–12.
- 21 De Leeuw ED. *Data quality in mail, telephone and face to face surveys*. Amsterdam: TT-Publikaties, 1992:59.
- 22 Catania JA. A Model for investigating respondent-interviewer relationships in sexual surveys. In: Bancroft J, ed. *Researching sexual behaviour: methodological issues*. Indianapolis: Indiana University Press, 1997:419.
- 23 Little RJA, Rubin DB. *Statistical analysis with missing data*. New York: John Wiley, 1987.
- 24 Brick JM, Kalton G. Handling missing data in survey research. *Stat Methods Med Res* 1996;5:215–38.
- 25 Craig CS, McCann JM. Item nonresponse in mail surveys: extent and correlates. *Journal of Marketing Research* 1978;15:285–89.
- 26 Messmer DJ, Seymour DT. The effects of branching on item nonresponse. *Public Opinion Quarterly* 1982;46:270–7.
- 27 Colsher PL, Wallace RB. Data quality and age: health and psychobehavioural correlates of item nonresponse and inconsistent responses. *J Gerontol* 1989;2:45–52.
- 28 Davis PB, Yee RL, Chetwynd J, et al. The New Zealand partner relations survey: methodological results of a national telephone survey. *AIDS* 1993;7:1509–16.
- 29 Slymen DJ, Drew JA, Wright BL, et al. Item nonresponse to lifestyle assessment in an elderly cohort. *Int J Epidemiol* 1994;23:583–91.
- 30 Dengler R, Roberts H, Rushton L. Lifestyle surveys—the complete answer? *J Epidemiol Community Health* 1997;51:46–51.
- 31 Sheikh K, Mattingly S. Investigating non-response bias in mail surveys. *J Epidemiol Community Health* 1981;35:293–6.
- 32 Vestbo J, Rasmussen FV. Baseline characteristics are not sufficient indicators of non-response bias in follow up studies. *J Epidemiol Community Health* 1992;46:617–19.
- 33 Prendergast MJ, Beal JF, Williams SA. An investigation of non-response bias by comparison of dental health in 5-year-old children according to parental response to a questionnaire. *Community Dent Health* 1993;10:225–34.
- 34 Melton LJ, Dyck PJ, Karnes JL, et al. Non-response bias in studies of diabetic complications: the Rochester diabetic neuropathy study. *J Clin Epidemiol* 1993;46:341–8.
- 35 Etter JF, Perenger TV. Analysis of non-response bias in a mailed health survey. *J Clin Epidemiol* 1997;50:1123–8.
- 36 Turrell G, Najman JM. Collecting food-related data from low socioeconomic groups: how adequate are our current research designs? *Aust NZ J Public Health* 1995;19:410–16.
- 37 Backlund E, Sorlie PD, Johnson NJ. The Shape of the relationship between income and mortality in the United States: evidence from the National Longitudinal Mortality Study. *Ann Epidemiol* 1996;6:12–20.
- 38 House JS, Kessler RC, Herzog AR. Age, socioeconomic status and health. *Milbank Q* 1990;68:383–411.
- 39 Australian Bureau of Statistics. *National Health Survey: Users Guide*. Canberra: Australian Government Publishing Service, 1996. Catalogue No. 4363.0.
- 40 Australian Bureau of Statistics. *Australian Standard Classification of Occupations: Dictionary*. Canberra: Australian Government Publishing Service, 1987.
- 41 Mathers C. *Health differentials among adult Australians aged 25–64 years*. Australian Institute of Health and Welfare: Health Monitoring Series No.1. Canberra: Australian Government Publishing Service, 1994.
- 42 Stokes ME, Davis CS, Koch GG. *Categorical data analysis using the SAS system*. Cary, NC: SAS Institute Inc, 1995.
- 43 Saunders P. Poverty and Health: exploring the links between financial stress and emotional stress in Australia. *Aust NZ J Public Health* 1998; 22:11–16.