Public Health Nutrition: 6(2), 181–189 DOI: 10.1079/PHN2002415

The socio-economic patterning of survey participation and non-response error in a multilevel study of food purchasing behaviour: area- and individual-level characteristics

Gavin Turrell*, Carla Patterson, Brian Oldenburg, Trish Gould and Marie-Andree Roy
Queensland University of Technology, School of Public Health, Victoria Park Road, Kelvin Grove, Brisbane, Queensland 4059, Australia

Submitted 10 April 2002: Accepted 5 September 2002

Abstract

Objective: To undertake an assessment of survey participation and non-response error in a population-based study that examined the relationship between socio-economic position and food purchasing behaviour.

Design and setting: The study was conducted in Brisbane City (Australia) in 2000. The sample was selected using a stratified two-stage cluster design. Respondents were recruited using a range of strategies that attempted to maximise the involvement of persons from disadvantaged backgrounds: respondents were contacted by personal visit and data were collected using home-based face-to-face interviews; multiple call-backs on different days and at different times were used; and a financial gratuity was provided.

Participants: Non-institutionalised residents of private dwellings ($n = 1003$), located in 50 small areas that differed in their socio-economic characteristics.

Results: Rates of survey participation – measured by non-contacts, exclusions, dropped cases, response rates and completions – were similar across areas, suggesting that residents of socio-economically advantaged and disadvantaged areas were equally likely to be recruited. Individual-level analysis, however, showed that respondents and non-respondents differed significantly in their sociodemographic and food purchasing characteristics: non-respondents were older, less educated and exhibited different purchasing behaviours. Misclassification bias probably accounted for the inconsistent pattern of association between the area- and individual-level results. Estimates of bias due to non-response indicated that although respondents and non-respondents were qualitatively different, the magnitude of error associated with this differential was minimal.

Conclusions: Socio-economic position measured at the individual level is a strong and consistent predictor of survey non-participation. Future studies that set out to examine the relationship between socio-economic position and diet need to adopt sampling strategies and data collection methods that maximise the likelihood of recruiting participants from all points on the socio-economic spectrum, and particularly persons from disadvantaged backgrounds. Study designs that are not sensitive to the difficulties associated with recruiting a socio-economically representative sample are likely to produce biased estimates (underestimates) of socio-economic differences in the dietary outcome being investigated.

Keywords
Survey participation
Non-response error
Socio-economic position
Food purchasing

During the last few decades, population-based surveys have been utilised widely by dietary researchers to examine food and nutrient intakes, to estimate the incidence and prevalence of diet-related disease, and to study associations between these factors and subgroups in the population (e.g. based on age, gender, socio-economic position).

An important issue confronting all population-based surveys is non-response error, generated when data on some of the units comprising the selected sample are not collected. Sample estimates derived from population-based studies are useful only if they are a reasonably close approximation of the population parameter[1]. If sample losses are both large and systematic (i.e. non-random), this is likely to introduce bias[2], which in turn will adversely affect the study's results and limit generalisability. Non-response error is a function of two factors: the survey response rate, and the nature and magnitude of the difference between respondents and non-respondents[1,3]. The response rate, usually defined as the number of

**Corresponding author:* Email g.turrell@qut.edu.au

completed cases as a proportion of the number of eligible cases[4], is often taken as a marker of a survey's quality. However, data quality need not necessarily be compromised by a less-than-high response rate if the respondents and non-respondents have similar sociodemographic characteristics, and are similar in terms of their proportionate distribution on the survey's main outcomes[5,6].

In this paper, we undertake a detailed assessment of survey participation and non-response error in a population-based study that examined the relationship between socio-economic position (SEP) and diet. The study, hereafter referred to as the Brisbane Food Study (BFS), used a multilevel research design to collect area- and individual-level data to estimate the relative contributions of environmental, economic and material, inter-personal and intra-personal factors to socio-economic variability in food purchasing behaviour. The first part of the paper examines five components of survey participation – non-contacts, exclusions, dropped cases, response rates and completions – in terms of whether and to what extent these components differ by SEP measured at the area level. The second part then uses individual-level data to compare respondents and non-respondents in terms of their sociodemographic and food purchasing characteristics, and the third part estimates the nature and magnitude of the bias in food purchasing behaviour that was due to non-response.

The investigation of these issues in the context of a population-based dietary survey is important, as evidence from studies of survey participation consistently show that people from socio-economically disadvantaged backgrounds are least likely to respond to, or participate in, survey research[7,8]. As a result, population-based samples typically under-represent the most socio-economically disadvantaged and over-represent the advantaged. The likely consequence of a socio-economically skewed sample is that the magnitude of socio-economic variability in the dietary issue being investigated will be underestimated. Given that socio-economic factors are the primary focus of the BFS, a detailed examination of survey participation and any associated bias was essential prior to any substantive analysis being undertaken. More broadly, however, the results of this study will be instructive for future population-based dietary surveys that use a similar research design and methodology to that described here.

Methods

Scope and coverage

The BFS was conducted in the Brisbane City Statistical Sub-Division and included households living in private dwellings. A household was defined as a group of two or more related or unrelated people who usually reside within the same private dwelling and who make common provision for food and other essentials for living; or a person living in a private dwelling who makes provision for his/her own food and other essentials for living without combining with any other person. A private dwelling was defined as a house, flat or home unit. Households within non-private dwellings such as hotels and motels, hostels and caravan parks, and persons in hospitals, nursing and convalescent homes, prisons and military establishments, were excluded.

Sample size, design and selection

The study sample consisted of 50 small areas (Census Collectors Districts, CCDs) and 1000 households. Decisions about appropriate sample sizes were underpinned by a range of considerations, including costs and operational constraints, the study's aims, the level of disaggregation and the accuracy of the survey estimates. We also took into account the 'pioneering' nature of the study and its emphasis on description and explanation rather than hypothesis testing, and that the results would be expressed using confidence intervals rather than *P*-values.

The study sample was selected using a stratified two-stage cluster design. Stratification consisted of grouping the 1517 CCDs comprising Brisbane City into 10 strata (deciles) based on each CCD's Index of Relative Socio-economic Disadvantage (IRSD) score. The IRSD is constructed by the Australian Bureau of Statistics using principal components analysis, and is derived from attributes such as low income, low educational attainment, high levels of public sector housing and high unemployment (among others)[9]. CCDs are deemed to be relatively homogeneous in terms of the socio-economic characteristics of the dwellings that they contain. Five CCDs were selected from each stratum using systematic without-replacement probability proportional to size sampling, and the socio-economic profiles of these are presented in Table 1. As would be predicted from the stratification process, there are strong (and often graded) associations between the strata and their social and economic characteristics. Sampling lists were then prepared for each of the 50 CCDs, which involved assigning an identification number to every dwelling in the area. Simple random sampling was then used to select the dwellings, with additional dwellings being identified as replacements for non-contacts, refusals and abandoned premises. A random selection procedure was not used to select individuals within dwellings; rather, given the focus of the study, we used a purposive approach and selected the person in the household who was primarily responsible for most of the food shopping.

Data collection

Data collection took place between September and December 2000, and was conducted on the basis of face-to-face interviews. Data collection proceeded as follows. First, letters of invitation were personally delivered to each of the selected dwellings. The letters

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press

Table 1 Comparing sample strata in terms of their socio-economic characteristics (%)*†

Socio-economic stratum	Unemployment rate	Dwellings with no motor vehicles	Single-parent families	Low-income families‡	Labourers	Left school under 15 years
1 (low)	18.5 (7.7)	32.0 (13.5)	16.4 (7.8)	50.6 (10.7)	15.8 (6.1)	55.2 (3.2)
2	11.5 (3.6)	21.0 (10.7)	10.1 (5.4)	42.7 (8.1)	11.3 (4.5)	44.1 (10.2)
3	11.9 (4.7)	14.0 (4.3)	12.1 (2.7)	36.4 (7.0)	11.0 (2.5)	44.5 (4.5)
4	13.8 (2.9)	19.1 (11.0)	7.7 (2.6)	30.4 (4.2)	7.2 (2.9)	30.6 (8.9)
5	6.8 (2.4)	13.2 (6.9)	11.0 (4.5)	27.4 (4.7)	6.7 (2.6)	35.9 (5.4)
6	6.7 (1.3)	13.4 (6.9)	11.0 (2.0)	28.4 (4.4)	7.5 (1.1)	35.1 (5.1)
7	6.4 (2.3)	12.5 (4.9)	8.2 (1.2)	23.6 (3.4)	4.3 (1.5)	35.2 (2.3)
8	5.5 (1.8)	9.0 (6.1)	8.7 (0.3)	26.1 (6.5)	4.4 (2.5)	29.1 (7.6)
9	5.4 (1.3)	5.2 (3.7)	7.2 (1.1)	15.3 (3.7)	5.0 (1.6)	27.8 (4.1)
10 (high)	5.1 (2.6)	6.3 (7.6)	7.2 (2.7)	16.5 (7.9)	3.2 (1.1)	15.7 (6.2)
Overall mean	9.2 (5.4)	14.6 (10.6)	10.0 (4.3)	29.7 (14.2)	7.7 (4.6)	35.3 (11.8)

*Based on 1996 Census data.
† Mean and standard deviation for the five Census Collectors Districts in each stratum.
‡ Families receiving AUS$20 000 per annum or less.

informed the residents about the study and how they had been selected, and advised that an interviewer would visit in the next two days to ascertain whether they were willing to participate, and if so, to arrange a suitable time to conduct the interview. The letters also offered each potential participant a small financial gratuity (AUS$10), and it provided a guarantee of confidentiality and anonymity. A total of 2123 letters of invitation were distributed throughout the data collection period.

If no household resident was contacted after three separate visits (24.3%, 517/2123), they were classified as a 'non-contact' and a replacement dwelling was selected. Considerable effort went into minimising non-contacts, including calling back at different times of the day (i.e. mornings, afternoons and evenings) and on different days (i.e. weekdays and weekends). Once contact was made, the interviewers were required to establish the potential participant's eligibility for inclusion in the study. Persons were deemed ineligible if they did not speak English well enough to complete an interview, if they exhibited an obvious cognitive/mental disorder or physical disability (e.g. blindness, deafness), if they represented a perceived safety risk to the interviewer (all of whom were female), or if they were very elderly and infirm. Of those persons who were contacted ($n = 1606$), 6.2% were excluded for one or more of these reasons, resulting in 1506 eligible contacts. Of this group, 32.4% ($n = 488$) subsequently refused to participate in the study. At this point, the interviewers were instructed to ascertain if the person refusing was willing to complete a short non-response card that consisted of four questions. Of the 488 people who refused to participate in the main study, 27.4% ($n = 134$) agreed to complete the non-response card. Finally, of those who agreed to participate in the study, a small number ($n = 18$, 1.7%) subsequently failed to keep their appointment with the interviewer. If after three separate call-backs contact was not re-established, these people were dropped and replacements selected. A diagrammatic summary of the data collection process is presented in Fig. 1.

Components of survey participation

As part of the management and administration of the BFS, detailed records were kept on the progress and outcome of every letter of invitation that was distributed. This was undertaken for the specific purposes of examining the socio-economic patterning of survey participation and determining whether and to what extent differential participation by SEP introduced non-response bias. The components of survey participation examined in this paper, and their operational definitions, are as follows:

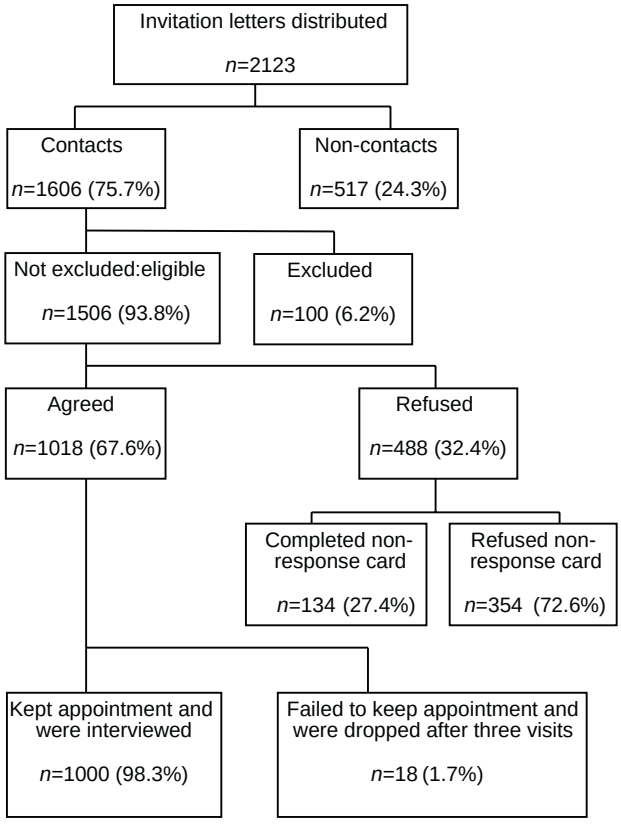

Fig. 1 Flow diagram of survey participation in the Brisbane Food Study

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press

- *Non-contact rate:* the number of persons in each CCD who could not be contacted after three visits, as a proportion of the number that received a letter of invitation.
- *Exclusion rate:* the number of persons in each CCD who when contacted was deemed ineligible to participate, as a proportion of the number of successful contacts.
- *Drop rate:* the number of persons in each CCD who initially agreed to participate, but who subsequently did not keep their interview appointment and could not be contacted after three visits, as a proportion of the number who agreed to participate.
- *Response rate:* the number of persons in each CCD who completed an interview, as a proportion of the number of eligible cases (excluding those who did not keep their appointment).
- *Completion rate:* the number of persons in each CCD who completed an interview, as a proportion of the number that received a letter of invitation.
- *Completed the non-response card rate:* the number of people in each CCD who completed the non-response card, as a proportion of the number who refused to participate in the study.

Measures

Examination of the six components of survey participation was undertaken on the basis of SEP measured at the stratum level. Data not presented here showed that each of the 10 strata was significantly different from the others when compared on their IRSD scores (reflecting each stratum's overall SEP). We can also see from Table 1 that each of the strata differed in terms of its socio-economic characteristics when measured using individual components of the IRSD, and many of these pair-wise differences were also statistically significant (data not presented).

The individual-level analysis of non-response was undertaken on the basis of the four questions that were included on the non-response card. The card sought information about the usual bread purchasing choice of the person in the household who did most of the food shopping, and their gender, age and education level. Bread was selected for the non-response card as it represented a commonly purchased and readily available food. More importantly however, different types of bread were able to be classified in terms of their 'healthiness', with the Australian Dietary Guidelines being used as the reference point for this assessment (e.g. wholemeal bread was considered healthier than white bread by virtue of its higher fibre content)[10]. Furthermore, based on earlier (unpublished) research, we found that the purchase of 'healthy' varieties of bread was indicative of a generally more healthy food purchasing pattern; thus bread choice represented a key indicator food, predictive of other food purchasing choices. As part of the non-response card, people were asked, 'When shopping, what type of bread do they usually buy?' (referring to the person who does most of the household's food purchasing). The response options included 'do not buy bread', 'white', 'wholemeal', 'multigrain', 'white high in fibre', 'rye', 'soy and linseed' and 'other (please specify)'. Multiple responses were permitted. Education level was originally measured using a 10-category question that asked about the person's highest qualification completed since leaving school. When undertaking bivariate analysis this measure was collapsed into four categories: no additional qualifications; vocational (trade or business certificate, or apprenticeship); diploma (associate or undergraduate); bachelor degree or higher (degree, postgraduate diploma, masters or doctorate). When undertaking multivariate analysis, cell size considerations necessitated that the re-coded education measure be further collapsed into two groups: those with or without educational qualifications since leaving school. Identically worded and formatted questions about bread choice, education, age and gender were included as part of non-response card and the face-to-face interviews.

Statistical analyses

In order to take account of the complex sample design and resultant cluster-correlated data, statistical analyses were undertaken using SUDAAN version 8.0[11]. Data pertaining to the six components of survey participation were coded as binary outcomes, with the value 1 being assigned to non-contacts, exclusions, dropped cases, responses and completions (including the non-response card), and 0 otherwise. Socio-economic differences in survey participation were examined by comparing the 10 strata on the basis of crude rates, and odds ratios (ORs) with 95% confidence intervals (CIs) (calculated using logistic regression).

The individual-level analyses involved comparing respondents and non-respondents on the basis of gender, education and bread purchasing using the chi-square test, and the *t*-test for independent samples was used to compare the mean ages of the two groups. The effect of non-response on socio-economic, gender and age differences in bread purchasing was assessed by calculating odds ratios using logistic models that combined both respondents and non-respondents, and then repeated for respondents only. Differences between the odds ratios for the two models were then expressed as a percentage, which served as an indicator of the over- or under-estimation in odds ratios that was due to non-response error. This approach has been used in other studies of survey participation[12].

Results

Table 2 compares the socio-economic strata in terms of the components of survey participation. Overall, there was little evidence that the socio-economic areas differed, and, where significant differences were found, the patterning

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press

Table 2 Comparing socio-economic strata in terms of the components of survey participation

Socio-economic stratum	Non-contacts*		Exclusions		Response		Completion		Completed non-response card	
	Rate (%)	OR (95% CI)	Rate (%)	OR (95% CI)	Rate (%)	OR (95% CI)	Rate (%)	OR (95% CI)	Rate (%)	OR (95% CI)
1 (low)	26.3	1.45 (0.93−2.25)	12.1	1.57 (0.77−3.21)	65.7	1.15 (0.72−1.83)	42.5	0.86 (0.59−1.25)	44.0	2.92 (1.22−6.98)
2	21.1	1.08 (0.67−1.72)	10.4	1.33 (0.64−2.77)	64.9	1.11 (0.70−1.76)	45.8	0.99 (0.68−1.44)	35.1	2.02 (0.84−4.82)
3	32.7	1.97 (1.27−3.03)	1.9	0.22 (0.06−0.80)	66.2	1.17 (0.73−1.87)	43.6	0.90 (0.62−1.31)	25.4	1.27 (0.51−3.18)
4	24.0	1.28 (0.80−2.03)	5.0	0.60 (0.24−1.48)	66.2	1.17 (0.73−1.87)	47.8	1.07 (0.73−1.57)	12.2	0.52 (0.17−1.53)
5	22.6	1.18 (0.74−1.88)	3.0	0.35 (0.12−1.02)	62.8	1.01 (0.64−1.60)	47.1	1.04 (0.71−1.52)	32.7	1.81 (0.76−4.30)
6	27.1	1.51 (0.96−2.37)	7.3	0.90 (0.39−2.05)	71.9	1.53 (0.94−2.50)	48.5	1.10 (0.75−1.61)	25.6	1.28 (0.48−3.42)
7	27.4	1.53 (0.97−2.39)	2.5	0.30 (0.09−0.94)	66.2	1.17 (0.73−1.87)	46.7	1.02 (0.70−1.49)	25.4	1.27 (0.51−3.18)
8	22.1	1.15 (0.71−1.85)	5.9	0.72 (0.30−1.72)	70.4	1.42 (0.88−2.31)	51.5	1.24 (0.84−1.83)	17.5	0.79 (0.27−2.26)
9	18.5	0.92 (0.56−1.51)	4.5	0.54 (0.21−1.38)	68.0	1.27 (0.79−2.04)	52.9	1.31 (0.88−1.94)	31.8	1.73 (0.69−4.36)
10 (high)	19.8	1.00	8.0	1.00	62.5	1.00	46.0	1.00	21.1	1.00
Overall	24.3		6.2		66.4		47.1		27.4	
	517/2123		100/1606		1000/1506		1000/2123		134/488	

OR – odds ratio; CI – confidence interval.
*Operational definitions for each of these components of survey participation are presented in the Methods section.

of effects did not suggest any directionality. In other words, individuals living in socio-economically disadvantaged areas appeared to be no less likely to have participated in the BFS. A number of specific findings, however, deserve brief mention. First, exclusion rates were highest in the two most disadvantaged areas, and although these rates were not significantly different when compared with the reference group (i.e. stratum 10), it was clear that these rates were considerably more elevated than those observed for the other strata. Second, results pertaining to dropped cases (not presented in Table 2) showed that the rates were 1.96 per 100 in stratum 1, and 7.41 per 100 in stratum 10. Again, although this difference was not statistically significant, it suggests that individuals living in socio-economically advantaged areas found it more difficult to keep pre-arranged appointments. Third, compared with the reference group, individuals in the most disadvantaged area were nearly three times as likely (OR = 2.92, 95% CI 1.22–6.98) to have completed the non-response card.

Table 3 presents data that compare respondents and non-respondents in terms of their sociodemographic characteristics and bread purchasing choices. Non-respondents were significantly older ($t = 4.88$, $P < 0.0001$) and less educated ($\chi^2 = 58.19$, df = 5, $P < 0.0001$) than their counterparts who participated in the main study; however, the gender distributions of the two groups were similar. In terms of bread choice, non-respondents were significantly more likely to purchase wholemeal bread ($\chi^2 = 10.40$, df = 1, $P < 0.001$), but less likely than respondents to purchase multigrain bread ($\chi^2 = 8.73$, df = 1, $P < 0.003$) and soy and linseed bread ($\chi^2 = 6.26$, df = 1, $P < 0.012$). The purchasing patterns of respondents and non-respondents were similar for the other types of bread.

Table 4 examines the associations between gender, age, education and bread purchasing choice, and provides an assessment of the mis-estimation in effect size that was due to non-response error. Compared with males, females were nearly 2.5 times more likely to purchase white high-fibre bread (OR = 2.48, 95% CI 1.49–4.11). Older persons exhibited a healthier bread purchasing profile: those aged 46 years or more were less likely than their younger counterparts to purchase white bread (OR = 0.51, 95% CI 0.40–0.66) and more likely to purchase wholemeal (OR = 1.82, 95% CI 1.42–2.33) and multigrain bread (OR = 1.51, 95% CI 1.16–1.97). These age differences were independent of the potentially confounding effects of education (i.e. on average, younger persons are more highly educated). Respondents who did not attain educational qualifications after leaving school were more likely to purchase white bread (OR = 1.51, 95% CI 1.17–1.94) and less likely to purchase multigrain bread (OR = 0.45, 95% CI 0.34–0.59). These results were independent of age.

Overall, the degree of mis-estimation of effects due to non-response error was small, ranging (in absolute

Table 3 Sociodemographic characteristics and bread purchasing behaviours of respondents and non-respondents

	Respondents (*n*=1003)*	Non-respondents (*n*=134)	*P*-value
Sex (%)			
Male	21.5	16.4	
Female	77.9	83.5	0.269
Age (years)			
Range	16–94	19–90	
Mean (SD)	45.6 (16.8)	53.2 (18.9)	0.0001
Highest educational qualification since leaving school (%)			
No additional qualifications	41.4	61.9	
Vocational	18.8	14.1	
Diploma	10.5	3.7	
Bachelor degree or higher	26.8	11.1	
Other (not easily classifiable)	2.2	4.4	
Missing or refused	0.3	4.4	0.0001
Bread purchase (%)			
Buys bread	96.6	97.7	0.480
White bread	57.6	51.1	0.162
Wholemeal bread	39.4	54.2	0.001
Multigrain bread	33.4	20.6	0.003
White bread high in fibre	15.9	12.2	0.274
Rye bread	7.3	3.8	0.137
Soy and linseed bread	11.8	4.5	0.012
Other bread	5.4	3.8	0.357

SD – standard deviation.
* After all data had been collected it was discovered that three extra interviews had been conducted. These three additional cases were not included in the analysis of survey participation (Table 2); however, they were retained for all other analyses.

Table 4 Associations between bread purchase, gender, age and education, for the total group and respondents only, and percentage of over- or underestimation in odds ratio due to non-response

Type of bread purchased	Total group (respondents and non-respondents)				Respondents only				% OR error*
	Gender†				Gender				
	Male	Female	OR	95% CI	Male	Female	OR	95% CI	
White‡	53.9	57.5	1.17	0.87–1.58	54.5	58.3	1.17	0.85–1.60	0.0
Wholemeal	44.7	40.4	0.81	0.60–1.10	44.5	38.2	0.76	0.55–1.04	−6.2
Multigrain	27.3	33.0	1.30	0.94–1.80	28.2	34.7	1.35	0.96–1.90	3.8
White high in fibre	8.2	17.4	2.48	1.49–4.11	8.1	18.1	2.60	1.52–4.44	4.8
	Age group (years)§				Age group (years)				
	≤45	≥46	OR	95% CI	≤45	≥46	OR	95% CI	
White	63.7	49.3	0.51	0.40–0.66	64.6	48.9	0.49	0.37–0.64	−3.9
Wholemeal	34.0	48.7	1.82	1.42–2.33	33.5	46.6	1.70	1.31–2.21	−6.6
Multigrain	29.3	35.3	1.51	1.16–1.97	29.5	38.3	1.64	1.25–2.17	8.6
White high in fibre	19.2	11.4	0.53	0.38–0.76	19.2	11.9	0.56	0.39–0.81	5.7
	Qualifications since leaving school†¶				Qualifications since leaving school				
	Yes	No	OR	95% CI	Yes	No	OR	95% CI	
White	54.1	60.5	1.51	1.17–1.94	54.1	62.3	1.60	1.22–2.09	6.0
Wholemeal	39.1	43.2	1.02	0.79–1.31	37.3	42.3	1.14	0.85–1.45	11.8
Multigrain	39.1	23.2	0.45	0.34–0.59	39.4	25.1	0.49	0.36–0.65	8.9
White high in fibre	15.9	15.1	1.07	0.76–1.50	16.1	15.7	1.06	0.74–1.52	−0.93

OR – odds ratio; CI – confidence interval.
* Percentage over- or under-estimation in odds ratios due to non-response: (OR for respondents minus OR for total group)/OR for total group.
† Logistic analysis, age-adjusted.
‡ Excludes respondents who reportedly never purchased bread ($n = 37$).
§ Logistic analysis adjusted for education (highest qualification).
¶ Excludes missing and refused ($n = 9$).

terms) from zero to 11.8%; moreover, the directionality of the mis-estimation was mixed. If we consider only the statistically significant differences, the data in Table 4 show that the impact of gender on bread purchase was slightly overestimated for white high-fibre bread, as was the impact of age on the purchase of multigrain bread and that of education on the purchase of wholemeal and multigrain bread. However, the impact of age on the

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press

purchase of white and wholemeal bread was slightly underestimated as a result of non-response error. The absolute average extent of mis-estimation for gender, age and education was 3.7%, 6.2% and 6.9% respectively, further confirming that the extent of error due to non-response was minimal.

Discussion

The first section of this study's analysis examined six components of survey participation – non-contacts, exclusions, dropped cases, response rates and completions – in terms of whether and to what extent these differed by SEP measured at the area level. Based on the results, there was little evidence that individuals living in socio-economically disadvantaged areas were less likely to be recruited into, or willingly participate in, the research. We need, however, to take a somewhat circumspect view of this result, and be mindful of the fact that this aspect of the study involved comparing individuals classified by their socio-economic area of residence and not on the basis of the socio-economic characteristics of the individuals themselves. In short, when interpreting the results of the area analyses, we need to be cautious of committing the ecological fallacy[13,14]. It is not necessarily the case that relationships (significant or otherwise) found between higher or aggregated units of analysis (e.g. CCDs) also apply to units at a lower or disaggregated level (e.g. individuals). If this study's analysis of the components of survey participation were undertaken on the basis of individual-level socio-economic characteristics, the results may have been different. Unfortunately, this was not possible, as we had no data on the socio-economic characteristics of those individuals who could not be contacted, who were deemed ineligible or refused any involvement with the study.

Despite finding little or no overall association between area-level SEP and the components of survey participation, a number of specific findings warrant brief discussion. First, as part of the letter of invitation, potential survey participants were offered a small financial gratuity (AUS$10), as research had shown that gratuities have a significant positive effect on response rates[15]. In this study, however, response rates were very similar across the 10 strata. This possibly suggests that the financial gratuity had no differential effect on the individuals living in the different socio-economic areas, or, alternatively, that it operated with greater effect in poorer areas, thus raising participation rates from a lower-than-average base level to a level equivalent to that of the more advantaged areas. Second, the exclusion rate was considerably raised in the two most disadvantaged strata. Information collated on sample exclusions showed that collecting data from socio-economically disadvantaged areas was a more difficult task than was the case with the more advantaged areas. Interviewers reported that CCDs comprising the two most disadvantaged strata had greater numbers of recent immigrants from non-English speaking backgrounds, more people with an apparent cognitive disorder, a greater number of households that were unapproachable owing to the presence of savage dogs, and more people who were assessed as being a potential safety risk. Third, contrary to what would have been predicted, residents in the most disadvantaged stratum (stratum 1) were significantly more likely than their counterparts in stratum 10 to agree to complete the non-response card. This counter-intuitive result is difficult to explain.

Comparing respondents and non-respondents in terms of their sociodemographic characteristics and bread purchasing behaviours was the focus of the second part of the analyses. Residents who completed the non-response card were found to be older and less educated than those who participated in the main study. Non-respondents also exhibited different bread purchasing behaviours, with this group being more likely to purchase wholemeal bread and less likely to purchase multigrain and soy and linseed bread. These results are consistent with the findings of other individual-level studies which show that non-respondents are qualitatively different from respondents[16–18] and that non-respondents are more likely to come from socio-economically disadvantaged backgrounds[19,20]. These individual-level findings, however, contradict some of the evidence pertaining to the area-level comparisons. Whilst the former approach suggests that individuals who refused to participate were more likely to be socio-economically disadvantaged, the latter suggested that refusal rates were similar across all areas, irrespective of their socio-economic profile. Misclassification bias is likely to explain this inconsistency. Research has shown that the socio-economic characteristics of areas represents an error-prone basis for classifying the socio-economic characteristics of individuals who reside in those areas, leading at best to the underestimation of socio-economic effects or, at worst, to the 'incorrect' finding of no association with SEP[21]. Sometimes, due to a lack of readily available individual-level data, policies and interventions directed at individuals are based on area-level evidence. Whilst ecological data are useful for these types of planning or decision making, the results of this present study, and earlier work by others, suggest caution when making inferences to individuals based on the characteristics of the areas or contexts they inhabit.

Given that the 'refusers' who completed the non-response card were disproportionately different from those who participated, it is reasonable to expect that the group who did not complete the card possibly differed in even more extreme ways. As a consequence, the study sample used for the BFS may under-represent the most socio-economically disadvantaged and those with the least healthful diets. This problem probably characterises most population-based dietary surveys[8].

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press

The third section of this study estimated the nature and magnitude of bias in food purchasing that was due to non-response. Importantly, we used bread purchasing as a marker for dietary quality more generally. Indeed, analyses conducted on other population-based dietary studies in Brisbane indicated that this approach was reasonable, as the purchase of 'healthy' bread (i.e. those varieties consistent with choices promulgated in the dietary guidelines) was predictive of healthy choices for other types of food. On reflection, this makes sense. People select 'healthy' varieties of bread for specific reasons: choosing a healthy option from a range of alternatives is a conscious, non-random act, underpinned by cognitive/rational factors (e.g. knowledge of the link between diet and health) and affective/emotional factors (e.g. food preference). These influences clearly do not pertain exclusively to a single food (in this case, bread), but also inform decisions about the choice of other foods[22]. Consistent with other studies, the results of this present work showed that women, older persons and socio-economically advantaged groups were more likely to choose 'healthy' food[22–25]. These results also accord with more broad-ranging investigations of the relationship between sociodemographic factors and compliance with dietary guideline recommendations[26,27].

Estimates of bias associated with non-response indicated that although respondents and non-respondents were qualitatively different, the magnitude of error associated with this differential was minimal. In the paper's introduction it was mentioned that a study's data quality is dependent in part on the overall response rate, the degree of difference between respondents and non-respondents in terms of their sociodemographic characteristics, and the extent to which these groups differed on the main outcome variables. The BFS achieved a moderate overall response rate of 66.4%, and while the respondents and non-respondents differed in terms of their age and education, the analysis of non-response suggested that any bias associated with this would be small. It would appear, therefore, that the methodology described in this paper produced data of acceptable quality, and that future findings can be cautiously generalised to the wider population.

Acknowledgements

Funding for this research was provided by a National Health and Medical Research Council Project Grant (No. 101217) and support from the Victorian Health Promotion Foundation (VicHealth). Dr Turrell is a Senior Research Fellow supported by a National Health and Medical Research Council/National Heart Foundation Career Development Award (CR 01B 0502).

References

1 Locker D. Response and non-response bias in oral health surveys. *J. Public Health Dent.* 2000; **60**: 72–81.
2 Andersen R, Kasper J, Frankel MR, eds. *Total Survey Error*. San Francisco, CA: Jossey-Bass, 1979.
3 Groves RM, Couper MP. *Non-response in Household Interview Surveys*. New York: Wiley, 1998.
4 Groves RM. *Survey Errors and Survey Costs*. New York: Wiley, 1989.
5 Dillman DA. The design and administration of mail surveys. *Annu. Rev. Sociol.* 1991; **17**: 225–49.
6 Locker D. Effects of non-response on estimates derived from an oral health survey of older adults. *Community Dent. Oral Epidemiol.* 1993; **21**: 108–13.
7 Turrell G. Income non-reporting: implications for health inequalities research. *J. Epidemiol. Community Health* 2000; **54**: 207–14.
8 Turrell G, Najman JM. Collecting food-related data from low socioeconomic groups: how adequate are our research designs? *Aust. J. Public Health* 1995; **19**: 410–6.
9 Australian Bureau of Statistics. *1996 Census of Population and Housing: Socioeconomic Indexes for Areas*. Information Paper, Catalogue No. 2039.0. Canberra: Australian Bureau of Statistics, 1998.
10 National Health and Medical Research Council. *Dietary Guidelines for Australians*. Canberra: Australian Government Publishing Service, 1991.
11 Research Triangle Institute. *SUDAAN User's Manual, Release 8.0*. Research Triangle Park, NC: Research Triangle Institute, 2001.
12 Hoeymans N, Feskens EJM, Van Den Bos GAM, Kromhout D. Non-response bias in a study of cardiovascular diseases, functional status and self-rated health among elderly men. *Age and Ageing* 1998; **27**: 35–40.
13 Blalock HM. *Conceptualization and Measurement in the Social Sciences*. Beverley Hills, CA: Sage, 1989.
14 Macintyre S, Ellaway A. Ecological approaches: rediscovering the role of the physical and social environment. In: Berkman LF, Kawachi I, eds. *Social Epidemiology*. New York: Oxford University Press, 2000; 332–48.
15 Roberts PJ, Roberts C, Sibbald B, Torgerson DJ. The effect of a direct payment or a lottery on questionnaire response rates: a randomised controlled trial. *J. Epidemiol. Community Health* 2000; **54**: 71–2.
16 Hill A, Roberts J, Ewings P, Gunnell D. Non-response bias in a lifestyle survey. *J. Public Health Med.* 1997; **19**: 203–7.
17 O'Neill TW, Marsden D, Silman AJ and the European Vertebral Osteoporosis Study Group. Differences in the characteristics of responders and non-responders in a prevalence survey of vertebral osteoporosis. *Osteoporosis Int.* 1995; **5**: 327–34.
18 Van den Akker M, Buntinx F, Metsemakers JFM, Knottnerus JA. Morbidity in responders and non-responders in a register-based population survey. *Fam. Pract.* 1998; **15**: 261–3.
19 Sheikh K, Mattingly S. Investigating non-response bias in mail surveys. *J. Epidemiol. Community Health* 1981; **35**: 293–6.
20 Melton LJ, Dyck PJ, Karnes JL, O'Brien PC, Service FJ. Non-response bias in studies of diabetic complications: the Rochester Diabetic Neuropathy Study. *J. Clin. Epidemiol.* 1993; **46**: 341–8.
21 Hyndman JCG, Holman CD, Hockey RL, Donovan RJ, Corti B, Rivera J. Misclassification of social disadvantage based on geographical areas: comparison of postcode and collector's districts analyses. *Int. J. Epidemiol.* 1995; **24**: 165–76.
22 Turrell G. Determinants of healthy food choice in a

population-based sample. *Am. J. Health Behav.* 1998; **22**: 342–57.
23 Patterson BH, Block G. Food choices and the cancer guidelines. *Am. J. Public Health* 1988; **78**: 282–6.
24 Turrell G. Determinants of gender differences in dietary behavior. *Nutr. Res.* 1997; **17**: 1105–20.
25 Shimakawa T, Sorlie P, Carpenter MA, Dennis B, Tell GS, Watson R, *et al.* Dietary intake patterns and sociodemographic factors in the Atherosclerosis Risk in Communities Study. *Prev. Med.* 1994; **23**: 769–80.
26 Worsley A, Crawford D. Who complies with the Australian Dietary Guidelines?. *Nutr. Res.* 1986; **6**: 29–34.
27 Turrell G. Compliance with the Australian Dietary Guidelines: have population-based health promotion messages been effective? *Nutr. Health* 1997; **11**: 271–88.

https://doi.org/10.1079/PHN2002415 Published online by Cambridge University Press