2012

The impact of complex survey design on prevalence estimates of intakes of food groups in the Australian National Children's Nutrition and Physical Activity Survey

Sandy Burden
University of Wollongong, sburden@uow.edu.au

Yasmine Probst
University of Wollongong, yasmine@uow.edu.au

David G. Steel
University of Wollongong, dsteel@uow.edu.au

Linda C. Tapsell
University of Wollongong, ltapsell@uow.edu.au

Publication Details

Burden, S., Probst, Y., Steel, D. G. & Tapsell, L. C. 2012, 'The impact of complex survey design on prevalence estimates of intakes of food groups in the Australian National Children's Nutrition and Physical Activity Survey', Public Health Nutrition, vol. 15, no. 8, pp. 1362-1372.

The impact of complex survey design on prevalence estimates of intakes of food groups in the Australian National Children's Nutrition and Physical Activity Survey

Keywords
physical, impact, activity, complex, survey, estimates, prevalence, design, intakes, food, groups, australian, national, children, nutrition

Disciplines
Arts and Humanities | Life Sciences | Medicine and Health Sciences | Social and Behavioral Sciences

Publication Details
Burden, S., Probst, Y., Steel, D. G. & Tapsell, L. C. 2012, 'The impact of complex survey design on prevalence estimates of intakes of food groups in the Australian National Children's Nutrition and Physical Activity Survey', Public Health Nutrition, vol. 15, no. 8, pp. 1362-1372.

 doi:10.1017/S1368980011003326

The impact of complex survey design on prevalence estimates of intakes of food groups in the Australian National Children's Nutrition and Physical Activity Survey

Sandy Burden[1,*], Yasmine Probst[2], David Steel[1] and Linda Tapsell[2]
[1]Centre for Statistical and Survey Methodology, University of Wollongong, Northfields Avenue, Wollongong, New South Wales 2522, Australia: [2]Smart Foods Centre, University of Wollongong, Wollongong, Australia

Submitted 3 February 2011: Accepted 9 November 2011

Abstract

Objective: To assess the impact of the complex survey design used in the 2007 Australian National Children's Nutrition and Physical Activity Survey (ANCNPAS07) on prevalence estimates for intakes of groups of foods in the population of children.
Design: The impacts on prevalence estimates were determined by calculating design effects for values for food group consumption. The implications of ignoring elements of the sample design including stratification, clustering and weighting are discussed.
Setting: The ANCNPAS07 used a complex sample design involving stratification, a high degree of clustering and estimation weights.
Subjects: Australian children aged 2–16 years.
Results: Design effects ranging from <1 to 5 were found for the values of mean consumption and proportion of the population consuming the food groups. When survey weights were ignored, prevalence estimates were also biased.
Conclusions: Ignoring the complex survey design used in the ANCNPAS07 could result in underestimating the width of confidence intervals, higher mean square errors and biased estimators. The magnitude of these effects depends on both the parameter under consideration and the chosen estimator.

Keywords
Design effects
Complex survey design
Dietary intake

The degree of complexity in survey design depends on the nature of the research question, just as the method of data collection influences the choice of sampling technique. For example, straightforward telephone interviews allow relatively simple sample designs, but comprehensive nutrition surveys tend to be longer and more complex. Moreover, collection of reliable anthropometric data involves face-to-face interviewing. To improve efficiency and reduce costs, nutrition and physical activity surveys often use complicated sample designs, involving stratified multistage sampling techniques. These designs can include the use of stratification, clustering and unequal probabilities of selection for different individuals. The resulting sample is not spread evenly throughout the population, but occurs in groups or clusters.

The 2007 Australian National Children's Nutrition and Physical Activity Survey (ANCNPAS07)[1] was undertaken to obtain food, nutrient, physical activity and anthropometric data on a national sample of children aged 2–16 years. The purpose of the survey was to enable food, beverage, supplement and nutrient intakes and physical activity levels among children to be assessed against relevant national guidelines. For these data, the use of clustering, unequal selection probabilities, stratification and sample weighting led to estimates having a sampling variance different from that which would have been obtained using a simple random sample (SRS). An SRS gives each possible sample the same chance of selection and means that the sample is spread approximately evenly through the population. For an SRS the calculation of estimates and associated standard errors is relatively straightforward. Standard methods of statistical analysis assume that an SRS has been obtained.

If the analysis of a nutrition survey ignores the complex design, the results will be methodologically unsound and subject to serious dispute. Typically confidence intervals will be too small, leading to inflation of type I error rates. That is, statistical significance is found when there is no real effect. The problem is not solved merely by using the sample weights which account for differences in selection probabilities, although this is often incorrectly assumed. In fact this view is implicitly encouraged if the survey data are released with, for example, no cluster information. Even when an analysis uses sample weights and the

*Corresponding author: Email sburden@uow.edu.au

contribution of the clustering to the overall design effect is low, use of standard analysis will not reflect the impact of the weights on variances. The use of sampling weights and the impact of complex sampling methods on survey analysis have received considerable attention over the past two decades[2–5].

The importance of properly accounting for sampling weights and the sample design is strongly emphasized in well-established surveys in the USA, for example the National Health and Nutrition Examination Survey (NHANES). Information on the NHANES website[6] states that:

> For NHANES datasets, the use of sampling weights and sample design variables is recommended for all analyses because the sample design is a clustered design and incorporates differential probabilities of selection. If you fail to account for the sampling parameters, you may obtain biased estimates and overstate significance levels.

Moreover the National Center for Health Statistics (NCHS) Analytic and Reporting Guidelines[7] state (p. 7) that 'Sample weights and the stratification and clustering of the design must be incorporated into an analysis to get proper estimates and standard errors of estimates' and that proper variance estimation procedures be used.

The aim of the study reported here was to assess the impact of complex survey design used in the ANCNPAS07 on prevalence estimates for intakes of groups of foods in the population of Australian children.

Methods

Data for the present analysis were obtained from the Australian Social Science Data Archive, Australian National University[1] under agreed conditions for research. The data made available for the analysis were de-identified and separate ethical approval was not required. The study reported here used the concept of design effects to quantify the effect of the sample design on prevalence estimates. For each estimate, design effects were used to measure the impact of stratification, clustering, unequal inclusion probabilities and other features of the sampling used. The design effect (*deff*) is the ratio of the sampling variance obtained using a complex survey design relative to the variance that would have been obtained from a simple random sample without replacement (SRSWOR) with the same expected sample size[8]. The *deff* for a parameter θ is calculated using the relationship $deff = \hat{V}(\hat{\theta}) \big/ \hat{V}_{\text{srswor}}(\tilde{\theta}_{\text{srs}})$, where $\hat{V}(\hat{\theta})$ is the design-based estimate of the variance for the parameter estimate $\hat{\theta}$ from a complex survey of size n, and $\hat{V}_{\text{srswor}}(\tilde{\theta}_{\text{srs}})$ is the variance estimate of the parameter $\tilde{\theta}_{\text{srs}}$ estimated from a similar hypothetical survey using SRSWOR and a sample size of n.

A design effect greater than one increases the width of confidence intervals, reduces the amount of disaggregation that is possible and reduces the power of analyses that are properly carried out. This limits the strength and value of the results. For example, suppose a survey has been designed using standard methods which assume an SRS to give a power of detecting important effects of 80 %. With a *deff* of 1·5 the power reduces to 65 %; for a *deff* of 2 it becomes 45 %; and for *deff* of 4 it is 35 %. Tests of statistical significance are also affected and a *deff* of 4 increases the conventional 5 % false positive rate used in hypothesis testing to 33 %.

A design effect can also be expressed as the effective sample size, $n_{\text{eff}} = n/deff$. For example, a sample of 4000 respondents has an effective sample size of 1000 if the *deff* is 4. So selecting several respondents within a cluster will be less efficient in terms of variance than using SRS. This has substantial implications for the way in which the survey data may be acceptable to the wider community and used in policy development.

ANCNPAS07 was conducted using a sampling scheme stratified by state/territory and by capital city statistical division/rest of state into thirteen strata. The number of children included from each state was proportional to the population of children in that state or territory. To collect physical activity data, anthropometric measurements and a 24 h diet history, an initial face-to-face interview was used. To facilitate the face-to-face interviewing and to help meet budget and time restrictions, fifty-four regions encompassing 246 postcodes were selected which were effectively primary sampling units (PSU). Initial selection and contact was made using random digit dialling (RDD) for telephone prefixes broadly corresponding to the selected postcodes and data were collected using computer-assisted personal interviewing (CAPI) and subsequent computer-assisted telephone interviewing (CATI).

The use of RDD led to the inclusion of individuals from 481 postcodes, due to both the overlap between telephone number prefixes and postcodes and telephone number portability. A location variable was created after selection so that participants in a close geographic proximity, based on their postcodes, were given the same location. These locations and the fifty-four PSU were both used to calculate design effects to demonstrate the impact of using incorrect sample information. In the sample design one stratum contained only one PSU, so to enable variance calculations the PSU was divided in half, resulting in fifty-five clusters. Similarly, within each stratum, locations with single observations were grouped, reducing the total number of locations from 210 to 194. We found the responding sample to be highly clustered with an average of 88 children per cluster (for the CAPI), varying from 19 to 161 (SD 29). Locations had an average size of 25 responding children, varying from 2 to 177 (SD 29).

In the sample, one child was selected per household leading to 4837 selected children and complete data for 4487 children. Within selected clusters, the probability of selection of a child depended on his/her location (stratum), age, gender and household composition. To account for the non-proportional sampling, weights were created based on age (divided into four groups), gender and stratum. A single weight, called the initial weight and denoted w_1, was produced for each child in the survey on the basis of a sample size of $n = 4837$. The initial weight for child i, $i = 1,\ldots,n$, equals $w_{1i} = N_h/n_h$, where n_h is the number of respondents in stratum h and N_h is the corresponding population. Separate weights were not included for respondents who did not complete all components of the survey, and household size and family structure were not included in the weights. The probability of selection was therefore only partially accounted for by the weights. Furthermore, as there were complete nutrient data from the CAPI for only 4826 of the 4837 participants, the population totals using the weights did not correspond to those for Australia available from the 2006 Census.

Due to the limitations in the weighting process, a final weight (w_2) was created by adjusting w_1 to fit population benchmarks and accounting for the probability of selection of each child in a household. Assuming that all children in a household had an equal probability of selection, the probability of selection for each subject was calculated as $\pi_i = 1/$(number of children aged 2–16 in household). The total effective sample size in stratum h using the final weight was $\hat{n}_h = \sum_{i\in h}\pi_i^{-1}$ and using initial weights was n_h. The final weight for each subject in stratum h was obtained using $w_{2i} = w_{1i}\pi_i^{-1}\left(n_h/\hat{n}_h\right)$.

Sample weights are used to produce an estimate which is less biased than its unweighted counterpart. However, the increased accuracy must be balanced against the increased design effect[9]. One approach to choosing the most efficient estimator is to examine the mean square error (MSE $=$ bias^2 $+$ variance) for each parameter[10]. For multiple variables, the relative mean square error (RMSE) can be used. Assuming the final weights produced unbiased estimates, the RMSE for the estimate of a mean ($\bar{Y}$) is $RMSE_{w_2} = \left[V\left(\bar{Y}_{w_2}\right)\right]/\bar{Y}_{w_2}$, where $V\left(\bar{Y}_{w_2}\right)$ is the variance of the estimate, calculated using the final weight w_2. Otherwise, $RMSE_{w_1} = \left[\left(\bar{Y}_{w_1} - \bar{Y}_{w_2}\right)^2 + V\left(\bar{Y}_{w_1}\right)\right]\Big/\bar{Y}_{w_1}$. The estimator with the smallest RMSE is preferred.

The coefficient of variation of the weights is given by $C_w = s_w/\bar{w}$, where s_w is the standard deviation of the weights and $\bar{w}$ is the mean. It measures the increased variance of the estimate due to the use of weights. When selection probabilities are not correlated with a variable, the design effect due to weighting is given by[8,11] $1 + C_w^2$. When correlation is present, approximations can be made[12].

Under some mild assumptions the contribution of sample clustering for the estimation of prevalence of a condition or risk factor is reflected in the relationship $deff = 1 + (\bar{n}-1)\delta$, where $\bar{n}$ is the average number of respondents per cluster and δ is a measure of the within-cluster homogeneity or intra-class correlation (ICC). Values of δ around 0·05 are common, which with $\bar{n} = 81$ gives $deff = 5$ and with $\bar{n} = 11$ gives $deff = 1{\cdot}5$. Hence the more clustered the design the higher the design effect. If the size of the clusters varies considerably, more complicated formulas apply. For applications where the clustering and weighting effects are multiplicative, the design effect is given by[11] $deff = [1 + (\bar{n}-1)\delta]\cdot\left(1 + C_w^2\right)$.

Statistical analysis

Design effects were estimated for the prevalence estimates of food consumption for ANCNPAS07 using the STATA statistical software package release 10 (2007; StataCorp LP, College Station, TX, USA). The variables chosen for analysis were the 120 three-digit sub-major groups used in the food categorization. The parameters chosen for analysis were mean consumption of each sub-major food group in grams and the proportion of the population consuming each food group. The CAPI 24 h recall diet history was used for all analyses.

Estimates and estimates of sampling variances were produced under a number of options for treating the weights and sample design features, including:

1. unweighted analysis assuming SRS;
2. weighted analysis assuming SRS;
3. weighted analysis incorporating stratification (thirteen strata) and clustering using the 194 locations in the data file; and
4. weighted analysis incorporating stratification (thirteen strata) and clustering using the fifty-five PSU.

Analysis under option 1 was the naïve analysis. The estimates and estimated variances were compared with analyses under option 4, which properly reflected the weighting and complex design. Option 2 accounts for the weighting but ignores the sample design and option 3 uses incorrect clusters.

Results

The effect of the survey design was highly variable (Fig. 1), with *deff* values for different food groups and estimators ranging from 0·3 to 5·1. These results are important for the analysis of nutrition surveys because an increase in the design effect affects the significance of the results. For example, a *deff* of 2 increases the width of the confidence interval of an estimator by 1·4 and a *deff* of 4 increases it by 2·0.

A common error is to regard the estimate with the lowest estimated standard error as the best. However, the standard error is only correct when all aspects of the weighting and design are accounted for. For the estimate

Public Health Nutrition

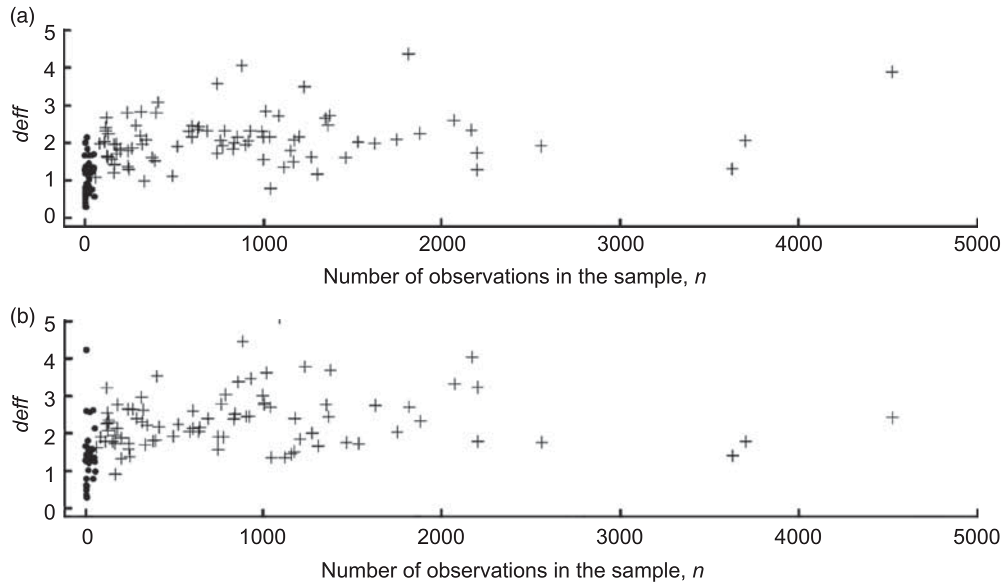

Fig. 1 Final design effects (*deff*) for (a) mean consumption of and (b) proportion of the population who consumed the three-digit food groups of ANCNPAS07 by number of observations, calculated using the final weights (w_2), stratification and clustering of the primary sampling units. Food groups with ≤55 observations are marked by ● and those with >55 observations are marked by + (ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey)

of mean consumption (proportion consuming), forty-five (fifty-seven) groups with *deff* > 2·0 are listed in Table 1 (Table 2). Most estimates were biased when an SRS was assumed and in all cases the confidence intervals were substantially wider when the correct sample design was used for estimation.

In most cases the complex sample design increased the design effect compared with an SRS, but for some food groups design effects less than one were obtained. The low design effects occurred when the number of children consuming foods from a group was less than fifty-five, the number of clusters in the sample design (which occurred in thirty-six food groups), so there was effectively no clustering. For these groups, the observed *deff* averaged only 1·1 (1·4) for the mean (proportion) estimator compared with 2·1 (2·3) for the other groups (Table 3 and Fig. 1) and 44 % of the groups had a final *deff* < 1.

As components of the design effect arise due to each element of the sample design, the impacts of each design element are considered separately in the following sections. In some sections, the notation ($n \le 55$) and ($n > 55$) will be used to split the results if required.

Weighting

The initial and final weights (w_1 and w_2) both had a similar right-skewed distribution with the same mean. The final weights had a higher standard deviation and a wider range due to the inclusion of the additional weighting component (Table 4).

The theoretically expected *deff* for the initial weights is $1 + C_{w_1}^2 = 1{\cdot}3$ and for the final weights is $1 + C_{w_2}^2 = 1{\cdot}6$. These correspond to the average observed design effect due to weighting for groups with $n > 55$. The 0·3 increase between w_1 and w_2 in both the theoretical and average observed *deff* for $n > 55$ occurs because neither set of weights is highly correlated with the response variables (Figs 2(a) and 3(a), Table 3). After weighting, design effects were generally below two and slightly higher for the final weights compared with the initial weights. Only six groups had *deff* > 2.

Weighted estimation increased the design effect, but it also reduced the bias of estimators in the survey. Assuming the estimate obtained using the final weights was unbiased, when an SRS was incorrectly assumed, the percentage bias for the mean$_{(n>55)}$ (prop$_{(n>55)}$) estimator was between −15 % and 16 % (−10 % and 22 %) for 95 % of groups. Using the initial weights the percentage bias was −9 %, 11 % (−4 %, 9 %; Fig. 4). For mean$_{(n \le 55)}$ (prop$_{(n \le 55)}$), the percentage bias had a much wider range for both an SRS, −23 %, 221 % (−43 %, 235 %), and the initial weights, −23 %, 83 % (−39 %, 91 %), and was generally positive.

The distribution of RMSE was similar for estimates obtained using no weights, initial weights and final weights (Fig. 5). On average, the final estimates had the lowest average RMSE (0·05), followed by the initial weighted estimates (0·07) and SRS estimates (0·12).

Stratification

For this survey stratification had very little impact on the design effects. All of the estimates showed no change

Table 1 Food groups with *deff* > 2·0 for mean consumption in grams of each three-digit food group in the ANCNPAS07

		SRS		Clustered		
deff	*N*	Mean	95 % CI	Mean	95 % CI	Name
2·02	1529	13·6	12·6, 14·6	12·9	11·6, 14·2	Other vegetables and vegetable combinations
2·02	109	12·2	9·68, 14·7	11·9	8·30, 15·6	Electrolyte, energy and fortified drinks
2·04	916	21·0	19·3, 22·6	21·3	18·9, 23·7	Poultry and feathered game
2·06	3697	67·1	65·4, 68·9	69·4	66·7, 71·8	Regular breads, and bread rolls (plain/unfilled/ untopped varieties)
2·06	760	5·71	5·18, 6·23	6·44	5·63, 7·27	Potato snacks
2·07	1172	13·7	12·7, 14·8	13·6	12·1, 15·2	Other fruiting vegetables
2·07	344	13·6	11·9, 15·4	13·8	11·2, 16·4	Mixed dishes where beef, veal or lamb is the major component
2·08	1746	5·47	5·08, 5·87	5·89	5·25, 6·52	Sugar, honey and syrups
2·14	13	0·001	0·00, 0·002	0·001	0·00, 0·003	Chemical raising agents and cooking ingredients
2·15	855	5·47	4·92, 6·02	5·70	4·89, 6·51	Leaf and stalk vegetables
2·15	999	17·5	16·2, 18·7	18·3	16·4, 20·3	Cakes, buns, muffins, scones, cake-type desserts
2·16	601	9·07	8·25, 9·90	9·51	8·22, 10·8	Sausages, frankfurters and saveloys
2·16	1036	48·8	45·3, 52·3	48·9	43·8, 54·0	Mixed dishes where cereal is the major ingredient
2·16	1199	9·42	8·68, 10·2	9·7	8·60, 10·8	Chocolate and chocolate-based confectionery
2·19	314	6·61	5·65, 7·56	7·27	5·85, 8·71	Dishes where vegetable is the major component
2·24	132	3·61	2·91, 4·31	4·18	3·05, 5·31	Other dishes where milk or a milk product is the major component
2·24	1875	63·1	60·2, 66·1	65·6	61·2, 70·1	Pome fruit
2·30	992	14·2	13·0, 15·4	14·1	12·3, 15·8	Cordials
2·31	583	1·27	1·14, 1·40	1·22	1·02, 1·41	Dairy blends
2·31	924	7·83	7·19, 8·47	7·71	6·71, 8·71	Peas and beans
2·31	784	25·2	22·9, 27·6	26·1	22·4, 29·8	Other fruit
2·32	117	2·59	2·00, 3·17	2·60	1·72, 3·49	Fin fish (excluding commercially sterile)
2·32	683	4·77	4·38, 5·15	4·90	4·31, 5·49	Cereal-, fruit-, nut- and seed-bars
2·34	2163	20·2	18·9, 21·6	21·2	18·8, 23·6	Gravies and savoury sauces
2·35	635	2·87	2·52, 3·22	2·94	2·39, 3·50	Nuts and nut products
2·41	117	0·67	0·52, 0·82	0·84	0·48, 0·95	Extruded or reformed snacks
2·43	640	17·3	15·5, 19·0	18·3	15·4, 21·3	Mixed dishes where poultry or game is the major component
2·46	601	8·52	7·63, 9·41	8·70	7·29, 10·1	Batter-based products
2·46	287	0·12	0·09, 0·14	0·13	0·06, 0·14	Multivitamin and/or mineral
2·47	1361	13·1	12·3, 14·0	13·0	11·7, 14·3	Processed meat
2·60	2067	53·5	50·9, 56·0	54·4	50·3, 58·6	Potatoes
2·66	1347	26·2	24·6, 27·9	27·2	24·4, 29·8	Muscle meat
2·67	122	7·24	5·63, 8·84	6·61	4·15, 9·06	Dairy milk substitutes, unflavoured
2·71	1085	2·70	2·01, 3·38	2·80	1·47, 4·12	Herbs, spices, seasonings and stock cubes
2·72	1371	124	116, 131	129	117, 142	Soft drinks, and flavoured mineral waters
2·80	400	0·51	0·43, 0·60	0·58	0·41, 0·74	Vegetable/nut oil
2·80	237	1·26	1·03, 1·49	1·21	0·83, 1·59	Cream
2·83	315	22·7	19·7, 25·6	23·0	17·8, 28·2	Soup (prepared, ready to eat)
2·84	1013	15·8	14·6, 17	16·5	14·4, 18·6	Tomato and tomato products
3·08	410	9·90	8·68, 11·1	11·9	9·25, 14·6	Dishes and products other than confectionery where sugar is the main component
3·49	1230	30·5	28·7, 32·4	28·9	25·5, 32·3	Tropical fruit
3·57	742	2·57	2·33, 2·82	2·88	2·31, 3·45	Jam and lemon spreads, chocolate spreads, sauces
3·89	4518	792	774, 810	826	789, 862	Mineral waters and water
4·05	878	29·6	27·0, 32·3	30·9	25·5, 36·2	Flours and other cereal grains and starches
4·37	1813	3·87	3·67, 4·07	3·76	3·34, 4·17	Margarine and table spreads

ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey.
For each group the design effect (*deff*), the number of observations (*N*) and, for both a simple random sample (SRS) and the clustered design (Clustered), the estimated mean consumption and the 95 % confidence interval limits are included in the table.

in *deff* when stratification was included (Figs 2(b) and 3(b), Table 1).

Clustering

The average increase in *deff* due to clustering for mean$_{(n>55)}$ (prop$_{(n>55)}$) was 0·4 (0·7; Table 1), although the change in *deff* was highly variable for different parameters and the different estimators (Figs 2(c) and 2(d), Figs 3(c) and 3(d)). The impact of clustering depended on both the pattern of responses for the variable and the location and size of clusters. When the correct clusters were not used – for example by treating location as the sampling unit – the variance was underestimated, decreasing the average *deff* for mean$_{(n>55)}$ (prop$_{(n>55)}$) by 0·1 (0·2) since the full cluster effect and the variation of locations within clusters was effectively ignored.

Table 2 Food groups with *deff* > 2·0 for proportion of the population who consumed the food group in the ANCNPAS07

		SRS		Clustered		
deff	*N*	Proportion	95 % CI	Proportion	95 % CI	Name
2·01	1268	0·26	0·25, 0·28	0·26	0·24, 0·28	Pasta and pasta products
2·04	1746	0·36	0·35, 0·38	0·36	0·34, 0·38	Sugar, honey and syrups
2·04	635	0·13	0·12, 0·14	0·13	0·12, 0·15	Nuts and nut products
2·05	583	0·12	0·11, 0·13	0·12	0·10, 0·13	Dairy blends
2·07	126	0·03	0·02, 0·03	0·03	0·02, 0·03	Single vitamin
2·13	50	0·01	0·008, 0·013	0·01	0·006, 0·014	Herbal and homeopathic supplements
2·13	601	0·12	0·12, 0·13	0·12	0·11, 0·14	Sausages, frankfurters and saveloys
2·14	175	0·04	0·03, 0·04	0·04	0·03, 0·05	Packed (commercially sterile) fish and seafood
2·16	640	0·13	0·12, 0·14	0·13	0·12, 0·15	Mixed dishes where poultry or game is the major component
2·17	410	0·08	0·08, 0·09	0·09	0·08, 0·11	Dishes and products other than confectionery where sugar is the major component
2·22	344	0·07	0·06, 0·08	0·07	0·06, 0·08	Mixed dishes where beef, veal or lamb is the major component
2·24	518	0·11	0·10, 0·12	0·11	0·10, 0·12	Eggs
2·25	117	0·02	0·02, 0·03	0·02	0·02, 0·03	Fin fish (excluding commercially sterile)
2·29	132	0·03	0·02, 0·03	0·03	0·02, 0·04	Other dishes where milk or a milk product is the major component
2·30	315	0·07	0·06, 0·07	0·06	0·05, 0·07	Soup (prepared, ready to eat)
2·33	1875	0·39	0·37, 0·40	0·39	0·37, 0·41	Pome fruit
2·35	129	0·03	0·02, 0·03	0·02	0·02, 0·03	Dishes where egg is the major ingredient
2·39	829	0·17	0·16, 0·18	0·16	0·14, 0·18	Cabbage, cauliflower and similar brassica vegetables
2·40	683	0·14	0·13, 0·15	0·15	0·13, 0·16	Cereal-, fruit-, nut- and seed-bars
2·40	287	0·06	0·05, 0·07	0·06	0·05, 0·07	Multivitamin and/or mineral
2·40	1172	0·24	0·23, 0·25	0·24	0·22, 0·26	Other fruiting vegetables
2·43	4518	0·94	0·93, 0·94	0·94	0·93, 0·95	Mineral waters and water
2·45	1361	0·28	0·27, 0·29	0·28	0·26, 0·30	Processed meat
2·45	900	0·19	0·18, 0·20	0·19	0·17, 0·21	Citrus fruit
2·45	916	0·19	0·18, 0·20	0·19	0·17, 0·20	Poultry and feathered game
2·51	835	0·17	0·16, 0·18	0·19	0·17, 0·20	English-style muffins, flat breads, and savoury and sweet breads
2·54	122	0·03	0·02, 0·03	0·02	0·02, 0·03	Dairy milk substitutes, unflavoured
2·57	24	0·005	0·003, 0·007	0·005	0·002, 0·008	Canned condensed soup (unprepared)
2·60	4	0·001	0·000, 0·002	0·001	0·000, 0·003	Stuffings
2·60	601	0·12	0·12, 0·13	0·13	0·11, 0·14	Batter-based products
2·62	42	0·009	0·006, 0·011	0·009	0·004, 0·013	Unspecified fats
2·63	322	0·07	0·06, 0·07	0·07	0·05, 0·08	Tea
2·65	266	0·06	0·05, 0·06	0·05	0·04, 0·05	Berry fruit
2·65	237	0·05	0·04, 0·06	0·05	0·04, 0·06	Cream
2·70	1036	0·21	0·20, 0·23	0·22	0·20, 0·24	Mixed dishes where cereal is the major ingredient
2·70	1813	0·38	0·36, 0·39	0·36	0·34, 0·38	Margarine and table spreads
2·74	1622	0·34	0·32, 0·35	0·32	0·30, 0·34	Carrot and similar root vegetables
2·76	178	0·04	0·03, 0·04	0·04	0·03, 0·04	Pickles, chutneys and relishes
2·77	1347	0·28	0·27, 0·29	0·28	0·26, 0·30	Muscle meat
2·77	999	0·21	0·20, 0·22	0·21	0·19, 0·23	Cakes, buns, muffins, scones, cake-type desserts
2·78	760	0·16	0·15, 0·17	0·18	0·16, 0·19	Potato snacks
2·81	1001	0·21	0·20, 0·22	0·22	0·20, 0·24	Other confectionery
2·96	314	0·07	0·06, 0·07	0·07	0·06, 0·08	Dishes where vegetable is the major component
3·01	992	0·21	0·19, 0·22	0·21	0·19, 0·23	Cordials
3·03	784	0·16	0·15, 0·17	0·16	0·14, 0·18	Other fruit
3·22	117	0·02	0·02, 0·03	0·03	0·02, 0·03	Extruded or reformed snacks
3·23	2195	0·45	0·44, 0·47	0·43	0·40, 0·45	Cheese
3·32	2067	0·43	0·41, 0·44	0·43	0·40, 0·45	Potatoes
3·37	855	0·18	0·17, 0·19	0·19	0·17, 0·21	Leaf and stalk vegetables
3·46	924	0·19	0·18, 0·20	0·18	0·16, 0·20	Peas and beans
3·53	400	0·08	0·08, 0·09	0·08	0·06, 0·09	Vegetable/nut oil
3·62	1013	0·21	0·20, 0·22	0·22	0·19, 0·24	Tomato and tomato products
3·69	1371	0·28	0·27, 0·30	0·30	0·27, 0·32	Soft drinks, and flavoured mineral waters
3·78	1230	0·25	0·24, 0·27	0·24	0·21, 0·26	Tropical fruit
4·03	2163	0·45	0·43, 0·46	0·45	0·42, 0·48	Gravies and savoury sauces
4·46	878	0·18	0·17, 0·19	0·18	0·16, 0·21	Flours and other cereal grains and starches
5·09	1085	0·22	0·21, 0·24	0·22	0·20, 0·25	Herbs, spices, seasonings and stock cubes

ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey.
For each group the design effect (*deff*), the number of observations (*N*) and, for both a simple random sample (SRS) and the clustered design (Clustered), the estimated proportion of the population who consumed the food group and the 95 % confidence interval limits are included in the table.

NS Public Health Nutrition

Table 3 Average design effects for the three-digit food groups for mean consumption of (in grams) and proportion of the population who consumed each food group in the ANCNPAS07

	Initial weight		Final weight		Stratification		Stratification & clustering 194 locations		Stratification & clustering 55 clusters	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
$\text{Mean}_{(n\le 55)}$*	1·1	0·32	1·1	0·45	1·1	0·45	1·1	0·44	1·1	0·48
$\text{Prop}_{(n\le 55)}$†	1·2	0·29	1·3	0·69	1·3	0·69	1·4	0·71	1·4	0·75
$\text{Mean}_{(n>55)}$‡	1·3	0·09	1·7	0·32	1·7	0·32	2·0	0·53	2·1	0·66
$\text{Prop}_{(n>55)}$§	1·3	0·05	1·6	0·11	1·6	0·11	2·1	0·53	2·3	0·75

ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey.
The results are split by the number of non-zero observations.
*$\text{Mean}_{(n\le 55)}$, mean estimator for groups with ≤55 observations.
†$\text{Prop}_{(n\le 55)}$, proportion estimator for groups with ≤55 observations.
‡$\text{Mean}_{(n>55)}$, mean estimator for groups with >55 observations.
§$\text{Prop}_{(n>55)}$, proportion estimator for groups with >55 observations.

Table 4 Distributional information for the initial weights (w_1) and final weights (w_2) and their correlation (Corr) with the survey variables

	Mean	Median	SD	Min	Max	C_w
Initial weights, w_1	727	596	419	114	1720	0·58
Final weights, w_2	727	541	577	62·1	5850	0·79
Corr (w_1, y_i)	0·00	0·00	0·03	−0·09	0·08	–
Corr (w_2, y_i)	−0·00	−0·00	0·03	−0·10	0·07	–

C_w, coefficient of variation.

Discussion

The results show that design effects obtained for prevalence estimates of food intakes are highly variable. They depend on many factors including the number of observations, sample weighting and the complex design. The magnitude of design effects can be large, indicating that confidence intervals are significantly wider than may be expected if the complex sample design and/or weighting is ignored.

The results also illustrate that design effects depend on both the chosen parameter and the estimator which is used. The main outputs from a survey frequently consist of prevalence estimates, such a means, proportions and population totals and the design effect for each of these will be different[13,14]. Complex sample design also has an impact on the estimates of parameters of statistical analysis, such as regression parameters from a linear or logistic regression and associated odds ratios, but they differ from the design effect on prevalence estimates and are not considered here. Further complexities are introduced when ratios or post-stratification are used, or if the design effect is calculated for both the population and for estimates for subgroups.

Considering the food groups with greater than fifty-five observations per cluster, the effect of weighting is generally similar to that expected by theory. Design effects due to weighting depend on the coefficient of variation of the weights and the correlation between the weights and the survey variables[9]. The inclusion of a component of weight due to the number of children per household increased the design effect by a relatively small and consistent amount for most food groups in the survey. Those with larger changes were food groups which have a different consumption pattern for households with small or large numbers of children, for example the groups 'dishes other than confectionery where sugar is the main ingredient' and 'jam and lemon spreads, chocolate spreads, sauces'.

A large proportion of the groups with less than fifty-five observations had a design effect of less than one. For these groups, most of the design effect arises due to weighting, with no appreciable change due to stratification and clustering. This occurs because the average number of non-zero observations per cluster is close to one. There is effectively no clustering, so the design effect is also close to one[13]. At face value, $deff < 1$ implies that the complex sampling scheme results in a lower variance than SRS. However, the variability around one is most probably due to estimation of the sample variance. As the effective degrees of freedom may be significantly less than the nominal degrees of freedom (= number of sampled PSU − number of strata = 41)[10] the stability of the variance estimator may be questionable. Sampling error can then cause the observed design effect to vary randomly above and below one[15]. Further investigation of this possibility is beyond the scope of the current paper.

The results also illustrate that choosing the appropriate estimator for a parameter entails a trade-off between bias

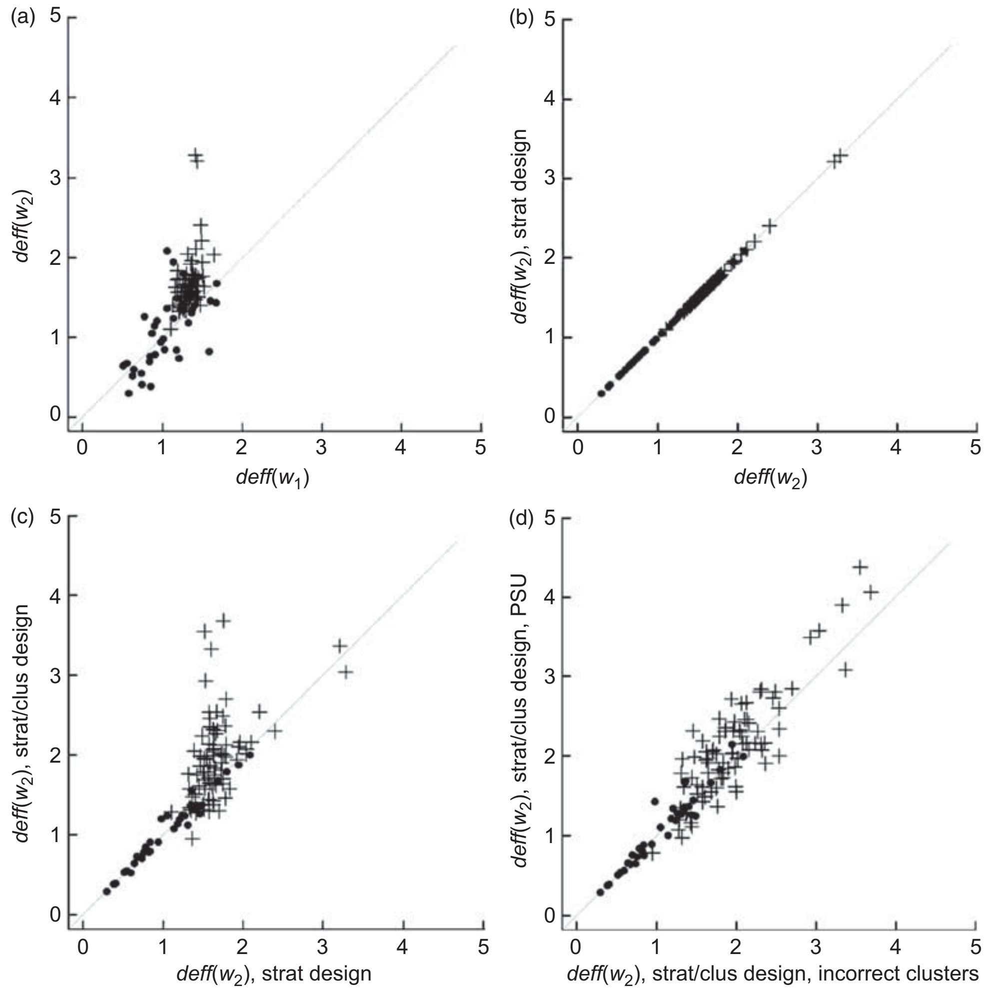

Fig. 2 Design effects (*deff*) for mean population consumption (g) of each three-digit sub-major food group from ANCNPAS07. The contribution of each sample design feature to the magnitude of *deff* is illustrated by the use of estimates obtained using (a) final weights compared with initial weights; (b) stratification; (c) clustering using provided locations (incorrect clusters); and (d) clustering using the original primary sampling units (PSU). w_1 and w_2 refer to the initial and final weights, respectively. Food groups with ≤55 observations are marked by ● and those with >55 observations are marked by + (ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey)

and design effect. Ignoring sample weights or using the wrong weight can result in a biased estimator. However, the use of weights may increase the design effect which affects the potential significance of the results. The effect of weighting also depends on the coefficient of variation of the weights and the correlation between the weights and the survey variables[9]. For the three-digit food groups in ANCNPAS07, if the clustering effect is ignored, the relatively small design effect from weighting means that in some cases an unweighted estimator has lower RMSE and may be preferred. However, it is not always possible to quantify all sources of bias. An alternative to a weighted estimator is to include survey design variables in a model for the variable of interest with unweighted regression estimation[10].

Considering the complex sample design, the effect of stratification is generally to decrease the design effect, because stratification removes one component of variance from the estimator. However, unless there is a large difference between strata the impact on the design effect is small, as it is in this case. The effect of clustering is determined by the number of sample units per cluster and the ICC within each cluster. As the ICC varies between 0 and 0·04 for different variables, the design effect due to clustering is highly variable and estimator specific.

In summary, design effects arise due to the interaction of the sample design and the population structure, so they will be low for universal items which do not vary geographically or by cluster such as the groups 'milk' or 'savoury biscuits'. They are higher when consumption varies by, for example, geographic location, age and/or gender. Design effects arise through both unequal selection probabilities and other elements of the sample

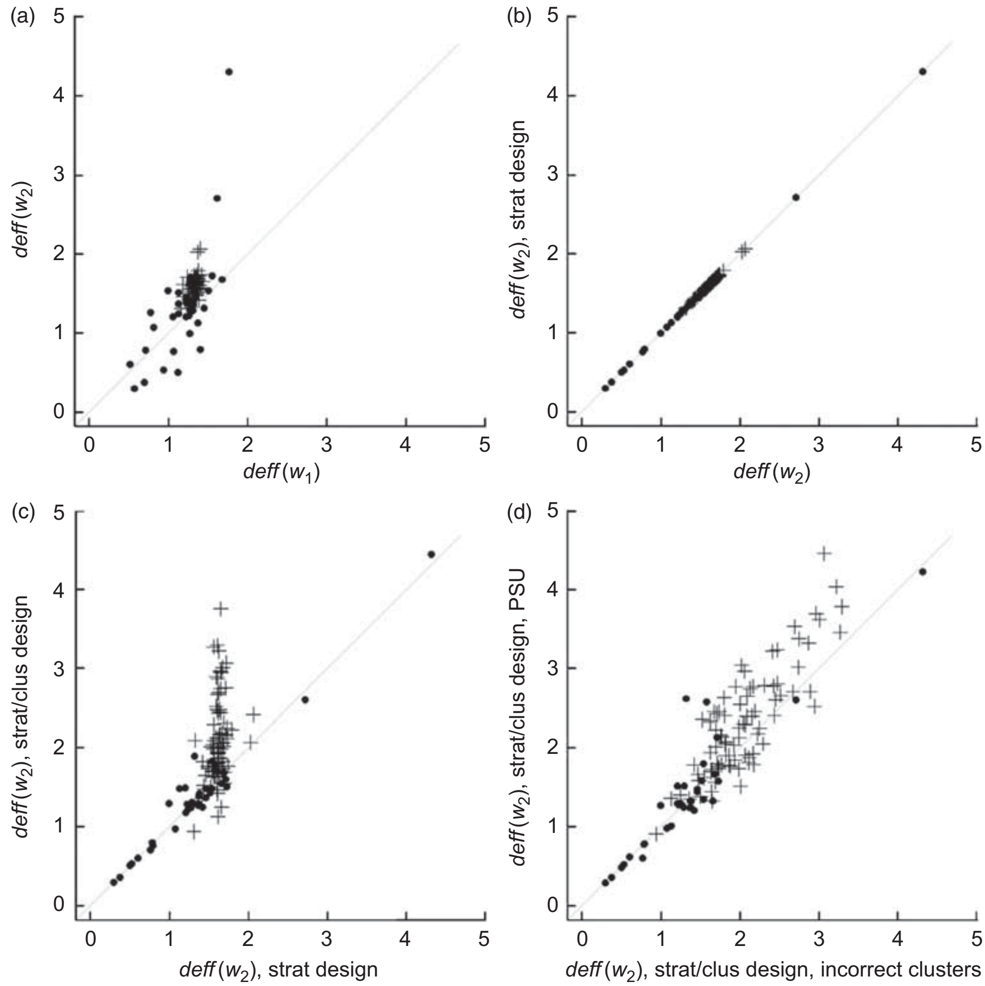

Fig. 3 Design effects (*deff*) for proportion of the population who consumed each three-digit sub-major food group from ANCNPAS07. The contribution of each sample design feature to the magnitude of *deff* is illustrated by the use of estimates obtained using (a) final weights compared with initial weights; (b) stratification; (c) clustering using provided locations (incorrect clusters); and (d) clustering using the original primary sampling units (PSU). w_1 and w_2 refer to the initial and final weights respectively. Food groups with ≤ 55 observations are marked by ● and those with >55 observations are marked by + (ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey)

design such as stratification and clustering. Hence, to obtain accurate standard errors, all of these elements need to be taken into consideration during analysis of complex survey data and stratum and cluster indicator variables must be included with the data.

Developing an appropriate design for a nutrition survey is difficult because there is considerable uncertainty about the values of relevant population characteristics such as the ICC and the likely number of observations. Also these parameters vary between variables and the type of analysis. A high degree of clustering can lead to a large design effect, but reducing the clustering increases costs.

Knowing the approximate magnitude of the design effect for an estimator is useful when designing future surveys. Design effects can be used to estimate the required sample size(s) for future surveys and the use of weighting, the choice of weights and also the degree of clustering and stratification in the survey can be tailored to achieve the desired standard error or power. Design effects are widely applicable and can be calculated for any estimator of a wide range of variables including macro- or micronutrient intakes for the population or for sub-populations. Estimators include means, totals, proportions or more complex estimators such as regression estimators – for which cluster effects are often lower[13].

In conclusion, ignoring complex survey design can result in underestimating the width of confidence intervals, higher mean square errors and biased estimators. The magnitude of these effects depends on both the parameter under consideration and the chosen estimator. It is important to consider the effects of complex sample design when both designing and analysing surveys.

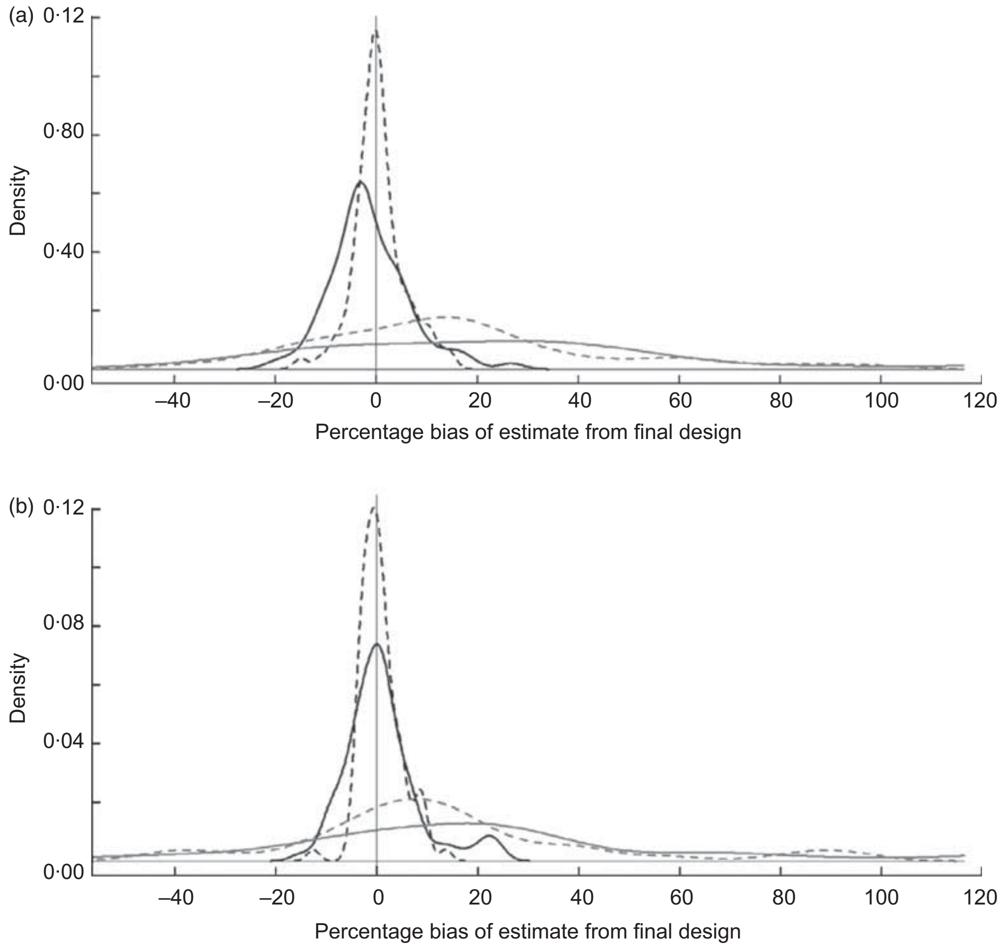

Fig. 4 Density plot of the percentage bias of estimates for (a) mean consumption of and (b) proportion of the population who consumed the three-digit food groups from ANCNPAS07. Solid lines show the simple random sample estimates compared with final weighted estimation (using w_2) and dashed lines show the initial weighted estimation (using w_1) compared with the final weighted estimation (using w_2). w_1 and w_2 refer to the initial and final weights respectively. Biases are separated into two groups representing food groups with $>$55 observations (shown in black) and those with $\leq$55 observations (shown in grey). The effect of complex survey design was not included in these calculations (ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey)

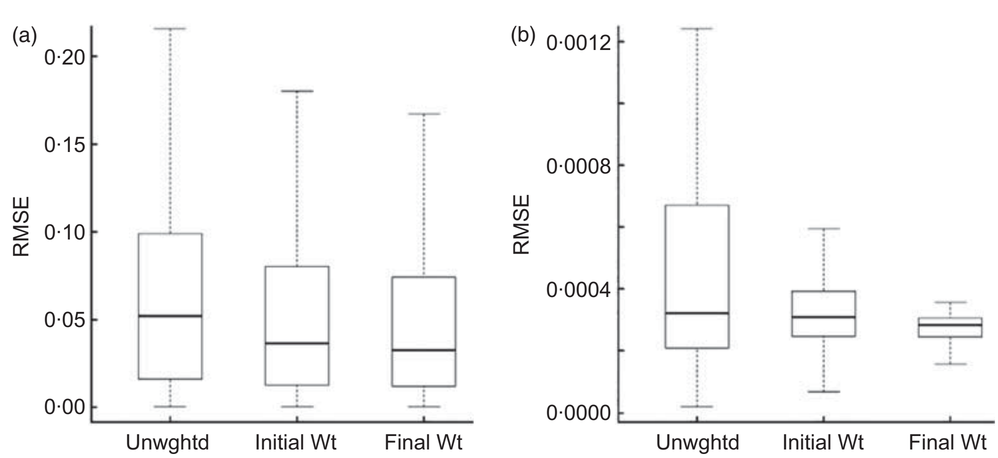

Fig. 5 Box-and-whisker plots of relative mean square error (RMSE) for (a) mean consumption of and (b) proportion of the population who consumed each food group from ANCNPAS07. Box-and-whisker plots are included for unweighted estimates (Unwghtd); estimates weighted using the initial weights, w_1 (Initial Wt); and estimates weighted using the final weights, w_2 (Final Wt). RMSE is calculated assuming that the final weights are unbiased (ANCNPAS07, 2007 Australian National Children's Nutrition and Physical Activity Survey)

Acknowledgments

The authors acknowledge the Australian Social Science Data Archive for use of the data. Those who carried out the original analysis and collection of the data bear no responsibility for the further analysis or interpretation of them. This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors. The authors have no conflicts of interest to declare. S.B. conducted the analyses and was the main author of the paper. Y.P. provided nutritional input and helped interpret the results. D.S. and L.T. conceived the idea for the study. D.S. provided statistical expertise, helped interpret the results, wrote sections of the manuscript and edited the manuscript. L.T. provided nutritional expertise, helped interpret the results and edited the manuscript.

References

1. Commonwealth Department of Health and Ageing *et al.* (2009) *The 2007 National Children's Nutrition and Physical Activity Survey [Computer file].* Canberra: Australian Social Science Data Archive, The Australian National University.
2. Clark R & Steel DG (2002) The effect of using household as a sampling unit. *Int Stat Rev* **70**, 289–314.
3. Skinner CJ, Holt D & Smith TMF (1989) *Analysis of Complex Surveys.* Chichester: Wiley and Sons.
4. Chambers RL & Skinner CJ (2003) *Analysis of Survey Data.* Chichester: Wiley and Sons.
5. Pfeffermann D (1996) The use of sampling weights for survey data analysis. *Stat Methods Med Res* **5**, 239–261.
6. National Center for Health Statistics (2010) Continuous NHANES Web Tutorial: Sample Design. http://www.cdc.gov/nchs/tutorials/Nhanes/SurveyDesign/SampleDesign/intro.htm (accessed August 2010).
7. National Center for Health Statistics (2005) Analytic and Reporting Guidelines for the National Health and Nutrition Examination Survey. http://www.cdc.gov/nchs/data/nhanes/nhanes_03_04/nhanes_analytic_guidelines_dec_2005.pdf (accessed August 2010).
8. Kish L (1965) *Survey Sampling.* New York: John Wiley and Sons.
9. Kish L (1992) Weighting for unequal Pi. *J Official Stat* **8**, 183–200.
10. Korn EL & Graubard BI (1999) *Analysis of Health Surveys.* New York: Wiley and Sons.
11. Gabler S, Haeder S & Lahiri P (1999) A model based justificationof Kish's formula for design effects for weighting and clustering. *Surv Methodol* **25**, 105–106.
12. Spencer BD (2000) An approximate design effect for unequal weighting when measurements may correlated with selection probabilities. *Surv Methodol* **26**, 137–138.
13. Gambino J (2009) Design effect caveats. *Am Stat* **63**, 141–146.
14. Park I & Lee H (2004) Design effects for the weighted mean and total estimators under complex survey sampling. *Surv Methodol* **30**, 183–193.
15. Curtin LR, Kruszon-Moran D, Carroll M *et al.* (2006) Estimation and analytic issues for rare events in NHANES. In *Proceedings of the Survey Research Methods Section, American Statistical Association.* http://www.amstat.org/sections/srms/proceedings/y2006/Files/JSM2006-000845.pdf (accessed August 2010).

Public Health Nutrition