

Estimation of population distributions of habitual nutrient intake based on a short-run weighed food diary

Clifton Gay*

Ministry of Agriculture, Fisheries and Food, Ergon House, 17 Smith Square, London SW1P 3JR, UK

(Received 16 November 1998 – Revised 13 August 1999 – Accepted 18 August 1999)

There have been many attempts to characterize day-to-day variation in nutrient intake. This variation has a fixed component, associated with particular days of the week, and a random component. Both components were studied for a range of nutrients, using 4 d weighed diary data from a large, nationally representative survey of people aged 65 years or over. Since day-to-day variation may distort the characterization of the population distribution of habitual nutrient intakes, especially when diets are studied over only a small number of days, a statistical method was developed to correct for this distortion. Results suggested that population distributions of habitual nutrient intake could be accurately constructed from 4 d weighed diary data and that the method might be successfully applied to studies based on as little as 2 d of observation. The method is particularly valuable for correcting estimates of extreme nutrient intakes for biases induced by uneven representation of days of the week and by within-person variation.

Nutrient intakes: Weighed diary: Variance components

The British National Diet and Nutrition Survey (NDNS) of people aged 65 years and over (Finch *et al.* 1998) is the second in an ongoing series of surveys conducted jointly by the Ministry of Agriculture, Fisheries and Food and the Department of Health. In common with the earlier NDNS of children aged 1½ to 4½ years (Gregory *et al.* 1995) fieldwork spanned a 12-month period, within which individual subjects were asked to complete a full weighed dietary record for four consecutive days. Subjects also participated in interviews designed to elicit important demographic details, crude food-frequency data, information on bowel movements, anthropometric measurements and (where consent was given) blood and urine samples. For each survey, a nutrient database, comprising more than 4000 foods, is used to translate daily records of foods consumed into estimated intakes of over fifty key nutrients. Information gathered is used to assess average and extreme dietary behaviour, potential intakes of contaminants and toxins, and to relate nutrient intakes to blood, health and demographic variables.

In the NDNS of people aged 65 years and over, complete 4 d food diaries were obtained for 1275 free-living subjects (632 males, 643 females) from an issued sample of 2172 eligible adults (a response rate of 59%). Response rates were considerably higher for completion of an initial questionnaire (75%), but only 61% agreed to keep a daily dietary record and only 59% completed the full 4 d. The length of the food

diary was chosen to be 4 d on the basis of a feasibility study (Hughes *et al.* 1995). A period of 4 d had also been adopted for the NDNS of children aged 1½ to 4½ years (Gregory *et al.* 1995), but the earlier Dietary and Nutritional Survey of British Adults (Gregory *et al.* 1990) used a 7 d diary. The implications of using a 4 d diary are considered here.

Objectives

When using the weighed diary approach adopted in the NDNS, one of the most important design issues is the number of diary days. Restricting the duration of the record may have a desirable impact on cost and response rate, but it is important to consider the overall effect on the accuracy of results. In the NDNS of people aged 65 years or over there were two main concerns associated with using a 4 d diary as opposed to a longer period.

- (1) Bias arising from uneven coverage of days of the week. The characterization of individuals is of particular concern, because, for any given individual, there are 3 d of the week for which no information is collected. However, even the population average may be affected, if, as in the NDNS of people aged 65 years or over, there is a tendency for some days to be more heavily sampled than others (e.g. Monday tends to be under-represented

Abbreviation: NDNS, National Diet and Nutrition Survey.

* **Corresponding author:** Dr Clifton Gay, fax +44 (0)171 238 6330, email c.gay@fscii.maff.gov.uk

while Friday is over-represented). Such bias can be avoided by having a diary which is a multiple of a full week (e.g. 7 d), or reduced by more sophisticated statistical analysis. However, the extent of the bias, and hence also the effort which should be expended to eradicate it, is unknown.

(2) The effects of within-person (day-to-day) variation. The

impact is twofold. It affects the characterization of the centre of the between-person intake distribution, which will be less precisely pin-pointed. It also affects the extremes, which will be pushed out further from the centre of the true between-person distribution of habitual intake. Extending the duration of the weighed diary has the advantage of diluting (averaging out) this

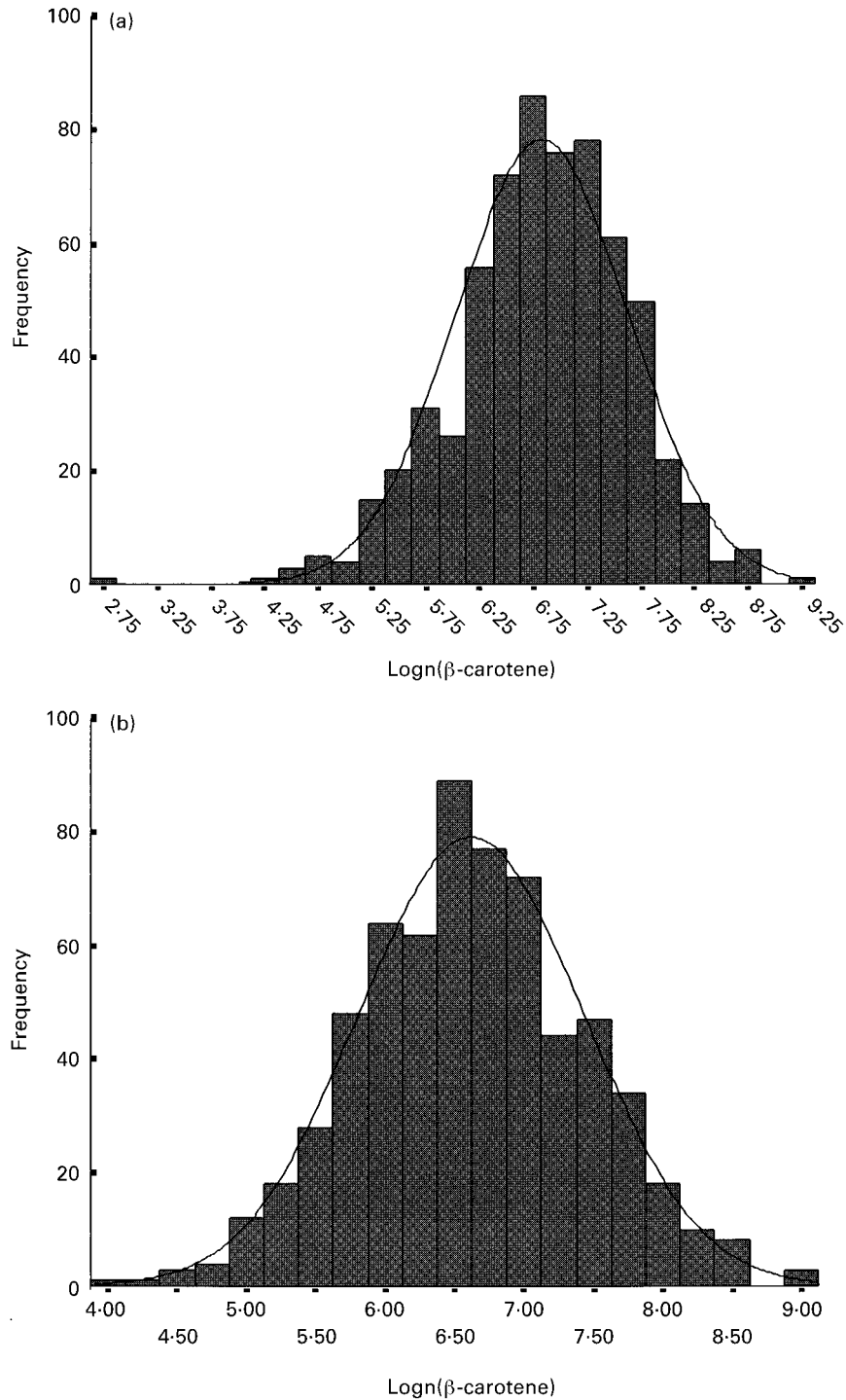


Fig. 1. Frequency histograms (with superimposed normal distribution curves) of log-transformed mean daily intakes of β -carotene for free-living (a) males and (b) females from the National Diet and Nutrition Survey of people aged 65 years and over (Finch *et al.* 1998).

Table 1. Mean nutrient intakes of free-living male and female subjects from the National Diet and Nutrition Survey of people aged 65 years or over (Finch *et al.* 1998) broken down by day of the week

Nutrient	Sun	Mon	Tues	Wed	Thur	Fri	Sat	F†	Average‡		Bias§ (%)	
									Weighted	Unweighted		
Free-living males (<i>n</i> 632, mean age 76 years)												
Energy (MJ)	7.8	7.7	7.7	7.8	7.9	8.0	8.0	1.49 NS	7.8	7.9	0.0 (0.2)	
Fat (g)	72.5	71.7	70.9	72.7	73.8	75.4	75.1	1.50 NS	73.2	73.4	0.2 (0.3)	
Protein (g)	71.5	68.8	69.4	70.4	68.9	69.2	67.9	1.44 NS	69.4	69.4	-0.0 (-0.1)	
Iron (mg)	10.8	10.8	10.6	11.1	10.8	10.9	11.1	0.41 NS	10.9	10.9	0.0 (0.1)	
Calcium (mg)	781	805	815	836	844	843	811	2.51*	819	822	2 (0.3)	
β -Carotene (μ g)	2377	1784	1682	1623	1780	1531	1497	0.83***	1753	1734	-19 (-1.1)	
Vitamin C (mg)	75.4	60.1	59.8	63.3	60.9	62.1	60.5	9.58***	63.1	63.0	-0.1 (-0.2)	
Frequency	316	244	292	347	438	477	414					
Free-living females (<i>n</i> 643, mean age 78 years)												
Energy (MJ)	5.9	5.9	5.8	5.9	5.8	5.9	6.0	0.81 NS	5.9	5.9	0.0 (0.1)	
Fat (g)	56.7	57.8	56.3	58.5	56.7	57.7	58.6	0.92 NS	57.5	57.5	0.0 (0.1)	
Protein (g)	57.1	55.3	54.1	55.9	53.5	53.9	54.1	2.79*	54.8	54.7	-0.1 (-0.3)	
Iron (mg)	8.6	8.1	8.7	8.5	8.6	8.4	8.3	1.58 NS	8.5	8.5	0.0 (0.1)	
Calcium (mg)	654	690	673	693	690	669	675	1.71 NS	678	678	-0 (-0.0)	
β -Carotene (μ g)	2040	1499	1357	1348	1317	1239	1443	8.88***	1463	1440	-24 (-1.6)	
Vitamin C (mg)	70.2	59.6	57.4	54.3	54.8	53.2	57.1	10.02***	58.1	57.5	-0.6 (-1.0)	
Frequency	331	274	325	361	415	463	403					

* $P < 0.05$, *** $P < 0.001$.† F test for statistical significance of differences between days of the week.

‡ Average daily intake, both weighted, to correct for uneven representation of different days of the week, and unweighted, which ignores the need for correction.

§ Bias represents the net effect of failing to weight according to day of the week.

|| Frequency is the number of occurrences of each day of the week among the completed weighed food diaries.

variation. However the size of the benefit needs to be quantified.

Since intakes of different nutrients may exhibit varying degrees and patterns of within-person variation, these concerns should be considered across a range of nutrients. Seven were studied, including energy, two macronutrients (fat and protein), two minerals (Fe and Ca) and two vitamins (β -carotene and vitamin C). β -Carotene was studied because its occurrence in a relatively narrow range of foods makes it likely to be unevenly distributed in the diets of human subjects, thus presenting a relatively harsh challenge for statistical analysis. Only nutrients derived from food sources were considered, the contribution of dietary supplements being excluded from all analyses.

Statistical methods

Statistical analysis and interpretation are greatly simplified

when the data under study follow a normal distribution. However, most, if not all, nutrient intakes are positively skewed in their distribution, both between and within individuals. To overcome this problem daily intake for each nutrient (X) was subjected to a transformation of the form:

$$Y = X^\lambda, \quad (1)$$

in which the special case $\lambda=0$ corresponds to a logarithmic transformation (Box & Cox, 1964). Appropriate λ -values for each nutrient were determined by an automated routine in which mean intake and within-person standard deviation were determined for each respondent, then log-transformed, before regressing the latter *v.* the former. Subtracting the estimated regression slope from unity suggests a value of λ consistent with stabilizing the within-person variance (e.g. Mathsoft, 1995) and will often simultaneously achieve approximate normality. The power transformations arising from this procedure are displayed in Table 2, together with the original units of measurement for each nutrient.

Table 2. Estimated variance components (between-person (σ_B^2) and within-person (σ_W^2)) for power-transformed daily intakes of seven nutrients, allowing for fixed day-of-the-week effects

Nutrient (X)	Original units	Transformed scale	Males		Females	
			σ_B^2	σ_W^2	σ_B^2	σ_W^2
Energy	MJ	$Y = X^{0.45}$	35.0	23.9	26.5	20.4
Fat	g	$Y = X^{0.40}$	0.41	0.43	0.38	0.38
Protein	g	$Y = X^{0.41}$	0.28	0.30	0.24	0.29
Iron	mg	$Y = X^{0.22}$	0.013	0.010	0.012	0.010
Calcium	mg	$Y = X^{0.31}$	0.62	0.45	0.59	0.39
β -Carotene	μ g	$Y = \log(X)$	0.41	0.93	0.42	0.94
Vitamin C	mg	$Y = X^{0.22}$	0.11	0.11	0.11	0.12

Kolmogorov–Smirnov one-sample tests were used to test the goodness-of-fit to a normal distribution of each power-transformed nutrient intake distribution. None of the tests provided any evidence (significant at the 5% level) to suggest that a normal distribution was not an adequate approximation. The worst-fitting nutrient was β -carotene, but even so the fit was tolerably good, particularly for females, as illustrated in Fig. 1.

After transforming the data, ANOVA was used to partition overall variation in nutrient intakes into components due to differences between individuals, systematic differences between days of the week and random within-person variation. This was done separately for free-living men and women. After analysis, results were back-transformed to the original units, where appropriate, for presentation purposes.

Estimates of differences in intake associated with particular days of the week were used to assess the biases that these might introduce. Estimates of within-person variance were used to assess the advantages of a longer recording period in terms of increasing the precision in characterizing the centre of the intake distribution and the accuracy of characterizing the extremes. The assessment relied on an assumption that consecutive days, measured on any individual, are as variable as non-consecutive days, and also that all respondents have equal within-person variance. These assumptions have been challenged by Tarasuk & Beaton (1992), but are not considered in detail. Their impact is likely to be small, particularly the latter, as the procedure used to select an optimal power transformation was designed to stabilize within-person variance.

Simulation was used to assess the impact of within-person variance on the characterization of the extremes of the intake distribution and the benefits of extending the

number of (consecutive) days comprising the food diary. Nutrient intake data were simulated, for three hypothetical designs, according to models implied by selected normalizing transformations, mean day-of-the-week effects and estimated variance components. Two procedures were used to derive estimates of extreme percentiles (the 2.5th and 97.5th) of the intake distribution. The most straightforward was to use sample percentiles corresponding to the percentile of interest, referred to as the non-parametric approach. The more sophisticated was to refit the model, based only on the data at any given simulation, and use this to generate a 52-week diary for each person. The rationale behind this parametric approach is that the simulated long-term diary for each respondent achieves an even allocation of days of the week and checks the tendency for within-person variation artificially to extend the tails of the intake distribution. What variation remains should reflect the true between-person distribution of habitual intakes and little else. The numbers of simulations performed were 1000 for the hypothetical 4 d design and 500 for the two remaining designs.

Results

Table 1 presents the results of an ANOVA to study the size and statistical significance of day-of-the-week effects. The *F* tests for the statistical significance of differences between days of the week highlight a number of effects which are too large to be attributable to chance. However, the most practically significant part of the table is that comprising the final three columns, since these measure the bias which results if the uneven representation of days of the week is ignored in the analysis. The bias was generally negligible, but approached a level at which it might be of concern for the two vitamins, β -carotene and vitamin C. There was a

Table 3. Confidence intervals (95% CI) for mean daily nutrient intakes under different survey designs, characterized by sample size (*n*) and number of diary recording days, for free-living male and female subjects

Nutrient	Survey design					
	Actual (4 d)		7 d		2 d	
	95% CI	Index*	95% CI	Index*	95% CI	Index*
Free-living males						
<i>n</i>		632		632		1264
Energy (MJ)	7.75, 8.05	100	7.75, 8.05	96.7	7.79, 8.01	75.8
Fat (g)	71.6, 75.2	100	71.7, 75.1	95.3	72.0, 74.8	77.8
Protein (g)	68.1, 70.7	100	68.1, 70.7	95.2	68.4, 70.4	78.1
Iron (mg)	10.6, 11.2	100	10.6, 11.2	95.9	10.7, 11.2	77.0
Calcium (mg)	800, 844	100	801, 843	96.6	805, 839	76.0
β -Carotene (μ g)	1632, 1836	100	1643, 1825	89.2	1646, 1821	86.0
Vitamin C (mg)	59.9, 66.1	100	60.1, 66.0	95.4	60.6, 65.4	77.8
Free-living females						
<i>n</i>		643		643		1286
Energy (MJ)	5.79, 6.01	100	5.79, 6.01	96.2	5.81, 5.99	76.6
Fat (g)	56.0, 59.0	100	56.1, 58.9	95.1	56.4, 55.4	78.2
Protein (g)	53.6, 55.8	100	53.7, 55.7	94.5	53.8, 55.6	79.0
Iron (mg)	8.27, 8.73	100	8.29, 8.71	95.4	8.32, 8.68	77.8
Calcium (mg)	659, 697	100	660, 696	97.0	664, 692	75.4
β -Carotene (μ g)	1345, 1535	100	1354, 1526	91.1	1361, 1519	83.5
Vitamin C (mg)	54.4, 60.6	100	54.5, 60.5	95.7	55.1, 59.9	77.4

* The index expresses the width of the 95% CI for both notional designs as a percentage of the width for the actual design (index = 100). Note that the index may be more informative than the confidence limits as these are only expressed to three significant figures or to the nearest integer (whichever is more accurate) and with large sample sizes (e.g. *n* 632) it is sometimes only the fourth significant figure which changes when one design is compared with another.

strong tendency for these vitamins to attain a high (or maximum) level of intake on Sunday and a low (or minimum) level on Friday. In both cases it resulted in a negative bias in estimated mean intake which may have exceeded 1% of the true value if days were not properly weighted. The reason for the downward bias for these vitamins is that their average intakes varied markedly between days and were higher on those days which were less well-represented (e.g. Sunday) than on those which were over-represented (e.g. Friday).

Intakes of two other selected nutrients (Ca and protein) exhibited statistically significant day-of-the-week variation for at least one sex group. In each case, as for β -carotene and vitamin C, Sunday was somewhat atypical, being associated with the lowest average intake of Ca, but the highest intake of protein.

The results in Table 2 demonstrate that the relative sizes of between- and within-person variance components differed markedly between nutrients. For intakes of Ca and energy the between-person component comfortably exceeded the within-person component, while for intake of β -carotene the reverse was true. Table 3 demonstrates the impact of the variance components on the precision of estimation of the mean intake, while Table 4 reveals their effect on the accuracy of estimates of extreme percentiles of the distribution, as represented by the 2.5th and 97.5th percentiles.

Table 3 suggests that increasing the recording period from 4 to 7 d gains only a small increase in the precision

with which mean intakes are estimated. If a 7 d diary were used, then for most nutrients the width of the CI for mean intake would be expected to fall to about 96% of its actual size (4 d diary). For intake of β -carotene the relatively high within-person variance means that increasing the length of the diary brings more reward, but even so it is clear that a greater increase in precision may be achieved by doubling the number of respondents, at the expense of reducing the recording period to 2 d.

The results in Table 4 were obtained by simulating new data, based on models fitted to the actual 4 d data. The most interesting comparisons were between survey designs, differentiated by duration of the food diary and number of respondents, and between the parametric and non-parametric methods of estimation. Comparisons should be based on point estimates of the extreme percentiles and their precision (95% CI). It is clear that non-parametric point estimates for the extreme percentiles were more extreme the lower the number of diary days. In fact, the bias associated with the non-parametric approach was so large for the 2 d design that even the 95% CI did not contain the corrected estimates generated by the parametric approach. The success of the parametric approach in neutralizing the effects of days of the week, and within-person variation, is evident from the consistency of the percentile estimates across designs. The non-parametric estimates converged towards the parametric estimates as the number of diary days increased.

Table 4. Comparison of estimated extreme percentiles of daily nutrient intake distributions, together with 95% confidence intervals for free-living male and female subjects aged 65 years and over, based on data simulated according to three different survey designs

Sex	Nutrient	Method*	Survey design					
			4 d (n 250)		7 d (n 250)		2 d (n 500)	
			Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Estimates of the 2.5th percentile								
Male	Energy (MJ)	P	4.7	4.1, 5.2	4.7	4.1, 5.2	4.7	4.3, 5.1
		N-P	4.4	3.9, 4.9	4.5	4.0, 5.0	4.2	3.9, 4.5
	Iron (mg)	P	5.5	4.6, 6.2	5.5	4.7, 6.2	5.5	4.9, 6.1
		N-P	5.1	4.4, 5.7	5.2	4.5, 5.9	4.7	4.2, 5.2
Vitamin C (mg)	P	14.7	10.2, 19.6	14.8	10.4, 20.1	14.6	11.3, 18.4	
	N-P	10.9	7.6, 14.5	12.4	8.9, 16.5	8.1	6.0, 10.3	
Female	Energy (MJ)	P	3.5	3.1, 3.9	3.5	3.1, 3.9	3.5	3.2, 3.9
		N-P	3.3	3.0, 3.7	3.4	3.1, 3.7	3.2	2.9, 3.4
	Iron (mg)	P	4.3	3.7, 4.9	4.3	3.7, 4.9	4.3	3.8, 4.8
		N-P	3.9	3.4, 4.4	4.1	3.6, 4.6	3.6	3.2, 4.0
Vitamin C (mg)	P	13.1	8.8, 17.7	13.1	8.8, 17.8	13.1	10.0, 17.1	
	N-P	9.4	6.5, 12.7	10.7	7.8, 14.1	6.9	5.0, 9.1	
Estimates of the 97.5th percentile								
Male	Energy (MJ)	P	11.7	10.9, 12.7	11.7	10.9, 12.7	11.7	11.1, 12.4
		N-P	12.1	11.3, 12.9	11.9	11.2, 12.7	12.4	11.7, 13.0
	Iron (mg)	P	18.7	16.8, 20.9	18.7	16.7, 21.1	18.7	17.2, 20.3
		N-P	19.5	17.7, 21.5	19.1	17.6, 20.9	20.3	18.9, 21.8
Vitamin C (mg)	P	164	134, 204	164	136, 203	162	141, 193	
	N-P	179	151, 216	172	146, 209	193	168, 221	
Female	Energy (MJ)	P	8.8	8.1, 9.5	8.8	8.2, 9.5	8.8	8.3, 9.4
		N-P	9.1	8.5, 9.7	8.9	8.4, 9.5	9.3	8.9, 9.9
	Iron (mg)	P	14.3	12.9, 16.1	14.3	12.9, 16.0	14.3	13.2, 15.7
		N-P	15.1	13.8, 16.7	14.7	13.4, 16.2	15.8	14.7, 17.1
Vitamin C (mg)	P	150	123, 186	152	125, 185	151	128, 177	
	N-P	167	140, 202	160	136, 189	182	158, 208	

P, parametric; N-P, non-parametric.

* The P method of analysis uses a theoretical distribution to estimate percentiles and corrects for both fixed day-of-the-week effects and random within-person variance. The N-P method makes neither of these adjustments.

The other main feature of Table 4 is that there was little loss of precision when the number of diary days was reduced. The greatest precision (tightest CI) was associated with the hypothetical 2 d design, although the assumption, implicit in this design, that doubling the number of subjects is a fair exchange for halving the number of diary days, is an optimistic one.

Discussion

The methods advanced here are applicable to data from a weighed food diary of at least 2 d duration and suppose that the main objective is characterizing the population distribution of daily nutrient intakes rather than classifying individuals. Increasing the length of the food diary is associated with a marginal improvement in the precision of estimated intakes (both mean and extreme intakes), but increasing the number of respondents may be more effective. If it is possible to trade off the cost of extending the duration of the diary (in terms of extra interviewer time and a decline in either response rate or the accuracy of record-keeping) against the cost of additional respondents, then a pragmatic basis for design might be to choose that combination of sample size and duration which maximizes the precision of results for a given overall outlay.

The statistical manoeuvre which facilitates valid characterization of nutrient intake distributions from short food records, by negating the impact of within-person variation, involves the following three key stages.

- Stage 1. Find a power transformation which, when applied to the daily nutrient intakes, stabilizes within-person variance and renders the between-person distribution approximately normal.
- Stage 2. On the transformed scale, separate the between-person and within-person components of variance and estimate the size of the fixed effects associated with particular days of the week.
- Stage 3. Use the fitted model to simulate food records of sufficient duration as to eliminate any unwanted within-person (day-to-day) variation, thereby focusing attention on appropriate aspects of the variation of intakes between people.

Identification of these stages makes it easier to highlight some of the assumptions involved in the proposed procedure and to discuss how it might be extended or modified. Stage 1 is fundamental since it underpins the assumed additivity of the fixed day-of-the-week effects and the ability to derive percentiles of the intake distribution. A power transformation successfully normalized the intake distribution for the seven nutrients investigated here, among each population sub-group considered, but it remains to be proved that it would work for all nutrients, or among all populations. A particular challenge might be presented by nutrients associated with specific foods which are not consumed (at least not by all people) on an everyday basis. In such circumstances the power transformation approach can only be applied sensibly to days when intake is non-zero for the nutrient concerned. Simulation of zero intakes necessitates a more complex model and might require additional food-frequency data, but was not developed further in the present study.

Separation of variance components at stage 2 also involves assumptions. Within-person variation is likely to be underestimated from data on consecutive days (Tarasuk & Beaton, 1992) and to a degree which increases as the duration of the food diary decreases. Relying on consecutive-day data may therefore result in overly extreme estimates of extreme percentiles, even after correction for within-person variation. Unfortunately, the extent of any bias cannot be assessed fully here, because of the nature of NDNS data, nor can the extent to which it is offset by the greater sample size, and hence precision, afforded by the economy of a consecutive-day design. Such topics might be worthy of future research. Were it to become available, an authoritative source of non-consecutive-day weighed diary data might be mined to provide correction factors with which to adjust the within-person variances provided by a typical consecutive day diary of given duration.

Another stage 2 assumption is that within-person variance is constant across individuals. If it is not constant, the resulting bias in estimates of extreme percentiles could be in either direction, but is unlikely to be serious. If the variance varies between people in a predictable way, such as increasing with mean nutrient intake, this information could be incorporated in the model used to simulate data at stage 3 and any bias might then be almost completely removed.

A further refinement to the model used at stage 2 might be to include additional fixed effects, such as a seasonal component or an adjustment factor for days of sickness.

In compiling Table 4 the decision taken at stage 3 was to simulate a 52-week diary. This was sufficient to diminish the contribution of within-person variation (to the variance of the intake distribution) to less than 1%. It also meant that any bias due to fixed day-of-the-week effects was completely eliminated. Selecting a long period focuses on the way habitual behaviour varies between people, but is not always appropriate. A useful feature of the proposed procedure is its ability to generate an intake distribution for the precise time frame required, be it a single day (for acute effects), several weeks (e.g. to study intakes during child-bearing) or longer term. This flexibility may encourage care in choosing the most appropriate scale of intake pattern in any given application.

A problem which lies beyond the remit of the present study is the accurate classification of individuals with respect to daily nutrient intake, an important task when attempting to relate nutrient intakes to health or nutritional status variables. Objectives of this kind heighten the value of increasing the duration of the weighed diary, because within-person variation undermines the ability to classify individuals and artificially dilutes correlations involving dietary variables (Gardner & Heady, 1973; Liu *et al.* 1978; James *et al.* 1981; Nelson *et al.* 1989). Since the procedure presented here is optimized for summarizing the population distribution, it only enhances the characterization of individuals to the extent of correcting for fixed day-of-the-week effects. When investigating relationships involving dietary variables there are likely to be more formidable problems (than that posed by within-person variation) which are not alleviated by increasing the recording period. Confounding variables, such as body size, shape, age, activity and morbidity, may distort or conceal true relationships, while the effect of any single

nutrient on status is likely to depend on the balance between it and a host of other nutrients in the diet. Controlled intervention studies are the only reliable basis for quantifying relationships between dietary and status variables (e.g. Chalmers *et al.* 1986; Ward *et al.* 1997), although dietary surveys are used in this way because they provide a convenient, less ethically sensitive vehicle for research.

Conclusion

This study has confirmed that day-to-day variations in nutrient intake, even within the same individual, may be large. The variation can be broken down into a random component and fixed effects associated with particular days of the week. An attempt has been made not only to estimate the size of effects, but also to assess their impact on important summaries of the data. It has been shown that, when analysed in an efficient manner, it may be possible to draw useful statistical inferences from weighed food diary data even when the recording period is as short as 2 d.

Acknowledgements

The National Diet and Nutrition Survey of people aged 65 years or over was funded jointly by the Ministry of Agriculture, Fisheries and Food and the Department of Health and conducted by Social and Community Planning Research in conjunction with the Medical Research Council's Dunn Nutrition Unit, the University of Newcastle, University College London and the University of Birmingham.

I would like to acknowledge the support received from members of the survey project board, especially the useful observations offered by Mike Day and Bob Wenlock.

References

- Box GEP & Cox DR (1964) An analysis of transformations. *Journal of the Royal Statistical Society B* **26**, 211–252.
- Chalmers J, Morgan T, Doyle A, Dickson B, Hopper J, Matthews J, Matthews G, Moulds R, Myers J, Nowson C, Scoggins B & Stebbing M (1986) Australian national health and medical research council dietary salt study in mild hypertension. *Journal of Hypertension* **4**, Suppl. 6, S629–S637.
- Finch S, Doyle W, Lowe C, Bates CJ, Prentice A, Smithers G & Clarke PC (1998) *National Diet and Nutrition Survey: People Aged 65 Years or Over*. Vol. 1: Report of the Diet and Nutrition Survey. London: The Stationery Office.
- Gardner MJ & Heady JA (1973) Some effects of within-person variability in epidemiological studies. *Journal of Chronic Disease* **26**, 781–795.
- Gregory JR, Collins DL, Davies PSW, Hughes JM & Clarke PC (1995) *The National Diet and Nutrition Survey of Children Aged 1½ to 4½ Years*. London: H.M. Stationery Office.
- Gregory JR, Foster K, Tyler H & Wiseman M (1990) *The Dietary and Nutritional Survey of British Adults*. London: H.M. Stationery Office.
- Hughes JM, Smithers G, Gay C, Clarke PC, Smith P, Lowe C, Prentice A, Bates C, Whitelaw M & Bingham S (1995) The British National Diet and Nutrition Survey of people aged 65 years or over: protocol and feasibility study. *Proceedings of the Nutrition Society* **54**, 631–643.
- Liu K, Stamler J, Dyer A, McKeever J & McKeever P (1978) Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *Journal of Chronic Disease* **31**, 399–418.
- James WPT, Bingham S & Cole TG (1981) Epidemiological assessment of dietary intake. *Nutrition and Cancer* **2**, 203–212.
- Mathsoft (1995) *S-Plus Guide to Statistical and Mathematical Analysis*, version 3.3. Seattle: StatSci.
- Nelson M, Black AE, Morris JA & Cole TJ (1989) Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *American Journal of Clinical Nutrition* **50**, 155–167.
- Tarasuk V & Beaton GH (1992) Statistical estimation of dietary parameters: implications of patterns in within-subject variation – a case study of sampling strategies. *American Journal of Clinical Nutrition* **55**, 22–27.
- Ward M, McNulty H, McPartlin J, Strain JJ, Weir DG & Scott JM (1997) Plasma homocysteine, a risk factor for cardiovascular disease, can be effectively lowered by physiological amounts of folic acid. *Quarterly Journal of Medicine* **90**, 519–524.