PROCEEDINGS OF THE NUTRITION SOCIETY

A Scientific Meeting was held at University College, Cork, Republic of Ireland on
13 September 1989

Symposium on
'Recent studies on food and nutrient intake in Ireland'

Dietary methodology: implications of errors in the measurement

BY WIJA A. VAN STAVEREN AND JAN BUREMA
Department of Human Nutrition, Wageningen Agricultural University, PO Box 8129, 6700 EV
Wageningen, The Netherlands

The assessment of food consumption of individuals and groups is extremely important for the science of nutrition. In the past decades several methods have been developed with various characteristics (Table 1) and adapted to serve widely different purposes of nutritional research. Although it is generally accepted that all methods have limitations and advantages, there are doubts about, for instance, the reliability of interview methods. To examine the validity of data obtained by such methods it is important to consider questions such as those phrased by Beaton, 1990 (presented at IUNS Conference Seoul, 1989): What does the estimated intake really represent? What is the nature and magnitude of the error in that estimate? What is the implication of the error for interpretation of analyses?

To answer these and similar questions, methods should be validated. The criteria for the validation procedure partly depend on the purpose of the study in which the method will be used.

EXPERIMENTAL V. OBSERVATIONAL STUDIES

First, it should be appreciated that there is a difference between the purposes of dietary experiments in metabolic wards and the aims of observational studies in so-called 'free-living' populations. Whereas nutritional experiments typically examine the effect of changing nutrient intake on indices of nutritional status in a fixed period of time, population-based observations are commonly used to examine the association between usual dietary intake of individuals or groups and other characteristics, such as disease status. There is a tendency to consider the results from metabolic studies as being more accurate than population-based observations. For the study of nutrition and health, however, both designs are useful since their purposes, limitations and advantages are different, as shown in Table 2 (Van Staveren & Burema, 1985).

In metabolic experiments the several-day weighed method is the preferred approach. This technique is being considered as the gold standard which is often used for other methods to be compared with. However, for observational studies in large populations this method is too cumbersome and time consuming. Thus, the 24 h recall method has often been used instead. In an excellent review Bingham (1987) gives the magnitudes of

Table 1. *Characteristics of dietary survey methods*

Collection of data	–Observation, record, interview
Time frame	–Incidental diet, usual diet
Portion sizes	–Weighed, estimated (models), frequency only
Conversion into nutrients	–Chemical analyses, nutrient data basis, food scores only

Table 2. *General purposes, limitations and advantages of an experimental and observational design for food consumption studies*

Design	General purpose	Limitations	Advantages
Food balance study (experimental design)	To examine the effect of changing nutrient intake during a fixed period of time on indices of nutritional status	Strong limited period of time Limited number of subjects Artificial conditions	Controllable data collection
Observations in free-living subjects (in general a cross-sectional design)	To examine the association between characteristics of diets of different individuals (groups) with other characteristics which these individuals (groups) exhibit	Data collection difficult to control Limited period of time	Large groups Data reflect real life situations

random errors in the energy intake measured by a 24 h recall from various studies: the coefficient of variation (CV) of differences between observations from a recall and from some reference methods (in most cases a 3 d record, but not necessarily a weighed record) was shown to vary from 7% to more than 40%. Bingham (1987) concluded: 'All of these studies in a variety of conditions and populations suggest that the daily recall method can be associated with unacceptably large errors, no matter whether food alone, or nutrients, are considered.' The question can be asked whether or not this conclusion holds for experimental as well as observational designs.

RANDOM V. SYSTEMATIC ERROR

The errors incurred in the measurement of diet may be random or systematic. Random errors, as denoted by their CV, affect the precision or reproducibility of a method. In contrast to systematic errors, the effect of random errors can be reduced by increasing the numbers of observations. Repeated measurements can, however, never remove an existing bias (i.e. a systematic error) from the observations.

The impact of random or systematic errors on the results of the study depends on the type of information needed. In epidemiological research, generally one or more of the following types of information might be aimed at (Cameron & Van Staveren, 1988); (1) assessment of mean energy (or nutrient) intake of a group, (2) assessment of the distribution of food consumption, more specifically the percentage of malnourished subjects in a population, without the purpose to identify them individually, (3) classification of individuals into extremes of the distribution of food consumption, or ordering them according to quintiles or tertiles for assessment of the association with

some other characteristic, (4) the absolute magnitude of the food consumption of an individual. The last type of information refers to clinical use, and will be omitted in the present paper.

Any systematic error will invalidate the results from type 1 and type 2 studies. However, a biased measurement might not affect the results from a type 3 study, unless this systematic error differs among subjects and is associated with the characteristic at issue. For instance, in the case of an energy balance study, only obese subjects underestimate energy intake, or rehabilitating anorexic patients overestimate energy intake. Since estimation of an association with the disease indicator is the ultimate goal of type 3 studies, ordering subjects correctly according to their dietary intake would be sufficient and a shift along the measurement scale may have no impact on the assessment of the effect, such as a relative risk.

On the other hand, a random error is always detrimental. However, there are ways to overcome this problem by special study designs. In type 1 studies, the measurement error will affect the precision of the estimation of the group mean to a slight extent, and this may be outweighed by increasing the sample size of the study group. The between-subject variation in a type 2 study will always be overestimated by the total variance if single dietary assessments are made. However, from duplicate observations the within-subject variation may be estimated and adjusted for in the analysis.

In type 3 studies, attenuated associations (biased towards the Null) will result from any misclassification due to random errors. The amount of misclassification, however, may be reduced by performing repeated observations on each subject, thus improving the effect estimate. In addition, as many investigators (Liu *et al.* 1978; Beaton *et al.* 1979; Sempos *et al.* 1985; Van Staveren *et al.* 1988; Nelson *et al.* 1989) have indicated, the magnitude of the unbiased measure of association can be estimated if within-subject variance and between-subject variance (or simply their ratio) are known. So, whereas the investigator should always be concerned about the impact of systematic errors, random errors should not be a concern if the design of the study makes it possible to account for within-subject variation. Depending on the nutrient at issue, many replications may be needed due to high day-to-day variations in intake.

Note that in a dietary history method the within-subject variation does not contain the day-to-day variation, since this method examines the usual diet in one interview. Consequently, random error is smaller and, thus, reproducibility is better. However, the validity of this method has often been questioned since the individual has to remember how frequently many different items of diet are eaten. We validated protein intake in Dutch adults, as assessed by dietary history, with a time frame of 1 month against the protein intake as assessed by 24 h N excretion in that month. On an aggregate level no difference was found between the two estimates. On an individual level, however, comparison is hardly possible, since one 24 h estimate of N excretion is not a good estimate of usual protein consumption (Van Staveren *et al.* 1985).

Other studies validated on the dietary history against records of food intakes. In most of these studies data obtained by dietary history seem to be overestimations as compared with data obtained from food records (Cameron & Van Staveren, 1988).

SUBJECT-ASSOCIATED BIAS

As has been stated for type 3 studies, a systematic error is only a matter of concern if it varies among subjects. It has been suggested earlier that systematic underestimation may

Table 3. *Mean daily energy intake, body-weight, body mass index (BMI), and change in body-weight over a period of 14 months for 123 young adult Dutch women*

Energy intake in quintiles	Energy intake		Body-wt (kg)	BMI (kg/m²)	Change in body-wt (kg)*
	kJ/d	kcal/d			
Very low	6113	1461	67·3	24·0	0·0
Low	7766	1856	62·3	22·7	0·1
Intermediate	8740	2089	59·6	21·6	1·2
High	9364	2238	58·9	21·2	0·9
Very high	10 958	2619	60·4	21·3	0·8

* Difference between the last three and the first three measurements.

in fact be associated with factors such as obesity. Also, the so-called flat slope syndrome (phenomenon of 'talking a good diet') is a manifestation of bias that varies between subjects. Such a bias of individual subjects does not necessarily have to be associated with the outcome variable at issue. For instance, in a follow-up study on the association between food intake and cancer, bias in the assessment of energy intake can be expected to be similar in future cases and non-cases. In such a situation, subject-associated bias will present itself in the data as disturbances due to the sampling procedure, and, thus, will be interpreted as random error!

In another example, a case-control study on gastric cancer, where the subjects are classified into quintiles of vitamin A intake as measured by a dietary history method, misclassifications will arise from over-reporters as well as under-reporters. Both kinds of subject-associated bias may occur, thus contributing to deflation of the observed association with gastric cancer. This attenuation of a correlation coefficient is commonly occurring due to random error, but may at least partly be due to subject-associated bias.

In the measurement of food intake itself, not only a bias towards the mean of the population, but also a bias away from the mean may occur. In a study in 114 Dutch women with fourteen 24 h recalls conducted monthly, some subjects appeared to be sustained under-reporters, whereas others persisted in over-reporting their energy intake (Van Staveren, 1985). Data obtained from those who reported a very low energy intake and those who reported a very high energy intake were inconsistent with the lack of changes in their body-weight (Table 3).

In five subjects who reported a very low energy intake we examined whether the inconsistency between their reported mean energy intake and the weight development was caused by a disturbed metabolic rate. Table 4 shows that in these subjects the mean resting metabolic rate as determined by indirect calorimetry was more than the reported mean total daily energy intake, which should have resulted in a loss instead of the observed gain in body-weight in four of these five subjects.

Although we did not document overestimation equally thoroughly, it may be expected that at least some of the subjects in the highest quintile of reported energy intake grossly overestimated their food consumption. This consistent and probably unconscious subject-associated under- or overestimation of intakes is very hard to overcome, but is not specific for the 24 h recall method.

In another study conducted in our institute, usual metabolizable energy intake of twenty-seven lean and eighteen overweight women was measured using the 7 d weighing

Table 4. *Body-weight, height, change in weight over a period of 14 months, and resting metabolic rate of five young adult women*

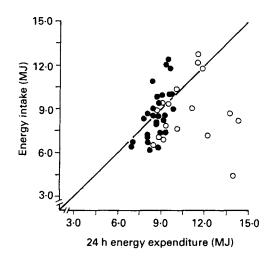| Subject	Body-wt (kg)	Body height (m)	Wt change (kg)	Resting metabolic rate		Energy intake	
				(kJ/24 h)	(kcal/24 h)	(kJ/24 h)	(kcal/24 h)
A	80·0	1·74	−0·3	7778	1859	4330	1035
B	65·1	1·68	3·8	6284	1502	5364	1282
C	62·8	1·74	0·3	6276	1500	3535	845
D	66·9	1·66	4·0	6540	1563	4862	1162
E	109·6	1·70	2·0	7807	1866	6740	1611

Fig. 1. Daily energy intake by 7 d weighing record, and the 24 h energy expenditure by indirect calorimetry, in twenty-seven lean (●) and eighteen overweight (○) women.

record method. The result was validated against energy expenditure as measured over three successive days in whole-body indirect calorimeters. Energy intake and 24 h energy expenditure were assumed to be estimates of energy requirements of the women. Subjects with a body mass index (BMI) of more than 25 kg/m^2 were classified as overweight. Mean BMI in the lean and overweight group was 20·7 (SD 1·9) and 33·5 (SD 6·9) respectively (De Boer, 1985).

Energy intake per d of the lean women, as estimated by the 7 d weighing record method was similar to the 24 h energy expenditure. The overweight women, however, reported a daily energy intake 1·9 MJ (SD 2·9) lower than their 24 h energy expenditure, so they under-reported their energy intake on average by approximately 18%.

In lean women, the energy intake was positively associated with energy expenditure (*r* 0·58, *P*<0·005). No significant association, however, was found in the overweight women. It was concluded that in lean women the 7 d weighing record method may give valid estimates of the usual (in contrast to incidental) energy intake, but in overweight women the validity of this method still remains questionable.

IMPLICATIONS FOR INTERPRETATION OF CVs

When such subject-associated systematic errors do occur when the weighed record is used, we just have to admit that in observational studies this method with a reputation for validity may produce less reliable results than most investigators think it does. Even the weighed record method may be associated with undesirably large errors and with CVs for random error that may be somewhat, but not much smaller, than those of the recall method. Note that in epidemiological studies, the individual differences between observed and true values are interpreted as random error in the observations!

As was mentioned earlier, the influence of within-subject random errors can be reduced by repeating 24 h measurements several times in the same individuals. That is why a 7 d record may prove to be a satisfactory dietary assessment, and it also explains why CVs from a 3 d recall are smaller than those associated with one 24 h recall.

Similarly, we may cope with random errors which originate from systematically under- or overestimating subjects in the study sample by increasing the sample size when the aim of the investigator is to estimate mean and standard deviation of a distribution (type 1 or 2 information). However, the proportion of misclassifications due to systematic under- or over-reporters cannot be reduced by increasing the number of observations, and this holds for observations from both interview and record methods. It is a challenging task for future research to determine characteristics of people and circumstances that are bound to consistently produce systematic errors. For instance, results from a study by Van Strien (1986) indicate that under-reporting food intake is probably related to restrained eating patterns. As a matter of fact, those patterns occur more often in obese than in non-obese women. Such information may be helpful to improve the selection of subjects in epidemiological studies.

In conclusion, the following remarks have to be borne in mind when we wish to make a choice for the dietary method to be used. The appropriateness of a dietary assessment method depends on the purpose of the investigation. For estimating mean and standard deviation of a distribution (type 1 and 2 information), systematic error is a matter of concern. Increasing the amount of observations may in some cases, but not all, allow for random errors. Responders who consistently under- or overestimate their food intake, may cause attenuated associations, and this is true not only for the recall but also for a record method.

REFERENCES

Beaton, G. H. (1990). Use of dietary data: approaches and issues. In *Proceedings of XIV International Congress of Nutrition*, Seoul 1989 [A. B. Smith, editor]. (In the Press).

Beaton, G. H., Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., De Ramos, M., Hewitt, D., Grambsch, P. V., Kassim, N. & Little, J. A. (1979). Sources of variance in 24-hour recall data: implications for nutrition study design and interpretation. *American Journal of Clinical Nutrition* 32, 2546–2559.

Bingham, S. A. (1987). The dietary assessment of individuals; methods accuracy, new techniques and recommendations. *Nutrition Abstracts and Reviews* 57, 705–742.

Cameron, E. M. & Van Staveren, W. A. (editors) (1988). *Manual on Methodology for Food Consumption Studies*. Oxford: Oxford University Press.

De Boer, J. O. (1985). Energy requirements of lean and obese women. PhD Thesis, Agricultural University, Wageningen, The Netherlands.

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.

Nelson, M., Black, A. E., Morris, J. A. & Cole, T. J. (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.

Sempos, C. T., Johnson, N. E., Smith, E. L. & Gilligan, C. (1985). Effects of intraindividual variation in repeated dietary records. *American Journal of Epidemiology* **121**, 120–130.

Van Staveren, W. A. (1985). Food intake measurements: their validity and reproducibility. PhD Thesis, Agricultural University, Wageningen, The Netherlands.

Van Staveren, W. A. & Burema, J. (1985). Food consumption surveys: frustrations and expectations. *Näringsforskning* **29**, 38–42.

Van Staveren, W. A., Burema, J., Deurenberg, P. & Katan, M. B. (1988). Weak associations in nutritional epidemiology: the importance of replication of observations on individuals. *International Journal of Epidemiology* **17**, Suppl., 964–969.

Van Staveren, W. A., De Boer, J. O. & Burema, J. (1985). Validity and reproducibility of a dietary history method estimating the usual food intake during one month. *American Journal of Clinical Nutrition* **42**, 554–559.

Van Strien, T. (1986). Eating behaviour, personality traits and body mass. PhD Thesis, Agricultural University, Wageningen, The Netherlands.