

Editorial

Food Frequency Questionnaires: Simple and cheap, but are they valid?

Estimating dietary intake is one of the most difficult tasks the physiologist can undertake (Garrow 1974). The obvious difficulties involved have led many to conclude that dietary surveys are worthless. Conversely, some surveys based on inappropriate methods, small and/or biased samples, and uncritical discussion or even acknowledgement of the errors involved, still find homes in the academic press. It is clear, however, that dietary intake (exposure) is related to health and disease either directly or as a co-factor, and consumption data is also useful for social and political purposes. Hence, there is a need for information about eating habits and dietary intake, but flawed data has been misleading.

Dietary survey methods are labour intensive and hence expensive, and yet yield data that lack precision. However, some nutritionists, alas, still talk about a 'gold standard' method (by which is usually meant the weighed inventory method applied over a 7-day period). The application of biomarkers, especially doubly labelled water (giving a usefully precise estimate energy expenditure even under field conditions), exposed a terrible truth; diet surveys were inaccurate. Worse than that, they were biased; people with 'large' body mass indexes (BMIs) tended to under-report their energy intakes (Macdiarmid & Blundell 1998). Hence, there is a need for a simple and cheap but improved method(s). The answer that has been widely accepted is the Food Frequency Questionnaire (FFQ), which has now been used in some huge epidemiological studies (Hansen *et al.* 2010) and also in some small studies (Prosser *et al.* 2010).

There are many variants of FFQ; almost every research group seems to design their own, often for good reason because they may be interested in a particular nutrient, food, or foods. Hence, an FFQ designed for one specific purpose (e.g. population group) should not be uncritically applied to another situation. Ideally, for every project involving a dietary survey, two questions must be addressed *de novo*: is it

valid and is it reliable? Two deceptively simple questions on which many projects have foundered.

The error due to bias has an unfortunate characteristic; no increase in sample size will reduce it, whereas random error can be reduced by recruiting more volunteers (Woolf 1954). Representative sampling, therefore, is the essence of a good study that can be generalized, and is more important than sample size per se. This is a powerful argument in favour of FFQs because a more representative sample may be willing to take part. Of course, they almost invariably have to be able to read (sometimes very lengthy and complex forms) and write (usually in English) and not be homeless or disabled, and so on, but such inconveniences are usually ignored. Thus, the most vulnerable and perhaps most nutritionally interesting sectors of the population are typically excluded and the resulting picture of nutritional health will be an overly rosy one.

The methods used to 'validate' vary as there are many forms of validity. I will mention only two here: relative (or comparative) and criterion related. The most common technique employed is to compare the results of one method with those of another. Correlations are often reported which are of very limited value (Bland Altman plots are more informative) and even coupled with excellent agreement between mean results do not answer the validity question; both methods may be equally poor and/or biased. The expression 'gold standard method' is now less often used, but in reality there is no dietary survey method that can serve this purpose. The weighed inventory method is fundamentally flawed – try weighing your restaurant meal or 'street' food. Furthermore, as Barnet Woolf (1954) pointed out, minimizing one source of error when others of similar magnitude are around (e.g. using food tables to calculate nutrient intake) is pointless.

The use of biomarkers (some factor which varies simply and preferably quickly with dietary intake) should be the norm, but no biomarker reflects all

aspects of dietary intake; doubly labelled waterreflect-energy intake (if 'weight' does not change), urine nitrogen more or less reflects protein intake, serum carotenoid levels – carotenoid intake and urine sugar(s) levels – sugar(s) consumption, and so on. But a method 'valid' for one of these need not be valid for the rest. Furthermore, collecting blood and urine samples is immensely problematic and will certainly lead to a very biased sample of volunteers. The late Sheila Bingham showed that collecting 24-h urine samples created its own problems of reliability and validity, and some eight collections (with completeness verified by a biomarker) would be necessary to 'validate' protein intake (Bingham & Cummings 1985). Impasse.

The question arises whether a study that 'validates' an FFQ is of much value to anyone but the project team. Certainly, no method can be 'validated' – methods are not valid or not, they lack validity to different degrees. A potential major influence on the data obtained in dietary surveys is the delivery method chosen. The presence of an interviewer may bias results (male/female, young/old, black/white, clinician/clerk) as may the setting chosen (e.g. clinic/home), but postal questionnaires typically have poor return rates and hungry people may overestimate portion sizes consumed (Beasley *et al.* 2004). There is evidence that data collected by computer is more valid than that collected by interviewer or paper questionnaire (Hackett *et al.* 1989). Thus, the method chosen cannot be divorced from its delivery – these two *together* constitute 'the method'.

A major problem with any dietary survey is portion size estimation (oddly, this includes studies that ask volunteers to weigh their own intake). Trying to weigh food intake reduces volunteer rates. In order to convert an estimate of frequency of consumption of food, some estimate of the portion size consumed is required. For example, I may report eating cornflakes 5 days each week. What weight of cornflakes was consumed each time? The fact that the portion size might vary each time is usually ignored and most often an 'average' figure is applied. More sophisticated FFQs may enable 'small', 'medium', or 'large' portions to be selected. This then means that the survey is doomed to rep-

licate (or at the least is biased by) whatever survey the 'average' (or other portion sizes) was based on. For example, a commonly used resource (including for surveys other than FFQ based) is the Food Atlas (Nelson *et al.* 1997), but the data on which this is based is now at least 20 years old and was collected from a biased sample of the population that excluded children, and so on.

Recently, there has been a large outpouring of papers 'validating' FFQs (Google Scholar gives some 4000–8000 hits for terms such as 'FFQ with validity' or 'FFQ with validation'). Should such studies be published? There is a catch-22-type of problem here: to publish results from an FFQ survey, questions will often have to be answered about whether it is 'valid' – a publication in a respected journal assures this (even if no one bothers to read the paper or critically evaluate the nature of the validation). Unfortunately, the results, even if 'valid', are probably not transferable to any other situation, which ideally would have to be proven. In addition, the term 'valid' is somewhat elastic, and energy intake-to-basal metabolic rate ratios of 1.1–1.6 or above have been claimed to indicate a 'valid' collection of dietary data.

Here is my opinion. If an FFQ is used, studies of validity and reliability should be carried out – the study team must have confidence of the method *as used in their hands*. I do not believe that publishing these studies is beneficial unless some novel feature or finding exists; for example, a new biomarker or combination might have been used or some unexpected finding has been discovered. I would suggest that a statement in the methods section that such studies were carried out with the most basic of findings (energy intake : BMR is probably as good as any other simple and cheap method) and an assurance that the authors would supply details on request should be sufficient to satisfy both reviewers and readers. This means relying on the integrity of researchers and having the confidence of science to expose (eventually) bad practice and even fraud.

The reliability and validity of dietary survey methods are universal problems that are inherent in trying to measure a complex dynamic behaviour, and the problems have been well known for a very long time but all too often ignored. FFQs have their place,

but the lure of cheapness and simplicity should not be accompanied by uncritical use.

Allan Hackett
Nutrition and Health Research Group
Liverpool John Moores University
IM Marsh Campus
Barkhill Road
Liverpool L17 6BD
Email: a.f.hackett@ljmu.ac.uk

References

- Beasley L., Hackett A.F., Maxwell S.M. & Stevenson L. (2004) The effect of satiety on perceptions of usual portion sizes. *Journal of Human Nutrition and Dietetics* **17**, 219–225.
- Bingham S. & Cummings J.H. (1985) Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. *American Journal of Clinical Nutrition* **42**, 1276–1289.
- Garrow J.S. (1974) *Energy Balance and Obesity in Man*. North Holland: London.
- Hackett A.F., Jarvis S.N. & Flanagan G.J. (1989) The feasibility of using school-based micro-computers to collect information on the health related behaviour of children. *Health Education Journal* **48**, 39–42.
- Hansen L., Dragsted L.O., Olsen A., Christensen J., Tjønneland A., Schmidt E.B. *et al.* (2010) Fruit and vegetable intake and risk of acute coronary syndrome. *British Journal of Nutrition* **104**, 248–255.
- Macdiarmid J. & Blundell J. (1998) Assessing dietary intake: who, what and why of under-reporting. *Nutrition Research Reviews* **11**, 231–253.
- Nelson M., Atkinson M. & Meyer J. (1997) *Food Portion Sizes – A Photographic Atlas*. MAFF: London.
- Prosser N.R., Health A.-L.M., Williams S.M. & Gibson R.S. (2010) Influence of an iron intervention on zinc status of young adult New Zealand women with mild iron deficiency. *British Journal of Nutrition* **104**, 742–750.
- Woolf B. (1954) Statistical aspects of dietary surveys. *Proceedings of the Nutrition Society* **13**, 82–94.