

Measuring nutritional exposures including biomarkers

BY PIETER VAN 'T VEER

*Department of Epidemiology and Public Health, Wageningen Agricultural University,
PO Box 238, 6700 AE Wageningen, The Netherlands*

Most epidemiological studies seek to explain the effects of environmental factors on health risk. In nutritional studies the environmental exposures can be described in terms of foods, nutrients and biomarkers. In order to determine the risk of nutritional exposures it is essential that the exposure is measured in an appropriate way. Most often risk is assessed across categories (thirds, fourths or fifths) of exposure and if correct classification cannot be made the wrong estimate of risk will be obtained; this applies no matter how well the study has otherwise been designed and executed. The development of a nutritional epidemiological study requires a clear hypothesis which identifies the relevant exposure and outcome. From this, the specific substance(s) of interest can be defined. Understanding how exposure may affect the outcome also clarifies the relevant window in time in which exposure needs to be measured. For example, in many chronic diseases the relevant time-frame is at some considerable distance in the past, and therefore skill is needed in identifying ways of assessing past diet. A clear study hypothesis also provides guidance for the selection of appropriate populations for study and the most effective design to be employed.

The effects of measurement error are particularly important for studies where there is a relatively weak association (Rothman & Poole, 1988). In the nutritional epidemiology of chronic diseases, weak associations are very common; that is, the estimate of risk associated with nutritional exposure is relatively small (for example, most cancers and heart disease). It is therefore a considerable challenge in nutritional epidemiology to reduce the errors associated with measuring exposure to maximize the possibility of measuring weak associations correctly (Kok & van 't Veer, 1989). This paper addresses the important issues in measuring nutritional exposures in epidemiological studies.

LEVELS OF MEASURING DIETARY EXPOSURES

Nutritional exposures can be defined in three different levels of measurement; as foods, nutrients and biomarkers. In principle, foods are the primary source of exposure. They are assessed in terms of frequency of consumption, number of portions and portion sizes, integrated to the average daily intake (w_i ; with the subscript indexing the food). Errors in the assessment of one of these components (frequency, number and size of portions) lead to errors in the observed average daily food intake. At the level of foods, associations with health outcomes may result from the food's content of nutrients, non-nutritive or toxic substances, its preparation, its association with (culturally) associated foods, or any interaction between them.

In nutritional epidemiology, it is common practice to convert, or to summarize, complex patterns of numerous foods in terms of nutrients, using food composition databases. This way, hypotheses formulated on the nutrient level can be evaluated. Thus, the average daily intake of a nutrient, for example fat, may be calculated as:

$$\text{average daily intake of a nutrient} = \sum (w_i c_i)$$

where w_i is the average daily intake of the food, and c_i is the fat content of the food obtained from food composition tables. Thus, average daily nutrient intake is a nutrient-weighted summary measure of the food pattern. For different nutrients, only the weighting factors are different. Results for various nutrients obtained from one single study are, therefore, always dependent on the same underlying food pattern, and cannot be regarded as independent. Furthermore, since nutrient intakes are calculated on the basis of the same food pattern (w_i), any two nutrients, for example energy and fat, tend to be correlated, leading to difficulties in discriminating between their independent effects and/or potential confounding. Finally, it is apparent that errors in average daily intake (w_i) and errors in nutrient contents (c_i) convert into errors in the level of the nutrient. Although values for nutrient contents given in food tables tend to remain the same for long periods of time, the actual contents in the foods do not, especially if we consider for instance some vitamins or trace elements. Moreover, food composition tables tend to be based on the need to evaluate established diet-related risk, not to address new exposures and research questions.

For biomarkers, the circumstances are quite different. Biomarkers are cellular, biochemical, or molecular alterations that are measurable in biological media, such as tissues, cells, or fluids; they may include any measurement in the sequence of events leading from exposure to disease, including markers of inherited or acquired susceptibility (Hulka *et al.* 1990). In the framework of measuring nutritional exposures, attention is directed towards markers of internal dose, as markers of dietary exposure. In comparison with the assessment of food and nutrient intake, biomarkers have the advantage that they integrate nutritional exposure from various sources, sometimes even over an extended period of time, and that they are not limited to substances tabulated in food composition tables; furthermore, they contain fundamentally different errors that are conceptually independent of those involved in the calculation of food and nutrient intake. Of course, it should be realized that biomarkers are often not solely determined by intake; in fact, the advantages of biomarkers may be reduced because biomarker levels are also influenced by digestion, absorption, metabolism, excretion, and exchange with other body pools. The degree to which this is disadvantageous, however, depends on the objective of biomarker use, for example, whether we want to use them as indicators of dietary intake, or to consider them as measures of internal exposure to be studied in relation to disease.

DETERMINANTS OF THE OBSERVED EXPOSURES

Major concepts regarding the measurement of nutritional exposure require a conceptual distinction between the true, but unknown, exposure, and the observed exposure, i.e. the measurement. In an ideal world, true exposure would be the unique determinant of the observed exposure, either in terms of foods, nutrients or biomarkers. In reality, the observed exposure depends on the sensitivity of the method of measurement to past and present true exposure, the ratio of within- to between-subject variance within a population, other determinants adding background 'noise' to the true exposure signal (not necessarily independent of exposure), and purely random or technical measurement errors (Fig. 1).

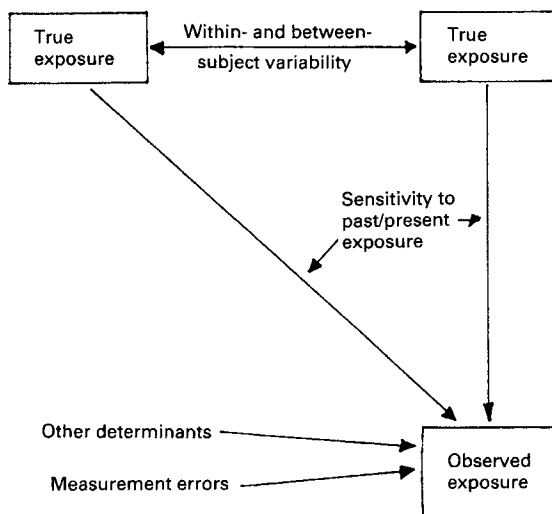


Fig. 1. Determinants of observed exposure in relation to true exposure of the population (within- and between-subject variance), time (sensitivity to past and present exposure), and extraneous sources of variation (other determinants, measurement errors).

In epidemiological investigations we are sometimes interested in the true exposure at the moment of measurement (e.g. surveillance programmes); under these circumstances we may use methods that are sensitive to present exposure only, and not to past exposure; here the disadvantage of often considerable within-subject variability is alleviated by increased sample size. If, as in many aetiological studies, research interest focuses on usual exposure of individual subjects, and not simply exposure at the moment of measurement, time integration, i.e. sensitivity to past and present exposure, becomes important. With regard to time-integrating properties, it is interesting to note that appropriate assessment of long-term exposure may result from constant exposure (with a low intra-subject variability:inter-subject variability ratio) coupled with sensitivity to present exposure only, or from a more variable exposure (higher intra-subject variability:inter-subject variability ratio) coupled with sensitivity to both past and present exposure. In terms of the validity of results in epidemiological studies, it is of course crucial that all determinants of the observed exposure are the same in the population groups being compared. This is especially important in retrospective studies. In order to meet this objective, both interviewer and responder biases should be minimized, which can be achieved by standardization of the measurement instruments and conditions. This includes standardization of the questionnaire or interview design (e.g. at item level, hierarchy of items, open-ended or fixed questions), setting of data collection (e.g. face-to-face or self-administered, home or hospital, index-subject or proxy, with or without partner or family), training of personnel (e.g. interviewing, data checking, cleaning) and study design (e.g. balance interviewers, seasons, measuring covariates).

TIME-FRAME OF MEASUREMENT INSTRUMENTS

Available methods of dietary assessment refer to different time-frames, i.e. actual diet (e.g. duplicate portion technique, weighing, observation, recording methods), usual diet

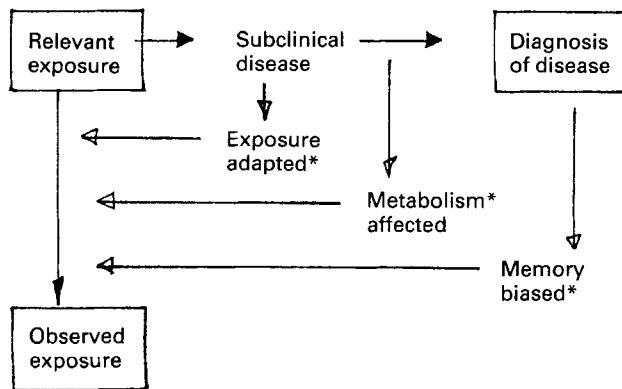
(repeated records, repeated 24 h recalls, ranging from dietary histories with cross-check to food frequency questionnaires), or remote diet in the distant past (memory-based dietary histories and food frequency methods only). In individual-oriented aetiological research, habitual diet is usually measured by memory-based methods; their validity is preferably assessed by comparison with methods that rely less heavily on memory and are considered to have reasonably independent measurement errors (repeated records, sometimes repeated 24 h recalls), or by comparison with truly memory-independent biomarkers of dietary intake (if available).

The relevant time-frame in aetiological studies depends on the underlying biological mechanism. This aetiological relevant time period, however, probably covers decades or even the whole of life, rather than the 1-year period that can be measured with memory-based methods or the few months that are covered by time-integrating biomarkers. For many chronic diseases, long-term or time-integrated diet is of interest, while specific periods in the distant past may be of special interest. Therefore, two approaches to dietary assessment have been explored to some extent. One approach, of particular relevance to retrospective studies, assumes that the diet at some moment in the distant past is aetiological relevant, and attempts have been made to assess the validity of measuring diet in the distant past. The other approach, more relevant to prospective investigations, assumes that diet is reasonably constant for long periods of time, and that one single dietary measurement might represent exposure over a considerable length of time. Of course, if dietary habits were really stable over time there would be no need to assess remote diet in the first place, since present and past diet would be closely correlated.

To assess validity of dietary assessment in the distant past, several studies have been conducted. In these studies diet was assessed at some moment in the past, and some 5–10 years later the investigators asked for recall of past food consumption; current diet was often assessed simultaneously. The results of these studies suggest that recalled diet is more strongly related to current diet than to the original (past) diet that was intended to be the subject of measurement (Bingham & Nelson, 1991). Thus, if remote diet is of interest, the price paid for obtaining the information is large in terms of potential bias towards current diet and added measurement error. Moreover, since it is not clear to what extent these phenomena may interfere with the comparability of information between cases and controls, measuring diet in the distant past does not seem very promising in retrospective study designs.

Observations from the Dutch prospective cohort study (van den Brandt *et al.* 1990) provide some insight into the assumption of stable dietary habits. In five consecutive years after a baseline dietary assessment the initial food frequency questionnaire was repeated in independent subsamples of the cohort. For each nutrient, the four Pearson coefficients of correlation with the baseline measurement were plotted on the time-axis, and a regression line was calculated. It appeared that the correlations tended to decrease only slightly, suggesting reasonably constant diets in this population (Goldbohm *et al.* 1993).

Although the approach described above may be considered generally applicable to dietary assessment for aetiological research on chronic diseases, it should be applied with considerable caution. First, the populations involved in the studies on validity of remote diet tended to be younger, probably with less stable dietary patterns, than the population involved in the Dutch cohort study. Moreover, the assumption of stable dietary patterns



*May be aggravated by treatment for disease

Fig. 2. Determinants of observed exposure in retrospective case-control studies. The further the disease has progressed the more discrepancies may develop between the relevant exposure and the observed exposure.

is seriously violated if interest is focused on adolescent diet (e.g. as related to breast cancer). Finally, life events may affect dietary habits to a considerable extent. For example, in colon cancer aetiology there is a debate on reproductive factors and age at first childbirth as risk factors, but it is unclear whether they should be regarded as physiological risk factors or as indicators of lifestyle and dietary habits.

DIETARY ASSESSMENT IN RETROSPECTIVE STUDIES

Like prospective cohort studies, most case-control studies on diet and chronic disease have assessed 'usual' diet, i.e. diet in the recent past. Both designs implicitly assume that dietary habits are reasonably constant in the recent aetiological time-window. The assessment of diet in case-control studies, however, has been the subject of substantial discussion. The major argument is that, even if the true or relevant exposure is adequately defined, the observed exposure may be biased, because it may be unintentionally adapted due to subclinical disease; the memory of cases may be elicited to give biased responses related to their knowledge of the disease, and, in the case of biomarkers, the metabolism may be affected (Fig. 2). As a consequence, the observed exposure in the cases has several important determinants that are lacking in the control group, i.e. lack of comparability of information and biased measures of association.

Although these potential pitfalls have been generally recognized, only recently has evidence been presented that seems to substantiate the extent to which they affect estimates of the effect of diet on risk of disease. Table 1 shows relative risks (RR) and odds ratios (OR) from both retrospective and prospective investigations. For alcohol and colorectal cancer the meta-analysis by Longnecker *et al.* (1990) indicated stronger associations in cohort studies than in case-control studies; for case-control studies the estimated RR values were similar in studies using community-based and hospital-based controls, which suggests that selection bias does not necessarily explain the discrepancy with cohort studies, and that cases may have under-reported their intake compared with controls. For dietary fat and breast cancer the case-control studies seemed to point to a weakly positive association, but the cohort studies that have appeared in the past few

Table 1. Association between dietary habits and chronic diseases in retrospective case-control studies v. prospective cohort studies

Study design	Alcohol and colorectal cancer*		Dietary fat and breast cancer†	
	RR‡	95% CI	RR§	95% CI
Case-control	1.07	1.02–1.12	1.14	1.09–1.19
Cohort	1.32	1.16–1.51	1.00	0.89–1.12

RR, relative risk; CI, confidence interval.

* Data from Longnecker (1990).

† Case-control data from Howe *et al.* (1990); cohort data from Michels & Willett (1993), (inverse variance weighted average RR of seven independent cohort studies).

‡ Relative risk or odds ratio per 24 g (two drinks) daily.

§ Relative risk or odds ratio per 14% energy as fat, approximating the (energy-adjusted) range between extreme quintiles of fat intake.

years have failed to confirm this. Recently, Giovanucci *et al.* (1993) confirmed these concepts in a case-control study on dietary fat and breast cancer, nested in the Nurses Health Study. They showed that selection and information bias did indeed lead to weaker associations between dietary fat and breast cancer in the case-control approach compared with prospective analysis. This is very much in line with the results published by Howe *et al.* (1990) and the combined evidence from prospective studies on this topic. Consequently it seems that there remains little room for retrospective dietary assessment in case-control studies. Moreover, it should be realized that if dietary exposure is assessed, no value should be attached to weak associations (e.g. OR between 0.7 and 1.5). Therefore, the relevance of case-control studies is in the domain of strong associations (e.g. OR above 2 or below 0.5), and not in the domain of weak associations relevant to public health and dietary recommendations to the general public. Apart from the nutrient approach, however, they may still yield valuable information on the level of food patterns, leading to identification of dietary exposures that cannot be derived from food tables (e.g. lactobacilli from fermented products, carcinogens produced during food preparation). Nevertheless, alternatives should be considered seriously.

ASSESSMENT OF DIET USING BIOMARKERS

Biomarkers of dietary exposure can provide an alternative to assessment of nutritional exposures. In the former examples on alcohol and dietary fat, biomarkers were not readily available, as may be the case for other nutrients as well. Nevertheless, biomarkers may make a valuable contribution, both to the validation of dietary methods and to prospective and retrospective investigations. Especially in prospective studies, biobanking of biological materials may add to the future value of the study, as they may be used to test as yet unforeseen hypotheses. The finite number of samples from cases in these cohorts, however, may necessitate initial screening of competing hypotheses in retrospective case-control studies, before analysis of the valuable cohort samples in a cohort-nested case-control design. In addition to exposure assessment, biomarkers may provide valuable information on early biological effects, early endpoints and disease characteristics.

The same general principles related to adequate exposure assessment (Fig. 1) and to biases in retrospective exposure assessment (Fig. 2) apply to biomarkers, but they are fundamentally different from food or nutrient assessment, because the errors are not memory-dependent (with subjectivity attached) and their assessment is mainly instrumental (objective). The potential disadvantage is that they usually refer to a more restricted time-frame than dietary methods, and their ability to integrate exposures from various dietary sources makes the translation to dietary recommendations more difficult.

The choice of a specific biomarker depends on biological aspects (sensitivity to intake, distribution over or exchange with other body pools, degree of time-integration required, other biomarker determinants, test-retest reliability) and the study design (retro- or prospective, validation, experimental) and study conduct (feasibility, logistics, ethics). The actual measurement of biomarkers includes many methodological and feasibility aspects related to collection (biological medium, completeness, contamination, additives), transport and storage (stability, temperature, oxygen, light, sample preparation, aliquots), analysis (sensitivity and specificity for analyte, amount of sample, detection limit, recovery, variability of various sources, reference and pool samples, laboratory exchange). Given the items in these lists, it is clear that the types of bias and error involved are very different from those in the assessment of habitual food and nutrient intake. This characteristic is of crucial importance in validation studies and has been discussed elsewhere (van 't Veer *et al.* 1993).

BIOMARKERS IN RETROSPECTIVE STUDIES

In addition to their potential use in validation studies and prospective cohort studies, biomarkers of dietary exposure may be of value in case-control studies. If nutrient databases are not available or are not considered sufficiently reliable for a specific research purpose, biomarkers may be the only option (e.g. in studies of trace elements, specific carotenes, phytoestrogens). Taking an example from diet and breast cancer, namely Se, the hypothesis of an inverse association between Se and breast cancer was formulated on the basis of ecological studies and animal experiments. Recognizing the impossibility of measuring individual Se intakes with dietary assessment methods, different biomarker media have been used, e.g. Se concentrations in plasma, erythrocytes and toenails. Table 2 summarizes the overall results in terms of case-control differences in biomarker levels, for retrospective and prospective study designs. On average, short-term markers of Se were lower in cases than in controls, whereas their levels were more or less similar in prospective designs (mostly cohort-nested case-control studies). Apparently, the presence of disease seems to modify the diet-biomarker association (Fig. 1). For time-integrated markers, however, the measured concentrations are stabilized by food habits and Se intake during a longer period of time, and they are likely to be less influenced by recent changes resulting from a growing tumour or recently modified eating habits. For concentrations of Se in erythrocytes and toenails the results of prospective and retrospective studies are much more in line with each other. This may imply that biomarkers of dietary exposure can indeed be applied in retrospective studies, provided they have sufficient time-integrating properties. In addition to toenail Se, subcutaneous tocopherols and carotenoids may be of special interest both to cardiovascular and malignant disease endpoints. Moreover, their combined assessment may give opportunities to study modification of effects by these functionally related antioxidant substances (Kardinaal *et al.* 1993).

Table 2. Range of case-control differences in Se concentrations in retrospective v. prospective studies for short-term (serum, plasma) and long-term (erythrocytes, toenails) biomarkers of Se status*

Study design	Short-term markers		Long-term markers	
	% Difference†	n‡	% Difference†	n§
Case-control	-36 to -2	230/420	-3 to 1	130/270
Cohort	-2 to 6	250/1370	-4 to 0	500/720

* For references see van 't Veer *et al.* (1994).

† Case-control difference divided by the mean of controls.

‡ Total number of cases/controls from five retrospective and nine prospective studies, rounded to the nearest ten.

§ Total number of cases/controls from two retrospective and two prospective studies, rounded to the nearest ten.

CONCLUSION

Adequate assessment of dietary exposures and biomarkers is highly relevant to nutritional epidemiology, both in validation studies and aetiological research. In validation studies biomarkers provide truly independent measures of exposure, whereas in aetiological studies they can either be complementary or provide measures of internal dose. Representing different levels of nutritional exposure, culturally determined food patterns may help to formulate goals for intervention and to elicit new hypotheses on potentially relevant food components. Nutrient intakes, calculated from food composition tables, add to comparability with other epidemiological studies at an international level and comparability with animal experiments. Biomarkers may be used for validation or calibration of nutrient intakes from different studies and to provide comparability with basic and *in vitro* nutrition research.

Ideally, an integrated approach to the measurement of nutritional exposures should eventually result in a biologically coherent view on the optimal diet for the reduction of chronic disease comparable with that achieved for nutritional deficiencies earlier this century.

REFERENCES

- Bingham, S. A. & Nelson, M. (1991). Assessment of food consumption and nutrient intake. In *Design Concepts in Nutritional Epidemiology*, pp. 153-191 [B. M. Margetts and M. Nelson, editors]. Oxford: Oxford University Press.
- Giovanucci, E., Stampfer, M. J., Colditz, G. A., Manson, J. E., Rosner, B. A., Longnecker, M., Speizer, F. E. & Willett, W. C. (1993). A comparison of prospective and retrospective assessments of diet in the study of breast cancer. *American Journal of Epidemiology* **137**, 502-511.
- Goldbohm, R. A., van 't Veer, P., van den Brandt, P. A., van 't Hof, M. A., Brants, H. A. M., Sturmans, F. & Hermus, R. J. J. (1993). Reproducibility of a food frequency questionnaire and stability of dietary habits determined from five annually repeated measurements. In *A Prospective Cohort Study on Diet and Cancer in the Netherlands. Design, Conduct, Analysis and First Results After 3.3 Years of Follow-up*, pp. 47-58 [P. A. van den Brandt and R. A. Goldbohm, joint thesis]. Maastricht: Datawyse University Press.
- Howe, G. R., Hirohata, T., Hislop, T. G., Iscovich, J. M., Yuan, J., Katsouyanni, K., Lubin, F., Marubini, E., Modan, B., Rohan, T., Toniolo, P. & Shunzhang, Y. (1990). Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies. *Journal of the National Cancer Institute* **82**, 561-569.

- Hulka, B. S., Wilcosky, T. C. & Griffith, J. D. (1990). *Biological Markers in Epidemiology*. Oxford: Oxford University Press.
- Kardinaal, A. F. M., van 't Veer, P., Kok, F. J., Kohlmeier, L., Martin-Moreno, J. M., Huttunen, J. K., Hallen, M., Aro, A., Delgado-Rodriguez, M., Gomez-Aracena, J., Kark, J. D., Martin, B. C., Mazaev, V. P., Riemersma, R. A., Ringstad, J. & Strain, J. J. (1993). EURAMIC study: antioxidants, myocardial infarction and breast cancer. Design and main hypotheses. *European Journal of Clinical Nutrition* **47**, Suppl. 2, 564–572.
- Kok, F. J. & van 't Veer, P. (1989). The strength of relationships which can be detected between diet and disease. In *Epidemiology Nutrition and Health. Proceedings of the First Berlin Meeting on Nutritional Epidemiology*, pp. 19–29 [L. Kohlmeier and E. Helsing, editors]. London: Smith Gordon & Co.
- Longnecker, M. P., Orza, M. J., Adams, M. E., Vioque, J. & Chalmers, T. C. (1990). A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. *Cancer Causes and Control* **1**, 59–68.
- Michels, K. B. & Willett, W. C. (1992). The women's health initiative: daughter of politics or science? *Principles and Practice of Oncology* **6**, 1–9.
- Rothman, K. J. & Poole, C. H. (1988). A strengthening program for weak associations. Meeting of the Netherlands Epidemiological Society, Wageningen, April 28–29, 1988. *International Journal of Epidemiology* **17**, S955–S959.
- van den Brandt, P. A., Goldbohm, R. A., van 't Veer, P., Volovics, A., Hermus, R. J. J. & Sturmans, F. (1990). A large-scale prospective cohort study on diet and cancer in the Netherlands. *Journal of Clinical Epidemiology* **43**, 285–295.
- van 't Veer, P., Kardinaal, A. F. M., Bausch-Goldbohm, R. A. & Kok, F. J. (1993). Biomarkers for validation. *European Journal of Clinical Nutrition* **47**, Suppl., S58–S63.
- van 't Veer, P., van den Brandt, P. A. & Kok, F. J. (1994). Selenium status and cancer: results from epidemiology. In *Proceedings of the Fourth International Congress on Trace Elements in Biology and Medicine* [A. Favier, editor] (In the Press).