

# Food photographs: practical guidelines I. Design and analysis of studies to validate portion size estimates

Michael Nelson<sup>1,\*</sup> and Jóhanna Haraldsdóttir<sup>2</sup>

<sup>1</sup>Department of Nutrition and Dietetics, King's College London, Campden Hill Road, London W8 7AH, UK; <sup>2</sup>Research Department of Human Nutrition, Royal Veterinary and Agricultural University, Rolighedsvej 30, DK-1958 Frederiksberg C, Copenhagen, Denmark

Submitted: 21 April 1998; Accepted: 8 August 1998

The use of photographs to assist subjects with estimation of food portion sizes is widespread. Only recently, however, has it been recognized that the format of their presentation may introduce errors into the measurement of food consumption. Several questions arise when photographs are used in dietary assessment:

- Do they help to improve estimates of portion size? (Intuitively we believe they do.)
- If so, by how much (in relation to other techniques)?
- Does their use introduce bias in the assessment of portion size?
- Is the bias the same in all subgroups (e.g. elderly, overweight)?
- Is it appropriate to use correction factors?
- How much does the design of the photograph series influence their usefulness?

Validation studies have been carried out in attempts to answer some of these questions, but there are wide variations in the approaches taken by different authors to both design and analysis. Some papers directly concerned with validation of food portion estimates using photographs are listed in Table 1. It is not the purpose of this paper to review such studies, but rather to present guidelines which may help to overcome some of the problems identified in previous studies. (For review papers on portion sizes see Young and Nestle, 1995; Cypel *et al.*, 1997<sup>1,2</sup>.)

The need for guidelines for photograph series became apparent following discussions on the validity of portion size estimates at a workshop organized by COST99 (EUROFOODS) in Norwich in April 1997. Subsequent discussions of draft papers circulated to members of the COST99 workshop and to members of the UK Nutritional Epidemiology Group in September 1997 and January 1998 led to the development of the present paper and its companion.

Before proceeding, it is useful to clarify a few terms that will be used throughout this paper:

- *Photograph series*: a set of photographs depicting different amounts of a particular food.

- *Photographic atlas*: a set of photograph series, usually bound together in a single volume.

- *Portion*: the amount eaten on any one occasion (first plus subsequent helpings).

- *Serving*: the amount of food served in a single helping.

## Theoretical concepts

In thinking about the design of photograph series and the development of validation studies, researchers must be aware of the psychological constructs which allow subjects to relate a photographic depiction of a given amount of food to an amount of food actually consumed.

1. *Perception of foods* in photographs in direct comparison with foods being shown to the subject.
2. *Conceptualization of foods*, i.e. the ability to translate an abstract mental impression of an amount seen or eaten into an amount depicted in a photograph.
3. *Memory of amounts eaten*, especially relevant to recall of diet using 24-h recall, diet history (DH) or food frequency questionnaires (FFQ).

Errors in perception, for example, will be of especial importance in studies which require subjects to provide descriptions of portion sizes of foods which are present in front of them (e.g. estimated records) (see Margetts and Nelson<sup>3</sup> for definitions of methods). In contrast, conceptualization and memory will be of especial importance in studies relating to recall of diet, but the contribution of the two components to error will vary according to the type of method, e.g. 24-h recall compared with FFQs.

## Sources of variance

In thinking about the design of photograph series and validation studies, it is necessary to identify the components of variance which may contribute to any disagreements between actual and reported consumption. Work in England, Denmark and elsewhere has begun to identify the main sources of

\*Corresponding author. E-mail: michael.nelson@kcl.ac.uk

**Table 1** References and details on validation studies of food photographs

Author and Reference	Design	Subjects
Haraldsdóttir <i>et al.</i> , 1994 <sup>15</sup>	<i>Usual portion sizes:</i> photo portions selected from FFQ vs. weighed portions from 14 day weighed food record	Men and women 40–64 years n = 144
Howat <i>et al.</i> , 1994 <sup>8</sup>	<i>Present portion size:</i> photo portions selected at presentation vs. weight of the presented portions	Women 20–49 years n = 44
Kirkcaldy-Hargreaves <i>et al.</i> , 1980 <sup>16</sup>	<i>Present portion size:</i> photo portions selected at presentation vs. weight of the presented portions	Men and women Mean age 22 years n = 60
Kuehnemann <i>et al.</i> , 1994 <sup>17</sup>	<i>Usual portion size:</i> photo portions selected in FFQ vs. 12 × 1 day weighed food intakes	Women* 22–50 years n = 22
Nelson <i>et al.</i> , 1994 <sup>18</sup>	<i>Present portion size:</i> photo portions selected at presentation vs. weight of the presented portions	Men and women 18–90 years n = 51
Nelson <i>et al.</i> , 1996 <sup>14</sup>	<i>Test meal portion sizes:</i> Photo portions selected after test meal vs. weight of the test meal portions	Men and women 18–90 years n = 136

\* The portion estimates concerned the intake of young children that the women took care of.

variance which need to be assessed. These can be categorized into five groups related to:

- The format and administration of photographs.
- The nature of the food being assessed.
- The characteristics of the subjects.
- The context of the administration of the dietary measuring instrument (clinical, educational or population survey).
- The reference measure.

These are the main topics addressed in this article.

### The purpose of validation studies

Validation studies are likely to have one of two major aims. These are in part dictated by the underlying purpose of the studies in which photographs are being used.

#### **(A) Assessment of errors related to determination of food portion sizes per se**

*Learn how food photographs affect the precision of portion size estimation.* In studies of this type the main aim will be to assess discrepancies between reported consumption based on the amounts of foods depicted in a photograph (the test measure) and the amounts actually present on a plate or consumed (the reference measure). Such studies are necessary for understanding the consequences of errors in terms of portion size estimates and the scope for improvements with training. They also have direct application in clinical and educational settings where the aim is to help subjects identify portions of a given size in relation to a clinical objective (e.g. weight loss).

#### **(B) Assessment of errors in estimates of food consumption and nutrient intake in population studies where the photos are used**

*Learn how food photographs contribute to misclassification of subjects according to estimates of food consumption or nutrient intake.* Misclassification may arise because of (a) errors in interpretation of photographic images and (b) subgroup differences in interpretation. Here the aim may be to assess how estimates of food consumption or nutrient intake based on a given dietary assessment tool (e.g. 24-h recall, FFQ) may differ according to whether or not photographs have been used to help subjects estimate portion size. Comparisons need to be made between three sets of observations based on:

- the test measure used with photographs;
- the test measure used without photographs; and
- a reference measure of diet.

In both types of study, the key elements are the same:

1. To have one or more reference measures which are free from bias or whose errors are known.
2. To administer the test instrument in the way it will be used in the main study.
3. To have a sample in the validation study which shares the demographic characteristics of the population from which the sample for the main study is to be drawn. Validation in one subgroup (e.g. females 35–54 years) does not imply validity in another (e.g. males 65 years and over).

The main difference between these studies is that type A aims is to provide information on the estimation error

for each food/photo series separately, whereas type B aims to provide information on the *net resulting error* introduced when the photos are included as part of the dietary instrument (e.g. as part of an FFQ where some foods are estimated by photos, other by standard units (slices of bread, number of apples)) in a population-based study. Type A studies, which are usually more simple and smaller than a type B study, are particularly useful in the process of developing photo series, and for comparing different alternatives and/or different types of aids such as household measures or food models (see companion paper in this issue). With appropriate design, information derived from type A studies can be extracted from type B studies.

There are numerous issues that need to be considered when designing a validation study:

- characteristics and number of the subjects
- the nature of the dietary assessment
- the context of administration of the measuring instrument
- the sources of error in the reference measure, and
- the types of analyses to be undertaken.

This section of the paper deals with each of these in turn.

### Characteristics and number of subjects (number of observations)

#### Characteristics of subjects

Previous validation studies have suggested that certain groups of subjects differ in their ability to make effective

use of photographs. Table 2 lists the characteristics and associated problems. When designing a validation study, it is necessary to assess performance in relation to all of these factors if they are likely to influence findings in the main study.

Other characteristics may also have an influence on perception and need to be addressed. These may be short term (e.g. whether or not a subject is hungry) or long term (e.g. level of food restraint, independent of factors such as body mass index (BMI) or weight).

#### Number of subjects

The number of subjects to be included in a study will depend upon the variance of the observations, and appropriate power calculations need to be carried out to ensure that sample size is adequate<sup>4</sup>. The power calculation will reflect the size of the differences between actual and estimated portion size (or food consumption or nutrient intake assessed using test and reference measures) which the study intends to demonstrate as being of statistical significance at  $P < 0.05$ .

For type A validation studies, the number of observations (as opposed to the number of subjects) required to demonstrate statistically significant differences between the actual and estimated portion sizes will be given by the expression:

$$n = \frac{2\sigma^2(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

where  $\sigma^2$  is the variance of the estimates of portion size and  $d$  is the size of the average difference

**Table 2** Subject characteristics which may need to be addressed in validation studies, the nature of the problem relating to the use of photographs, and the likely influence on outcomes relating to estimates of portion size, food consumption or nutrient intake

Characteristic	Nature of problem	Consequence
Age	Children under 12 years and older people less able to create the necessary mental constructs	Children more likely to be misclassified; older people more likely to overestimate intake
Gender	Men and women differ in perception of portion size using photographs	Comparison of diet–disease risk in men and women confounded by differences in dietary assessment
Weight and BMI	Heavier subjects with higher BMI tend to underestimate portion size	Misclassification of subjects due to under-reporting
Education	Level of education may influence perception and conceptualization skills	Comparison of diet–disease risk in different educational strata confounded by differences in dietary assessment
Occupation	Subjects who work in food-related environments (restaurants, academic departments of nutrition, dietetics departments, etc.) may have better perception of portion size than other subjects	Apparent benefits of using photographs may be overestimated if validation studies rely on people in food-related occupations
Culture	Food culture and exposure to systems of measurement result in differences in the utilization of photographs in different social class or ethnic groups	Comparison of diet–disease risk in different social class or ethnic groups confounded by differences in dietary assessment

between actual and estimated portion sizes. The variance (expressed as a coefficient of variation) typically varies from 25 and 80%, while average differences (expressed as a percentage) range from as low as 5% to as high as 100% (of actual weight) or more. If the aim is to have differences demonstrated to be statistically significant at the 5% level with 80% power, then

$$(Z_{\alpha/2} + Z_{\beta})^2 = (1.96 + 0.84)^2 = 7.85$$

Say CV% = 50% and per cent difference is 20%, then *n* (the number of observations required) will be approximately 100. This can be achieved by having relatively few subjects making multiple measurements at different portion sizes. As a rule of thumb (in order to obtain adequate diversity in subject characteristics and abilities) we recommend having at least 25 men and 25 women in a type A validation study, ensuring that the cross-section of the subjects corresponds to that of the population in which the photograph series are ultimately intended for use.

A type B validation study might set out to investigate whether estimates of vitamin C intake are more accurate (closer to estimates based on validated weighed inventories of diet, the reference measure) when photographs of fruits and vegetables are used in conjunction with an FFQ, compared with descriptions of amounts in household measures. Suppose the aim is to show that a higher proportion of subjects have their vitamin C intake classified to within ±20% of the reference measure. Say that in the photograph group, the proportion of subjects classified to within ±20% is 80%, and that in the household measures group the proportion classified to within ±20% is only 50%. The number of subjects per group (*n*) is given by the formula:

$$n = \frac{2\bar{p}\bar{q}(Z_{\alpha/2} + Z_{\beta})^2}{d^2}$$

where:  $\bar{p}$  = the average proportion in the two groups;  $\bar{q} = 1 - \bar{p}$ ; *d* is the difference in the proportion; and  $(Z_{\alpha/2} + Z_{\beta})^2$  reflects values for *P* (type 1 error  $\alpha$ ) and power.

In this example,

$$\bar{p} = (80\% + 50\%)/2 = 65\%$$

$$\bar{q} = 0.35,$$

$d = 80\% - 50\% = 30\%$  (or expressed as a difference between proportions,  $d = 0.3$ ), and

$$(Z_{\alpha/2} + Z_{\beta})^2 = (1.96 + 0.84)^2 = 7.85$$

(where *P* = 0.05 and power = 80%)

so *n* = 40 per group. If the proportion classified to within ±20% was only 70%, then the study would require 95 subjects per group to demonstrate a statistically significant benefit of using the photographs.

If the impact of factors such as gender, age, BMI, etc., are to be investigated, subjects have to be selected for that purpose, and the number of subjects has to be sufficiently large to allow analyses to be undertaken in each subgroup. If there is no intention to investigate the independent effects of these factors, the sample for the validation study should have equal numbers of men and women and a distribution of age and other characteristics which reflects the population within which the main study is to be undertaken.

### The design of validation studies

Essentially, there are three formats of dietary assessment which relate to the validity of assessment of portion size (Table 3). Each of these formats is associated with particular dietary assessment methods, and requires application of different psychological skills. Assessment of *yesterday's* portion size is appropriate for use in 24-h recall, and utilizes functions of perception, conceptualization and memory. Assessment of *usual* portion size (in FFQ and DH) will make far greater demands on these skills. Strong conceptualization and memory skills, especially, will be needed to integrate information over meals, days and seasons. For the assessment of *present* portion size, used with any form of current record, skills of perception will be in greatest demand, although conceptualization and

**Table 3** Relationship between type of assessment of portion size, dietary method and psychological construct needed to provide appropriate responses

Characteristic of portion size assessment	Associated dietary assessment method	Associated psychological constructs		
		Perception	Conceptualization	Memory
Yesterday's portion size	24-h recall	✓	✓✓	✓✓
Usual portion size	FFQ DH	✓	✓✓✓	✓✓✓
Present portion size	Estimated food record Food checklist	✓✓	✓	✓

memory skills will be needed occasionally for foods not recorded at the time of consumption.

In planning a validation study, researchers should consider the contribution that each of the psychological constructs makes to any errors in reporting of portion size (which may differ between foods or food groups). These constructs (or skills) should be related to the dietary survey methods for which photographs are to be used. Ideally, a validation study would be able to identify the psychological components that were contributing to errors in estimation of portion size (and hence contributions to errors in estimates of food consumption or nutrient intake). This would facilitate improvements in the dietary assessment tools being used.

If it is not possible to have a detailed design which addresses the psychological component separately, a validation study should, at the very least, replicate the context of the main study to determine the size of any errors. If the photographs are to be used primarily in a 24-h recall, for example, then they should be tested in that context. If the photographs are to be used in a different context later, every effort should be made to secure the necessary resources to undertake an appropriate validation for that context. It must also be recognized that a validation study which addresses only one of the psychological constructs (e.g. perception) will not be sufficient to determine the validity of the photographs in a wider context (e.g. assessment of diet using an FFQ). If resources cannot be made available to carry out an appropriate validation exercise, authors must acknowledge when they are using an unvalidated tool.

For validation of portion sizes *per se* (study type A) the comparisons will be between estimates of amounts based on photographs (the test measure) and actual quantities of food (the reference measure). In validation studies including the entire dietary assessment instrument (study type B) two test measures may be relevant, one with the photographs and another without the photos (e.g. using household measures or standard portions instead).

The following sections outline the components of design for validation studies relating to each of the formats listed in Table 3. For each of these outlines, there will no doubt be variations which may be appropriate for specific research objectives.

### **Study type A: validation of portion sizes *per se***

#### ***Assessing yesterday's portion size***

Validations of this type compare the recall of food consumed with actual amounts eaten on the previous day.

● **Test measure:** 24-h recall of diet, or recall of specific meals or foods from the previous day. Quantities recalled using photographs.

● **Reference measure:** valid weighed record of all foods (or selected foods or meals) consumed on the previous day. The record must be kept at the time of consumption.

The key components of the study design are shown in Box 1.

#### *Box 1*

1. The test measure is applied the day after the reference measure.
2. Presentation of the foods should be in a realistic setting, i.e. foods served as a realistic meal that the subjects consume.
3. Subjects must be able to serve themselves, either by doing this directly or by selecting one of several pre-weighed portions.
4. Food portions must be weighed.
5. Plate wastage must be measured and subtracted.
6. Subjects must be asked to recall their diet the following day, using photographs to help them indicate portion sizes consumed.

Unrealistic settings should be avoided, i.e. asking subjects to select portions from a large buffet containing more foods than they would typically consume at a single meal.

An effective design would involve subjects attending a dining room where they can serve themselves while the foods served can be weighed covertly. Although this means taking subjects away from their usual environment this is preferable to letting subjects weigh their own food at home, as that might draw their attention inappropriately to the portion size.

#### ***Assessing usual portion size***

Validations of this type compare estimates of usual consumption based on FFQ or DH with records which reflect food consumption or nutrient intake over a corresponding period<sup>1, \*</sup>.

● **Test measure:** estimates of usual portion sizes using photographs, in the context of FFQ, DH or as a separate list of questions on this issue.

● **Reference measure:** weighed record of all foods, or specific foods, consumed in a period representative of the period covered by the FFQ or DH.

\* As a general principle, in validation studies relating to FFQ or DH, subjects should complete the FFQ or DH *twice* before the administration of the reference measure (to facilitate assessment of reproducibility) and once after to assess validity<sup>5</sup>.

The key components of the study design are shown in Box 2.

*Box 2*

1. The test measure must be applied before the reference measure.
2. Subjects must keep weighed records of food consumption.
3. Records must be kept long enough to include an adequate number of entries for foods for which an estimate of 'usual' portion size is being assessed (minimum number of entries, e.g. three or four, depending upon the requested precision).
4. Plate wastage must be weighed and subtracted.

In this type of study, it is necessary to ensure that the food under scrutiny appears a sufficient number of times in the record to provide an estimate of usual portion size with a known confidence interval. The number of occurrences of the food in the record is likely to vary from food to food and from subject to subject. Thus, at the outset of the study, it will be necessary to determine the minimum number of days of record that will provide enough observations for relatively infrequently eaten foods for which information on portion size is needed.

If the food records do not include all foods but are limited to a few foods where photos are used for portion size estimation in the FFQ/DH, this may make it easier to obtain a large number of eating occasions per food. However, this advantage must be balanced against a potential bias introduced in consumed portion sizes because of the special attention drawn to those specific foods with this kind of design.

**Assessing present portion size**

Validations of this type compare estimates of present food consumption with amounts estimated using photographs, and they form the basis for assessment of errors related specifically to perception.

- *Test measure*: estimate of food portion size based on photographs.
- *Reference measure*: pre-weighed portions of food presented to subjects.

The key components of the study design are shown in Box 3.

*Box 3*

1. The test measure is applied at the same time as the reference measure is presented.
2. The weighed food portions may be consumed by the subject as a meal, or just presented to them without being consumed.

The most realistic setting for this type of test is to have subjects serve themselves with foods, which are covertly weighed, and then let them estimate the portion size, using the photos, prior to consumption.

An alternative design is to present subjects with pre-weighed portions of foods without later consumption. This is a less realistic setting, but has the advantage that a larger number of foods may be tested in one session. However, care must be taken not to present so many foods that a fatigue effect is introduced. The deliberate matching (or deliberate lack of matching) of the pre-weighed portions and the photo portions must be systematically planned in advance. The same is true regarding the order of presentation.

The answers given by the subjects may be self-administered, i.e. the subject fills in a form, or they may be collected by an interviewer who writes the answers down. Self-administration comes more close to the situation where subjects keep an estimated food record with photos for portion estimates. This type of study can readily be carried out in a Departmental kitchen.

**Study type B: evaluation of the effect of photographs on the estimation of food consumption or nutrient intake**

In this type of validation study it may be relevant to include two test measures, one with the photographs and another without the photos but using a relevant alternative instead (e.g. household measures or standard portions). As there will always be an error linked to assessment of dietary intake, the purpose of introducing a second test measure is to evaluate whether the net resulting error may be reduced by using photos for portion estimations, instead of an alternative such as household measures.

**Assessing yesterday's intake**

Validations of this type compare the recall of foods consumed with actual amounts eaten on the previous day.

- *Test measure 1*: 24-h recall of diet, or selected meals, from the previous day. Quantities recalled using photographs.
- *Test measure 2*: 24-h recall of diet, or selected meals, from the previous day. Quantities recalled in household measures.
- *Reference measure*: valid weighed record of all foods, or selected meals, consumed on the previous day. The record must be kept at the time of consumption.

The key components of this study design are much the same as shown in Box 1.

The weighing of food intake may be carried out at home, with subjects weighing their own foods. This

introduces some problems, however, as it may draw subjects' attention to portion size, and affect their ability to describe food portions with or without photographs in a way which is different from subjects who are asked to complete a 24-h recall with no training. There may also be doubt about the validity of the weighed records kept by the subject.

Alternatively, this kind of study can be carried out in a metabolic kitchen or cafeteria where subjects can serve themselves while the foods can be weighed covertly. This takes subjects away from their usual environment, however, and this may bias their perception.

### **Assessing usual intake**

Validations of this type compare estimates of usual consumption based on FFQ or DH with records which reflect food consumption or nutrient intake over a corresponding period.

- *Test measure 1*: FFQ or DH. Quantities estimated using photographs (where appropriate).
- *Test measure 2*: FFQ or DH. Quantities estimated in household measures.
- *Reference measure*: validated weighed record of all foods consumed in a period representative of the period covered by the FFQ or DH.

The key components of the study design are shown in Box 4.

#### *Box 4*

1. Subjects must keep validated weighed records of food consumption.
2. The records must be representative of the period over which the FFQ or DH is to be evaluated.
3. The records must include an adequate number of entries for foods for which an estimate of 'usual' portion size is being assessed.
4. Plate wastage must be weighed and subtracted.

In studies of this type the focus of the analysis will in part dictate the design of the study. If the aim is simply to see if photographs improve the level of agreement between the test measure and the reference measure in terms of nutrient intake, the design can remain relatively uncomplicated. Conversely, the aim might be to identify the component of error in any disagreement between test and reference measure that can be attributed to errors in the estimate of portion size related to the use of photographs (as opposed to error related to frequency of consumption of particular foods). In this case it will be necessary to

ensure that the food under scrutiny appears a sufficient number of times in the record to provide an estimate of usual portion size with a known confidence interval (see section above on characteristics and number of subjects and observations).

### **Assessing present intake**

Validations of this type compare estimates of present food consumption with estimated intake using photographs. It looks at how perception may contribute to error in the assessment of food consumption or nutrient intake based on prospective records of consumption or prospective food check lists.

- *Test measure 1*: diet record or food check list. Quantities estimated using photographs (where appropriate).
- *Test measure 2*: diet record or food check list. Quantities estimated in household measures.
- *Reference measure*: validated weighed record of all foods consumed.

The key components of the study design are shown in Box 5.

#### *Box 5*

1. There must be weighed records of food intake, kept either by the researcher in a controlled setting, or by the subject.
2. Assessment of food portion size by the subject should be made at the time the food is consumed and not retrospectively.
3. In a controlled setting subjects record amounts using photographs (and, if possible, in household measures).
4. Where subjects are keeping their own weighed record of intake, they should weigh all foods and also (for those foods for which photographs are available) record amount using the photos (and, if possible, using household measures).
5. Plate wastage must be weighed and subtracted.

The preferable design for this type of study is a controlled setting, e.g. a metabolic unit or an institution, where subjects consume *all* their meals in a setting in which the foods can be weighed covertly, and only record their estimated intake using the photos. This is a useful technique for assessing a wide range of foods for which data on perception are needed. An alternative strategy in this context is to present subjects with pre-weighed portions and ask them to use photographs to assess the portion size prior to consumption.

### The context of administration of the measuring instrument

In any validation study, it is necessary to ensure that the context replicates faithfully the context in which the photographs are to be used in the main study. The context or the purpose of the assessment will introduce further elements of variance above and beyond those already listed. The subject must be absolutely clear regarding the information that is being requested. There may be additional elements of design that are useful to consider.

If the decision has been made to use an existing atlas:

- Is the number and range of foods depicted appropriate to the study population?
- If there are fewer photographs than you would like, is the potential loss of precision in estimates of intake acceptable?

Do subjects:

- report on food viewed or food consumed? (Having to consume the food may influence the portion sizes being considered.)
- report serving size or portion size? (Clarification of this is *vital!*)

Are subjects:

- given clear instructions for self-selection of foods and self administration of questionnaires?
- presented with photographs by the interviewer or allowed to administer the photographs to themselves?
- trained in the use of photographs before being asked to assess portion size?
- able to feed back information on ease of use?

With regard to the study implementation:

- Are interviewers adequately trained in the administration of the photographs?
- Is the mode of reporting appropriate to the purpose of the study (visual analogue scale versus selection of individual images) considering possible fatigue effects?

Several additional points warrant mention:

- Subjects should be blind to the purpose of the validation study, and should not be told that it is their ability to estimate portion sizes that is being tested.
- Photographs may be either self-selected and self-administered, or presented in an interview. Does it make a difference if the subject is presented with a set of photographs by the researcher or if they select the images themselves? The possible differences in

estimates related to self- or interviewer administration need to be kept in mind when thinking of the applicability of the results of a validation study.

- Responses to a dietitian may differ from those given to a research assistant. This also needs to be considered in the design of a validation study. The role of interviewer training (in presenting the photographs) and potential subject-interviewer interactions in a validation study needs to reflect the reality of the administration of photographs in the main investigation.

- Is it worth considering the possibility of training subjects to use the photographs more effectively? Time taken to provide a short period of training (learning to improve perception) may yield later benefits in terms of correct subject classification according to food consumption or nutrient intake.

Finally, it is important to consider the context in which the study is to take place. For example: the recall of portion size information during the completion of an FFQ sent to a subject by post will be that of a person who is unlikely to have had any training in the assessment of portion size. This may differ in character from the assessment of portion size by a subject who is taking part in a validation study which involves repeated measures of diet over an extended period. Bolland *et al.*<sup>6,7</sup> and Howat *et al.*<sup>8</sup> have demonstrated that training improves subjects' ability to use photographs to describe amounts consumed.

### The sources of error in the reference measure

In studies in which the primary aim is to understand the nature of the errors associated with the assessment of portion sizes of specific foods, the reference measure should be foods which are weighed accurately. The weighing may be either overt (e.g. when a subject records a weighed inventory of consumption) or covert (e.g. when an investigator prepares dishes of known weights and asks the subject to identify a photograph which they believe most closely depicts the amount of food shown on the plate). In this context, the principal source of error in the reference measure will be the mechanical accuracy of the weighing scales in use, although where subjects do their own weighing, other inaccuracies could be introduced due to technical errors in weighing or recording.

In population surveys, the sources of error in the reference measures have been described in detail by Nelson<sup>5</sup>. The work of Bingham *et al.*<sup>9,10</sup> has highlighted the need to have independent markers of the validity of dietary measures whose errors do not correlate with those of the test measure (e.g. doubly-labelled water, urinary nitrogen). It is clear, however, that the markers relating to the validity of dietary assessments almost



always relate to a composite entity (e.g. total energy expenditure as a marker for energy intake, urinary nitrogen as a marker for protein intake), and not to individual foods or groups of foods. Lutein in serum has been suggested as a useful marker for dark green vegetable consumption<sup>11</sup> and poly-aromatic hydrocarbons in serum as a marker for charred grilled foods in non-smokers<sup>12</sup>.

### Analysis of validation studies

There are several types of analyses which can be undertaken relating to either portion size assessment or population estimates of food and nutrient intake.

#### Portion size estimates

There are several ways in which results from comparisons of test measures (photographic estimates of portion size) and reference measures (actual weights) can be compared.

*Difference between test and reference measures.* These can be analysed according to whether subjects are asked to indicate which photograph most nearly represents the actual portion size or to use a visual analogue scale. In the former case, there will be a minimum theoretical error not greater than one half of the increment between portion weights depicted in the photographs. In the latter, a subject with perfect judgement could estimate portion size with no error. The standard deviation of the mean difference between actual and estimated portion sizes (and hence the standard error and coefficient of variation) will inevitably be greater for the former type of assessment. In both types of analysis, results can be expressed in terms of absolute differences (*g*):

$$\text{Error (g)} = \frac{\text{test measure} - \text{reference measure}}{\text{reference measure}}$$

or in terms of per cent error:

$$\% \text{ Error} = \frac{\text{test measure} - \text{reference measure}}{\text{reference measure}} \times 100$$

An alternative way of expressing the error is to describe it in terms of the difference between the number of the selected photograph and the number of the photograph which most closely represents the actual portion weight. Table 4 shows results relating to the same findings presented in these three ways.

*Correlation and regression.* Pearson and Spearman correlation coefficients describe the level of agreement between two sets of measures across the entire range of measures. They do not, however, reflect the level of agreement between individual pairs of observations. They lend themselves to regression analyses. Simple regression can be used to generate regression lines which when plotted may help to reveal biases which occur at different levels of portion size (e.g. small, medium, large). Multiple regression analyses can be used to identify potential sources of influence on agreement (e.g. age, BMI).

Useful alternative measures are the intra-class correlation coefficient ( $r_1$ ) and the *kappa* statistic ( $\kappa$ ), which take into account both the order of agreement and the differences between two measures. While  $r_1$  and  $\kappa$  are difficult to interpret on their own, they lend themselves very well to situations in which two different approaches to estimating portion size are being assessed (e.g. use of household measures compared with photographs both validated against weighed measures).

*Classification by quantiles (e.g. tertiles, quartiles).* Classification by quantiles is more generally useful than the correlation coefficient because it describes the extent of the agreement between actual and reported measures in different segments of the distribution. Table 5 shows the level of agreement in nutrient intake (by thirds of the distribution) for intakes based on portion sizes estimated using photographs versus actual amounts, and for intakes based on 'standard portions' versus actual amounts. The table shows the extent of misclassification and the Pearson correlation coefficient  $r$ . It reveals the extent to which misclassification increases as  $r$  decreases.

**Table 4** Comparison of actual and reported portion sizes, expressed in three ways: error (*g*), error (%) and difference between the number of the selected photograph and the number of the photograph which most closely represents the actual portion weight

Food item	<i>n</i>	Mean actual portion size ( <i>g</i> )		Mean reported portion size ( <i>g</i> )		Error ( <i>g</i> )		% Error		Percentage of subjects with a selection error of <i>n</i> photographs				
		Mean	SD	Mean	SD	Mean	SE	Mean	SE	0	±0.5	±1.0	±1.5	±2 or more
Spaghetti	32	173	64	149	69	-24*	9	-14*	5	47	28	25	0	0
Vegetables (mixed boiled)	32	86	39	89	41	+3	7	+4	9	44	35	16	3	3
Meat stew	32	210	95	226	117	+16	13	+8	6	34	41	25	0	0

\*  $P < 0.05$ .

**Table 5** Classification of subjects according to thirds of the distribution of the energy and nutrient contents of meals based on estimates of portion size using photographs versus actual portion size and standard portion sizes versus actual portion size, based on 18 foods eaten at one test meal, for 136 subjects (adapted from Nelson *et al.*<sup>14</sup>)

Nutrient	Photographs vs. actual amounts (per cent classified in:)				'Standard portions' vs. actual amounts (per cent classified in:)			
	Same third	Adjacent third	Opposite third	<i>r</i> *	Same third	Adjacent third	Opposite third	<i>r</i> *
Energy (kcal)	79	20	1	0.92	71	29	1	0.80
Fat (g)	83	16	2	0.93	62	37	1	0.85
Fatty acids (g)								
saturated	80	19	1	0.93	72	27	1	0.84
monounsaturated	78	22	0	0.91	64	36	0	0.84
polyunsaturated	86	14	0	0.96	79	21	0	0.90
Iron (mg)	68	32	0	0.84	60	40	0	0.79
Calcium (mg)	85	15	0	0.94	79	21	1	0.85
Vitamin C (mg)	75	25	0	0.91	63	35	2	0.74
Non-starch polysaccharides (g)	72	28	0	0.84	61	35	4	0.78

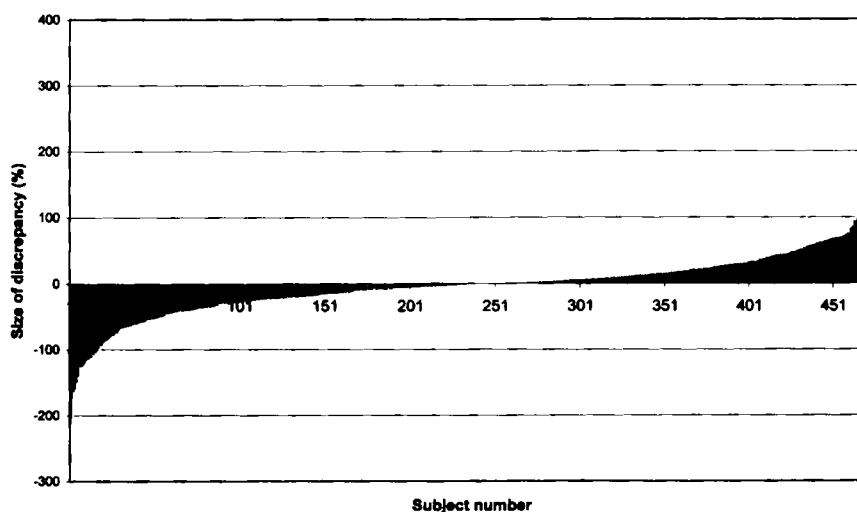
\*All values for '*r*' significant at  $P < 0.01$ .

*Cumulative error plots.* Figure 1 shows the size of the discrepancy between estimated and actual portion sizes for all foods consumed in the study by Nelson *et al.*<sup>14</sup>. The plot makes clear the full range of the errors and the numbers of subjects with minimal or extreme errors.

*Cumulative per cent error.* This is extremely useful for comparing errors associated with different approaches to assessment of portion size, or between different foods or between subgroups within the study. Figure 2 shows the plot of cumulative errors described in Table 4 relating to the agreement between photographs selected by subjects and the photograph most nearly representing the actual portion weight. Subjects were allowed to point to a specific photograph or between two photographs to indicate the nearest equivalent. An error relating to a point between two photographs corresponds to a discrepancy of half a photograph. The

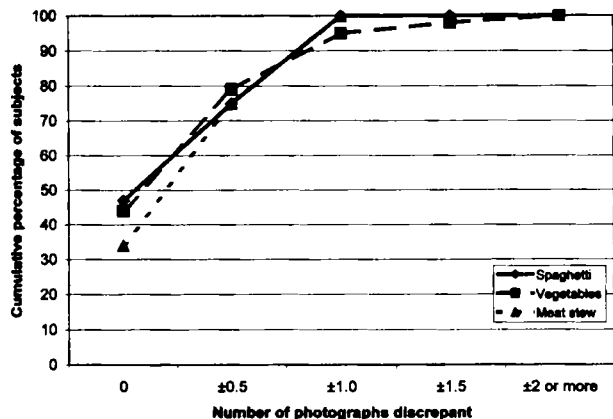
graph shows that subjects were more successful at identifying the correct photograph for vegetables (44%) than for meat stew (34%), but that the cumulative error is greater for vegetables than for meat stew.

Figure 3 shows the cumulative percentage of subjects ( $n = 136$ ) plotted against the size of the difference between test (photographic estimate) and reference (weighed) measures, expressed as a percentage of the reference measure according to portion size. There is a more rapid accumulation of accurate results for the large and medium portions. The implication is, for example, that it may be easier to demonstrate an association between consumption of a particular foodstuff and a specific disease outcome in men than in women because men are more likely to eat larger portions (and assuming, of course, that men and women are equally good at using photographs to estimate portion size).

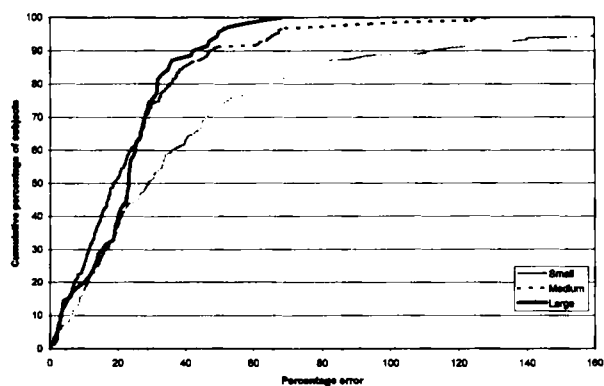


**Fig. 1** Size of discrepancy (1%) between estimated and actual portion size for all foods ( $n = 475$ )

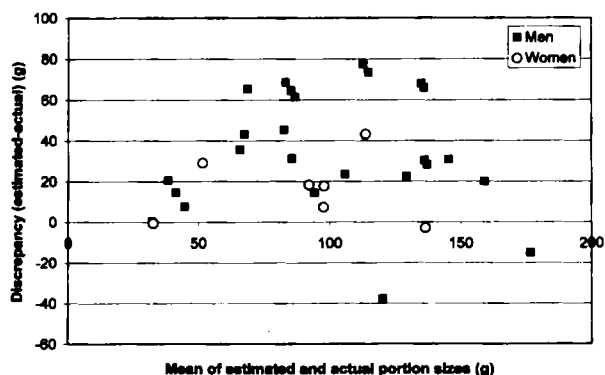
**Bland–Altman plots.** Bland and Altman<sup>13</sup> suggest plotting the difference between two measures against their mean (or sum). This reveals biases which may be associated with measures at particular points along the distribution. It also lends itself well to plotting results



**Fig. 2** Cumulative percentage of subjects whose choice of photograph in relation to the actual food portion weight was correct or discrepant



**Fig. 3** Cumulative percentage of subjects ( $n = 136$ ) plotted against the difference between test (photographic estimate) and reference (weighed) measures expressed as a percentage of the reference measure, according to portion size



**Fig. 4** Mean portion size estimate of broccoli and size of discrepancy (g), by gender

from different subsets of the sample (e.g. according to body size or age) in order to see if biases are present in particular subgroups. Figure 4 shows the relationship between mean portion size estimate of broccoli and size of discrepancy, by gender. The implication from this figure is that errors tend to be over- rather than underestimates, that men are more likely to have larger errors than women, and that the size of the error increases as the portion size increases – especially for men.

**Population studies**

The analyses in population studies can be similar in character to those undertaken relating to specific portion sizes. The scope of the analyses can, however, embrace groups of foods and nutrients. Errors can be expressed in terms of absolute differences in estimates of food ( $\text{g day}^{-1}$ ) or nutrient (e.g.  $\text{mg day}^{-1}$  calcium intake) assessed using different approaches to the determination of portion size (e.g. photographs versus household measures versus a reference measure). It may also be possible to use quantile analysis (seeing how well subjects are classified in groups ranked according to estimated level of intake)<sup>14</sup>.

**Conclusions**

A weakness of all dietary assessment methods is the inability to identify which foods are well measured and which are poorly measured. Photographic validation studies help to:

- quantify the level of precision of measurement for a wide variety of foods;
- identify those foods which cannot be measured reliably; and
- understand the extent to which photographs improve estimates of nutrient intake.

There are three outcomes likely to accrue from validation studies of the use of photographs in the assessment of portion size.

1. It will be possible to identify subgroups in the population for whom photographs are inappropriate.
2. It may allow for the introduction of correction factors for the population as a whole or when making comparisons between subgroups. If there are systematic errors that differ between subgroups, then it may be possible to introduce correction factors specific to those subgroups.
3. Where photographs are seen to be inappropriate as an aid to the assessment of portion sizes, it will be possible to use the techniques outlined in these guidelines to make comparisons using other approaches (e.g. computer images on screen).

There is a persistent danger that the errors inherent in any nutritional investigation may undermine the ability to identify diet–disease relationships. It is in the interests of researchers to understand the nature of these errors and wherever possible correct them. Validation studies constitute an essential element of this process.

### Acknowledgements

This paper has been written in consultation with many colleagues from the European COST99 group and the UK Nutritional Epidemiology Group, too numerous to mention individually but whose comments are greatly appreciated. The work was supported by the European Concerted Action COST99 'Food consumption and composition data (EUROFOODS)'.

### References

- 1 Young LR, Nestle M. Portion sizes in dietary assessment: issues and policy implications. *Nutr. Rev.* 1995; **53**: 149–58.
- 2 Cypel YS, Guenther PM, Petot GJ. Validity of portion-size measurement aids: A review. *J. Am. Diet. Assoc.* 1997; **97**: 289–92.
- 3 Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*, second edition. Oxford: Oxford University Press.
- 4 Cole TJ. Sampling, study size and power. In: Margetts BM, Nelson M, eds. *Design Concepts in Nutritional Epidemiology*, second edition. Oxford: Oxford University Press, 1997.
- 5 Nelson M. The validation of dietary assessment. In: Margetts BM, Nelson M. *Design Concepts in Nutritional Epidemiology*, second edition. Oxford: Oxford University Press, 1997.
- 6 Bolland JE, Yuhas HA, Bolland TW. Estimation of food portion sizes: effectiveness of training. *J. Am. Diet. Assoc.* 1988; **88**: 817–21.
- 7 Bolland JE, Ward JY, Bolland TW. Improved accuracy of estimating food quantities up to 4 weeks after training. *J. Am. Diet. Assoc.* 1990; **90**: 1402–4.
- 8 Howat PM, Mohan R, Champagne C, Monlezun C, Wozniak P, Bray GA. Validity and reliability of reported dietary intake data. *J. Am. Diet. Assoc.* 1994; **94**: 169–73.
- 9 Bingham SA, Gill C, Welch A, Day K, Cassidy A, Khaw KT, Sneyd MJ, Key TJA, Roe L, Day NE. Comparison of dietary assessment methods in nutritional epidemiology: weighed records v. 24 h recalls, food-frequency questionnaires and estimated-diet records. *Br. J. Nutr.* 1994; **2**: 619–43.
- 10 Bingham SA, Cassidy A, Cole T, Welch A, Runswick S, Black AE, Thurnham D, Bates CE, Cassidy A, Khaw KT, Day NE. Validation of weighed records and other methods of dietary assessment using the 24 h urine technique and other biological markers. *Br. J. Nutr.* 1995; **73**: 531–50.
- 11 Smith SA, Campbell DR, Elmer PJ, Martini MC, Slavin JL, Potter JD. The University of Minnesota Cancer Prevention Research Unit vegetable and fruit classification scheme (United States). *Cancer Causes and Control* 1995; **6**: 292–302.
- 12 Crawford FG, Mayer J, Santella RM, Cooper TB, Ottman R, Tsai W-Y, Simon-Cerejido G, Wang M, Tang D, Perera FP. Biomarkers of environmental tobacco smoke in preschool children and their mothers. *J. Natl. Cancer Inst.* 1994; **86**: 1398–402.
- 13 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**: 307–10.
- 14 Nelson M, Atkinson M, Darbyshire S. Food photography 2. Use of food photographs for estimating portion size and the nutrient content of meals. *Br. J. Nutr.* 1996; **76**: 31–49.
- 15 Haraldsdóttir J, Tjønneland A, Overvad K. Validity of individual portion size estimates in a food frequency questionnaire. *Int. J. Epidemiol.* 1994; **23**: 787–96.
- 16 Kirkcaldy-Hargreaves M, Lynch GW, Santor C. Assessment of the validity of four food models. *J. Can. Diet. Assoc.* 1980; **41**: 102–10.
- 17 Kuehnemann T, Stanek K, Eskridge K, Angle C. Comparability of four methods for estimating portion size during a food frequency interview with caregivers of young children. *J. Am. Diet. Assoc.* 1994; **94**: 548–50.
- 18 Nelson M, Atkinson M, Darbyshire S. Food photography 1: The perception of food portion size from photographs. *Br. J. Nutr.* 1994; **72**: 649–63.