Towards Transparent Development of Food Frequency Questionnaires

Scientific basis of the Dutch FFQ-TOOL™: a computer system to generate, apply and process FFQs

Marja Molag

Thesis Committee

Thesis supervisors
Prof. dr. ir. P. van 't Veer
Professor of Nutrition and Epidemiology, Division of Human Nutrition
Wageningen University

Prof. dr. ir. P.A. van den Brandt
Professor of Epidemiology, Department of Epidemiology
Maastricht University

Thesis co-supervisor
Dr. J.H.M. de Vries
Assistant professor, Division of Human Nutrition
Wageningen University

Other members
Prof. dr. C.M.J. van Woerkum
Wageningen University

Dr. C. van Rossum
RIVM, Bilthoven

Prof. dr. R.P. Stolk
University Medical Centre Groningen

Prof. dr. R.P. Mensink
Maastricht University

This research was conducted under the auspices of the Graduate School VLAG (Food Technology, Agrobiotechnology, Nutrition and Health Sciences).

Towards Transparent Development of Food Frequency Questionnaires

Scientific basis of the Dutch FFQ-TOOL™: a computer system to generate, apply and process FFQs

Marja Molag

Thesis

submitted in partial fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof.dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Doctorate Board
to be defended in public
on Friday 15 January 2010
at 1.30 PM in the Aula

Marja Molag
Towards Transparent Development of Food Frequency Questionnaires
Scientific basis of the Dutch FFQ-TOOL™: a computer system to generate, apply and process FFQs
160 pages

Thesis Wageningen University, Wageningen, the Netherlands (2010).
With references, with summaries in Dutch and English.

ISBN 978-90-8585-539-2

ABSTRACT

Introduction
A well-designed food frequency questionnaire (FFQ) aims to assess habitual intake of foods or nutrients of interest in the target population. Therefore, the food list of an FFQ has to be adapted to the prevailing dietary habits of the target population and updated when re-used some time after initial development. However, due to lack of expertise, time, or finances, FFQs are often re-used without adaptations. To simplify and standardize the development of new FFQs, we developed a computer system, the *Dutch FFQ-TOOL™*, to generate, apply, and process FFQs. The aim of this thesis was to provide the scientific basis for development of this system.

Methods
We first characterized FFQ design by a systematic review of validation studies. For selecting food items in the system, we evaluated a simple procedure, called *MOM2*, which selected food items on the basis of explained variance in nutrient intake without taking covariance from other food items into account. To improve questions in FFQs, we conducted focus group discussions to investigate problems encountered by adults when filling out FFQs. Using the information from these studies; we developed a prototype of the system, which used the Dutch National Food Consumption Survey 1998 for the selection of food items. This prototype was used to generate an FFQ, which was subsequently validated against biomarkers and 3-day food records.

Results
In the systematic review, we observed that FFQs with more items (>200 items) were better able to rank people according to their intake of most nutrients than shorter FFQs (<100 items). *MOM2* appeared suitable to select food items that contributed importantly to variance in nutrient intake, leading to only a few more food items than regression analysis. Focus group discussions showed that 36 out of 40 respondents were confused by examples that were meant to clarify questions and 31 out of 40 respondents had difficulties in identifying consumed foods, because categorization of foods was not logical to them. This information was used to develop standard questions and answering categories which were included in the *Dutch FFQ-TOOL™*. The validation of a first FFQ generated by the system showed that validity coefficients between FFQ and true intake were 0.50 for protein, 0.36 for potassium and 0.50 for PUFA, as assessed by the methods of triads applied to the FFQ, food records, and biomarkers of intake. These results are similar to those from validation studies of other FFQs.

Conclusion
We have developed a flexible data-based computer system that can generate FFQs using standardized procedures for multiple nutrients of interest, which is suitable to assess habitual food and nutrient intake of adults with sufficient validity.

CONTENTS

1 Introduction

INTRODUCTION

Assessment of dietary intake with FFQs

In epidemiologic research, food frequency questionnaires (FFQs) are often used to study the relationship between dietary intake and disease in large studies [1-3]. FFQs are the only instrument that can be self-administered and used to assess usual dietary intake. They may be used to assess intake of one nutrient or food group such as calcium or dairy products or they may provide a more general assessment of dietary intake focusing on energy, macronutrients, and important micronutrients. Basically, FFQs consist of questions on the consumption frequency of food items required to assess the usual intake of nutrients of interest at an individual level. Each major food item in the FFQ represents many variations of that food or even a group of related foods. For example, pizza, a typical FFQ item, represents all the varieties of pizza that are available [4]. To improve the description and accuracy of dietary intake estimates of respondents, further questions about consumed portion sizes and preparation methods are often included. Finally, intakes of foods are converted to food group and nutrient intakes using food composition databases.

Popularity of FFQs

The popularity of FFQs stems from their ease of administration, ability to assess habitual dietary intake over an extended period of time, and relatively low costs [3]. An important advantage of dietary assessment by FFQs, as compared to multiple food records or 24-hour dietary recalls, is that response rates are relatively high. The reason is that FFQs can be filled out at home and respondent burden is usually low [5-8]. Moreover, food consumption of respondents is not affected by filling out FFQs [8], whereas keeping food records may lower intake during the recording days [3]. In addition, a limitation of 24-hour dietary recalls is that underreporting of intake increases for second and further administrations, as compared to the first [9]. An additional argument in favour of FFQs is that they may be less susceptible to BMI-related underreporting than other dietary assessment methods such as 24-hour dietary recalls [10].

Drawbacks

There is some concern that FFQs are too much affected by measurement error to detect weak associations between diet and health outcomes [11]. This conclusion was mainly based on the OPEN study, a very large validation study in which valid reference biomarkers for energy (doubly labelled water) and protein (urinary nitrogen), together with a FFQ and 24-hour dietary recalls, were observed in 484 healthy volunteers [9]. However, this finding is open to question because there was no realistic assessment of within-person variability in the biomarkers [12]. Doubly labelled water was measured twice in a period of two weeks and supposed to be representative of a reference period of one year. This is not very likely as Black and Cole estimated that the within-person variation was approximately twice as great at an interval of one year compared with having no interval between the replicates [13,14]. Furthermore, the results from the OPEN study cannot be readily generalized to other FFQs because the values for total energy intake by their FFQ were substantially lower than for many other FFQs used in current studies [14].

Another study that caused concern about measurement error in FFQs was a small cohort study from the United Kingdom. A very strong association between total and saturated fat and risk of breast cancer was observed, if fat intake was assessed with food records, whereas a weaker and non-significant relation was seen using an FFQ. This finding may be explained by misclassification of respondents according to their fat intake by FFQs [15-17]. However, in a population that was characterized by a wider range in dietary fat intake, dietary fat intake assessed by FFQ was associated with breast cancer [18]. This indicates that FFQs are especially useful in populations that are characterized by a wide range in intake [19]. In addition, statistical power gained by including many more respondents in FFQ-based studies, could outweigh the loss from greater measurement error by FFQs [20]. Thus, although measurement error in FFQs may be a problem in some studies, FFQs will continue to be an important method to assess usual dietary intake in nutritional epidemiology.

Development of FFQs

An FFQ must be developed for the aim of a study, assessing the foods and nutrients of interest in the target population. Until now, development of FFQs is very labour intensive and asks for specific expertise. First, informative foods for the food list must be selected. Foods are considered to be informative if they are consumed regularly, contribute substantially to the nutrient(s) of interest, and vary in intake between persons [3]. As a source to select informative foods, it is preferable to have food intake data of the population collected by food records or 24-hour dietary recalls. Foods may be selected from the database based on their contribution to the level of intake in the population [21], or on their ability to rank respondents according to their intake [3]. In epidemiologic studies, it is very important to select foods that are able to rank participants according to their intake. Forward regression is very suitable for this, selecting foods that explain most variance, taking covariances between nutrient intakes of different foods into account [3].

The above mentioned procedures select foods at the level of food codes in a food composition table. For development of an FFQ, related foods must be combined into food items. For this, decisions about aggregation of food items are made. To assess the level of intake of an individual or population, the level of intake is often overestimated when many detailed food items are included [22,23]. Thus, for monitoring purposes highly aggregated food items may be more suitable. However, in epidemiology it is important to rank individuals accurately according to their intake. For this purpose, food items at a detailed level are more suitable, because they better capture the between-person variation in intake. However, a disadvantage of selecting food items at a detailed level is that it results in a longer food list.

Next, an FFQ is composed from the selected food items by transposing selected food items into questions about frequency of consumption. Answering categories are added and further questions about portion sizes and preparation methods may be added, depending on the aim of the study. Open-ended questions add little to nutrient intake estimates, because these are often left blank by respondents [24]. After developing the FFQ, a nutrient database is created for the items on the food list, which is complicated and requires much consideration [4]. This database must be accurate and represent current foods and nutrients of interest [4]. The value for each food item must be based on the weighted mean composition of all foods represented in the item, applying weights from the food consumption data used as a source for item selection.

The above procedures create FFQs targeted to assess intake of nutrients of interest or foods in a population with a specific food pattern. Accordingly, food lists of FFQs have to be adapted to food consumption habits if used for another population. Also, food lists must be updated to include new foods that were recently introduced at the market, when re-used a few years after their initial development [25]. However, it is very labour intensive to update food lists, nutrient databases and analysis programs in a standardized way. As a consequence, FFQs are often re-used without updating or adapting to other populations because researchers do not have the expertise, time, or resources to do so [26].

Validation of FFQs

After development, the FFQ needs to be validated in the target population in order to assess the extent of dietary misreporting, as FFQs cannot measure dietary intake without errors [9]. For nutrients that have no biomarkers of intake, ranking by FFQ is generally compared to ranking by replicates of food records, 24-hour dietary recalls or a dietary history method. In general, correlation coefficients between FFQ and the reference method, from 0.50 and higher are considered as sufficiently valid [3]. Preferably, energy and nutrient intake assessed by an FFQ is validated against recovery biomarkers, for example, energy intake by doubly labelled water, protein by urinary nitrogen, and potassium by urinary potassium [27,28]. Recovery biomarkers enable researchers to evaluate both ranking of participants according to their intake and the level of intake provided by the FFQ. Urinary nitrogen and urinary potassium are relatively cheap and easily available. Unfortunately, doubly labelled water is expensive and therefore not generally applied.

Moreover, all three markers depend on collection of 24-hour urines which is a tedious procedure, potentially limiting validity of these biomarkers [27].

In addition to recovery markers, concentration markers are often determined to evaluate the intake of several other nutrients. Concentration markers are suitable to evaluate ranking of participants according to their intake, but are not suitable to estimate the level of intake. Examples of concentration markers are polyunsaturated fatty acids (PUFA) determined in serum cholesteryl esters [29] to evaluate PUFA intake, and carotenoids determined in serum or plasma to evaluate intake of fruit and vegetables [30]. Unfortunately, for most nutrients no biomarkers of intake are available. In addition, biomarkers can also not be considered as a gold standard of intake because they suffer from within person variation in metabolic processes, laboratory variation and day to day variation in intake. All these factors may result in attenuated observed associations between intake by FFQ and biomarkers.

Because of the above limitations of biomarkers, it is recommended to include a triangular comparison between the FFQ, the reference method and a biomarker. This approach, called the method of triads gives a quantitative estimate of the questionnaire's validity coefficient. It assumes that correlations between the three measurements are explained entirely by the fact that all are linearly related to true intake and that their random measurement errors are mutually independent. The method of triads, a basic estimating technique in factor and path analysis, is based on fitting a theoretical to an observed correlation matrix [31,32]. However, a major disadvantage of this comparison is a possible overestimation of the validity of FFQs because of correlated errors between FFQs and 24-hour dietary recalls or food records [33,34]. Therefore the questionnaire's validity coefficient is generally interpreted as an upper limit of the true validity, whereas correlations between biomarker and questionnaire are interpreted as a lower limit.

A computer system to develop FFQs

Because development of FFQs requires specific expertise and costs a lot of time, we aimed to develop a computerized system for generation of FFQs using food consumption data as a basis. Previously, a computer system to devise and process FFQs was developed in the UK [35]. However, developing an FFQ within this system was still quite labour intensive as the selection of food items was not automated. Another example was Vofrex, a program developed in the Netherlands by TNO Quality of Life (formerly TNO Nutrition and Food Research). This system standardized the selection of food items, on the basis of their percentage contribution to nutrient intake in a population from food consumption data, and it included processing of completed FFQs. It decreased the time that researchers needed to develop and process FFQs and it was relatively quick and cheap. A major drawback was that the system could not be updated with new food consumption data or food items. Also, updating of the databases was done manually item by item. Thus, computer systems for development and processing of FFQs are still quite limited.

Therefore, we needed a new and more flexible computer system, in which selection of food items for ranking of respondents according to their intake is automated, and databases are updated as soon as new databases become available. For this purpose, experts in dietary assessment from the Netherlands joined their expertise. These experts were employed by Wageningen University, Wageningen; National Institute for Public Health and the Environment (RIVM), Bilthoven; TNO Quality of Life, Leiden; and Department of Epidemiology, Maastricht University.

As a first step, we will describe the general requirements for the system, and will further specify requirements from an evidence-based, a respondent's and a technical perspective.

GENERAL REQUIREMENTS

The most important requirement is that the data-based system allows different researchers in the Netherlands to develop and process food frequency questionnaires (FFQs) in a standardized way for assessing dietary intake for different research purposes, and for diverse populations. We specified this in scientific requirements, requirements from a respondents' perspective, and technical requirements.

Scientific requirements

The system must be able to generate FFQs from representative food consumption data using transparent, valid, and reproducible procedures.

Respondent's perspective

The system must be able to compose FFQs with questions about the consumption of food items that are familiar to respondents and that are asked in a comprehensive way with clear answering categories.

Technical requirements

The system must be able to develop FFQs by automatically selecting informative foods and subsequently generating standard questions using the selected food items. In addition, it must be able to compute nutrient intakes for the selected food items.

RATIONALE AND OUTLINE OF THIS THESIS

The overall aim of this thesis was to describe the development of a flexible data-based computer system to generate, apply, and process tailored FFQs for nutrients of interest and target groups. To achieve this, we further characterized FFQ design by providing an overview of FFQ validation studies and an evaluation of the validity of FFQs for classifying subjects relative to their intake in relation to number of items in the FFQ, use of portion-size questions, origin of the FFQ, and administration mode (chapter 2). Then, we started developing of the computer system. However, as computations of regression analyses were complex and demanding we tested a simple method that selected food items based only on variance in intake, without taking their covariances into account [36]. In chapter 3, the new procedure, that explained most of the variance in nutrient intake without taking other items into account, was compared to forward regression analyses. To fulfil requirements for respondents, we held a series of focus group discussions with respondents to improve FFQ questions (chapter 4). We investigated problems encountered by younger and older respondents when filling out food frequency questionnaires. Thereby, we focused on three aspects: (1) comprehension of questions, (2) identification of foods, and (3) problems in assessing frequencies of consumption. Subsequently, we described the development of a prototype of this system by weighing information from the above described studies and practical aspects. We further described the requirements the system needed to fulfil from a scientific, respondents and technical perspective (chapter 5). In a next step, we developed a first FFQ using the computer system. To fulfil the scientific requirements, we validated an FFQ generated by the system against biomarkers and 3-day food records (chapter 6). The main findings from these studies were summarized and discussed in chapter 7.

REFERENCE LIST

1. Thompson FE, Byers T. Dietary assessment resource manual. J Nutr 1994;124(11 Suppl):2245s-317s.

2. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr 2002;5(4):567-87.

3. Willett W. Nutritional Epidemiology. second edition ed. New York, NY: Oxford University Press, 1998.

4. Pennington JA, Stumbo PJ, Murphy SP, McNutt SW, Eldridge AL, McCabe-Sellers BJ, Chenard CA. Food composition data: the foundation of dietetic practice and research. J Am Diet Assoc 2007;107(12):2105-13.

5. Burley V, Cade J, Margetts B, Thompson R, Warm D. Consensus Document on the development, validation and utilisation of food frequency questionnaires. United Kingdom: Nuffield Institute for Health, University of Leeds and Institute of Human Nutrition, University of Southampton, 2000;63.

6. Drewnowski A. Diet image: a new perspective on the food-frequency questionnaire. Nutr Rev 2001;59(11):370-2.

7. Barrett Connor E. Nutrition epidemiology: how do we know what they ate? Am J Clin Nutr 1991;54(1 Suppl):182s-7s.

8. Biro G, Hulshof KF, Ovesen L, Amorim Cruz JA. Selection of methodology to assess food intake. Eur J Clin Nutr 2002;56 Suppl 2:S25-32.

9. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard Barbash R, Sunshine J, Schatzkin A. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003;158(1):1-13.

10. Lissner L, Troiano RP, Midthune D, Heitmann BL, Kipnis V, Subar AF, Potischman N. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. Int J Obes (Lond) 2007;31(6):956-61.

11. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, Ferrari P, Carroll RJ. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr 2002;5(6a):915-23.

12. Willett WC, Hu FB. Not the time to abandon the food frequency questionnaire: Point. Cancer Epidemiol Biomarkers Prev 2006;15(10):1757-8.

13. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique: implications for validating reported dietary energy intake. Eur J Clin Nutr 2000;54(5):386-94.

14. Willett W, Kipnis V, Subar AF, Schatzkin A, Midthune D, Troiano RP, Schoeller DA, Bingham S, Freedman LS. Invited commentary: OPEN questions. Am J Epidemiol 2003;158(1):22-6.

15. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet 2003;362(9379):212-4.

16. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiol Biomarkers Prev 2005;14(12):2826-8.

17. Freedman LS, Potischman N, Kipnis V, Midthune D, Schatzkin A, Thompson FE, Troiano RP, Prentice R, Patterson R, Carroll R, Subar AF. A comparison of two dietary instruments for evaluating the fat-breast cancer relationship. Int J Epidemiol 2006;35(4):1011-21.

18. Thiebaut ACM, Kipnis V, Chang SC, Subar AF, Thompson FE, Rosenberg PS, Hollenbeck AR, Leitzmann M, Schatzkin A. Dietary fat and postmenopausal invasive breast cancer in the national institutes of health - AARP diet and health study cohort. Journal of the National Cancer Institute 2007;99(6):451-62.

19. Freedman LS, Schatzkin A, Thiebaut ACM, Potischman N, Subar AF, Thompson FE, Kipnis V. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis [1]. Cancer Epidemiol Biomarkers Prev 2007;16(6):1321-2.

20. McNeill G, Masson L, Macdonald H, Haggarty P, Macdiarmid J, Craig L, Kyle J. Food frequency questionnaires vs diet diaries. Int J Epidemiol 2009;38(3):884.

21. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124(3):453-69.

22. Bogers RP, Dagnelie PC, Westerterp KR, Kester AD, van Klaveren JD, Bast A, van den Brandt PA. Using a correction factor to correct for overreporting in a food-frequency questionnaire does not improve biomarker-assessed validity of estimates for fruit and vegetable consumption. J Nutr 2003;133(4):1213-9.

23. Serdula M, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. Epidemiology 1992;3(6):503-8.

24. Block G, Mandel R, Gold E. On food frequency questionnaires: the contribution of open-ended questions and questions on ethnic foods. Epidemiology 2004;15(2):216-21.

25. Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: The women's health trial feasibility study in minority populations. Am J Epidemiol 1997;146(10):856-69.

26. Subar AF. Developing dietary assessment tools. J Am Diet Assoc 2004;104(5):769-70.

27. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. J Nutr 2003;133 Suppl 3:921s-4s.

28. Tasevska N, Runswick SA, Bingham SA. Urinary potassium is as reliable as urinary nitrogen for use as a recovery biomarker in dietary studies of free living individuals. J Nutr 2006;136(5):1334-40.

29. Poppitt SD, Kilmartin P, Butler P, Keogh GF. Assessment of erythrocyte phospholipid fatty acid composition as a biomarker for dietary MUFA, PUFA or saturated fatty acid intake in a controlled cross-over intervention trial. Lipids Health Dis 2005;4:30.

30. Al-Delaimy WK, Ferrari P, Slimani N, Pala V, Johansson I, Nilsson S, Mattisson I, Wirfalt E, Galasso R, Palli D, Vineis P, Tumino R, Dorronsoro M, Pera G, Ocke MC, Bueno-de-Mesquita HB, Overvad K, Chirlaque M, Trichopoulou A, Naska A, Tjonneland A, Olsen A, Lund E, Alsaker EH, Barricarte A, Kesse E, Boutron-Ruault MC, Clavel-Chapelon F, Key TJ, Spencer E, Bingham S, Welch AA, Sanchez-Perez MJ, Nagel G, Linseisen J, Quiros JR, Peeters PH, van Gils CH, Boeing H, van Kappel AL, Steghens JP, Riboli E. Plasma carotenoids as biomarkers of intake of fruits and vegetables: individual-level correlations in the European Prospective Investigation into Cancer and Nutrition (EPIC). Eur J Clin Nutr 2005;59(12):1387-96.

31. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr 1997;65(4 Suppl):1232s-9s.

32. Ocke MC, Kaaks RJ. Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr 1997;65(4 Suppl):1240s-5s.

33. Kipnis V, Midthune D, Freedman LS, Bingham S, Schatzkin A, Subar A, Carroll RJ. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153(4):394-403.

34. Willett W. Invited commentary: a further look at dietary questionnaire validation. Am J Epidemiol 2001;154(12):1100-2; discussion 1105-6.

35. Wise A, Birrell NM. Design and analysis of food frequency questionnaires--review and novel method. Int J Food Sci Nutr 2002;53(3):273-9.

36. Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. Am J Epidemiol 1996;143(5):514-21.

2

Design characteristics of food frequency questionnaires in relation to their validity

Marja L. Molag[1], Jeanne H.M. de Vries[1], Marga C. Ocké[2], Pieter C. Dagnelie[3], Piet A. van den Brandt[3], Margje C.J.F. Jansen[4], Wija A. van Staveren[1], Pieter van `t Veer[1]

[American Journal of Epidemiology, 2007 166(12): 1468-78]

[1] Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands

[2] National Institute for Public Health and the Environment, Bilthoven, the Netherlands

[3] Department of Epidemiology, Maastricht University, Maastricht, the Netherlands

[4] TNO Quality of Life, Zeist, the Netherlands

ABSTRACT

The authors investigated the role of food frequency questionnaire (FFQ) design, including length, use of portion-size questions, and FFQ origin, in ranking subjects according to their nutrient intake. They also studied the ability of the FFQ to detect differences in energy intake between subgroups and to assess energy and protein intake. In a meta-analysis of 40 validation studies, FFQs with longer food lists (200 items) were better than shorter FFQs at ranking subjects for most nutrients; results were statistically significant for protein, energy-adjusted total fat, and energy-adjusted vitamin C. The authors found that FFQs that included standard portions had higher correlation coefficients for energy-adjusted vitamin C (0.80 vs. 0.60, $p < 0.0001$) and protein (0.69 vs. 0.61, $p = 0.03$) than FFQs with portion-size questions. However, it remained difficult from this review to analyze the effects of using portion-size questions. FFQs slightly underestimated gender differences in energy intake, although level of energy intake was underreported by 23% and level of protein intake by 17%. The authors concluded that FFQs with more items are better able to rank people according to their intake and that they are able to distinguish between subpopulations, even though they underestimated the magnitude of these differences.

INTRODUCTION

Food frequency questionnaires (FFQs) are widely used to assess the dietary intake of large populations. The popularity of FFQs stems from their ease of administration, ability to assess dietary intake over an extended period of time, and low costs [1]. They are therefore often used in epidemiologic studies to investigate the relation between diet and disease. For some purposes, information about level of intake is very important, for example, to set recommendations for nutrient intake. For most epidemiologic studies, FFQs must be able to classify individuals correctly according to their dietary intake. However, Bingham and others have argued that, probably because of misclassification, FFQs are not always able to detect weak associations [2,3]. Because of this debate and their established role in epidemiologic research, FFQs need to be further characterized and subsequently improved.

FFQs differ in the way they are developed and show large variations in design characteristics, such as number of items or inclusion of portion-size questions. Such variations could affect reported intakes [4-6].

In this study, we first aimed to provide an overview of FFQ validation studies and the validity of FFQs in classifying subjects relative to their intake in relation to number of items in the FFQ, use of portion-size questions, origin of the FFQ, and administration mode. A second aim was to provide an overview of the validity of FFQs in assessing absolute energy and protein intakes as determined in studies using recovery biomarkers and to establish whether FFQs can detect known differences in energy intake between men and women.

MATERIALS AND METHODS

Search strategy and study selection

We searched MEDLINE (National Library of Medicine, Bethesda, Maryland) for validation studies of FFQs that assessed respondents' habitual dietary intake and were published between 1980 and December 2006. An FFQ was defined as a questionnaire with a food list and a frequency-response section where subjects report how often each food item is consumed [7]. Search terms used were the Medical Subject Headings (MeSH) "nutrition assessment," "questionnaires," "evaluation studies" OR "reproducibility of results" AND keywords "food frequency questionnaire" OR "FFQ" AND "validity" OR "validation" OR the publication type "validation studies."

Studies that met all of the following inclusion criteria were included in the review:

1) Describing FFQs developed for epidemiologic purposes
2) Addressing habitual diet by reporting at least energy intake but preferably also intake of other nutrients including total fat, carbohydrates, protein, alcohol, calcium, and vitamin C

3) Studying adult populations in the age range 18-82 years with a westernized diet but not FFQs validated exclusively among those older than age 60 years
4) Validating FFQs with one of the following reference methods: 24-hour dietary recalls, food records, diet history interview, or recovery biomarkers

Studies that assessed only specific nutrients or food groups such as fruit and vegetable consumption were excluded from this review.

Data extraction and classification

The following design characteristics of the FFQ were extracted: number of items in the FFQ (ranging from 44 to 350), use of portion-size questions versus predefined standard portion sizes, and "origin" of the FFQ, for example, the Willett type [8], the Block type [9]; the European Prospective Investigation into Cancer and Nutrition (EPIC) type [10] and "other FFQs." We extracted the following validation study characteristics: study size (N), gender (males, females), FFQ administration mode (interview or self-administration), reference method (food record, 24-hour dietary recall, or diet history interview), and total number of days over which the reference method was applied—categorized as a short period of 1-7 days, a medium period of 8-14 days, and a long period of 15 days or more. We used 8-14 days as the reference category in the meta-regression analyses because the 24-hour dietary recall method was not applied for 15 days or more.

We analyzed nutrients that covered different aspects of a habitual diet. We included energy and all energy-providing macronutrients (i.e., protein, carbohydrates, fat, and alcohol). We added vitamin C and dietary fiber to represent fruit and vegetables and some other food components, for example, calcium, to represent other specific foods.

Statistical analysis

Ranking

We extracted crude and energy-adjusted correlation coefficients and, if available, gender-specific and/or gender-adjusted correlation coefficients between FFQs and reference methods to compare studies with respect to ranking of subjects. Correlation coefficients were first converted into a standard normal metric by using Fisher's r-to-Z transformation, expressed in equation [1] [11], in which r_i is the correlation coefficient from study i.

$$Zr_i = \frac{1}{2} Log_e\left(\frac{1+r_i}{1-r_i}\right) \qquad [1]$$

The transformed effect sizes were then used to calculate an initial pooled average, mean Zr_i, in which each correlation coefficient was weighted by the variance, $1/(n_i - 3)$. Then, to identify confounders to be used for adjustments in meta-regression analyses, we stratified these correlation coefficients by sex, type of reference method, category of number of days that this method was applied, and administration method. The weighted averages, mean Zr, were also used to perform the Cochrane Q test for heterogeneity between studies for the assessed nutrients, expressed in equation [2] [12].

$$Q = \sum_{i=1}^{k}(n_i - 3)(Zr_i - meanZr) \qquad [2]$$

This test had a p value of <0.0001 for all nutrients, which indicated that the correlation coefficients from the validation studies were heterogeneous.

Following the heterogeneity test, random-effects meta-regression was conducted to explain this heterogeneity by FFQ design characteristics using the restricted maximum likelihood approach, as per Thompson and Sharp [13]. For each nutrient, number of items as a continuous variable, use of portion-size questions or standard portions, FFQ origin —Willett, Block, EPIC, or other— and several potential confounders (gender, reference method, and number of days over which this method was applied) were regressed on the transformed correlation coefficients Zr_i. Weights were assigned based on the variance ($1/(n_i - 3)$). Results of the meta-regression are presented as predicted values of Z, retransformed to r, using a model that included an intercept, a reference period of 8-14 days, an average value of 0.5 for the indicator variable for sex, and similarly so for the reference method. All data were analyzed by using STATA 8 software (Stata Corporation, College Station, Texas).

Validity of absolute intake

To assess energy intake, validity was defined as the difference between the mean levels of energy intake assessed by the FFQ minus the mean levels of energy expenditure determined by the doubly labeled water method. For protein intake, it was defined as the difference between protein intake assessed by the FFQ and protein intake estimated from 24-hour urinary nitrogen excretion. If only nitrogen excretion was reported in the paper, we estimated protein intake by assuming that urinary nitrogen was excreted as a constant proportion of 80 percent of total nitrogen intake [14], and 16 percent of protein is nitrogen [15]. Thus, protein intake was estimated from the following formula:

$$\text{protein} = 6.25 \times (\text{urinary } N/0.80). \qquad [3]$$

Validity of gender differences in energy intake

To evaluate the extent to which gender differences in energy intake could be detected by FFQs, we extracted gender-specific mean energy intake, including standard deviations (if not available, we assumed it was 3 MJ because this mean of standard deviations

Table 2.1 Overview of the 40 studies included in the review by first author, year of publication, and "origin" of the FFQ[a]

First author, year [reference no.]	Origin[b]	No. of subjects	No. of FFQ items	FFQ administration method	Use of portion-size questions	FFQ reference period	Reference method	Repetitions reference method
Andersen, 1999 [26]	Other	125 men	180	Self	Yes	Unknown	FR[a]	14 d[a] in 5 w[a]
Andersen, 2003 [53]	Other	17 women	180	Self	Yes	1 y[a]	DLW[a]	1 (10 d)
Bingham, 1997 [27]	EPIC[a]	156 women	130	Self	No	Unknown	FR	4 × 4 d in 1 y
Block, 1990 [28]	Block	102 women	60	Self	Yes	Unknown	FR	2 × 7 d
Block, 1990 [9 c]	Block	102 women	94	Self	Yes	6 m[a]	FR	3 × 4 d in 1 y
Block, 1992 [29]	Block	85 men and women	98	Interview	Yes	1 y	FR	4 × 3 d in 1 y
Bohlscheid-Thomas, 1997 [56]	EPIC	49 men, 55 women	104	Self	Yes	1 y	24-HR[a]	12× in 1 y
Boucher, 2006 [17]	Block	166 women	126	Self	Yes	Unknown	24-HR	2×
Brunner, 2001 [30]	Other	457 men, 403 women	127	Self	No	1 y	FR	7 d
Callmer, 1993 [31]	Other	57 men, 50 women	250	Self	Yes	1 y	FR	6 × 3 d in 1 y
Engle, 1990 [32]	Other	16 men, 34 women	120	Self	Yes	3 m	FR	7 d
Feunekes, 1993 [50]	Other	95 men, 96 women	104	Interview	Yes	1 m	DH[a]	1 × 1 m
Fidanza, 1995 [33]	Other	11 men, 35 women	93	Self	Yes	1 y	FR	7 d
Flagg, 2000 [18]	Block	216 men, 223 women	114	Self	Yes	1 y	24-HR	4× in 1 y
Goldbohm, 1994 [34]	Other	59 men, 50 women	150	Self	Yes	1 y	FR	3 × 3 d in 1 y
Hartwell, 2001 [35]	Other	16 men, 9 women	162	Self	No	1 y	FR	2 × 4 d
Jain, 2003 [19]	Other	151 men, 159 women	166	Self	Yes	1 y	24-HR	3× in 1 y
Johansson, 2002 [20]	EPIC	96 men, 99 women	84	Interview	Yes	Unknown	24-HR	10× in 1 y
Katsouyanni, 1997 [21]	EPIC	42 men, 38 women	190	Self	Yes	1 y	24-HR	12× in 1 y
Kroke, 1999 [51]	EPIC	75 men, 59 women	146	Self	Yes	1 y	DLW[d], 24-HR	14 d, 12× in 1 y
Larkin, 1989 [36]	Other	228 men and women	116	Interview	Yes	1 y	FR	16 d in 1 y
Longnecker, 1993 [37]	Willett	64 men, 74 women	116	Self	No	1 y	FR	3 × 2 d or 2 × 2 d
Mannisto, 1996 [38]	Other	152 women	110	Self	No	1 y	FR	2 × 7 d in 3 m
Martin-Moreno, 1993 [39]	Other	147 women	118	Self	Yes	1 y	FR	4 × 4 d in 1 y
McKeown, 2001 [40]	EPIC	58 men, 88 women	130	Self	No	1 y	FR	2 × 7 d

Munger, 1992 [41]	Willett	44 women	126	Self	No	Unknown	FR	3 d
Ocke, 1997 [22]	EPIC	63 men, 58 women	178	Self	Yes	1 y	24-HR	12× in 1 y
Patterson, 1999 [42]	Other	113 women	122	Self	Yes	3 m	FR	4 × 4 d
Pietinen, 1988 [43]	Other	190 men	203	Self	Yes	1 y	FR	12 × 2 d in 6 m
Pietinen, 1988 [44]	Other	297	44	Self	No	1 y	FR	12 × 2 d in 6 m
Pisani, 1997 [23]	EPIC	47 men, 150 women	47	Self	Yes	Unknown	24-HR	8-14× in 1 y
Riboli, 1997 [45]	EPIC	57 men, 50 women	350	Self	Yes	1 y	FR	6 × 3 d in 1 y
Rimm, 1992 [46]	Willett	127 men and women	131	Self	No	1 y	FR	2 × 7 d in 6 m
Schroder, 2001 [47]	Other	44 men and women	157	Self	Yes	Unknown	FR	3 d
Subar, 2001 [24]								
Willett	Willett	254 men, 293 women	126	Self	No	1 y	24 HR	4× in 1 y
DHQ[a]	Other	501 men, 560 women	124	Self	Yes	1 y	24-HR	4× in 1 y
Block	Block	247 men, 267 women	106	Self	Yes	1 y	24-HR	4× in 1 y
Subar, 2003 [52]	Other (DHQ)	261 men, 223 women	124	Self	Yes	1 y	DLW, 24-HR	14 d, 2× in 3 m
Tjonneland, 1991 [48]	EPIC	59 men, 85 women	92	Self	Yes	1 y	FR	2 × 7 d
van Liere, 1997 [25]	EPIC	123 women	238	Self	Yes	Unknown	24-HR	12× in 1 y
Willett, 1985 [8]	Willett	173 women	61	Self	No	1 y	FR	4 × 7 d in 1 y
Willett, 1988 [49]	Willett	150 women	116	Self	No	1 y	FR	4 × 7 d, 3-4 y before

[a] FFQ, food frequency questionnaire; FR, food record; d, days; w, weeks; y, year; DLW, doubly labeled water; EPIC, European Prospective Investigation into Cancer and Nutrition; m, months; 24-HR, 24-hour dietary recall method; DH, diet history; DHQ, dietary history questionnaire. [b] Refers to FFQs developed from, for example, the Willett type [8], the Block type [9], the EPIC type [10], and "other FFQs." [c] Only women from the usual diet group were included in our analyses. [d] DLW was used in only a subgroup of 28 subjects.

was reported in 31 other included studies) and N, or the standard error (SE) for mean energy intake. We subtracted the mean level of energy intake of the women estimated by each FFQ (FFQ_{women}) from that of the mean level of the men (FFQ_{men}), and we did the same for the reference method (Ref_{women} and Ref_{men}).

Gender difference FFQ:

$$FFQ_{men} - FFQ_{women} \pm \sqrt{SE^2_{FFQ_{men}} + SE^2_{FFQ_{women}}} \qquad [4]$$

Gender difference reference method:

$$Ref_{men} - Ref_{women} \pm \sqrt{SE^2_{Ref_{men}} + SE^2_{Ref_{women}}} \qquad [5]$$

We tested whether there was a difference between both results by using an independent t test.

RESULTS

Description of studies

The search procedure resulted in 40 papers (table 2.1) describing 42 FFQs that matched the inclusion criteria. The majority of FFQs were validated against 24-hour dietary recalls [16-25] or food records [8,9,26-49]. One FFQ was validated against a diet history method [50], two FFQs against 24-hour dietary recalls and doubly labeled water [51,52] , and one FFQ against doubly labeled water only [53]. Six FFQs [8,24,37,41,46,49] were developed from the Willett FFQ [8]. For this FFQ, an extensive food list was shortened by removing infrequently eaten items and including items contributing most to between-person variance using data from Nurses' Health Study participants. Willett FFQs included on average 13 (range, 61-131) items and asked respondents to report their frequency of consumption of a given reference portion size in a table format. Another six FFQs [9,17,18,24,28,29] were developed from the Block FFQ [9]. This FFQ was developed by using food items that contributed over 90 percent of the total population intake of energy and several nutrients in the Second National Health and Nutrition Examination Survey database [54]. These Block FFQs consisted of an average of 100 items (range, 60-126), and all asked portion-size questions.

Within the EPIC, project country-specific FFQs were developed including items that cumulatively contributed most to between-person variance [21,23,40,48,51,55,56] or to total nutrient intake [22]; of three FFQs, the method of development was not described [20,25,45]. For the EPIC FFQs, we found 11 validation studies performed in nine countries [16,20-23,25,27,40,45,48,51,56]; two were conducted in the United Kingdom [27,40] and two in Germany [16,51,56]. We analyzed them as separate studies. The FFQs validated in the EPIC studies consisted of an average of 154 items (range, 47-350). Nine of these FFQs included portion-size questions, and two assigned standard portion sizes. Although FFQ design between EPIC FFQs varied a lot, the design of their validation studies [57] was carefully standardized, except for the United Kingdom, Denmark, and Sweden because they joined the EPIC project at a later stage [27,45,48,57]. Three other EPIC validation studies did not match the inclusion criteria [58-60].

The "other FFQs" were also developed by including items contributing most to between-person variance or to total population intake. We included 19 FFQs as "other FFQs" [19,24,26,30-36,38,39,42-44,47,50,52,53]. They consisted of an average of 139 items (range, 44-250); 15 of them included portion-size questions, and four assigned standard portion sizes.

Between study differences in ranking subjects

For all nutrients, pooled correlation coefficients between FFQ and reference methods ranged from 0.45 for energy and protein to 0.74 for alcohol (table 2.2), and energy-adjusted correlation coefficients were 0.02-0.08 higher for most nutrients, except for vitamin C (0.05 lower). There were differences between studies due to gender and the reference method used, although they were not statistically significant (table 2.2). As expected, for most nutrients, correlation coefficients were significantly higher when the reference method was used for 8-14 days than for 1-7 days (table 2.2). Correlation coefficients did not increase further when the reference method was used for 15 days or more. For all nutrients, we also looked at the number of consecutive days on which a food record was kept and found that correlation coefficients were lower when the reference method consisted of food records kept for more than 5 days consecutively. After energy adjustment, these differences became less pronounced or even reversed.

We observed no statistically significant differences in correlation coefficients between the interviewer- [19,20,29,36] and self-administered FFQs regarding the nutrients considered (table 2.2).

Heterogeneity by FFQ design characteristics

In the meta-regression analyses, we observed that FFQs with longer food lists (200 items) had 0.01-0.17 higher correlation coefficients than FFQs with shorter food lists (100 items) for most nutrients (table 2.3). Correlation coefficients were even higher for longer food lists for crude protein (0.56 for 200 items vs. 0.46 for 100 items, p = 0.002), energy-adjusted protein (0.68 vs. 0.51, $p < 0.001$), energy-adjusted total fat (0.68 vs. 0.59, p = 0.02), and energy-adjusted vitamin C (0.68 vs. 0.51, p = 0.001; table 2.3). The diet history method was used in only one study and therefore was excluded from main analyses.

FFQs with portion-size questions had much higher correlation coefficients for energy-

Table 2.2 Pooled correlation coefficients[b] for energy and nutrients between FFQs and reference methods stratified by characteristics of the validation studies

	No. of FFQs[c]	Energy intake	Total fat	Protein	Carbohydrate	Alcohol	Calcium	Vitamin C	Dietary fiber
No. of FFQs[c]	40	40	35	30	32	21	25	27	32
Crude *r* (range)	40	0.45 (0.16-0.77)	0.53 (0.18-0.88)	0.45 (0.14-0.70)	0.53 (0.25-0.77)	0.74 (0.29-0.90)	0.55 (0.20-0.75)	0.58 (0.16-0.82)	0.46 (0.25-0.74)
Validation study characteristics									
Reference method									
Diet history	1	0.77	NA[d]	NA	NA	NA	NA	NA	NA
24-HR[d]	13	0.42	0.49	0.46	0.56	0.69	0.55	0.55	0.48
FR[d]	26	0.53	0.56	0.47	0.51	0.78	0.56	0.59	0.44
FR <6 consecutive days	16	0.56	0.59	0.47	0.53	0.78	0.59	0.63	0.49
FR ≥6 consecutive days	10	0.42	0.51	0.46	0.50	0.78	0.53	0.55	0.36
No. of days that reference method was applied									
1-7	13	0.37	0.59	0.33	0.45	0.69	0.54	0.64	0.41
8-14	17	0.48[a]	0.54	0.51[a]	0.58[a]	0.76	0.57	0.56	0.46
≥15	10	0.46[a]	0.44	0.49[a]	0.55	0.80	0.59	0.52	0.56
Population characteristics									
Gender									
Men	19	0.44	0.52	0.49	0.57	0.66	0.56	0.53	0.44
Women	26	0.38	0.47	0.45	0.52	0.75	0.57	0.60	0.52
Both, unadjusted	10	0.48	0.46	0.45	0.48	0.82	0.45	0.46	0.42
Administration method									
Interviewer	4	0.50[e]	0.50[e]	0.44[e]	0.58[e]	0.54[e]	0.51[e]	0.51[e]	0.54[e]
Self	35	0.43	0.54	0.45	0.53	0.75	0.56	0.59	0.46

No. of FFQs[c]	22	19	17	18	13	14	17	16
Energy-adjusted *r* (range)		0.56 (0.27-0.82)	0.49 (0.16-0.76)	0.61 (0.36-0.75)	0.76 (0.32-0.90)	0.61 (0.39-0.71)	0.53 (0.19-0.79)	0.54 (0.24-0.74)
Validation study characteristics								
Reference method								
24-HR	7	0.54	0.47	0.57	0.74	0.65[e]	0.55	0.48
FR	15	0.57	0.51	0.57	0.78	0.59	0.52	0.59
FR <6 days consecutive	7	0.57	0.46	0.57	0.80	0.63	0.42	0.59
FR ≥6 consecutive days	8	0.57	0.53	0.58	0.76	0.58	0.54	0.51
No. of days that reference method was applied								
1-7	6	0.61[e]	0.22[e]	0.50	0.43[e]	0.59[e]	0.54[e]	0.39[e]
8-14	10	0.55	0.55[a]	0.62	0.81[a]	0.64[a]	0.54	0.55
≥15	6	0.55	0.45	0.53	0.84[a,e]	0.62	0.49	0.61
Population characteristics								
Gender								
Men	8	0.51	0.50	0.60	0.72	0.64	0.49	0.39
Women	13	0.58	0.45	0.55	0.78	0.61	0.59	0.54
Both, unadjusted	5	0.62[e]	0.59[e]	0.60[e]	0.75[e]	0.57[e]	0.35[e]	0.61[e]

[a] $p < 0.05$ for this characteristic by meta-regression analyses using a model that contained only this characteristic. [b] Results pooled as *Z*-transformed values, weighed by variance, retransformed to correlation coefficients between food frequency questionnaires (FFQs) and reference methods. [c] Maximum number of FFQs in this category; not all FFQs included all nutrients. [d] NA, not available; 24-HR, 24 hour dietary recall method; FR, food record. [e] Based on <5 FFQs.

Table 2.3 Correlation coefficients[b] calculated from predicted values in a meta-regression model for energy and nutrients between FFQs and reference methods by characteristics of the FFQ

	No. of FFQs[c]	Energy intake	Total fat	Protein	Carbohydrate	Alcohol	Calcium	Vitamin C	Dietary fiber
Crude correlation coefficients									
No. of FFQs[c]	39	39	35	30	32	21	25	27	32
Design characteristics									
FFQ size[d]									
100 items		0.52	0.51	0.46	0.55	0.80	0.58	0.56	0.52
150 items		0.52	0.54	0.51[a]	0.57	0.80	0.60	0.60	0.53
200 items		0.53	0.56	0.56[a]	0.58	0.80	0.63	0.64	0.53
Portions[e]									
Standard portions	11	0.46	0.52	0.55	0.56	0.72	0.67	0.68	0.45
Portion-size questions	28	0.52	0.54	0.51	0.57	0.80	0.60	0.60	0.53
Origin[f]									
Willett	6	0.51	0.42	0.46	0.49	0.36[g]	0.64[g]	0.67[g]	0.47
Block	6	0.58	0.53	0.50[g]	0.56	0.42[g]	0.56[g]	0.63[g]	0.58
EPIC	11	0.48	0.53	0.49	0.55	0.82[a]	0.52[g]	0.53	0.44
Other	16	0.54	0.48	0.44	0.52	0.83[a]	0.53	0.48[a]	0.55
Energy-adjusted correlations									
No. of FFQs[c]	22		19	17	18	13	14	17	16
Design characteristics									
FFQ size[d]									
100 items			0.59	0.51	0.59	0.68	0.68	0.51	0.54
150 items			0.64[a]	0.61[a]	0.63	0.76	0.69	0.60[a]	0.57
200 items			0.68[a,g]	0.68[a,g]	0.66[g]	0.81[g]	0.70[g]	0.68[a,g]	0.59[g]
Portions[e]									
Standard portions	9		0.66	0.69	0.63	0.61	0.68	0.80	0.50
Portion-size questions	12		0.64	0.61[a]	0.68	0.76	0.69	0.60[a]	0.57
Origin[f]									
Willett	5		0.52	0.47	0.61	0.36	0.68	0.69	0.56
Block	2		0.62[g]	NA[g,h]	0.63[g]	0.52[g]	NA[g]	NA[g]	0.62[a,g]
Epic	5		0.64	0.61	0.59	0.82[a]	0.63	0.50	0.55
Other	10		0.42	0.56	0.58	0.85[a]	0.66	0.52[a]	0.61

[a] $p < 0.05$ in meta-regression analyses. [b] Results of the meta-regression are presented as predicted values of Z, retransformed to correlation coefficients between food frequency questionnaires (FFQs) and reference methods, using a model that included the intercept and a reference period of 8-14 days, and at an average value of 0.5 for the indicator variable for gender and also at an average value of 0.5 for the reference method. [c] Maximum number of FFQs in this category; not all FFQs included all nutrients. [d] Including the use of portion-size questions in the meta-regression model; 100 items was used as the reference. [e] Including 150 items on the FFQ in the meta-regression model; standard portions were used as the reference. [f] Refers to FFQs developed from, for example, the Willett type [8], the Block type [9], the EPIC type [10], and "other FFQs"; the Willett FFQ was used as the reference. [g] Based on <5 FFQs. [h] NA, not available.

adjusted alcohol than FFQs with predefined standard portions (0.76 vs. 0.61), although they were not statistically significant ($p = 0.23$). On the contrary, FFQs with portion-size questions had significantly lower correlation coefficients for energy-adjusted protein (0.61 vs. 0.69, $p = 0.03$) and for energy-adjusted vitamin C (0.60 vs. 0.80, $p < 0.001$) than FFQs with predefined standard portions. For other nutrients, correlation coefficients were 0.08 lower to 0.08 higher for FFQs with portion-size questions compared with standard portions.

Regarding origin of the FFQ, we observed that correlation coefficients for most crude macronutrients were higher for Block and EPIC FFQs than for Willett and "other" FFQs. For calcium and vitamin C, Willett FFQs performed better, and, after energy adjustment, other correlation coefficients improved for Willett FFQs.

Table 2.4 Mean energy intake reported by the FFQs[a] compared with mean energy expenditure determined by doubly labeled water, and mean protein intake reported by the FFQs compared with protein intake calculated from nitrogen excreted in urine

First author, year [reference no.]	No. of subjects	Mean intake reported by the FFQ	Reference method	Repetitions reference method	Underestimation in FFQ (%)
Doubly labeled water		Energy intake (MJ/day)	Energy expenditure (MJ/day)		
Andersen, 2003 [53]	17 women	8.28	9.23	1 (10 d[a])	11
Kroke, 1999 [51]	28 men and women	9.05	11.23	1 (14 d)	19
Subar, 2003 [52]	245 men, 206 women	7.90 for men, 6.11 for women	11.92 for men, 9.53 for women	1 (11-14 d)	34 for men, 36 for women
Protein		Protein intake (g/day)	Protein excretion (g/day)		
Kroke, 1999 [51]	75 men, 59 women	75	97	4× in 1 y[a]	23
Ocke, 1997 [22]	46 men, 43 women	92 for men, 71 for women	99 for men, 81 for women	4× in 1 y	7 for men, 12 for women
McKeown, 2001 [40]	57 men, 77 women	89 for men, 78 for women	100 for men, 77 for women	6× in 9 m[a]	18 for men, 12 for women
Riboli, 1997 [45]	29 men, 24 women	95 for men, 76 for women	78 for men, 62 for women	8× in 1 y	-18 for men, 25 for women
Bingham, 1997 [27]	156 women	82	77	4 × 2 d in 1 y	-6
Subar, 2003 [52]	202 men, 150 women	71 for men, 55 for women	104 for men, 77 for women	2× in 14 d	32 for men, 29 for women

[a] FFQs, food frequency questionnaires; d, days; y, year; m, months.

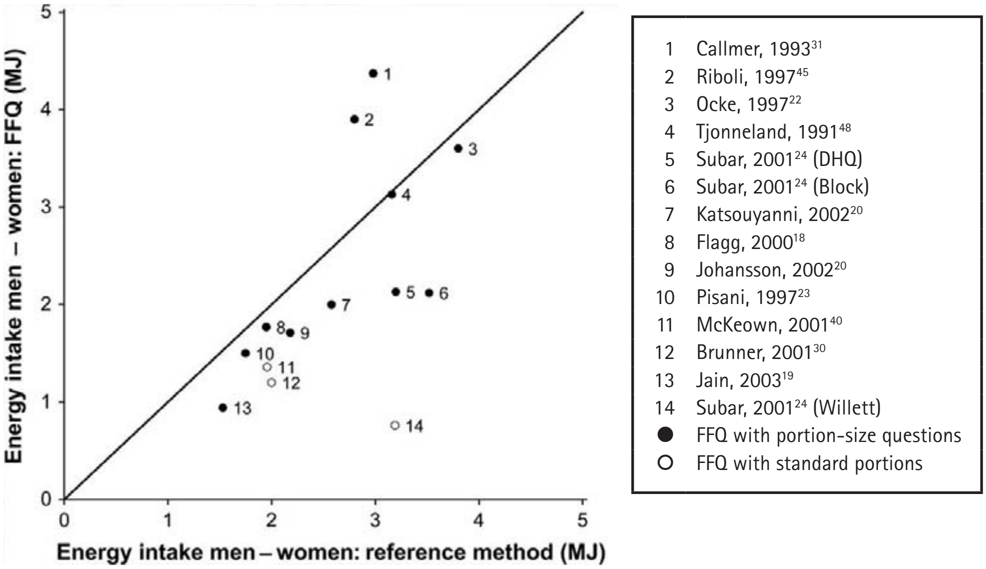


Figure 2.1 Mean energy intake of men minus mean energy intake of women according to the food frequency questionnaire (FFQ) and reference method including the 24-hour dietary recall method or food records. The boxed key lists the first author's name, year of publication, and reference number of the study corresponding to the numbered circles in the figure. DHQ, diet history questionnaire.

Validity in absolute intake

Absolute energy intake estimated by FFQs was validated against energy expenditure estimated with the doubly labeled water method [51-53]. Two FFQs in small European studies underestimated energy intake by 11 percent and 19 percent [51,53]. In one large study conducted in the United States, energy intake was underestimated by 34 percent for men and 36 percent for women [52], and protein intake was underestimated by 32 percent for men and 29 percent for women. Five EPIC FFQs were also validated against level of protein intake estimated from urinary nitrogen excretion [22,27,40,45,51,52] (table 2.4). In these studies, estimation of protein intake varied from an underestimate of 23 percent to one study that overestimated protein intake by 18 percent for men and 25 percent for women [61]. In the latter study, the longest FFQ with 350 items was applied.

Gender-differences in energy intake

Because the goal of FFQs is to distinguish subpopulations that differ with respect to nutrient intake, we tested whether FFQs are able to detect a "known" difference in energy intake between men and women. On average, the gender difference in energy intake was smaller according to FFQs (2.09 MJ for men minus women, 95 percent confidence interval: 1.62, 2.56) than according to the reference methods (2.62 MJ, 95 percent confidence interval: 2.24, 3.00). Thus, on average, FFQs underestimated the gender differences in energy intake compared with the reference methods (figure 2.1) by -0.53 MJ (95 percent confidence interval: -1.13, 0.07, p = 0.09 using an independent t test). Exceptions were the two longest FFQs [31,45], consisting of 250 and 350 items, respectively; they overestimated the difference in energy intake between men and women. The three FFQs that did not include portion-size questions [24,30,40] found on average a smaller gender difference than FFQs that asked portion-size questions [18-24,31,45,48].

DISCUSSION

This quantitative review of studies validating FFQs shows that the number of items in the food list is the major determinant in ranking subjects with respect to their intake (for 100 extra items, correlation coefficients increased by 0.01-0.13). In general, portion-size questions and FFQ origin influenced ranking of subjects only slightly.

An important point of discussion is comparability of the studies included in this review. We aimed to address differences in characteristics of FFQ design and not of study design. We increased comparability by restricting our analyses to FFQs developed to cover the complete diet and validated among adults with Western food habits, and we adjusted for potential confounders such as the reference method and the total number of days this method was used. There were differences in the design of the validation study: correlation coefficients were lower when food records were used for more than 5 days consecutively. However, this finding did not influence our results because there were also numerous studies included with another design.

In this study, we were not able to adjust for differences in energy needs or underreporting related to body weight and physical activity because these data were available for only four studies. To account for this limitation, we evaluated both crude and energy-adjusted correlation coefficients because energy adjustment leads to a focus on the relative composition of the dietary pattern [62], and it has been suggested that it reduces correlated errors between the FFQ and reference methods [63].

In addition, variation in FFQ design was limited: the FFQs varied in the number of items and the use of portion-size questions, but differences in the reference period and the administration method were limited, prohibiting conclusions regarding the latter. Finally, we accounted for unknown between-study differences originating from study design and population by using a random-effects model in the meta-regression.

Our analyses showed that FFQs with a longer food list (200 items) were better at ranking people than FFQs with a shorter food list (100 items). These findings were clearest for protein and total fat, which are calculated from many different food sources. In the development phase of an FFQ, similar items are grouped together into items whose composition can be heterogeneous; an example is 20 different meat items combined into two items on the FFQ. Sometimes, items that contribute not much to total intake are omitted although they were important in explaining between-person variance. In summarizing, our results regarding the number of items should be used as an argument not to reduce the length of the food list too much when developing FFQs to rank persons according to nutrient intake.

Results of the meta-regression analyses showed that inclusion of portion sizes did not consistently affect the ranking of different nutrients. Ranking was worse for protein and vitamin C determined by FFQs that used portion-size questions instead of standard portions, and ranking improved for alcohol when FFQs used portion-size questions. An explanation for this unexpected finding might be that, for some foods such as vegetables, it is difficult to indicate how much was eaten, especially when they are part of mixed dishes [64]. It might be easier to quantify the number and amount of alcoholic drinks; alcohol intake is ranked relatively well compared with other nutrients [65].

An important disadvantage of using standard portions is that interindividual variance decreases [66,67]. However, two validation studies in Denmark and the Netherlands found only small differences when analyzing FFQs using information from the portion-size questions compared with analyzing the same FFQs using standard portions [68,69]. These small differences may reflect that quantification of portion sizes is of minor importance compared with frequency, that the relevant individual portion sizes were not estimated correctly [69], or that portion sizes listed do not correspond well with portions actually consumed. For example, actual portion sizes (e.g., super size) are probably much larger than standard portions used by US Department of Agriculture and the Food and Drug Administration [70,71]. Portion sizes were also estimated in different ways in the FFQs analyzed by including photographs, descriptions, and household measures such as spoons. Thus, it must be taken into account that portion-size questions do not always improve the performance of FFQs or that methods to estimate portion sizes should be improved.

The novelty of this review compared with previous reviews of FFQs [7,67,68,72] is that we specifically analyzed the association between design characteristics of FFQs and their validity. Three other studies specifically compared the validity of Block- and Willett-type FFQs [24,73,74]. A limitation of our review is that we could not disentangle the effects of type of questionnaire —Block or Willett FFQs— from the effects of use of portion-size questions and number of items. We did not have enough power to do so because only six Block and six Willett FFQs were included in the models. In general, we found that the Block FFQ performed better than the Willett FFQ, but, after energy adjustment, results regarding the different types were more comparable. This finding was also observed previously [24].

Apart from ranking subjects for etiologic studies, FFQs are sometimes used to assess absolute level of intake, for example, to calculate the percentage of a population that meets recommended dietary intake guidelines. We found that FFQs validated against recovery biomarkers underestimated the level of energy intake on average by 20 percent and the level of protein intake by 11 percent; thus, they are not suitable to assess levels accurately. However, FFQs are able to distinguish between subpopulations, as indicated by the analyses of the gender differences in energy intake. This difference was very similar to the gender difference found in a review that used doubly labeled water to estimate mean energy expenditure [75]. These results showed that the average difference in energy intake was much smaller when standard portions were used. This is an argument to use at least sex-specific standard portions.

Our review shows that the number of items on the FFQ should not be reduced just because of the length of the food list; doing so might reduce the validity of the FFQ. In addition, portion-size questions do not improve FFQs for all nutrients. We should pay attention to this factor in the development process or by improving methods to estimate portion size. In addition, our review shows that FFQs are able to distinguish between subpopulations, although the magnitude of these differences is underestimated.

ACKNOWLEDGEMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

REFERENCE LIST

1. Subar AF. Developing dietary assessment tools. J Am Diet Assoc 2004;104(5):769-70.
2. Schatzkin A, Kipnis V, Carroll RJ, Midthune D, Subar AF, Bingham S, Schoeller DA, Troiano RP, Freedman LS. A comparison of a food frequency questionnaire with a 24-hour recall for use in an epidemiological cohort study: results from the biomarker-based Observing Protein and Energy Nutrition (OPEN) study. Int J Epidemiol 2003;32(6):1054-62.
3. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet 2003;362(9379):212-4.
4. Wolk A, Bergstrom R, Adami HO, Ohlander EM, Bruce A, Holmberg L, Bergkvist L. Self-administered food frequency questionnaire: the effect of different designs on food and nutrient intake estimates. Int J Epidemiol 1994;23(3):570-6.
5. Jain M, McLaughlin J. Validity of nutrient estimates by food frequency questionnaires based either on exact frequencies or categories. Ann Epidemiol 2000;10(6):354-60.
6. Kuskowska Wolk A, Holte S, Ohlander EM, Bruce A, Holmberg L, Adami HO, Bergstrom R. Effects of different designs and extension of a food frequency questionnaire on response rate, completeness of data and food frequency responses. Int J Epidemiol 1992;21(6):1144-50.
7. Willett W. Nutritional Epidemiology. second edition ed Oxford university press, 1998.
8. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122(1):51-65.
9. Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol 1990;43(12):1327-35.
10. Riboli E. Nutrition and cancer: background and rationale of the European Prospective Investigation into Cancer and Nutrition (EPIC). Ann Oncol 1992;3(10):783-91.
11. Kleinbaum. Applied regression analysis and multivariable methods. 3 ed. Pacific Grove, CA: Brooks/Cole Publishing Company, 1998.
12. Field AP. Is the meta-analysis of correlation coefficients accurate when population correlations vary? Psychol Methods 2005;10(4):444-67.
13. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. Stat Med 1999;18(20):2693-708.
14. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. J Nutr 2003;133 Suppl 3:921s-4s.
15. Matthews DE. Modern nutrition in health and disease. Proteins and amino acids. Vol. 9th ed. Baltimore, MD: Williams & Wilkins Company, 1999.
16. Bohlscheid Thomas S, Hoting I, Boeing H, Wahrendorf J. Reproducibility and relative validity of food group intake in a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S59-70.
17. Boucher B, Cotterchio M, Kreiger N, Nadalin V, Block T, Block G. Validity and reliability of the Block98 food-frequency questionnaire in a sample of Canadian women. Public Health Nutr 2006;9(1):84-93.
18. Flagg EW, Coates RJ, Calle EE, Potischman N, Thun MJ. Validation of the American Cancer Society Cancer Prevention Study II Nutrition Survey Cohort Food Frequency Questionnaire. Epidemiology 2000;11(4):462-8.

19. Jain MG, Rohan TE, Soskolne CL, Kreiger N. Calibration of the dietary questionnaire for the Canadian Study of Diet, Lifestyle and Health cohort. Public Health Nutr 2003;6(1):79-86.

20. Johansson I, Hallmans G, Wikman A, Biessy C, Riboli E, Kaaks R. Validation and calibration of food-frequency questionnaire measurements in the Northern Sweden Health and Disease cohort. Public Health Nutr 2002;5(3):487-96.

21. Katsouyanni K, Rimm EB, Gnardellis C, Trichopoulos D, Polychronopoulos E, Trichopoulou A. Reproducibility and relative validity of an extensive semi-quantitative food frequency questionnaire using dietary records and biochemical markers among Greek schoolteachers. Int J Epidemiol 1997;26 Suppl 1:S118-27.

22. Ocke MC, Bueno de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.

23. Pisani P, Faggiano F, Krogh V, Palli D, Vineis P, Berrino F. Relative validity and reproducibility of a food frequency dietary questionnaire for use in the Italian EPIC centres. Int J Epidemiol 1997;26 Suppl 1:S152-60.

24. Subar AF, Thompson FE, Kipnis V, Midthune D, Hurwitz P, McNutt S, McIntosh A, Rosenfeld S. Comparative validation of the Block, Willett, and National Cancer Institute food frequency questionnaires: the Eating at America's Table Study. Am J Epidemiol 2001;154(12):1089-99.

25. van Liere MJ, Lucas F, Clavel F, Slimani N, Villeminot S. Relative validity and reproducibility of a French dietary history questionnaire. Int J Epidemiol 1997;26 Suppl 1:S128-36.

26. Andersen LF, Solvoll K, Johansson LR, Salminen I, Aro A, Drevon CA. Evaluation of a food frequency questionnaire with weighed records, fatty acids, and alpha-tocopherol in adipose tissue and serum. Am J Epidemiol 1999;150(1):75-87.

27. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, Lubin R, Thurnham DI, Key TJ, Roe L, Khaw KT, Day NE. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. Int J Epidemiol 1997;26 Suppl 1:S137-51.

28. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. Epidemiology 1990;1(1):58-64.

29. Block G, Thompson FE, Hartman AM, Larkin FA, Guire KE. Comparison of two dietary questionnaires validated against multiple dietary records collected during a 1-year period. J Am Diet Assoc 1992;92(6):686-93.

30. Brunner E, Stallone D, Juneja M, Bingham S, Marmot M. Dietary assessment in Whitehall II: comparison of 7 d diet diary and food-frequency questionnaire and validity against biomarkers. Br J Nutr 2001;86(3):405-14.

31. Callmer E, Riboli E, Saracci R, Akesson B, Lindgarde F. Dietary assessment methods evaluated in the Malmo food study. J Intern Med 1993;233(1):53-7.

32. Engle A, Lynn LL, Koury K, Boyar AP. Reproducibility and comparability of a computerized, self-administered food frequency questionnaire. Nutr Cancer 1990;13(4):281-92.

33. Fidanza F, Gentile MG, Porrini M. A self-administered semiquantitative food-frequency questionnaire with optical reading and its concurrent validation. Eur J Epidemiol 1995;11(2):163-70.

34. Goldbohm RA, van den Brandt PA, Brants HA, van't Veer P, Al M, Sturmans F, Hermus RJ. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994;48(4):253-65.

35. Hartwell DL, Henry CJ. Comparison of a self-administered quantitative food amount frequency questionnaire with 4-day estimated food records. Int J Food Sci Nutr 2001;52(2):151-9.

36. Larkin FA, Metzner HL, Thompson FE, Flegal KM, Guire KE. Comparison of estimated nutrient intakes by food frequency and dietary records in adults. J Am Diet Assoc 1989;89(2):215-23.

37. Longnecker MP, Lissner L, Holden JM, Flack VF, Taylor PR, Stampfer MJ, Willett WC. The reproducibility and validity of a self-administered semiquantitative food frequency questionnaire in subjects from South Dakota and Wyoming. Epidemiology 1993;4(4):356-65.

38. Mannisto S, Virtanen M, Mikkonen T, Pietinen P. Reproducibility and validity of a food frequency questionnaire in a case-control study on breast cancer. J Clin Epidemiol 1996;49(4):401-9.

39. Martin Moreno JM, Boyle P, Gorgojo L, Maisonneuve P, Fernandez Rodriguez JC, Salvini S, Willett WC. Development and validation of a food frequency questionnaire in Spain. Int J Epidemiol 1993;22(3):512-9.

40. McKeown NM, Day NE, Welch AA, Runswick SA, Luben RN, Mulligan AA, McTaggart A, Bingham SA. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. Am J Clin Nutr 2001;74(2):188-96.

41. Munger RG, Folsom AR, Kushi LH, Kaye SA, Sellers TA. Dietary assessment of older Iowa women with a food frequency questionnaire: nutrient intake, reproducibility, and comparison with 24-hour dietary recall interviews. Am J Epidemiol 1992;136(2):192-200.

42. Patterson RE, Kristal AR, Tinker LF, Carter RA, Bolton MP, Agurs Collins T. Measurement characteristics of the Women's Health Initiative food frequency questionnaire. Ann Epidemiol 1999;9(3):178-87.

43. Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. I. A self-administered food use questionnaire with a portion size picture booklet. Am J Epidemiol 1988;128(3):655-66.

44. Pietinen P, Hartman AM, Haapa E, Rasanen L, Haapakoski J, Palmgren J, Albanes D, Virtamo J, Huttunen JK. Reproducibility and validity of dietary assessment instruments. II. A qualitative food frequency questionnaire. Am J Epidemiol 1988;128(3):667-76.

45. Riboli E, Elmstahl S, Saracci R, Gullberg B, Lindgarde F. The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake. Int J Epidemiol 1997;26 Suppl 1:S161-73.

46. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135(10):1114-26.

47. Schroder H, Covas MI, Marrugat J, Vila J, Pena A, Alcantara M, Masia R. Use of a three-day estimated food record, a 72-hour recall and a food-frequency questionnaire for dietary assessment in a Mediterranean Spanish population. Clin Nutr 2001;20(5):429-37.

48. Tjonneland A, Overvad K, Haraldsdottir J, Bang S, Ewertz M, Jensen OM. Validation of a semiquantitative food frequency questionnaire developed in Denmark. Int J Epidemiol 1991;20(4):906-12.

49. Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE. The use of a self-administered questionnaire to assess diet four years in the past. Am J Epidemiol 1988;127(1):188-99.

50. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. Am J Clin Nutr 1993;58(4):489-96.

51. Kroke A, Klipstein Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R, Boeing H. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. Am J Clin Nutr 1999;70(4):439-47.

52. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard Barbash R, Sunshine J, Schatzkin A. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003;158(1):1-13.

53. Andersen LF, Tomten H, Haggarty P, Lovo A, Hustvedt BE. Validation of energy intake estimated from a food frequency questionnaire: a doubly labelled water study. Eur J Clin Nutr 2003;57(2):279-84.

54. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124(3):453-69.

55. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. Am J Clin Nutr 1997;65(4 Suppl):1130S-7S.

56. Bohlscheid-Thomas S, Hoting I, Boeing H, Wahrendorf J. Reproducibility and relative validity of energy and macronutrient intake of a food frequency questionnaire developed for the German part of the EPIC project. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S71-81.

57. Kaaks R, Riboli E. Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S15-25.

58. Lietz G, Barton KL, Longbottom PJ, Anderson AS. Can the EPIC food-frequency questionnaire be used in adolescent populations? Public Health Nutr 2002;5(6):783-9.

59. Bartali B, Turrini A, Salvini S, Lauretani F, Russo CR, Corsi AM, Bandinelli S, D'Amicis A, Palli D, Guralnik JM, Ferrucci L. Dietary intake estimated using different methods in two Italian older populations. Arch Gerontol Geriatr 2004;38(1):51-60.

60. Anonymous. Relative validity and reproducibility of a diet history questionnaire in Spain. II. Nutrients. EPIC Group of Spain. European Prospective Investigation into Cancer and Nutrition. Int J Epidemiol 1997;26 Suppl 1:S100-9.

61. Riboli E, Toniolo P, Kaaks R, Shore RE, Casagrande C, Pasternack BS. Reproducibility of a food frequency questionnaire used in the New York University Women's Health Study: effect of self-selection by study subjects. Eur J Clin Nutr 1997;51(7):437-42.

62. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol 1986;124(1):17-27.

63. Kipnis V, Subar AF, Midthune D, Freedman LS, Ballard Barbash R, Troiano RP, Bingham S, Schoeller DA, Schatzkin A, Carroll RJ. Structure of dietary measurement error: results of the OPEN biomarker study. Am J Epidemiol 2003;158(1):14-21.

64. Subar AF, Heimendinger J, Patterson BH, Krebs Smith SM, Pivonka E, Kessler R. Fruit and vegetable intake in the United States: the baseline survey of the Five A Day for Better Health Program. Am J Health Promot 1995;9(5):352-60.

65. Feunekes GI, van 't Veer P, van Staveren WA, Kok FJ. Alcohol intake assessment: the sober facts. Am J Epidemiol 1999;150(1):105-12.

66. Noethlings U, Hoffmann K, Bergmann MM, Boeing H. Portion size adds limited information on variance in food intake of participants in the EPIC-Potsdam study. J Nutr 2003;133(2):510-5.

67. Burley V, Cade J, Margetts B, Thompson R, Warm D. Consensus Document on the development, validation and utilisation of food frequency questionnaires. Nuffield Institute for Health, University of Leeds and Institute of Human Nutrition, University of Southampton, 2000;63.

68. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires – a review. Public Health Nutr 2002;5(4):567-87.

69. Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. Int J Epidemiol 1992;21(4):770-7.

70. Young LR, Nestle M. Expanding portion sizes in the US marketplace: implications for nutrition counseling. J Am Diet Assoc 2003;103(2):231-4.

71. Young LR, Nestle MS. Portion sizes in dietary assessment: issues and policy implications. Nutr Rev 1995;53(6):149-58.

72. Wise A, Birrell NM. Design and analysis of food frequency questionnaires – review and novel method. Int J Food Sci Nutr 2002;53(3):273-9.

73. Wirfalt AK, Jeffery RW, Elmer PJ. Comparison of food frequency questionnaires: the reduced Block and Willett questionnaires differ in ranking on nutrient intakes. Am J Epidemiol 1998;148(12):1148-56.

74. Tylavsky FA, Sharp GB. Misclassification of nutrient and energy intake from use of closed-ended questions in epidemiologic research. Am J Epidemiol 1995;142(3):342-52.

75. Black AE, Cole TJ. Within- and between-subject variation in energy expenditure measured by the doubly-labelled water technique: implications for validating reported dietary energy intake. Eur J Clin Nutr 2000;54(5):386-94.

3

Selecting informative food items for compiling food frequency questionnaires: comparison of procedures

Marja L. Molag[1], Jeanne H.M. de Vries[1], Niels Duif[1], Marga C. Ocké[2], Pieter C. Dagnelie[3], R. Alexandra Goldbohm[4], Pieter van `t Veer[1]

[Submitted]

[1] Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands

[2] National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

[3] Department of Epidemiology, Maastricht University, Maastricht, the Netherlands

[4] TNO Quality of Life, Leiden, the Netherlands

ABSTRACT

The authors automated the selection of foods in a computer system that compiles and processes tailored FFQs. For the selection of food items, several methods are available. The aim of this study was to compare food lists made by *MOM2*, which identifies food items explaining most of the variance in intake of the nutrients of interest without taking other items into account with food lists made by forward regression and to compare both procedures. Food items were selected for the nutrients of interest from two days of recorded intake in 3524 adults aged 25-65 y. Food lists by 80% *MOM2* were compared to those by 80% explained variance for regression on differences between the number and type of food items, and were evaluated on 1) the percentage of explained variance and 2) percentage contribution to population intake computed for the selected items on the food list. *MOM2* selected the same food items for calcium, a few more for fat and vitamin C, and a few less for carbohydrates and dietary fiber than forward regression. Food lists by *MOM2* based on 80% of variance in intake covered 75% to 87% of explained variance for different nutrients by regression and contributed 53% to 75% to total population intake. Concluding, selections by variance in nutrient intake or *MOM2* is a practical procedure for developing food lists for FFQs and its selections are comparable to those by regression analysis.

INTRODUCTION

Food frequency questionnaires (FFQs) are commonly used to assess dietary intake in large epidemiologic studies [1]. Despite comments on their validity FFQs will continue to be used as they can be distributed in much larger populations than food records [2] and are able to assess intake with a longer reference period than for example food records or 24-hour dietary recalls.

Basically an FFQ consists of a food list enumerating the most informative food items for the purpose of a study. A food list should be as short as possible because long lists are less cost and time efficient, may bore respondents and make them less motivated to fill out an FFQ [3]. Informative food items need to fulfill three general characteristics. The food must be consumed regularly, contribute substantially to the nutrient(s) of interest, and be able to rank individuals according to their intake, i.e. varying in use between persons [3].

Willett [3] describes three different procedures to select informative food items. A simple procedure identifies food items with a high nutrient content from published food composition tables. However, this approach might lead to inclusion of food items that are consumed infrequently. The second procedure uses open-ended food intake data from a population, such as those obtained by food records or 24-hour dietary recalls. Food items are selected on the basis of their percentage contribution to nutrient intake in a population [4]. This selection procedure is simple and suitable if the purpose of the FFQ is to estimate the absolute level of intake in a population. A third approach, forward regression, uses similar data but predicts the food items that explain most variance, taking covariance between nutrient intakes of food items into account. This selection procedure is very suitable if the purpose of the FFQ is to rank or classify individuals according to their intake e.g. in epidemiologic studies.

It is essential that the food list of an FFQ is adapted to relevant new foods introduced on the food market, food patterns in the target population and to nutrients of interest [5]. For this reason, it is recommended to develop new FFQs for each study; however, as this process is highly labor-intensive, often existing 'old' questionnaires are reutilized or modified [6]. To facilitate the development of tailor-made FFQs, the authors devised a new computer system in which food lists will be automatically generated and updated in a standardized way. For item selection, this computer system will use food consumption databases of the population of interest and food composition tables.

For this computer system, we needed a feasible approach to select relevant food items automatically from the databases. The second approach described by Willett, which includes food items highly contributing to the level of nutrient intake of the total population could be incorporated relatively simply into our computer system.

Incorporation of the third approach, described by Willett, using forward regression is much more complicated. Forward regression evaluates many different combinations of food items and their estimated regression coefficients in order to provide the combination that explains the highest variance in nutrient intake based on their variance and covariance. This process overloads the computer system, because large databases are used and regression analysis tests all possible combinations in search of the most optimal combination of food items. An alternative method simply selects food items based on variance in nutrient intake only. Since this method does not take covariance in nutrient intake of different food items and their estimated regression coefficients into account, and tests only one combination of food items [7], it does not overload the computer system. This procedure, refers to the second moment (the variance) of the nutrient intake distribution and was therefore previously [7] called *MOM2*.

As the above mentioned selection procedures select foods at the level of food codes in a food composition table, related foods are combined into food items after the procedure. Because this grouping is often quite arbitrary [3], we developed standardized hierarchical levels of food groups for use in the computer system.

The aim of the present study was to compare food lists made by *MOM2*, which identifies food items explaining most of the variance in nutrient intake without taking other items into account with food lists made by forward regression and to compare both procedures.

METHODS

Data

Food consumption data of the Dutch National Food Consumption Survey of 1997/1998 were used for selecting food items. This dataset comprised 6250 non-institutionalized persons aged 1-97 y in 2564 households, representative of the Dutch population according to sociodemographic characteristics [8]. For the present illustration of the method, data of a subpopulation of 3524 adults between 25 and 65 years of age were used for analyses.

Information on food consumption in this dataset was obtained with a two day food record and the average intake on these two consecutive days was used. The foods consumed at home were recorded in a household diary for all individual members of the household by the person usually engaged in preparation of the meals. Consumption out of home was recorded by every participant in a personal diary. Food consumption data were collected during 40 wk/y and for the total population evenly distributed over the seasons and all days of the week.

Nutrient intake was calculated using the Dutch food composition table, NEVO 1996 [9].

Choice of nutrients

In order to study the suitability of the selection procedures, various nutrients that represent different aspects of the food pattern were incorporated. We focused on carbohydrates and fat to represent the energy-yielding macronutrients. To these, we added vitamin C as a representative nutrient for vegetables, fruit and vitamins, dietary fiber for vegetables, fruit and cereals, and calcium for dairy foods and minerals.

Grouping of food items

Food items may be enquired at different aggregation levels depending on the level of detail required for the purpose of the study. Therefore, we divided foods into several subgroups at different levels of aggregation. Foods in the Dutch food composition table are combined into 24 food groups such as 'bread', 'fruit' and 'vegetables'. These food groups comprise a large number of foods, and are not suitable for being used as items in an FFQ. Therefore, these 24 food groups, regarded as hierarchical level one, were further subdivided into smaller food groups at four hierarchical levels of aggregation of food items. This was done by two dieticians based on similarity in eating occasions, portion sizes and nutrient contents. In total 87, 237, and 356 food items were present at aggregation levels two, three and four, respectively. An illustration of this subdivision is given for dietary fiber in bread (figure 3.1).

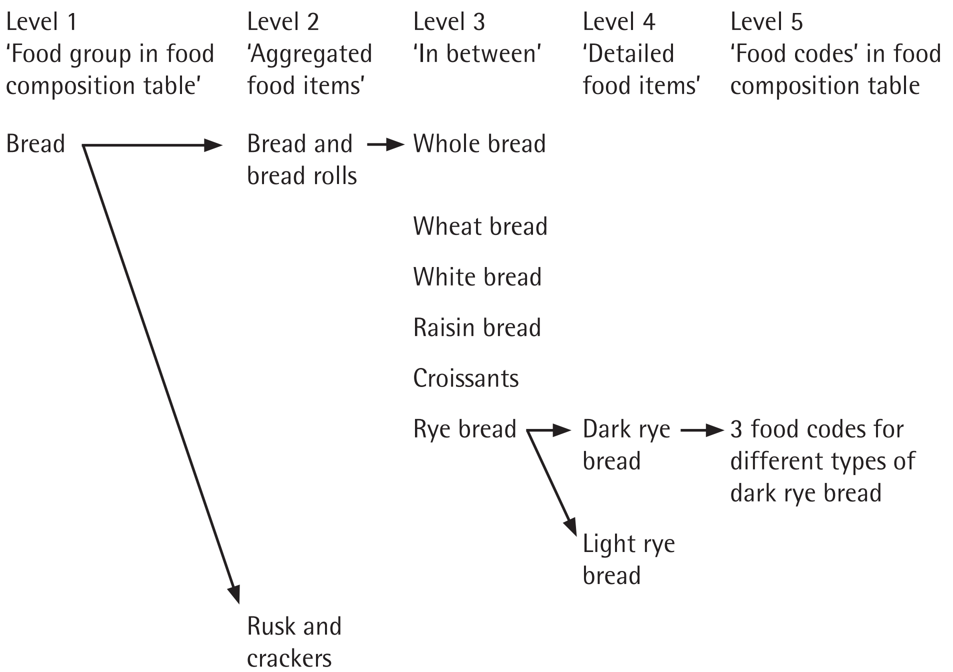


Figure 3.1 Example of aggregation levels for the food group ' Bread'.

Statistical methods

Selection of food items by *MOM2*

We selected food items using *MOM2*, i.e. a procedure that identifies food items starting with those that explain the largest variances in nutrient intake [7]. This method does not take covariance in nutrient intake of different food items into account, and tests only one combination of food items. In order to select food items, individual nutrient scores were calculated for all food items. This score F_{ij} is defined as the amount of nutrient consumed by individual i (i=1,...n) in food item j (j=1,...k). Nutrient scores over k subsequently selected food items add up to W_i, i.e. $\sum_{j=1}^{k} F_{ij} = W_i$. For the selection process, foods are ranked on the basis of the variance in their nutrient score var $F_j = \sum_{i=1}^{n} (F_{ij} - \bar{F}_j)^2$, with $\bar{F}_j$ the mean nutrient score of food item j over all individuals. For the selection of food items $R^2{}_w$ was computed i.e. the ratio of the variance in nutrient intake in the selected subset of food items W_i to the variance in the total nutrient intake Z_i over all food items in the dataset i.e. $\sum_{j=1}^{\text{total}} F_{ij} = Z_i$. The selection of food items was stopped when it exceeded a preset criterion, i.e. 80% of W_i over Z_i; thus

$$R^2{}_w = \left(\sum_{i=1}^{n} (W_i - \bar{W})^2 \,/\, \sum_{i=1}^{n} (Z_i - \bar{Z})^2\right) * 100\% \qquad [1]$$

In this formula $\bar{W}$ represents the mean nutrient intake for all individuals from selected food items and $\bar{Z}$ represents the mean nutrient intake for all individuals from all food items in the dataset.

Food items were selected per nutrient and for three different aggregation levels of food items separately. These were aggregation levels two, three and four, an illustration of this subdivision is given for dietary fiber in bread (figure 3.1). For an overview of selection procedures see table 3.1.

Selection of food items by forward regression analysis

Forward regression analysis was used as reference method to identify food items that explain most of the variance for each nutrient. Total nutrient intake (Z_i) obtained by all food items for all individuals i (=1,...,n) in the dataset was regressed on nutrient score F_{ij}

$\sum_{i=1}^{n} Z_i = \alpha_0 + \sum_{i=1}^{n} \alpha_i F_{ij} + \varepsilon_i$ [10]. Forward regression analysis adds food items to the selection

Table 3.1 Overview of selection procedures and evaluation criteria to select important food items for a food frequency questionnaire that explain variance or contribute to nutrient intake of the total population

Method	Principle	Selection procedure	Stopping value	Evaluation criterion
MOM2[a]	Selects food items on the basis of variance in nutrient scores, not taking covariance between nutrient scores of different food items into account.	$\sum_{i=1}^{n}(F_{ij} - \overline{F}_j)^2$ [b]	This procedure was stopped when the cumulative percentage of total variance R^2_w equaled at least 80%.	$R^2_w = (\sum_{i=1}^{n}(W_i - \overline{W})^2 / \sum_{i=1}^{n}(Z_i - \overline{Z})^2) * 100\%$ with $W_i = \sum_{j=1}^{k} F_{ij}$ [c]
Forward regression	Selects food items that contribute to explained variance accounting for covariance in intake of the nutrient scores of these food items.	$(\hat{Z}_i - \overline{\hat{Z}})^2$ [d]	This procedure was stopped when the percentage of explained variance predicted by regression analysis $R^2_{regression}$ equaled at least 80%.	$R^2_{regression} = (\sum_{i=1}^{n}(\hat{Z}_i - \overline{\hat{Z}})^2 / \sum_{i=1}^{n}(Z_i - \overline{Z})^2) * 100\%$ with $\hat{Z}_i = \alpha_0 + \sum_{j=1}^{k}\alpha_i F_{ij} + \varepsilon_i$ [e]
Percentage contribution	Selects food items that contribute most to the mean (first moment) of the intake of a nutrient for the population.	F_{ij}	In this study used for comparison only.	Percentage contribution $= \sum_{i=1}^{n} W_i / Z_i * 100\%$ with $W_i = \sum_{j=1}^{k} F_{ij}$

[a] *MOM2* refers to variance or the second moment of the nutrient intake distribution. [b] k = the number of selected food items, F_{ij} = nutrient score computed for individual i ($i = 1, \ldots, n$) in a food item j ($j = 1, \ldots k$).

$\overline{F}_j$ = the mean nutrient score of a food item over all individuals.

[c] R^2_w = the percentage of total variance covered, W_i = the nutrient score computed for k selected food items for an individual i, $\overline{W}$ = the mean nutrient score for k selected food items over all individuals Z_i = nutrient intake by all food items in the dataset for individual i, $\overline{Z}$ = the average nutrient score from all food items in the dataset over all individuals.

[d] $\hat{Z}_i$ = the predicted nutrient intake for individual i, $\overline{\hat{Z}}$ = the average predicted nutrient intake for all individuals n.

[e] $R^2_{regression}$ = explained variance calculated by regression, α_0 = intercept in regression formula, α_1 = regression coefficient that accompanies predicted nutrient score of a food item, ε_i = error term in regression formula for individual i. Note that the error terms over all individuals sum to zero.

starting with those with the highest predicted explained variance based on total nutrient intake and their estimated regression coefficients. This method does take covariance in nutrient intake of different food items into account, and tests all possible combinations of food items. The addition of food items stopped if the predicted variance based on the k selected food items exceeded 80% of the total variance in Z_i.

$$R^2_{regression} = (\sum_{i=1}^{n} (\hat{Z}_i - \bar{\hat{Z}})^2 / \sum_{i=1}^{n} (Z_i - \bar{Z})^2) * 100\% \quad [2]$$

In which $\hat{Z}_i$ represents the predicted nutrient intake for individual i and $\bar{\hat{Z}}$ represents the mean predicted nutrient intake over all n individuals.

Evaluation of selected foods by MOM2

Food items selected by *MOM2* were evaluated for each nutrient separately on the following characteristics:

- Differences in types and number of food items compared to selected food items by forward regression. A shorter food list was preferred.
- The percentage of explained variance by regression computed for food lists developed by *MOM2* and entered in regression analysis. This was obtained by squaring the correlation between nutrient intake by food items on the food list of the FFQ and nutrient intake by the total dataset. This provides an estimate of explained variance [11] and is easy to process because of the limited food list.

For comparison, we used the food items selected by *MOM2* and forward regression to calculate the 'percentage contribution' [4] to evaluate the level of population intake covered of the nutrients of interest, important for practice, see formula [3].

For determining 'percentage contribution' to nutrient intake, the nutrient scores were added up for the subset of selected items $\sum_{j=1}^{k} F_{ij} = W_i$ for each individual and divided by the cumulative nutrient score over the total of all food items, $\sum_{j=1}^{total} F_{ij} = Z_i$; subsequently 'percentage contribution' to population intake was calculated by

$$\text{Percentage contribution} = \sum_{i=1}^{n} W_i / \sum_{i=1}^{n} Z_i * 100\% \quad [3]$$

With W_i nutrient intake of subject i from a selected subset of k food items and Z_i = nutrient intake of subject i over all items in the dataset, i.e. the reference value for total nutrient intake.

Order of selections for different nutrients

New food items selected for the second and further nutrients were added to the list of foods selected for the first nutrient. We studied whether the order of selections for different nutrients influenced the final food list for different combinations of nutrients. In the first approach, the order was from the nutrient of which 80% variance was explained by the lowest number to that explained by the highest number of food items. In the alternative approach the reverse order was used. To study these effects, food lists developed by *MOM2* with food items defined at aggregation level two were used. Analyses were done for the following sets of nutrients: vitamin C and carbohydrates, fat and carbohydrates, dietary fiber and carbohydrates. Finally, we compared the order of selections for all nutrients in this study: vitamin C, calcium, fiber, total fat and carbohydrates.

All analyses were performed in SAS, version 9.1 (SAS Institute Inc., Cary, N.C., USA).

RESULTS

Selection by the MOM2 procedure

Selection at aggregation level two

For the five nutrients of interest, table 3.2 shows the results for the three evaluation criteria, comparing the *MOM2* procedure to the regression approach. An important result was that food items selected by *MOM2*, covering 80% of variance R^2_w, also covered at least 80% of explained variance by regression analysis (table 3.2). *MOM2* selected one or two food items more for macronutrients than forward regression and the same number of items for dietary fiber, vitamin C, and calcium. As an example, differences between the 15 food items included by *MOM2* and the 13 included by forward regression for carbohydrates at aggregation level two were studied in more detail (table 3.3). Eleven selected food items were identical for both procedures, though the order of their inclusion differed, whereas *MOM2* included 4 food items that were not included by forward regression. Specifically both procedures included 'rice', but *MOM2* included 'cooked potatoes', 'pasta', and 'ready-to-eat meals' whereas forward regression did not. This is possibly due to the negative correlation between 'cooked potatoes and 'rice' r = -0.18, 'pasta' r = -0.19 and 'ready to eat meals' r = -0.14.

Food items selected by *MOM2* also resulted in slightly higher percentages contribution to population intake (ranging from 57% for vitamin C to 75% for carbohydrates) than those by regression analysis (ranging from 57% for vitamin C to 68% for total fat and dietary fiber). In summary, *MOM2* performed similarly to regression analysis, although *MOM2* selected a few more food items than forward regression.

Table 3.2 Food items selected based on 80% of variance in nutrient intake, *MOM2*[a], compared to selections that explain 80% of variance by forward regression for carbohydrates, total fat, fiber, vitamin C and calcium evaluated for food items at three aggregation levels

Aggregation level	Level 2 'Aggregated food items'		Level 3 'Intermediate'		Level 4 'Detailed food items'	
Selection procedure	*MOM2*[a]	Regression	*MOM2*[a]	Regression	*MOM2*[a]	Regression
Nutrient Evaluation criteria						
Carbohydrates						
No. of selected food items	15	13	20	24	24	29
Explained variance by regression analysis (%)	82	81	76	81	75	81
Contribution to population intake (%)	75	67	68	68	67	70
Total fat						
No. of selected food items	11	10	28	25	42	37
Explained variance by regression analysis (%)	83	81	80	80	82	80
Contribution to population intake (%)	69	68	66	63	66	63
Dietary Fiber						
No. of selected food items	7	7	13	14	17	19
Explained variance by regression analysis (%)	81	82	77	80	75	80
Contribution to population intake (%)	70	68	68	61	66	60
Vitamin C						
No. of selected food items	3	3	7	6	12	11
Explained variance by regression analysis (%)	87	87	84	82	82	81
Contribution to population intake (%)	57	57	60	51	53	51
Calcium						
No. of selected food items	3	3	5	5	9	10
Explained variance by regression analysis (%)	87	87	80	80	79	80
Contribution to population intake (%)	61	61	56	56	53	54

[a] *MOM2* selects food items that explain variation in nutrient intake in a population.

Selection at aggregation level three

At aggregation level three, food items selected by *MOM2*, covering 80% of variance R^2_w, reached 76% of explained variance by regression analysis for carbohydrates to 84% for vitamin C (table 3.2). *MOM2* selected 20 food items for carbohydrates which was less than the 24 by forward regression. In contrast, *MOM2* included 28 food items for total

Table 3.3 Food items selected to explain 80 percent of *MOM2* compared to selections by forward regression for carbohydrates at aggregation level 2

	MOM2 selection[b]	Cumulated explained variance by regression (%)	Forward regression	Cumulated explained variance by regression (%)
1	Bread and bread rolls	30	Bread and bread rolls	30
2	Sugar, honey or dessert sauce	46	Sugar, honey or dessert sauce	46
3	Soft drinks including light and sports drinks	53	Soft drinks including light and sports drinks	53
4	French fries	55	Cakes and large cookies	59
5	Rice	58	Fresh fruit	62
6	**Cooked potatoes**[a]	59	Chocolates, chocolate and candy bars[b]	65
7	Large cookies	64	Rice	68
8	Fresh fruit	68	Milk and other dairy drinks	70
9	Alcoholic drinks	69	French fries	73
10	Milk and other dairy drinks	72	Small cookies and biscuits[b]	75
11	**Pasta**[a]	73	Cake and pie	77
12	**Fruit and vegetable juice**[a]	76	Desserts	79
13	**Ready-to-eat meals**[a]	77	Alcoholic drinks	81
14	Desserts	79		
15	Cake and pie	82		

[a] Food items included by *MOM2*, but not by forward regression are printed in bold. [b] Food items included by regression, but not by *MOM2* are underlined.

fat compared to 25 by forward regression. Food items selected by *MOM2* on aggregation level three contributed to a slightly higher 'percentage contribution' of population intake (56% for calcium to 68% for carbohydrates) than those by forward regression (51% for vitamin C to 68% for carbohydrates).

Selection at aggregation level four

At aggregation level four, food items selected by *MOM2* covering 80% of variance R^2_w, reached 75% of explained variance by regression for carbohydrates to 82% for total fat and vitamin C (table 3.2). *MOM2* included 42 food items for total fat, five more than forward regression, reaching 82% of explained variance by regression. In contrast to this, *MOM2* included fewer food items for carbohydrates than forward regression, 24 food items instead of 29. Results for selection of food items at aggregation level 4 are also shown in figure 3.2. The number of food items included by *MOM2* was slightly higher than by forward regression analysis to reach a similar level of explained variance by regression. Figure 3.2 also shows that the *MOM2* procedure was much more efficient at selecting food items that explained variance than the 'percentage contribution' procedure. With the same number of selected food items, *MOM2* covered even 20 to 30% more explained variance

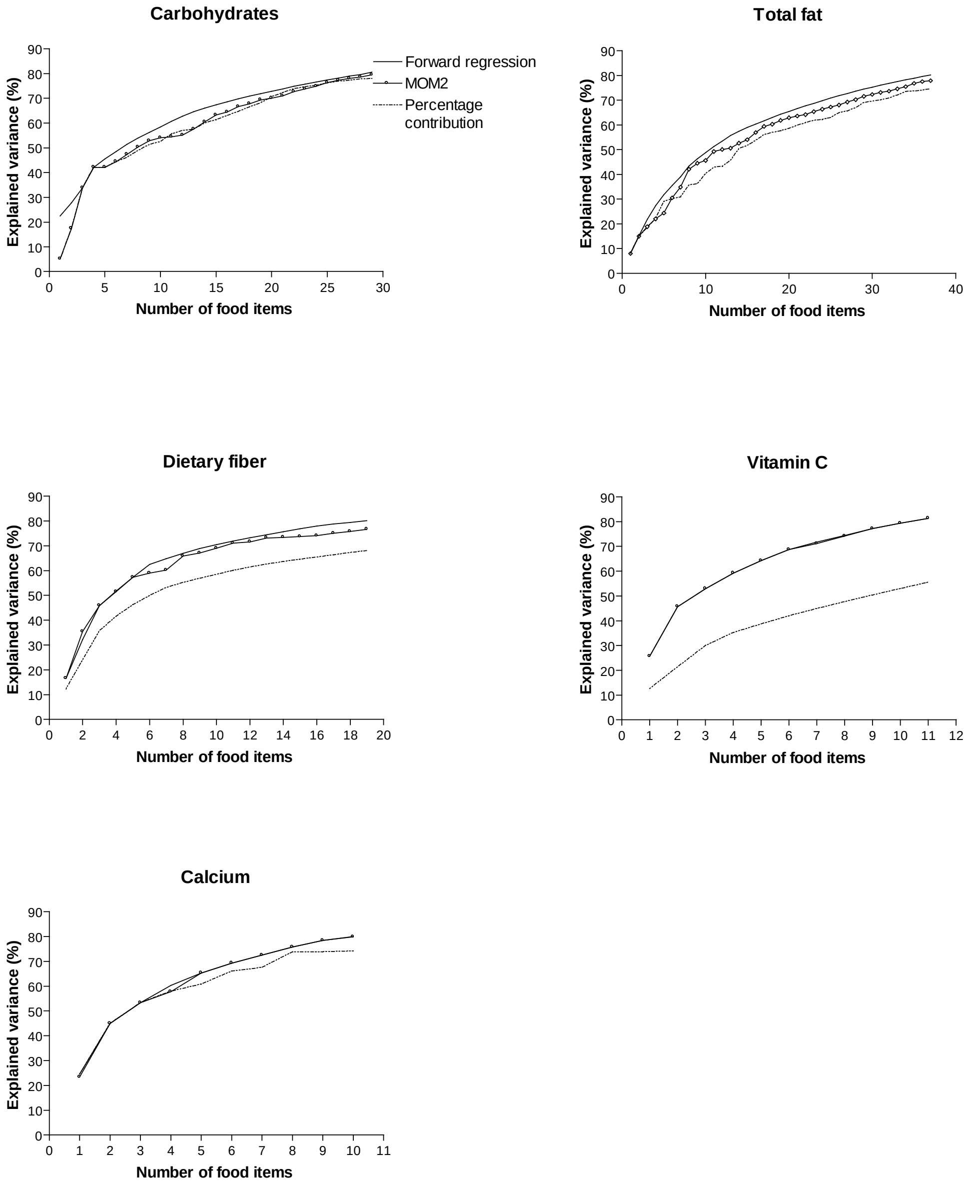


Figure 3.2 Explained variance by regression for forward regression, *MOM2* and 'percentage contribution' to population intake, by the number of included food items at aggregation level four.

for dietary fiber and vitamin C than 'percentage contribution'. In addition, percentage contribution to total population intake was similar for food items selected by *MOM2* (53% for calcium and vitamin C to 67% for carbohydrates) and forward regression (51% for vitamin C to 70% for carbohydrates).

Order of selecting nutrients

Regarding the order of selections for different nutrients, the total number of food items selected by *MOM2* differed by at most two food items. When foods were first selected for vitamin C followed by carbohydrates, 14 food items were selected, and when the order was reversed 16 items were selected. For the combination of vitamin C, calcium, fiber, total fat and carbohydrates 23 food items were included, and 22 for the opposite order. The percentage of explained variance reached by regression analysis for both selections did not differ by more than 3%. Selections made first for nutrients of interest explained by the lowest number of food items, followed by those explained by a high number of food items, led to lowest total number of food items at the highest level of explained variance.

Table 3.4 Effect of order selections for different nutrients on the number of selected food items, explained variance of the nutrients for two opposing orders

Order	Order of starting selections: nutrients with the lowest number of selected foods to that with the highest number		'Reverse order'		
Nutrients	No. of food items	Explained variance by regression (%) *	No. of food items	Explained variance by regression (%) *	No. of identical food items
Vitamin C and	14	90	16	90	14
Carbohydrates		73		74	
Carbohydrates and	21	84	22	83	20
Total fat		85		87	
Dietary fiber and	17	86	18	87	17
Carbohydrates		80		82	
Vitamin C and	23	94	22	94	20
Calcium		91		91	
Dietary fiber		91		90	
Total fat		83		86	
Carbohydrates		83		80	

[a] Explained variance was calculated by regression analysis for all selected food items.

DISCUSSION

For automated selection of informative food items for the food list of an FFQ, we evaluated a simple selection procedure, *MOM2*, which selects food items explaining the highest degree of variance in intake of selected nutrients. This simple approach was compared to food lists derived from forward regression that also takes covariance in nutrient intake into account. Food lists developed by *MOM2* and forward regression were similar. Because *MOM2* did not take covariance into account it included a few more food items in order to reach a similar level of explained variance by regression. As a consequence the percentage contribution to total population intake of the nutrients was slightly higher.

A novel aspect of our study is that we evaluated *MOM2* for the selection of food items for a food list of an FFQ using food consumption data collected with an open method, whereas previously *MOM2* was used to shorten the food list of an existing FFQ [7]. MAX_r, another selection procedure tested by the same authors, was not feasible to include in the present study, as it requires testing of all possible combinations of food items, even more than regression analysis that already overloaded the new computer system [12]. *MOM2* was considered feasible and differences relative to regression depend on the dataset from which food items were selected. An important advantage of our dataset was that many different food items were included. An important limitation of the dataset used in our study was that it was not optimal for regression analysis. Since only two subsequent food record days were available of each subject, between person variance in this dataset was artificially high since this also contains part of day-to-day variation within persons [13]. Also, the dataset included multiple persons from the same household, lowering between person variance in food intake and increasing correlations between foods. With more recording days, the dataset would have better reflected the usual dietary pattern of individuals and been more suitable to select informative food items for an FFQ by regression analysis. Without covariances in food consumption, *MOM2* and forward regression would have selected identical sets of food items [7]. Although covariances between food items exist [7], variances (reflecting between-subject variation) for many food items are much larger than covariances, and therefore dominate the selection process resulting in comparable food lists by *MOM2* and regression. This justifies the use of a much simpler method such as *MOM2*, which does not optimize the selection like forward regression.

To compare the performance of FFQs developed by *MOM2* with those developed using other procedures, we included food lists developed by *MOM2* in regression analysis and computed the explained variance. Explained variance, computed for food lists developed by *MOM2*, was only 0 to 5% less than for food lists developed by forward regression. The percentage contribution tended to be higher for *MOM2* than for regression, especially at the most aggregated levels of food items, because food lists by *MOM2* contained more food items due to the fact that covariances were not taken into account.

An advantage of automatically generating food lists is that this approach urges scientists to make the selection process and further decisions explicit a priori, such as the grouping of food items. In the present study, we compared food lists by *MOM2* and forward regression at different levels of aggregation. To develop an FFQ, food lists generated at different levels of aggregation need to be put together, for example it has to be decided whether it is more informative to assess fiber intake by consumption of 'bread' than consumption of 'whole wheat bread', 'white bread', 'croissants' and 'all other types of bread' as separate items. Food items at a high aggregation level cover the 'percentage contribution' of a nutrient better and result in a short list. This may result in a good list if the interest of the study is to assess the level of intake. However, in epidemiology we are more interested in accurately ranking individuals according to their intake and for this purpose food items at a detailed level are more suitable because they better capture the between-person variation of intake. However, a disadvantage of selecting food items at a detailed level is a longer food list. Currently decisions of aggregating items are based on experience. Optimization processes such as linear programming may help to decide on the most informative level of aggregation in future. However, a major limitation is that linear programming does not allow us to combine foods into new food groups during the process, for example including 'whole bread' and create a new food group to assess 'all other types of bread'. This new group may be important because all types of bread that are on their own not relevant enough for assessing fiber intake, may be relevant in their combination. Grouping of food items is often not explicitly described in literature, but for future research it would be important to automate this process and make it more transparent.

We focused on the statistical methods relevant to the selection of food items in an automated system, however, other factors are also important in developing FFQs. The way in which food items are grouped may also affect responses. For example, respondents may underestimate their food intake if fewer food items are included in the FFQ [14], whereas increasing the number of items may lead to overreporting [15]. The order in which food items are listed in the FFQ also influences responses [16], for example putting specific items in the FFQ prior to general items was shown to increase reported intake [17]. Moreover, it is important that a food list is comprehensible for respondents, and it may be desirable to add extra food items if this increases the face validity of FFQs. Modifications that increased comprehensibility improved validity of estimates of nutrient intake [18]. All these factors support using a simpler method such as *MOM2*, because the precision gained by using forward regression is limited compared to the impact of above mentioned factors.

A problem in generalizing our findings to developing food lists for complete FFQs may be that selection procedures were evaluated for only five nutrients. However, as they represent largely independent components of the habitual diet, we expect that these selection procedures behave similarly for other nutrients. We observed most differences for the macronutrients carbohydrates and fat, though these were minor regarding the number and type of items selected. For the micronutrients vitamin C and calcium differences were very small. Another problem is that the order in which selections for different nutrients were made slightly affected the total number of food items included in the food list. When starting with the inclusion of foods for a specific nutrient, explained by a limited number of food items, foods added to explain further nutrients also add to explained variance for the earlier selected nutrients. To further evaluate the *MOM2* selection procedure we plan to develop an FFQ using *MOM2* in order to select informative food items and validate this FFQ against another dietary assessment method and biomarkers of exposure.

We conclude that for developing food lists for FFQs, it is not necessary to take covariance in nutrient intake into account; it seems sufficient to selects food items based on the highest variance in nutrient intake.

ACKNOWLEDGEMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

The authors thank Saskia Meyboom and Henny Brants for their dietetic expertise in providing important background information for the development of the computer system such as grouping of food items at hierarchical aggregation levels.

REFERENCE LIST

1. Thompson FE, Byers T. Dietary assessment resource manual. J Nutr 1994;124(11 Suppl):2245s-317s.
2. McNeill G, Masson L, Macdonald H, Haggarty P, Macdiarmid J, Craig L, Kyle J. Food frequency questionnaires vs diet diaries. Int J Epidemiol 2009;38(3):884.
3. Willett W. Nutritional Epidemiology. second edition ed Oxford university press, 1998.
4. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124(3):453-69.
5. Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: The women's health trial feasibility study in minority populations. Am J Epidemiol 1997;146(10):856-69.
6. Subar AF. Developing dietary assessment tools. J Am Diet Assoc 2004;104(5):769-70.
7. Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. Am J Epidemiol 1996;143(5):514-21.
8. The Dutch Nutrition Centre. Zo eet Nederland: Resultaten van de Voedselconsumptiepeiling 1997-1998. Results of the Dutch National Food Consumption Survey 1997-1998; in Dutch. Voedingscentrum, den Haag. Den Haag: Voedingscentrum, 1998.
9. NEVO. Nederlands Voedingsmiddelentabel; in Dutch; Dutch Food Composition Table. 1996.
10. Rosner B, Willett WC, Spiegelman D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. Stat Med 1989;8(9):1051-69.
11. Byers T, Marshall J, Fiedler R, Zielezny M, Graham S. Assessing nutrient intake with an abbreviated dietary interview. Am J Epidemiol 1985;122(1):41-50.
12. Thomas DG, Mark SD. Max_r: an optimal method for the selection of subsets of foods for the measurement of specific nutrient exposures. Comput Methods Programs Biomed 1997;54(3):151-6.
13. Lambe J, Kearney J, Leclercq C, Zunft HF, De Henauw S, Lamberg Allardt CJ, Dunne A, Gibney MJ. The influence of survey duration on estimates of food intakes and its relevance for public health nutrition and food safety issues. Eur J Clin Nutr 2000;54(2):166-73.
14. Serdula M, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. Epidemiology 1992;3(6):503-8.
15. Bogers RP, Dagnelie PC, Westerterp KR, Kester AD, van Klaveren JD, Bast A, van den Brandt PA. Using a correction factor to correct for overreporting in a food-frequency questionnaire does not improve biomarker-assessed validity of estimates for fruit and vegetable consumption. J Nutr 2003;133(4):1213-9.
16. Kuskowska Wolk A, Holte S, Ohlander EM, Bruce A, Holmberg L, Adami HO, Bergstrom R. Effects of different designs and extension of a food frequency questionnaire on response rate, completeness of data and food frequency responses. Int J Epidemiol 1992;21(6):1144-50.
17. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr 2002;5(4):567-87.
18. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, Mittl B, Gibson JT, Ziegler RG. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102(2):212-25.


Halfvolle
Melk
Eerlijk en puur

4 Options for improvement of FFQs using cognitive interviewing

Marja L. Molag[1], Reint Jan Renes [2], Saskia Meijboom[1], Henny A.M. Brants [3, 4], Pieter C. Dagnelie [5], Jeanne H.M. de Vries[1]

[Submitted]

[1] Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands
[2] Department of Communication Science, Wageningen University, Wageningen, the Netherlands
[3] National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands
[4] TNO Quality of Life, Zeist, the Netherlands
[5] Department of Epidemiology, Maastricht University, Maastricht, the Netherlands

ABSTRACT

Background

Food frequency questionnaires (FFQs) are often used to assess dietary intake in clinical and epidemiologic studies and also for dietary counselling. However, filling out an FFQ is a complex cognitive task, because a variety of cognitive processes is involved in filling out an FFQ. Therefore problems experienced by Dutch respondents when filling out an FFQ were studied to improve questions for FFQs.

Methods

Based on think-aloud interviews and literature, categories of problems were identified and discussed in seven focus group meetings with 40 respondents aged 25-40 y or 40-65 y. Example questions from FFQs used in the meetings had been improved according to principles of cognitive science. Results of discussions were transcribed and analyzed by means of thematic coding.

Results

Respondents experienced problems in comprehension of questions resulting in selective answers because they tended to restrict their reports to the given examples. They also had difficulties in identifying foods, because categorization of foods was not logical to them; in aggregating frequencies of consumption of single foods into one food category; and in assessing relative frequencies of foods.

Conclusions

This study confirms that filling out questions of an FFQ is a complex task for respondents. Questions from an FFQ improved by principles of cognitive science may still encounter problems relating to question comprehension, food identification and reporting of frequencies. FFQs may be further improved by removing or changing examples to the questions, and by including cues that explain how to answer questions about relative frequencies.

INTRODUCTION

Food frequency questionnaires (FFQs) are often used to assess dietary intake in clinical and epidemiologic studies and for dietary counselling. Food frequency questionnaires (FFQs) have become the primary method to assess dietary intake in large studies, especially because they are easily administered to large populations at relatively low costs. FFQs are considered accurate for ranking individuals according to their intake, but are less suitable for providing a measure of the level of intake [1].

Despite their popularity, dietary intake cannot be assessed without errors using FFQs. Some errors are linked to the questionnaire itself. For example, dietary information may be missed because not all foods are assessed in an FFQ [1-3]. Errors are also due to misreporting by respondents because filling out an FFQ is a complex task. This task can be divided into distinct stages. The first stage is question comprehension, in which the respondent interprets the meaning of a question [4]. The second is retrieval, in which the respondent searches long-term memory for relevant information. It is especially difficult to recall specific episodes of food intake [5]. The third stage is estimation or judgment, in which the respondent evaluates the information retrieved from memory and its relevance to the question. In this stage, respondents encounter difficulties in recalling frequencies and portion sizes of consumed foods [6]. Fourth is the response stage, in which the respondent decides what answer to provide and may under- or over report consumption of specific foods based on e.g. socially desirable answers [7,8]. Reporting is also difficult because respondents are obliged to describe their intake according to predefined categories of foods. Thus, a variety of cognitive processes involved in filling out an FFQ explains why it is difficult to accurately report dietary intake by this method.

Accuracy of reporting may be improved by making FFQs cognitively easier for respondents. Subar et al used cognitive interviewing to identify problems that respondents encounter when filling out an FFQ [9]. They applied solutions to these problems for developing the Diet History Questionnaire [10]. Solutions included changes to improve comprehension, such as the order of food items, computing average frequencies for aggregated categories and seasonal items, inclusion of portion size ranges and adding a separate response category for foods that were never eaten [9]. These modifications improved validity of estimates of intake [10].

Despite all this effort, underreporting was still a significant problem for the Diet History Questionnaire [8]. Therefore, a qualitative study was performed in order to improve new FFQs.

Problems encountered by younger and older adults when filling out food frequency questionnaires were investigated in focus group discussions. Thereby three aspects were important: (1) comprehension of questions, (2) identification of foods, and (3) assessing frequencies of consumption. In addition, solutions to these problems were explored.

METHODS

Design

As a preparation to the focus group discussions, respondents were asked to fill out selected questions from an FFQ. Subsequently, they participated in one of seven focus group meetings together with other respondents from the same age group. Two different age groups were included, because it was assumed they had different eating patterns, which could affect the encountered problems. In general, younger adults have a more multi-cultural eating pattern with more foreign dishes and ready-to-eat meals [11], whereas older adults have a higher consumption of traditional foods, eat smaller portion sizes, and have a more organized, traditional eating pattern [12].

In October and November 2007, all focus group discussions were conducted by three trained researchers using an interview guide with discussions topics.

Respondents

Respondents were recruited in the surroundings of Wageningen (the Netherlands), via flyers, approaching them personally near schools, through advertisements in local papers or by email. Respondents who were able to read and speak Dutch were included and who had a Dutch food consumption pattern, because these characteristics are important for filling out the FFQ.

Food frequency questionnaire

Questions written in full and inquiring about 68 food items of a main meal from an existing 121-item FFQ were selected. These questions were considered as most important for total dietary intake and representative of the other questions. The FFQ was self-administered and had been developed at the Human Nutrition Division of the Wageningen University to estimate the intake of energy and macronutrients of the previous 4 weeks. It was based on the validated FFQ by Feunekes et al. [13], updated, and adapted according to a cognitive-based nested approach 9 and used in several studies [14-16]. Using this FFQ, first frequency of food intake was questioned by frequency categories starting at 'never' and then ranging from 'once a month' to 'six to seven times per week'. This was followed by more specific questions about number of portions and portion sizes. After that, questions about specific types of foods included in a food item, for example skimmed milk in the food item milk, were questioned as relative frequencies by asking whether the food was consumed either always, often, sometimes or never as compared to the other types in the food item (figure 4.1 and figure 4.2).

The selection of the FFQ in the study questioned consumption of soup, some ready-to-eat meals, pasta, rice, legumes, potatoes, vegetables, fish, meat, vegetarian meat substitutes, gravy, additions to meals such as sauces, and desserts. Three types of ready-to-eat meals

1a How often did you consume meat during the past month? Please do not count meat that was added to soup, snacks or cold cuts consumed as a sandwich filling.

this month not	1 day a month	2–3 days per month	1 day a week	2–3 days a week	4–5 days a week	6–7 days a week
❍1	❍2	❍3	❍4	❍5	❍6	❍7
Go to the next questions						

1b Each day you ate meat how many **portions** did you consume? One portion is a hamburger, a drumstick, a steak or another piece of meat of about 100 grams.

½	1	1½	2 or more
❍1	❍2	❍3	❍4

1c What **types of meat** did you eat?

	always	often	sometimes	never
Beef: steak, cut of veal, roast beef, braising steak, stewing meat, rolled beef, steak tartar **Pork:** pork steak, schnitzel, pork tenderloin **Other:** chicken-/ turkey without skin, rolled chicken, hare, goat's meat, horse's meat, venison, ostrich	❍1	❍2	❍3	❍4
Beef: olives of beef, marbled braising steak, rib-eye, rib of beef, tongue **Pork:** fillet of pork, cutlet, ham, shawarma, shoulder chop **Other:** chicken with skin, lamb's ground meat, lamb, pheasant, rabbit, partridge, Turkish ground meat	❍1	❍2	❍3	❍4
Pork: hamburger, smoked sausage, slice of bacon, bacon, sausages, spare-ribs, bratwurst **Other:** minced meat (beef and pork), chicken burger, mutton, mutton shawarma, cold cuts	❍1	❍2	❍3	❍4
Liver- and kidney products	❍1	❍2	❍3	❍4
Snack products	❍1	❍2	❍3	❍4
Unknown type of meat	❍1	❍2	❍3	❍4
Other, ______________________	❍1	❍2	❍3	❍4

1d How often did you eat the above mentioned types of meat **in breadcrumbs**?

always	often	sometimes	never
❍1	❍2	❍3	❍4

1e In which type of cooking fat did you fry your **meat**? Please indicate at maximum two types.

- ❍ Butter
- ❍ Margarine in a paper wrap
- ❍ Margarine in a tub
- ❍ Margarine enriched with unsaturated fatty acids
- ❍ Margarine with plantensterols or -stanols
- ❍ Fluid margarine
- ❍ Solid cooking fat
- ❍ Fluid cooking fat
- ❍ Olive oil, arachnids oil and frying oil
- ❍ Sunflower oil and other types of oil
- ❍ Lard/bacon
- ❍ Unknown type of cooking fat
- ❍ Other, ______________________
- ❍ I do not use any type of cooking fat, for example because I grill my meat

Figure 4.1 Example of a question about meat consumption included in the FFQ.

2a	How often did you consume **pasta** such as penne, spaghetti or Chinese noodles during the past month? Please do not count ready-to-eat meals, these have already been addressed.						
	this month not	1 day a month	2-3 days per month	1 day a week	2-3 days a week	4-5 days a week	6-7 days a week
	❍1	❍2	❍3	❍4	❍5	❍6	❍7
	Go to question X						

2b	Each day you ate pasta, how many **serving spoons** did you take on average?					
	1-2	3-4	5-6	7-8	9-10	11-12
	❍1	❍2	❍3	❍4	❍5	❍6

Figure 4.2 Example of a question about pasta consumption included in the FFQ.

were included in this FFQ, i.e. Chinese/Indonesian food, pizza, and ready-to-eat pasta meals. Questions about meat consumption were divided into seven categories: meat with a low, moderate, or high fat content, liver and kidney products, snacks, unknown, and other (figure 4.1). The low and moderate fat categories contained different types of beef and pork. The FFQ also contained a list of different brands of cooking fats and spreads consumed in the Netherlands to help respondents identify these fats. Questions of other FFQs commonly used in the Netherlands [17,18] were used during the focus group discussions as alternative examples for comparison.

Procedures

The tasks between the three researchers conducting the focus group interviews were divided: one chaired the interview, one monitored the process, and one made notes. The respondents were seated around a table to allow interaction, eye contact and free flow of discussion. Sessions lasted between 120 to 150 minutes. Four focus group discussions were conducted among the younger age group, each consisting of two to ten adults, and three discussions were conducted among the older age group, each with eight adults.

Discussion topics were identified based on literature and think aloud interviews conducted in five respondents. In the interviews, the respondents read the questions out loud and verbalized their thoughts while answering these questions. In 30% of the FFQ questions, one or more of the five respondents encountered a problem. Of these problems, 45% were related to comprehension, 10% to identification of foods or the order of questions, and another 45% to assessment of amounts and frequencies.

A standardized interview guide was used during the focus group discussions and consisted of three parts: an introduction of interviewers and respondents, a group discussion about topics of interest, and a closure in which respondents could indicate problems that had not been mentioned before. The first topics discussed were comprehension of questions including use of cues and examples, and the format of the FFQ. Cues are hints that describe for example how to fill out a certain question, or remind respondents of certain foods. For the format, an alternative FFQ with all questions in a table was shown to respondents and differences between both formats were discussed. This is an often used format to considerably reduce length of a FFQ. Second, difficulties in identification of foods consumed were discussed. Based on the problems encountered in the think-aloud interviews, two examples of questions were selected, one about meat consumption (figure 4.1) and one about sauce consumption. Ability of respondents to report consumption of mixed dishes and ready-to-eat meals were also discussed. Third, difficulties in assessing frequencies and relative frequencies of consumption were discussed. As alternative to the relative frequencies of 'always-often-sometimes-never', it was asked whether an item, such as the previous example of skimmed milk was used *most often* in relation to the other type(s).

The study protocol was not handed in to the medical ethical committee, according to the Dutch law this was not required because this study had a non-invasive character and no sensitive data of respondents was collected. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Data analysis

Focus group discussions were transcribed and analyzed by means of thematic coding. Responses were coded according to the categorized topics and the opinion of respondents about the topic, for example 'suggestions to improve examples'. The transcripts of the results of two focus group discussions (one with older respondents and one with younger respondents) were analyzed independently by the three researchers who were present during the focus group interviews. After the analyses of these discussions, the results of the other five were analyzed by two of the three researchers. Agreement between the two researchers in coding was 72%. Differences in analyses were discussed, agreement was reached on all codes, and the problems were categorized.

RESULTS

Sixteen respondents aged 25 to 40 y, median age 30 y, and 24 respondents aged 40 to 65 y, median age 59 y were included. All respondents in the younger age group that applied were included; and of those in the older age group 24 respondents out of 67 applications were randomly selected in this study. Seven respondents had a low, 11 a medium and 22 a high educational level (table 4.1).

Not many differences between the younger and older age groups in the problems encountered when filling out the FFQ were observed. Therefore, the results are presented for both age groups together and will be only specified per age group where relevant.

Table 4.1 Characteristics of respondents participating in the focus group discussions

Age group	Young (25-40 y)		Older (41-65 y)		Total
	Men $n = 6$	Women $n = 10$	Men $n = 12$	Women $n = 12$	$N = 40$
Age (median) [range]	30 [26-39]	32 [26-40]	61 [49-64]	56 [41-65]	51 [26-65]
Education level[a]					
Low (primary education, low vocational training, low technical training)	0	1	2	4	7
Medium (secondary education, medium technical training, medium vocational training)	0	4	5	2	11
High (higher vocational education and university)	6	5	5	6	22

[a] According to Standardized Educational Levels (Central Bureau of Statistics, The Netherlands).

Comprehension of questions

Respondents encountered several problems in filling out FFQs related to comprehension of questions (table 4.2). Some respondents were confused by unfamiliar terms used in the FFQ, such as "pasta", "home-made" and "being on a diet". Six respondents used some foods in a different way than was questioned in the FFQ. For example, respondents did not consider consumption of cheese added to *cold* dishes as part of a *hot* meal, such as salads and desserts. Seventeen respondents provided suggestions of terms or foods that were more familiar to them.

Another problem related to comprehension applied to examples that were added to the questions to make them clearer. Only four respondents thought the examples were clear and 11 respondents found them only useful. Criticism about the examples used was provided by 12 younger and 22 older respondents. An example, of a question they criticized was about sauce consumption. In this question, pasta sauce, which is consumed as part of a dish, was combined with other sauces, such as mushroom sauce, that are served separately to dishes. Also, the respondents found it hard to estimate the amount of sauces that are part of a dish. Their suggestion was to split up this question into two separate questions, one about pasta sauces and one about other types of sauces that are typically added separately.

Cues that were added to clarify questions increased comprehension according to 36 respondents, and 12 respondents thought cues were clear. Respondents made use of cues to know which foods they were supposed to include in their answer and to make them feel confident of filling out the FFQ in a correct way. On the other hand, two respondents considered cues disturbing and three thought they were superfluous. Also, ten respondents provided suggestions to improve their wording.

Table 4.2 Number of respondents mentioning specific opinions related to comprehension of questions discussed in focus group meetings

Topic	Opinion	Number of respondents (Younger, 25-40 y) N = 16	Number of respondents (Older, 41-65 y) N = 24
Familiarity of terms and foods	Familiar foods	2	4
	Unfamiliar foods	7	9
	Terms clear	3	2
	Unfamiliar terms	8	11
	Suggestions to improve terms	7	6
	Different use of foods than assessed	2	4
	Suggestions to improve description of foods	2	2
Examples to the questions	Examples clear	4	0
	Examples useful	8	3
	Criticism about examples used	12	22
	Examples superfluous	2	2
	Examples missed	6	3
	Examples directive	2	5
	Suggestions to improve examples	8	5
Cues[a]	Cues useful	12	22
	Cues clear	8	4
	Cues not disturbing	3	8
	Cues disturbing	2	2
	Cues superfluous	1	2
	Formulation of cues	1	3
	Suggestions to improve formulation of cues	4	6
Format of the FFQ	Length of FFQ too long	10	3
	Time spent on filling out FFQ not so bad	4	6
	Overview of FFQ questions in tables improves	9	2
	FFQ with questions in tables is better	3	7
	Negative about FFQ with questions in tables	12	15
	Suggestion to improve lay-out of FFQ questions in tables	5	1
	FFQ with questions written in full is better	6	13
	Suggestion to improve lay-out of FFQ with questions written in full	4	0

[a] Cues = hints how to fill out a certain question, reminders of certain foods, etc.

Thirteen respondents, especially younger respondents, thought the total length of the FFQ was too long, even though it was only half the length of the complete FFQ. Surprisingly, the time spent on filling out the FFQ was less than they had expected beforehand. The alternative FFQ with all questions put into tables, was more positively evaluated by younger than older respondents. The younger respondents indicated that this format provided them with a good overview of the number of meals they had filled out over the past month. In contrast, the older respondents thought the format was overwhelmed with information and they often felt in doubt of checking the right answer.

Identification of foods

Respondents mentioned problems with identifying foods such as meat, cooking fats, mixed dishes, and ready-to-eat meals (table 4.3). In total 31 out of the 40 respondents reported problems with the question about meat consumption, that was divided into seven categories based on type and differences in fat content. Respondents did not understand the differences between the categories, which repetitively included beef and pork. They suggested improving this question by restructuring categories according to type of meat thus leaving beef or pork in only one category.

Respondents also had problems in identifying cooking fats used. For example, five respondents did not know what type of cooking fat they consumed because they did not do the shopping or preparation of the meals. All other respondents thought they knew what type of cooking fat was used, although three of them actually filled out the wrong category. None of the respondents had used the list with brand names of cooking fats added to the FFQ to clarify what type of cooking fat they consumed. The reason for not using the list was that the purpose of this list was not clear or that it was thought to be unnecessary. Respondents suggested to put the list with cooking fats at a prominent position in the FFQ and to add a clear description.

Eighteen respondents thought they could easily fill out questions about foods consumed as part of mixed dishes, whereas four respondents had difficulties in describing these. Four respondents also had problems in indicating the portion sizes of ingredients of mixed dishes. Three respondents provided suggestions to improve reporting of mixed dishes, for example by assessing the type of mixed dishes they usually consumed such as Italian or Mexican dishes.

Only two respondents had consumed ready-to-eat meals. Six respondents thought that ready-to-eat meals could be reported in the FFQ easily, whereas three respondents thought that it would be difficult to report their consumption of these meals. Finally, ten respondents gave suggestions to improve the reporting of ready-to-eat meals by adding more detailed questions.

Table 4.3 Number of respondents mentioning specific opinions related to identification of foods discussed in focus group meetings

Topic	Opinion	Number of respondents (Younger, 25-40 y) N = 16	Number of respondents (Older, 41-65 y) N = 24
Problems in identifying foods	Difficulties in identification of meat	9	22
	Difficulties in identification of sauces	9	0
Problems in identifying cooking fats	Problems in categories of cooking fats	3	4
	Cooking fat unknown	3	2
	Choosing wrong type of cooking fat	2	1
	Not consulting list of cooking fats	16	24
	Place list of cooking fats	4	1
	Suggestions to improve list of cooking fats	7	10
Mixed dishes	Easy to fill out mixed dishes	7	11
	Difficulties in filling out mixed dishes	1	3
	Problems with portion sizes of ingredients mixed dishes	3	1
	Suggestions to improve reporting of mixed dishes	2	1
Ready-to-eat meals	Reporting consumption of ready-to-eat meals	1	1
	Easy to report ready-to-eat meals	2	4
	Unknown composition of ready-to-eat meals	3	3
	Difficulties in reporting ready-to-eat meals	3	0
	Suggestions to improve reporting of ready-to-eat meals	7	3

Assessing frequencies and relative frequencies

Ten respondents had difficulties in aggregating frequencies of consumption over the past month, because they tried not to exceed the number of total meals or foods consumed in a month. Categories used to assess usual frequencies of consumption were logical according to 22 respondents, whereas 12 thought they were not (table 4.4). Thirteen respondents thought they could accurately describe their intake according to the provided categories, whereas three thought these categories were not precise enough. Three respondents suggested to improve lay out of the frequency categories in the FFQ, by optically separating the frequency categories of once a week and 2-3 times a month.

Table 4.4 Number of respondents mentioning specific opinions related to assessing frequencies of consumption discussed in focus group meetings

Topic	Opinion	Number of respondents (Younger, 25-40 y) N = 16	Number of respondents (Older, 41-65 y) N = 24
Frequency answering options	Frequency answering categories satisfying	9	13
	Frequency answering categories not satisfying	7	5
	Frequency categories accurate	6	7
	Frequency categories not precise enough	1	2
	Difficulties in aggregating frequencies	5	5
	Suggestions to improve frequency categories	3	3
Relative frequencies	Criticism formulation of relative frequencies	7	6
	Relative frequencies preferred above choosing foods used 'most often'	10	6
	Choosing foods used 'most often' preferred above using relative frequencies	5	15
	Difficulties in aggregating relative frequencies	5	4

Most respondents were able to use the relative frequencies 'always-often-sometimes-never' to describe intake of specific types of food such as low-fat types in relation to other types. Thirteen respondents commented on the terms used to describe relative frequencies. They had difficulties in filling out that a type of food had been eaten 'always', when they had consumed a food once during the past month. They also experienced difficulties in aggregating relative frequencies. In total 20, mostly older respondents preferred a format in which they chose the foods used 'most often', above the relative frequencies of food use as 'always-often-sometimes-never'. In contrast, younger respondents were more pleased with relative frequencies, because these allowed them to describe variation in intake.

DISCUSSION

Although the FFQ used in this study was developed according to principles of cognitive research [19] in order to make filling out the FFQ as easy as possible, respondents still encountered problems. They had problems in comprehension of questions due to examples which confused them, although the examples were added to clarify questions and to help them. The respondents also found it difficult to identify foods, mainly because divisions in categories of foods e.g. types of meat according to fat content were not clear. Finally, they reported problems in assessing frequencies because they found it difficult to aggregate them over the past month and to assess relative frequencies of foods consumed.

It is important to evaluate whether problems that respondents encounter in filling out FFQs were found. Think-aloud interviews were used to identify important problems. These interviews are considered as a good and objective method for this purpose [20]. Subsequently, the focus group discussions made it possible to further study and validate the identified problems in a broader age group and a larger number of respondents. An advantage of focus group discussion is also that they permitted respondents, who did not understand the questions, to listen and contribute when they felt comfortable [21], thus allowing all respondents to articulate their problems. The number of responses in the results section gives an indication of the number of respondents with a certain opinion. Other respondents may have agreed or disagreed but might not have expressed their view, and are therefore not included in this number. Agreement between scoring of researchers was 72%, which may be considered as quite good for agreement between qualitative judgements [22].

It is important to consider whether findings may be generalized. Women and higher educated respondents were overrepresented in the focus group discussions. The accuracy with which respondents fill out FFQs depends on knowledge about their own long-term dietary pattern, which is often gained by purchase and preparation of one's own food [23]. Nowadays, women are probably still more often responsible for meal preparation than men. However, in this study, respondents were instructed to consult the person usually involved in preparing the meals if any specifics of food preparation were unknown. As for education level, it is important to note that filling out an FFQ is a complex cognitive task. For this reason, misreporting of intake is more common in lower educated respondents [24,25]. Thus, in the general population there may even be more problems in filling out FFQs than found in this study, making it even more important to make questions as simple and clear as possible.

Problems due to cognitive difficulties encountered by respondents in filling out FFQs were studied. Though findings from previous cognitive research were used to improve the FFQ, it is important to consider in which aspects the FFQ was different from those of others. Results show that respondents had difficulties with the question about meat consumption because they felt too many different types of meat were included in one category. The FFQ of Matt et al also grouped many different types of meat, and because of respondents' difficulties in estimating frequencies of grouped items, meat and some other items were separated into multiple questions [26]. The Diet History Questionnaire included each food in a separate question and included only a few examples [10]. It may be easier for respondents to split up a complex question such as about meat consumption into multiple simple questions about specific types of pork and beef. This increases clarity and may even compensate for the increased questionnaire length [27].

This study provided important insights for developing new FFQs. Examples to questions may confuse respondents especially because they often do not cover all food items that a question inquires about. Also, difficulties may be encountered, if different examples to questions can be eaten at multiple meals or as snacks, or at a non-routine frequency of consumption [10]. Thus, it is recommended to carefully consider whether including examples as a clarification is really necessary.

An important problem of FFQs is that this method relies on memory of respondents, and that it is very difficult to recall repeated behaviour such as dietary intake [28]. Some researchers consider the FFQ to be merely a measure of attitude towards foods instead of assessing food consumption [29]. Therefore, it seems very important to include cues that emphasize respondents to enter all *consumed* foods and not for example the foods that they like. Also, cues make it possible to make respondents aware of what foods to include in their answer, as was shown in this study, and cues stimulate recall of specific episodes as was shown by literature [26]. In addition, this study showed that it is important to improve identification of foods by using terms that are familiar and categories of food that are understood by respondents. For this purpose, it is important to pre-test new FFQs in target populations.

To assess frequencies of usual consumption, categories used in this study appeared to be adequate. Relative frequencies of consumption of foods within a general category e.g. of low-fat milk may provide important information, but may also be difficult to fill out. An alternative and probably easier option especially for older respondents is to fill out the foods that were used 'most often'. For younger respondents relative frequencies may be used because they indicated to prefer describing variation in their usual dietary pattern. Thompson et al also applied questions on relative frequencies in their Diet History Questionnaire and considered them to be accurate [10]. Thus, relative frequencies may be used in an FFQ, however, as they are difficult to be filled out, inquiring about no more than five or six food items within one category is recommended.

A difficult decision in designing FFQs is whether or not to include portion size questions. This topic was not addressed during the focus group discussions as it is too complex to discuss during a short time, and cognitive strategies to report portions have been studied by others [30]. Portion size questions may contribute to better capturing the between-person variation of intake. However, respondents often skip these questions [9], and they find it difficult to estimate consumed portion sizes. Even when pictures are included in an FFQ, or when portion sizes were estimated from food records, ranking of respondents according to their intake hardly improved compared to the use of standard portion sizes. However, assessing the level of intake did improve by including portion size questions [18,31]. Thus, whether it is good to include portion size questions depends on the aim of the FFQ.

Implementing the above mentioned improvements in a new food frequency questionnaire is interesting, especially when evaluating it against an independent method in a validation study. It is also recommend using web-based FFQs because they probably cause less problems. Comprehension of web-based FFQs may improve filling out questionnaires by including tailored pop-up. Also, a web-based FFQ may include pictures of unfamiliar foods to improve identification of foods. Assessing frequencies of consumption may be improved by including more frequency categories and by providing information about the relative frequency categories. Also, reporting will be more accurate, as respondents can be compelled to fill out questions, and will be more convenient because questions that do not apply to them can automatically be skipped. Thus, self-reports by web-based FFQs will reduce the amount of data cleaning, but ask for some computer skills of the respondents.

Concluding, FFQs can be further improved by carefully considering the use of examples, by only combining foods into categories that are similar for respondents and not based on their nutritional composition only, and by including cues that explain how to answer questions, for example about relative frequencies of food use within a larger category of foods.

ACKNOWLEDGEMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

The authors thank Marieke van Bakel for performing the think aloud interviews, Miranda Ruijgers for transcribing the focus group interviews and Karin Borgonjen for organizing the focus group interviews and helping with data analysis.

REFERENCE LIST

1. Thompson FE, Byers T. Dietary assessment resource manual. J Nutr 1994;124(11 Suppl):2245s-317s.
2. Barrett Connor E. Nutrition epidemiology: how do we know what they ate? Am J Clin Nutr 1991;54(1 Suppl):182s-7s.
3. Kohlmeier L, Bellach B. Exposure assessment error and its handling in nutritional epidemiology. Annu Rev Public Health 1995;16:43-59.
4. Jobe JB, Mingay DJ. Cognitive research improves questionnaires. Am J Public Health 1989;79(8):1053-5.
5. Johnson Kozlow M, Matt GE, Rock CL. Recall strategies used by respondents to complete a food frequency questionnaire: an exploratory study. J Am Diet Assoc 2006;106(3):430-3.
6. Godwin SL, Chambers Et, Cleveland L. Accuracy of reporting dietary intake using various portion-size aids in-person and via telephone. J Am Diet Assoc 2004;104(4):585-94.
7. Johansson L, Solvoll K, Bjorneboe GE, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 1998;68(2):266-74.
8. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, Sharbaugh CO, Trabulsi J, Runswick S, Ballard Barbash R, Sunshine J, Schatzkin A. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003;158(1):1-13.
9. Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, Schatzkin A, Hartman A, Swanson C, Kruse L, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95(7):781-8.
10. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, Mittl B, Gibson JT, Ziegler RG. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102(2):212-25.
11. van Kreyl CF. Ons eten gemeten. In: National institute of public health RIVM, ed, 2004;1-364.
12. Shahar D, Fraser D, Shai I, Vardi H. Development of a food frequency questionnaire (FFQ) for an elderly population based on a population survey. J Nutr 2003;133(11):3625-9.
13. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. Am J Clin Nutr 1993;58(4):489-96.
14. Kroeze W, Oenema A, Campbell M, Brug J. The Efficacy of Web-based and Print-delivered Computer-tailored Interventions to Reduce Fat Intake: Results of a Randomized, Controlled Trial. J Nutr Educ Behav 2008;40(4):226-36.
15. Werkman A, Schuit AJ, Kwak L, Kremers SPJ, Visscher TLS, Kok FJ, Schouten EG. Study protocol of a cluster randomised controlled trial investigating the effectiveness of a tailored energy balance programme for recent retirees. BMC Public Health 2006;6.
16. Verheijden MW, Van Der Veen JE, Bakx JC, Akkermans RP, Van Den Hoogen HJM, Van Staveren WA, Van Weel C. Stage-matched nutrition guidance: Stages of change and fat consumption in Dutch patients at elevated cardiovascular risk. J Nutr Educ Behav 2004;36(5):228-37.
17. Grootenhuis PA, Westenbrink S, Sie CM, de Neeling JN, Kok FJ, Bouter LM. A semiquantitative food frequency questionnaire for use in epidemiologic research among the elderly: validation by comparison with dietary history. J Clin Epidemiol 1995;48(7):859-68.

18. Ocke MC, Bueno de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.

19. Subar AF, Heimendinger J, Patterson BH, Krebs Smith SM, Pivonka E, Kessler R. Fruit and vegetable intake in the United States: the baseline survey of the Five A Day for Better Health Program. Am J Health Promot 1995;9(5):352-60.

20. Schwarz N, Oyserman D. Asking questions about behavior: cognition, communication, and questionnaire construction. Am J Eval 2001;22(2):127-160.

21. Pasick RJ, Stewart SL, Bird JA, D'Onofrio CN. Quality of data in multiethnic health surveys. Public Health Rep 2001;116(SUPPL. 1):223-43.

22. Perreault WDW, Leigh LE. Reliability of nominal data based on qualitative judgments. J Marketing research 1989;26(2):135-48.

23. Willett WC. Accuracy of food-frequency questionnaires. Am J Clin Nutr 2000;72(5):1234-6.

24. Tooze JA, Subar AF, Thompson FE, Troiano R, Schatzkin A, Kipnis V. Psychosocial predictors of energy underreporting in a large doubly labeled water study. Am J Clin Nutr 2004;79(5):795-804.

25. Bedard D, Shatenstein B, Nadon S. Underreporting of energy intake from a self-administered food-frequency questionnaire completed by adults in Montreal. Public Health Nutr 2004;7(5):675-81.

26. Matt GE, Rock CL, Johnson Kozlow M. Using recall cues to improve measurement of dietary intakes with a food frequency questionnaire in an ethnically diverse population: an exploratory study. J Am Diet Assoc 2006;106(8):1209-17.

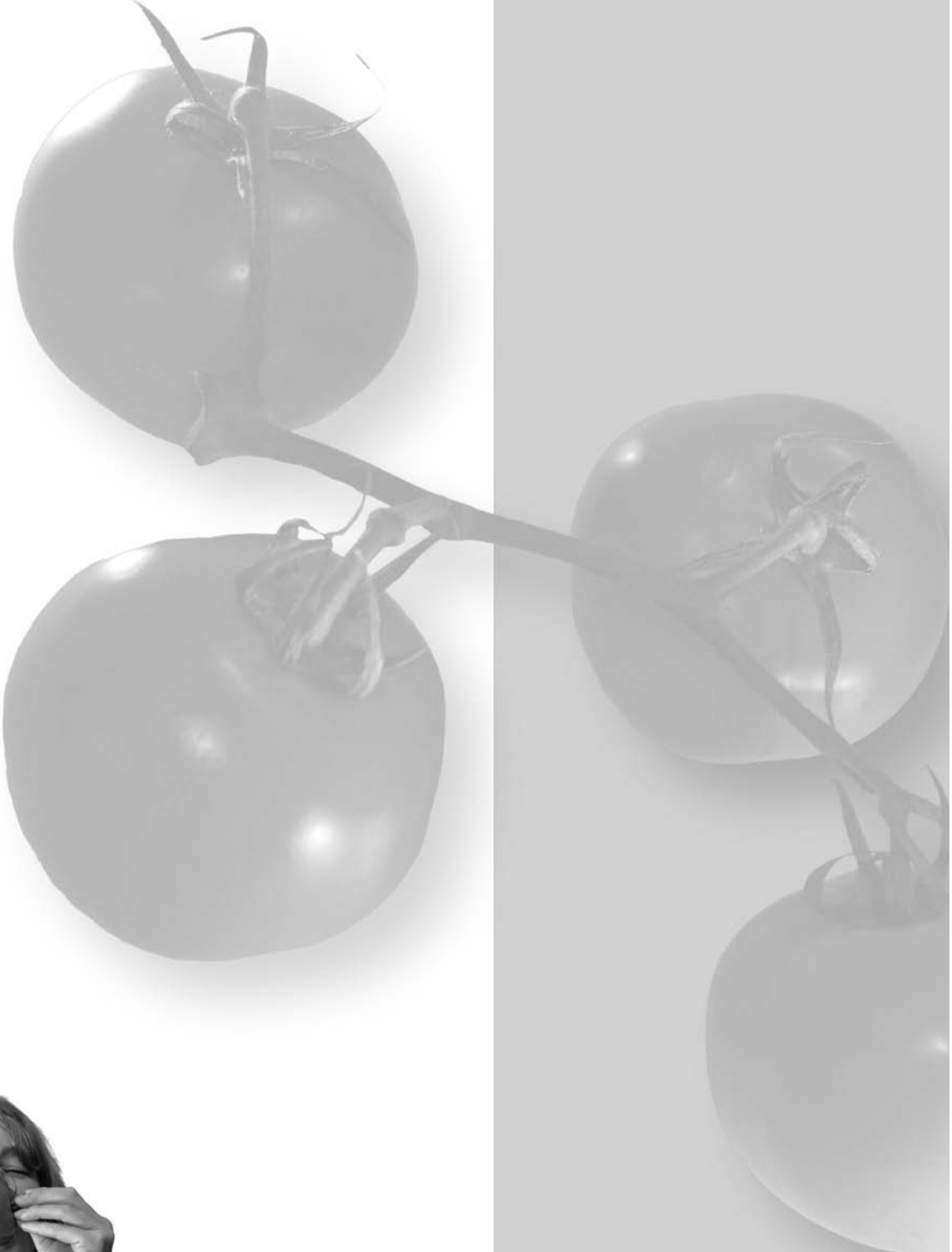
27. Subar AF, Ziegler RG, Thompson FE, Johnson CC, Weissfeld JL, Reding D, Kavounis KH, Hayes RB. Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. Am J Epidemiol 2001;153(4):404-9.

28. Schwarz N. How the questions shape the answers. Am Psychologist 1999;54(2):93-105.

29. Drewnowski A. Diet image: a new perspective on the food-frequency questionnaire. Nutr Rev 2001;59(11):370-2.

30. Chambers Et, Godwin SL, Vecchio FA. Cognitive strategies for reporting portion sizes using dietary recall procedures. J Am Diet Assoc 2000;100(8):891-7.

31. Nothlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. J Nutr 2007;137(12):2781-6.

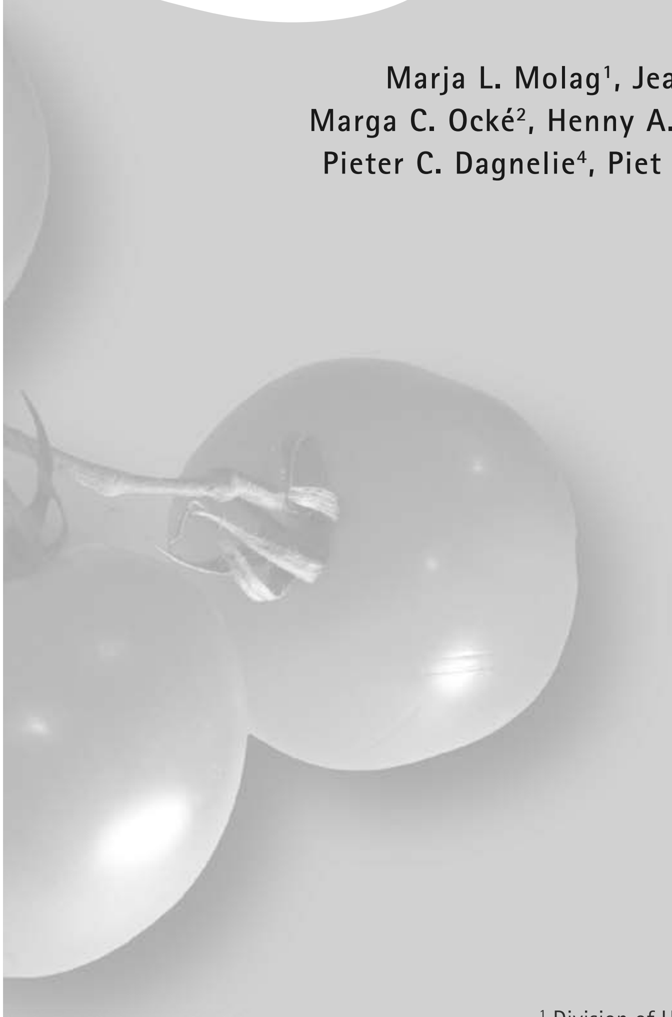


5

The Dutch FFQ-TOOL™: development and use of a computer system to generate and process FFQs

Marja L. Molag[1], Jeanne H.M. de Vries[1], Saskia Meijboom[1], Marga C. Ocké[2], Henny A.M. Brants[2, 3], R. Alexandra Goldbohm[3], Pieter C. Dagnelie[4], Piet A. van den Brandt[4], Pieter van `t Veer[1]

[Submitted]



[1] Division of Human Nutrition, Wageningen University, the Netherlands

[2] National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

[3] TNO Quality of Life, Leiden, the Netherlands

[4] Department of Epidemiology, Maastricht University, the Netherlands

ABSTRACT

Background

Food frequency questionnaires (FFQs) are used to assess usual dietary intake of participants often in large studies. They have to be revised and updated regularly because of new research questions and a continuously, rapidly changing food market and altering food patterns. However, developing new FFQs or updating old versions is difficult because time, finances or expertise are often not available. We developed a computer system, the so-called *Dutch FFQ-TOOL™*, that can quickly generate and process FFQs tailored for specific research questions or populations.

Objective

We describe the structure and contents of the system and show, as an example, how an FFQ for assessing vitamin C and dietary fibre intake is developed by the system. This process is evaluated by comparison with predefined requirements.

Design

Using the *Dutch FFQ-TOOL™*, which is an internet-based system, FFQs are developed using food consumption data of the target population. For use of FFQs in epidemiologic studies, foods may be selected based on explained variance in nutrient intake, to rank individuals according to their intake. For use of FFQs to monitor populations, foods may be selected based on the contribution to the level of nutrient intake in a population. After selection, foods are automatically transposed into standard questions and included in the FFQ. Finally, food and nutrient intake of completed FFQs is computed by the *Dutch FFQ-TOOL™*.

Results

We describe the development of an example FFQ of 30-items, for estimating vitamin C and dietary fibre intake that fulfilled most requirements from a scientific, respondent's, and technical perspective. The only requirement not met was to completely automate the selection of foods.

Conclusion

The prototype of the *Dutch FFQ-TOOL™* is able to generate, apply, and process tailored FFQs for multiple nutrients and target groups by reproducible and valid procedures.

INTRODUCTION

Food frequency questionnaires (FFQs) are the most often used instrument to assess usual dietary intake in large epidemiologic studies. Although the accuracy of FFQs has been debated [1,2], FFQs will continue to be used as they are more feasible than other methods to assess long-term intake in large populations [3-6]. A well-designed FFQ is targeted to assess habitual intake of foods or nutrients of interest in a specific population. This implies that the food list of an FFQ has to be adapted to food consumption habits of the target population. An appropriate food list can be obtained by selecting foods from food consumption data collected by food records or 24-hour dietary recalls from a similar population [7]. In this way, each FFQ is tailor-made to assess intake of selected foods or nutrients of the population of interest. Unfortunately, many researchers do not have time, money, or expertise to develop new FFQs, as development of questionnaires is a labor-intensive and specialized task [8]. Moreover, this task is difficult because updated food consumption data is often not available to the researcher [8], and requirements from a scientific, respondent's and technical perspective must be carefully balanced. Therefore, automating the development of FFQs by experts in the field of dietary assessment could help other researchers in developing and updating these questionnaires more easily and efficiently.

Automating development of FFQs

To facilitate FFQ development, researchers have previously tried to automate FFQ development. Researchers in the UK developed Windiets [9] in which nutrient computations of completed FFQs were automated. However, foods were manually selected from food consumption data from weighted food records reported by a representative sample of the population (ref personal communication Alan Wise dd 19-04-2005). Finally, selected foods were transposed into questions to develop an FFQ.

Another example is the Vofrex/Vovris system, developed in the Netherlands by TNO Quality of Life (formerly TNO Nutrition and Food Research) [10,11]. In Vofrex, foods contributing most to population intake were automatically selected, and in Vovris, processing of completed FFQs was automated. As a source for selecting foods, the system used food consumption data from food records, collected in the Dutch National Food Consumption Survey 1992 [12]. The Vofrex/Vovris system had a few limitations. The most important was lack of flexibility with regard to including new versions of food composition databases and new foods. In addition, no subpopulations could be defined. Also, foods could only be selected based on their percentage contribution to level of population intake, and not by cumulative variance in nutrient intake which is more important for epidemiologic studies [13].

Therefore, there was a need to device a new and more flexible computer system that supports the development and processing of FFQs. This system should automate FFQ development, for different subpopulations. The selections in the system should not only be based on percentage contribution but also on variance in nutrient intake to generate FFQs that are suitable for use in epidemiologic studies.

The aim of this paper is to give a short description of the prototype of the so-called '*Dutch FFQ-TOOL™*', in which FFQ development is automated and researchers can develop and process tailored FFQs in a standardized way. We illustrate this by applying the system to an example, i.e. development of an FFQ that ranks adults according to their vitamin C and dietary fiber intake. This process is evaluated by comparison with requirements the system had to fulfill.

MAIN OBJECTIVES AND REQUIREMENTS OF THE DUTCH FFQ-TOOL™

The general aim of the *Dutch FFQ-TOOL™* was to develop a data-based system that generates valid FFQs by the use of transparent and reproducible procedures, which are suitable for assessing intakes of energy and nutrients of interest for different populations.

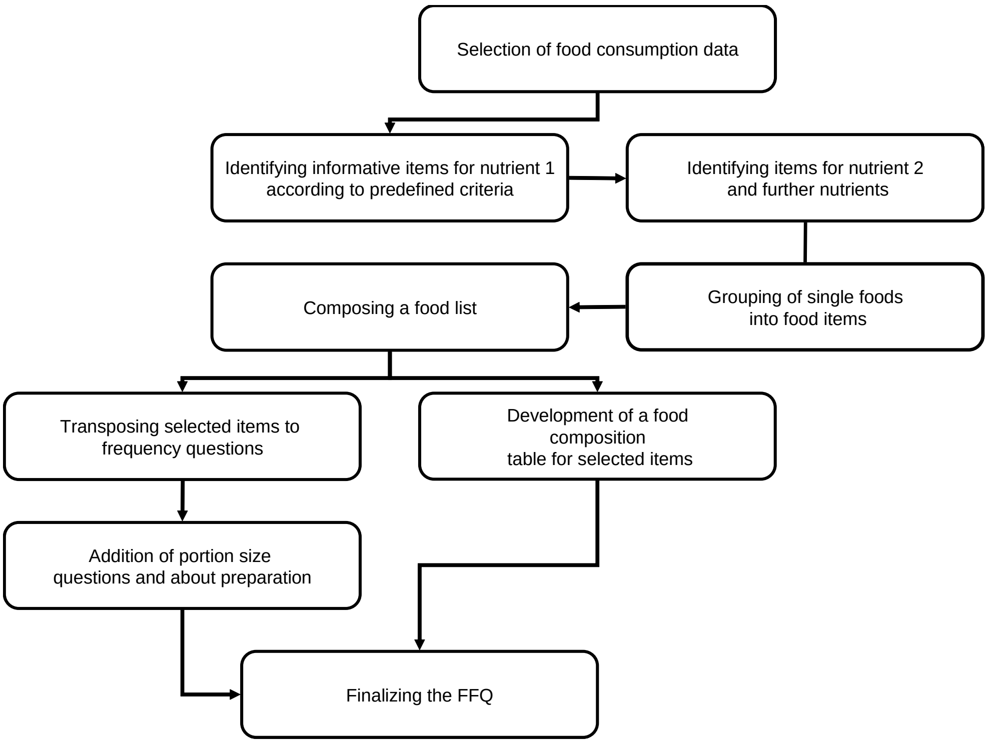


Figure 5.1 Overview of the *Dutch FFQ-TOOL™* to develop and process FFQs.

First, we explored which processes had to be automated to develop and process an FFQ. The identified processes were: selection of food consumption data of the target population, identification of important foods for all nutrients of interest, aggregation of these foods to food items, composition of a food list from these food items, transposition of the food list into standard questions with answering categories for frequencies, portion sizes and preparation methods, quality control of completed FFQs, construction of a nutrient database for processing of completed FFQs, calculation of nutrient intakes, and finally quality control of processed FFQs (see figure 5.1). These processes needed to fulfill several requirements from a scientific, respondent's and technical perspective. We identified the following requirements:

1. *Scientific requirements*
    a. Inclusion of representative food consumption data sampled with valid methods in a population closely resembling the target population.
    b. Selection of foods by valid and reproducible procedures from the food consumption data, based on variance in intake for ranking of respondents according to their intake or on contribution to population intake per nutrient of interest.
    c. Composition of a food list for multiple nutrients at the most informative aggregation level of food items.
    d. Evaluation of the food list by computing the percentage of explained variance or percentage contribution in the food consumption data.
    e. Estimation of consumed amounts based on portion size questions that incorporate serving units and amounts representing the range in amounts consumed by the target population.
    f. Evaluation of completed FFQs by checks in the *Dutch FFQ-TOOL*™ indicating missing values, and exceeding of predefined minima and maxima per item and per FFQ, allowing the researcher to mutate these values.
    g. Computation of nutrient intake in the *Dutch FFQ-TOOL*™ for the selected food items using a valid and complete national food composition database.

2. *Requirements from a respondent's perspective*
    a. Recognition of food items.
    b. Clear, unambiguous and simple questions.
    c. Indication of consumption frequencies in the FFQ using categories that represent the frequency range in the population.

d. Indication of consumed amounts using portion size question that are understood and easily filled out.
e. Limitation of the length of the FFQ to achieve maximal response rates.
f. Adaptation of lay-out of FFQs to the target population.
g. Face validity of FFQs.

3. *Technical requirements*
   a. Automation of food selection from food consumption data at acceptable speed in the system.
   b. Addition of foods must be possible if the initial selection of food items does not meet the required percentage of explained variance or coverage of the level of intake. It must also be possible to add extra foods to increase face validity or new foods that are not included in the food consumption data.
   c. Transposition of selected food items into standard questions.
   d. Possibility to create exceptions to standard questions if needed.
   e. Computation of nutrient intake in the *Dutch FFQ-TOOL™* for food items in the FFQ of which consumption is reported by respondents.
   f. Providing a clear overview of the different steps involved in composing the FFQ, thus providing a user-friendly system for researchers developing FFQs.
   g. Possibility to update food consumption data as well as food composition data, with flexibility in format of new databases.

METHODS

Structure and contents of the Dutch FFQ-TOOL™

The *Dutch FFQ-TOOL™* is an internet-based system for researchers devising new FFQs. The system consists of two parts: 'selection of food items', and 'questions generation' in which selected foods are transposed into standard questions and the FFQ is further finalized. The system makes use of flexible relational databases. For an overview of the databases and software included in the *Dutch FFQ-TOOL™*, see figure 5.2.

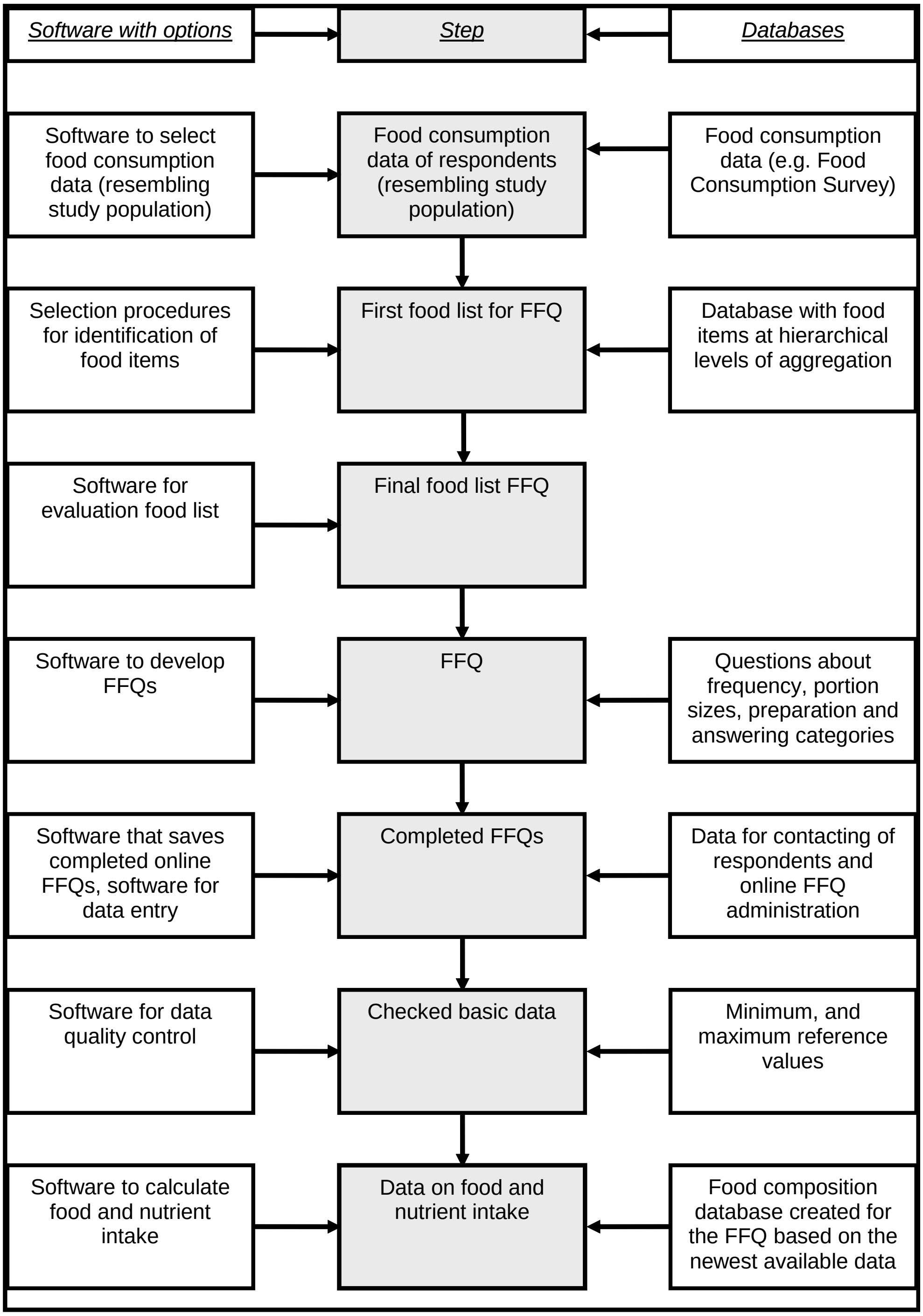


Figure 5.2 Software and databases in the *Dutch FFQ-TOOL*™ to generate and process FFQs.

Selection of food items

Selection of food consumption data

For the selection of food items, the system uses food consumption data and food composition tables. In the prototype of the *Dutch FFQ-TOOL™*, we used the Dutch National Food Consumption Survey 1998 [14] as a food intake database. This database consists of estimated food records filled out on two consecutive days in 1997/1998 by 6250 individuals aged between 1 and 97 years. The Dutch Food Composition Table of 1996 was used for nutrient computations of the Dutch National Food Consumption Survey 1998. We used the1998 database because it is the most extensive and recent database covering all age groups that was available to us.

Identifying informative items

Two procedures have been included in the *Dutch FFQ-TOOL™* for the selection of foods from food consumption data. For use of FFQs in epidemiologic studies, foods can be selected based on explaining variance in intake. Regression analysis is very suitable to select food items for this purpose, but overloads the computer system, because large databases are used and regression analysis tests all possible combinations of food items [15]. Therefore, we incorporated a simple procedure that selected food items based on variance in intake, *MOM2*, testing only one combination of food items, without taking other co-varying items into account [15] (chapter 3) (requirement 1b). If the FFQ is used to estimate the absolute level of intake, foods are selected based on their contribution to the level of nutrient intake in a population [7] (requirement 1b).

Grouping of single foods into food items

As the above mentioned selection procedures select foods at the level of single foods, as represented by food codes in a food composition table, related foods must be combined into food items after the procedure. We developed standardized hierarchical levels of food groups for use in the computer system. Foods in the Dutch food composition table are combined into 24 food groups such as ‘bread’, ‘fruit’ and ‘vegetables’ that comprise too many foods to use as items in an FFQ. Therefore, the 24 groups, regarded as hierarchical level one, were further subdivided into smaller food groups at four hierarchical levels of aggregation. This was done by two dieticians who based their choices on similarity in eating occasions, portion sizes, and nutrient contents. The dieticians took also into account whether respondents could possibly answer questions about usual consumption of each food item (requirement 2a). Their choices resulted in 87, 237, and 356 food items at aggregation levels two, three and four, respectively. An illustration of this subdivision is given for fresh fruit in table 5.1.

Table 5.1 Example of food items that explain variance in vitamin C and dietary fiber intake for 'Fresh fruit' in the food group 'Fruit'

Level 1 'Food group in Food composition table'	Level 2 'Aggregated food items'	Variance Vitamin C (%)	Variance Dietary fiber (%)	Level 3 'In between'	Variance Vitamin C (%)	Variance Dietary fiber (%)	Level 4 'Detailed food items'	Variance Vitamin C (%)	Variance Dietary fiber (%)
Fruit	Fresh fruit	34.3	10.9	Citrus fruit	21.1	1.1	Orange	17.1	0.74
							Tangerine	2.2	0.28
							Grapefruit	1.25	0.05
							Other	0.3	0
				Non-citrus fruit	10.9	6.5	Apple	0.7	2.65
							Strawberries	4.5	0.58
							Kiwi	3.4	0.10
							Banana	0.48	1.11
							Pear	0.08	0.72
							Grapes	0.01	0.26
							Peaches	0.02	0.06
							Other	0.78	0.22

Question generation

To compose a food list for an FFQ, a choice must be made on the most appropriate level of aggregation for each food item, depending on the level of detail required for a study and the nutrients of interest. After deciding on the level of aggregation, selected food items are automatically transposed into standard questions. For this, several supportive tables have been compiled and incorporated. These tables contain information needed for automated generation of questions, such as standard frequency questions, standardized portion size questions, questions about preparation of foods, answering categories, usual daily order of consuming food items, serving units, and serving sizes per unit of food items. Respondents need clear questions, because it improves accuracy of FFQs [16]. Therefore, we used findings from cognitive research and dietetic expertise to develop clear questions [17,18] (chapter 4) (requirement 2b). Questions were formatted according to a nested approach [17]. In this approach, frequency of intake is first assessed, followed by more specific questions about number of portions, portion sizes, and preparation methods. Default answering categories for frequencies started at 'never' and then ranged from 'once a year' to 'six to seven times per week' (requirement 2c). Questions may be clarified by including cues that stimulate memory of respondents or remind them of consumed food items [18]. An example of a cue may be 'think about consumption for breakfast, lunch, dinner and as in-between snack'. In the *Dutch FFQ-TOOL™*, we did not include standard cues for specific food items.

The researcher developing the FFQ decides which reference period is most suitable for an FFQ depending on the aim of the study and the nutrients of interest. A reference period of one month is suitable for assessing macronutrient intake, whereas one year is more suitable for assessing seasonally consumed food items and micronutrients [17]. Moreover, it is even possible to include a few questions with a different reference period, for example to assess seasonally consumed food items.

Addition of portion size questions

For reporting portions, it is easiest to define consumed portions in ranges of household measures such as one cup or two cups [17] or in units such as a slice of bread. Thus, we included this possibility in the *Dutch FFQ-TOOL™* (requirement 2d) and portion sizes, were based on common Dutch portion sizes [19] (requirement 1e). We did not include pictures for the estimation of portion sizes. Ranking of respondents hardly improved when portion sizes were estimated using pictures compared to standard portions [20-22]. It was also possible to add questions to adjust portion sizes, for example to indicate whether the amount of margarine spread on a slice of bread was large or small.

Development of a food composition table and processing of FFQs

Processing of completed FFQs is also automated in the *Dutch FFQ-TOOL™*, based on the selected food items. The nutrient composition of selected food items is derived by computing the weighted mean of amounts of the single foods of which the food items are composed using the Dutch Food Composition Table 2006 [23] (requirement 1g). In addition, minima and maxima of consumed amounts and missing values of completed FFQs are shown to the researcher who evaluates these values and adapts the data were needed (requirement 1f). Thus, procedures for quality control were also semi-automated and included in the *Dutch FFQ-TOOL™*.

To illustrate how tables and procedures can be used to develop an FFQ, we here describe an example of generating an FFQ for assessing vitamin C and dietary fiber intake, and we evaluate whether the requirements are met.

RESULTS

Illustration of development of an FFQ

In this example, we describe the development of an FFQ that aims to rank adults between 25 and 65 years according to their vitamin C and dietary fiber intake. First, food intake data of individuals between 25 and 65 years was selected from the Dutch National Food Consumption Survey 1998, as they resembled our target population (requirement 1a). This resulted in a database that contained recorded dietary intake of two days by 3524 individuals.

Identifying food items

We started with the selection of food items, and set a criterion to explain at least 80% of variance in vitamin C and dietary fiber intake. First, we selected food items for assessing vitamin C, because 80% of variance for this nutrient is explained by a smaller number of food items than dietary fiber. We started selecting food items at the highly aggregated level two, because this results in a short food list. The food items 'fresh fruit', 'fruit and vegetable juice' and 'cooked or fried vegetables' explained 85% of variance in vitamin C intake, thereby meeting the criterion. For dietary fiber the food items: 'Bread', 'Cooked potatoes', 'Fruit', 'Cooked or fried vegetables', 'Soup', 'Pulses', 'French fries', 'Peanuts, nuts and seeds', and 'Crisps and salty biscuits' explained 81% of variance in dietary fiber intake.

Composing a food list

We decided for each of the above mentioned food items whether it had to be assessed at a more detailed level of aggregation (requirement 1c). As an example, we describe this further process for one food item, 'fresh fruit', as it was selected for both vitamin C and dietary fiber. We based our choice for the right level of aggregation on the values of explained variance for vitamin C and dietary fiber provided by the system (table 5.1). Variation in vitamin C was mainly derived from 'citrus fruit' in the food item 'fresh fruit'. Within 'citrus fruit' both vitamin C and dietary fiber were mainly explained by 'oranges'. Therefore, it was thought to be most informative to include 'oranges' and a new food item 'other types of citrus fruit' in the food list. From the food item 'non-citrus fruit' the food items 'strawberries' and 'kiwis' explained most variance in vitamin C intake, whereas for dietary fiber 'apples' and 'bananas' were most important. Therefore, 'strawberries', 'kiwis', 'apples', 'bananas' and a new food group that contained 'other types of non-citrus fruit' were included in the food list. An initial food list was composed after repeating decisions for all food items at aggregation level two and this resulted in an initial food list containing 30 food items.

The initial food list was evaluated in the *Dutch FFQ-TOOL™* by computing explained variance using regression analysis for the 30 selected food items. According to regression analyses, the selected food items explained 88 percent of variance for vitamin C and 78 percent for dietary fiber in the Dutch National Food Consumption Survey 1998 (requirements 1d). As the explained variance of dietary fiber did not meet the criterion of 80%, we added three food items which increased explained variance to 81% (requirement 3b).

Question generation

Subsequently, selected food items were automatically transposed into standard predefined frequency questions and a draft FFQ was generated in the *Dutch FFQ-TOOL™* (requirement 3c and 2b). We chose a reference period of one year for this FFQ because vitamin C consumption is known to have a large day to day as well as seasonal variation. Adaptations to improve comprehensiveness were made (requirement 3d). For example, the question about 'strawberries' was restricted to the summer only. Then the FFQ was finalized and lay-out of the FFQ was adapted to the target population (requirement 2f).

DISCUSSION

We described a prototype of a flexible data-based system, the '*Dutch FFQ-TOOL™*', in which development of FFQs was automated from food consumption data. Compared to the Vofrex/Vovris system and Windiets [9], a major improvement of the *Dutch FFQ-TOOL™* is its ability to select food items that explain variance in nutrient intake and increased flexibility to include new databases and select subpopulations. Other major advantages of the *Dutch FFQ-TOOL™* include standardization of procedures and increased transparency of selection of food items. Researchers may use this system to develop and process tailored FFQs in a standardized way. We illustrated this by giving an example of development of an FFQ by the system. The results of this process showed that almost all predefined requirements for the system were met.

Although fully automating selection of food items was an important requirement, we were unable to achieve this. The reason was that it appeared to be difficult to automatically select the most informative level of aggregation of a food item. The most optimal choice depends on the level of detail required for the aim of a study and the nutrients of interest. Therefore, we created a screen in the *Dutch FFQ-TOOL™* that shows values of variance in intake or percentage contribution for food items for all nutrients of interest at different levels of aggregation. The researcher, developing the FFQ, can use this information and expertise to decide which level of aggregation is most appropriate. In a next step, the selected food items are evaluated in the system by computing the percentage of explained variance or percentage contribution of the selected food items in the food consumption data. In future, we will try to further automate selection and thus increase transparency by exploring optimization techniques for the selection of food items on the most informative aggregation level.

A challenge during development of FFQs is balancing requirements that might be contradictory. An example is that generating standard questions may conflict with the requirement of developing clear questions. Some food items require questions that are not standard. For example sugar should not be assessed in the predefined standard question: 'How often did you eat sugar?', but it must be assessed as 'How often did you use sugar in your coffee or tea'. We solved this problem by providing the possibility to adapt standard questions where needed.

A current problem of the prototype of *Dutch FFQ-TOOL™* is that until now only the Dutch Food Consumption Survey 1997/1998 has been included. To generate updated FFQs, databases with more recent food consumption databases are needed. Another limitation of the dataset used in our study was that only two subsequent days of intake were collected. Thus, for foods infrequently eaten, such as fish, it is difficult to estimate true between-person variation in intake. Ideally, we would have included more recent food consumption data including more reports per person, but this information was not yet available. Recently, a (semi-)continuous food consumption surveillance has started pertaining to the general population aged 4 to 69 years, which consists of two non-consecutive 24-hour dietary recalls in combination with a self-administered food propensity questionnaire [24]. This data will be used to update food consumption data in the *Dutch FFQ-TOOL™*, and will thus solve for the greater part the above described problems.

After development, the FFQ needs to be validated in the target population. We have validated a first FFQ developed by the system (chapter 6). Validity of this FFQ was very comparable to other FFQs [25]. However, for practical and financial reasons it will be impossible to validate each new FFQ generated by the system. Therefore, it would be very helpful to have a first indication of the validity provided by the construct validity of the newly developed FFQ. Construct validity is defined as the degree to which the data collected reflect or measure the variable of interest [3]. Thus, in the *Dutch FFQ-TOOL™*, first estimates of the construct validity of FFQs are provided by explained variance and level of intake covered by the selected food items [6]. Thus, we urge researchers using FFQs developed in the system to provide this information in their papers.

Adapting the Dutch FFQ-TOOL™ for use abroad

An interesting question is whether the *Dutch FFQ-TOOL™* could be adapted for use in other countries than the Netherlands. The *Dutch FFQ-TOOL™* makes use of a Dutch food consumption database, Dutch food composition tables, and questions and answering categories based on a Dutch food pattern. This means that the software with options would be suitable (see left side figure 5.2) whereas the databases should be replaced by country-specific data (see right side figure 5.2). Thus, although it would be basically possible to adapt the *Dutch FFQ-TOOL™* for use in other countries than the Netherlands, it depends on the availability of databases tailored for the population of interest with regard to food habits and language whether the system can be used.

Further development of the Dutch FFQ-TOOL™

This paper described a prototype of the *Dutch FFQ-TOOL™* which is able to develop and process FFQs, but still needs fine-tuning to increase clarity, user-friendliness, and speed. To further increase transparency, selection of food items at different hierarchical levels of aggregation and for multiple nutrients of interest needs to be further optimized. Also, developing FFQs for different target populations than adults is possible if appropriate food consumption data is imported, although a few adaptations may be necessary to standard questions and answering categories.

Future work will involve developing web-based FFQs because these are more convenient for respondents, cause less misreporting, and considerably reduce processing time of

FFQs. Important advantages of web-based FFQs over paper versions include increased comprehension by including tailored pop-up. Also, web-based FFQs may include pictures to improve identification of foods and estimation of portion sizes [26-28].

We conclude that the prototype of the flexible data-based *Dutch FFQ-TOOL™* is able to generate, apply and process tailored FFQs for multiple nutrients and by evidence-based procedures. Importantly, the software of the *Dutch FFQ-TOOL™* simultaneously generates a database for processing of completed FFQs. As future FFQs can be filled out online as a web-based version, processing time of the FFQs is considerably reduced.

ACKNOWLEDGEMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

REFERENCE LIST

1. Bingham SA, Luben R, Welch A, Wareham N, Khaw KT, Day N. Are imprecise methods obscuring a relation between fat and breast cancer? Lancet 2003;362(9379):212-4.
2. Kristal AR, Peters U, Potter JD. Is it time to abandon the food frequency questionnaire? Cancer Epidemiol Biomarkers Prev 2005;14(12):2826-8.
3. Kristal AR, Potter JD. Not the time to abandon the food frequency questionnaire: counterpoint. Cancer Epidemiol Biomarkers Prev 2006;15(10):1759-60.
4. Freedman LS, Schatzkin A, Thiebaut ACM, Potischman N, Subar AF, Thompson FE, Kipnis V. Abandon neither the food frequency questionnaire nor the dietary fat-breast cancer hypothesis [1]. Cancer Epidemiol Biomarkers Prev 2007;16(6):1321-2.
5. McNeill G, Masson L, Macdonald H, Haggarty P, Macdiarmid J, Craig L, Kyle J. Food frequency questionnaires vs diet diaries. Int J Epidemiol 2009;38(3):884.
6. Willett W. Nutritional Epidemiology. second edition ed. New York, NY: Oxford University Press, 1998.
7. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124(3):453-69.
8. Subar AF. Developing dietary assessment tools. J Am Diet Assoc 2004;104(5):769-70.
9. Wise A, Birrell NM. Design and analysis of food frequency questionnaires--review and novel method. Int J Food Sci Nutr 2002;53(3):273-9.
10. Brants HAM, Bouman M. Voedsel Frequentie Expert, de ontwikkeling van een geautomatiseerd systeem voor het maken van voedselfrequentievragenlijsten. in Dutch; Food frequency expert, the development of an automated system for the development of FFQs.. Deel 4: De gebruikershandleiding; in Dutch Part 4: The manual. Zeist: TNO Nutrition, 1998.
11. Kistemaker C. VOVRIS - Voedselfrequentie Vragenlijst Interaktief Verwerkings Systeem; in Dutch; Interactive Processing System for Food Frequency Questionnaires. In: Nutrition T, ed. Een toelichting op het gebruik. Zeist: TNO Nutrition, 1991.
12. Zo eet Nederland, 1992: resultaten van de Voedselconsumptiepeiling 1992. in Dutch; Results of the Dutch National Food Consumption Survey 1992. Voorlichtingsbureau voor de Voeding. Den Haag, 1993.
13. Willett WC, Reynolds RD, Cottrell Hoehner S, Sampson L, Browne ML. Validation of a semi-quantitative food frequency questionnaire: comparison with a 1-year diet record. J Am Diet Assoc 1987;87(1):43-7.
14. Rapportage voedselconsumptiepeilingen VCP-3: 10 jaar trend; in Dutch; Results of the Dutch National Food Consumption Survey 3; 10 year trends Zeist: TNO Voeding, 1998.
15. Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. Am J Epidemiol 1996;143(5):514-21.
16. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, Mittl B, Gibson JT, Ziegler RG. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102(2):212-25.
17. Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, Schatzkin A, Hartman A, Swanson C, Kruse L, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95(7):781-8.

18. Matt GE, Rock CL, Johnson Kozlow M. Using recall cues to improve measurement of dietary intakes with a food frequency questionnaire in an ethnically diverse population: an exploratory study. J Am Diet Assoc 2006;106(8):1209-17.

19. Donders-Engelen M, Van Der Heyden L, Hulshof K. Maten en Gewichten en Codenummers 1992; In Dutch; Food portion sizes and coding instructions. In: Donders-Engelen, Heyden VD, Hulshof, eds. Wageningen: Wageningen University and TNO Nutrition, Zeist, 2003.

20. Ocke MC, Bueno de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.

21. Haraldsdottir J, Tjonneland A, Overvad K. Validity of individual portion size estimates in a food frequency questionnaire. Int J Epidemiol 1994;23(4):786-96.

22. Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. Int J Epidemiol 1992;21(4):770-7.

23. NEVO. Nederlands Voedingsmiddelentabel; in Dutch; Dutch Food Composition Table. 2006.

24. Ocké M, Hulshof KF. Naar een nieuw Nederlands voedingspeilingsysteem; in Dutch; Towards a new system of dietary monitoring in the Netherlands. Voeding Nu 2005;12:26-8.

25. Molag ML, de Vries JH, Ocke MC, Dagnelie PC, van den Brandt PA, Jansen MC, van Staveren WA, van't Veer P. Design characteristics of food frequency questionnaires in relation to their validity. Am J Epidemiol 2007;166(12):1468-78.

26. Kumanyika SK, Tell GS, Shemanski L, Martel J, Chinchilli VM. Dietary assessment using a picture-sort approach. Am J Clin Nutr 1997;65(4 Suppl):1123s-9s.

27. Wengreen HJ, Munger RG, Wong SS, West NA, Cutler R. Comparison of a picture-sort food-frequency questionnaire with 24-hour dietary recalls in an elderly Utah population. Public Health Nutr 2001;4(5):961-70.

28. Ocke MC, Bueno de Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol 1997;26 Suppl 1:S37-48.

6

Validation of a food frequency questionnaire developed with a new tool for automated development and processing of FFQs against biomarkers and 3-d food records

Marja L. Molag[1], Pieter van `t Veer[1], Marga C. Ocké[2], Saskia Meijboom[1], Paul J.M. Hulshof[1], R. Alexandra Goldbohm[3], Pieter C. Dagnelie[4], Jeanne H.M. de Vries[1]

[Submitted]

[1] Division of Human Nutrition, Wageningen University, Wageningen, the Netherlands

[2] National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands.

[3] TNO Quality of Life, Leiden, the Netherlands

[4] Department of Epidemiology, Maastricht University, Maastricht, the Netherlands

ABSTRACT

Introduction

We developed a food frequency questionnaire (FFQ) using a computer system (the *Dutch FFQ-TOOL*^{TM}) that generates and processes FFQs tailored for research questions and populations.

Objective

To validate an FFQ that aimed to rank participants according to their intake of energy and selected nutrients developed with this system.

Methods

An FFQ of 118 food items was generated. Intakes according to FFQs completed by 46 men and 63 women of 25-65 y were compared with those by 3-d estimated food records, by nitrogen and potassium excretions in 24-hour urine checked with PABA, and by concentrations of polyunsaturated fatty acids (PUFA) determined in serum cholesterylesters. We applied the methods of triads to estimate validity coefficients between FFQ and true intake for protein, potassium and PUFA.

Results

Deattenuated correlation coefficients between nutrient intake by FFQ and food records ranged from 0.35 for PUFA to 0.73 for carbohydrates. Lower limits of the validity coefficient, estimated by correlation coefficients between protein, potassium, and PUFA intakes by FFQ with nitrogen and potassium in urine and PUFA in serum were 0.24, 0.26 and 0.33 for men, and 0.08, -0.05 and 0.27 for women, respectively. Upper limits of the validity coefficient for the total population were estimated according to the method of triads and were 0.50, 0.36 and 0.50 for protein, potassium and PUFA, respectively.

Conclusion

The FFQ developed by the *Dutch FFQ-TOOL*^{TM} performed rather well, although improvement of generated FFQs is possible by selecting food items explaining a higher level of variance and including new foods.

INTRODUCTION

Food frequency questionnaires (FFQs) are often used in epidemiological studies to assess dietary intake of populations. A well-designed FFQ is targeted to assess habitual intake of foods or nutrients of interest in a specific population. This implies that the food list of an FFQ has to be adapted to food consumption habits of the target population and updated, with new foods at the market, when re-used some time after initial development [1]. Unfortunately, FFQs are often re-used without adaptations due to lack of expertise, time or finances [2]. To make the development of new FFQs easier and more standardised, we developed the so called '*Dutch FFQ-TOOL™*'. This is a prototype of a computer system that can efficiently generate and process FFQs tailored for a specific research questions and subgroup of interest of the Dutch population.

Basically an FFQ consists of a food list enumerating the most informative food items for a study. Foods on the food list must be consumed regularly, contribute substantially to the nutrient(s) of interest, and be able to rank individuals according to their intake, i.e. to discriminate intakes between persons [3]. Two procedures are often used to select informative food items for the FFQ. The first procedure selects food items on the basis of their percentage contribution to nutrient intake of a population from open-ended food consumption data [4]. This selection procedure is simple and suitable if the purpose of the FFQ is to estimate the absolute level of intake of a population. A second procedure uses forward regression analyses to predict the food items that explain most variance in nutrient intake taking covariance between nutrient intakes of food items into account [3]. This procedure also uses food consumption data and is preferred if the FFQ is intended to be used for ranking or classifying individuals according to their intake e.g. in epidemiological studies.

In the *Dutch FFQ-TOOL™*, informative food items may be identified by both procedures. For the procedure identifying food items based on explaining variance in nutrient intake, we use a procedure called *MOM2*, which does not take covariance of nutrient intakes from other items into account [5], as regression analyses does. Our selection procedures use open-ended food consumption data, derived from 24-hour dietary recalls or food records, of the relevant population. Subsequently, an FFQ is constructed by transposing the selected most informative food items into predefined questions. Finally, nutrient compositions of filled out FFQs are computed within the *Dutch FFQ-TOOL™* using the Dutch food composition table.

To test a prototype of the *Dutch FFQ-TOOL™*, we constructed a first FFQ for an adult population and validated it for its ability to rank adults according to their intake of energy, macronutrients, fatty acids, potassium, and dietary fibre by comparison with biomarkers of protein, potassium and fatty acid intake and a 3-day estimated food record.

METHODS

Participants

Men and women between 25 and 65 y were recruited in Wageningen and its surroundings (the Netherlands), via flyers, through advertisements in local papers or by email. We included subjects that were able to read and write Dutch. Exclusion criteria were: employment at the division of Human Nutrition of Wageningen University, pregnancy, lactation, participating in other studies, donating blood at the blood bank, illness, diabetes, renal failure, oversensitivity to sulphonamides, or using antibiotics containing sulphonamides.

All participants gave written informed consent. The study was approved by the Medical Ethics Committee of Wageningen University and conducted in accordance with the principles of the Declaration of Helsinki.

Study design

The study was conducted from February until April 2008. Participants filled out a three day food diary, followed after two to nine weeks by the FFQ All participants collected one 24-hour urine sample on a day assigned randomly between February and April, and underwent a venipuncture just before filling out the FFQ.

Measurements

Participants were instructed about filling out the FFQ, the 3-day food record, and collection of the 24-hour urine sample at a meeting just before the first measurements started.

Dietary assessment

Development of the FFQ Using the prototype of the *Dutch FFQ-TOOL™*, we identified food items that explained at least 80% of variance in intake of energy, macronutrients, fatty acids, potassium and dietary fibre of a representative sample of the Dutch population according to data from the Dutch National Food Consumption Survey 1998 [6]. These data comprised estimated food records filled out on two consecutive days in 1997/1998 by 3524 adults aged between 25 and 65 y. Nutrient intakes were computed using the Dutch food composition database of 1996 [7]. Because food intake data had been recorded as single foods, we developed standardized food groups at three hierarchical levels of aggregation based on similarity in nutrient contents, portion sizes, and eating occasions. A final food list was composed by combining food items identified for each nutrient of interest and choosing the best level of aggregation for each food item based on explained variance and experience. After compilation, the final food list consisted of 118 items and was evaluated by regression analysis. The food list accounted for 87% of explained variance for dietary fibre up to 93% for fatty acids, according to regression analysis. With respect to the level of intake [4], selected food items covered 70% of protein intake, 77% of energy intake and at least 80% of intake of all other nutrients of interest, as computed by the *Dutch FFQ-TOOL™*.

The food frequency questionnaire The 118-item FFQ was self-administered and developed to assess habitual intake during the past month. The questionnaire was structured by a meal-based pattern. The questions were organized according to a nested approach that is cognitively easier for respondents [8]. First, frequency of food intake was assessed in categories starting at 'never' and then ranging from 'once a month' to 'six to seven times per week'. This was followed by more specific questions about portion sizes and preparation methods.

To evaluate the completeness and user-friendliness of the FFQ by respondents, we asked the respondents to complete an evaluation questionnaire, which included questions about use of ready-to-eat meals, supplements and any food items that respondents had missed in the FFQ. Nutrient intakes of FFQ-reports were calculated using the Dutch food composition database of 2006 [9]. The nutrient composition of each food item was derived by computing the weighted mean of consumed amounts of the single foods of which the food item consisted, the weighted mean as reported in the Dutch Food Consumption Survey [6].

Food records

For food records, participants were asked to record each food or drink immediately after consumption, describing the type of food consumed in as much detail as possible and the amount consumed in household units or in estimated weight, and to record recipes of composite dishes. Food records were structured according to breakfast, lunch, dinner and three in-between meals. The week and days of recording were assigned at random. For the total group, all days of the week were equally represented. After completion, the food records were checked for clarity and completeness by a trained dietician. Models of cups, glasses and sandwiches were used to let participants indicate their usual serving sizes. Again, nutrient intakes from food records were calculated using the Dutch food composition database 2006 [9].

Biomarkers

24-hour urine collection

Each participant collected one 24-hour urine sample on a day assigned at random on Sundays to Thursdays, as it was not feasible to immediately process urines collected on Fridays and Saturdays. Participants received a safety pin to attach to their underclothing as a reminder, two 2-liter containers that contained 3 grams boric acid as a preservative, and a small bottle and for women urination funnels. To evaluate the completeness of the 24-hour urine collections, participants were given three 80-mg tablets of PABA to take at mealtimes on the day of the urine collection [10-12]. Participants were asked to discard their first urine specimen on the collection day and subsequently collect all specimens for 24-hours including the first morning specimen the next day. During the urine collection period, participants recorded in a diary: the time of beginning and ending of collections, time that the PABA tablets were taken, lost specimens if any and medications or supplements taken. Within one day of sampling, all 24-hour urine collections were weighed and 10 ml aliquots were stored at -20°C.

Venipuncture

Participants were allowed to take a low-fat breakfast in the morning before the blood sampling. From each participant, a trained lab technician drew a (non-fasting) blood sample (9 ml) from the antecubital vein. Blood was collected in vacuum serum tubes (Terumo, venosafe), clotted for 45 minutes at room temperature (20-22˚C) and centrifuged for 15 min at 1200 g at 4˚C. Serum samples from each participant (~ 3 ml) were aliquoted into two cryo-tubes for storage at -80°C until analysis.

Chemical analyses

PABA

For analyses, aliquot urine samples were thawed and mixed. PABA recovery was measured colorimetrically [10] at the Laboratory of Division of Human Nutrition of Wageningen University, The Netherlands. The analytical variation for this assay was <4%. Urine samples that recovered <50% of the PABA were excluded from analysis as they were considered incomplete [11]. Samples containing 50-85 percent PABA were also considered incomplete, but the content of analytes was proportionally adjusted to 93 percent PABA recovery [13]. Urines with more than 120% recovery of PABA were excluded from analysis as this indicated interference of the colorimetric assay with other aromatic amines [11].

Nitrogen

Nitrogen was determined with the Kjeldahl method using a semi automated analysis system (Tecator, Höganäs, Sweden). The analytical variation for this assay was <4%. To convert nitrogen excretion into protein intake the following formula was used:

(Urinary nitrogen/0.81) * 6.25 [14].

Potassium

Potassium was measured using an ion-selective electrode on a Beckman Synchron LX20 analyzer. Analytical variation was <1%. Potassium excretion was adjusted for faecal excretion using the following formula:

Potassium excretion = urinary potassium / 0.77 (15).

Fatty acids in serum

Serum samples were thawed and mixed. Fatty acid distributions (PUFA) were determined as previously described [16] using gas chromatography (Agilent Technologies, Palo Alto, CA, USA). This is considered as a good biomarker of PUFA intake during the previous 2-3 weeks [16]. Fatty acids were identified by comparison with known standards. CV for components present at level >5% were ~3%. For components present at levels <5%, CV was 3-5%.

Anthropometry

Height and weight were measured without coat and shoes and with empty pockets when participants visited the department for venipuncture. BMI of each participant was computed by dividing weight (kg) by squared height (m²).

Statistical analysis

Intake data for energy as well as most nutrients and biomarkers were normally distributed. Exceptions were fatty acids, mono- and disaccharides, and total carbohydrates, and these were log-transformed. Random errors were studied by analyzing the association of participants according to nutrient intakes between FFQ and food record. To account for within-person variation in nutrient intake, deattenuated correlation coefficients were calculated using the following formula:

$$r_t = r_0 \left\{ 1 / \left[1 + \left(S^2{}_w / S^2{}_b \right) / k \right] \right\}^{0.5}$$

where r_t is the deattenuated Pearson correlation coefficient, r_0 is the observed correlation coefficient, ($S^2{}_w$ / $S^2{}_b$) is the within-person variance divided by the between-person variance for a nutrient and k = 3, for recording food intake during 3 days [3]. In addition, energy-adjusted correlation coefficients according to the residuals method were calculated [17]. Confidence intervals for the observed correlations were calculated using Fisher's Z transformation [18].

For nutrients for which biological markers were determined (i.e. protein, fatty acids and potassium), the method of triads [19] was applied to calculate the validity coefficient between intakes by FFQ and the unknown 'true' dietary intake. For this analysis, the following formula was used:

$$r_t = \sqrt{r_{FFQ\text{-}Biomarker} * r_{FFQ\text{-}FoodRecord} / r_{Biomarker\text{-}FoodRecord}}$$

The estimate of r_t can be interpreted as the upper limit, because errors in estimates of intake by FFQ and food record are probably correlated [20], whereas the correlation coefficient between the FFQ and biomarker is considered as the lower limit of the true validity coefficient. Confidence intervals were estimated using bootstrap sampling where 200 samples of equal size (n = 105) were obtained by random sampling with replacement [21].

To evaluate the level of intake, mean protein and potassium excretions estimated from the biomarker were compared with mean protein and potassium intakes. These excretions provide a reliable marker of their respective intakes if a participant is in balance for these nutrients.

All analyses were performed in SAS, version 9.1 (SAS Institute Inc., Cary, N.C., USA).

RESULTS

Participants

We included 46 men, median age 60 [range 25-64] y, and 63 women median age 55 [25-65] y (see table 6.1). Ten of them had a low, 47 a medium and 53 a high educational level. Participants varied in BMI, ranging from 18 to 38 kg/m^2, with a median of 25 kg/m^2. All 109 participants filled out the FFQ and food records and donated blood. For analyses of urinary biomarkers, information of four participants was not available; two because they did not collect urine due to illness and two because their PABA recoveries were >120%. Concentrations of urinary biomarkers of two other participants, who had taken only two instead of three PABA tablets, were included because these participants reported to have collected their urine completely, but had forgotten to take one of the PABA-tablets.

Table 6.1 Characteristics of study participants

Characteristics	Males (n = 46)	Females (n = 63)	Total group (n = 109)
Median age [range] y	60 [25-64]	55 [25-65]	59 [25-65]
Weight (kg) *SD*	83.4 ± 11.6	69.3 ± 10.3	75.2 ± 12.9
Height (m) *SD*	1.80 ± 0.05	1.68 ± 0.08	1.73 ± 0.09
BMI kg/m^2			
<20 (n)	0	3	3
20-25 (n)	21	35	56
25-30 (n)	21	20	41
>30 (n)	4	5	9
Education			
Low (n)	7	3	10
Intermediate (n)	18	29	47
High (n)	21	31	52

Table 6.2 Crude, deattenuated and energy-adjusted and deattenuated correlation coefficients between nutrient intake by FFQ and by estimated 3-day food diaries in men and women

	Males (n = 46)			Females (n = 63)		
	Crude	Deattenuated	Energy-adjusted and deattenuated[a]	Crude	Deattenuated	Energy-adjusted and deattenuated[a]
Nutrient						
Energy	0.52 [0.28;0.71]	0.61 [0.39;0.77]		0.50 [0.31;0.68]	0.63 [0.45;0.76]	
Protein	0.41 [0.14;0.63]	0.55 [0.30;0.72]	0.47 [0.21;0.67]	0.45 [0.16-;.58]	0.50 [0.29;0.66]	0.62 [0.45;0.76]
Total fat[b]	0.51 [0.26;0.70]	0.66 [0.46;0.80]	0.44 [0.17;0.64]	0.44 [0.22;0.62]	0.56 [0.36;0.71]	0.41 [0.18;0.59]
Monounsaturated fatty acids	0.33 [0.04;0.56]	0.49 [0.23;0.68]	0.10 [-0.20;0.38]	0.35 [0.09;0.53]	0.45 [0.23;0.63]	0.40 [0.17;0.59]
Polyunsaturated fatty acids	0.41 [0.13;0.62]	0.57 [0.34;0.74]	0.49 [0.24;0.69]	0.26 [0.02;0.48]	0.35 [0.11;0.55]	0.34 [0.10;0.54]
Saturated fatty acids[b]	0.54 [0.30;0.72]	0.69 [0.50;0.82]	0.51 [0.26;0.70]	0.41 [0.18;0.60]	0.50 [0.29;0.67]	0.53 [0.32;0.69]
Total Carbohydrates [b]	0.66 [0.47;0.80]	0.73 [0.56;0.84]	0.54 [0.29;0.72]	0.62 [0.44;0.75]	0.70 [0.54;0.80]	0.36 [0.12;0.56]
Mono- and disaccharides[b]	0.62 [0.40;0.77]	0.69 [0.50;0.81]	0.35 [0.07;0.58]	0.51 [0.30;0.67]	0.56 [0.37;0.71]	0.51 [0.30;0.67]
Dietary fibre	0.40 [0.12;0.62]	0.46 [0.20;0.66]	0.41 [0.13;0.62]	0.49 [0.26;0.65]	0.54 [0.34;0.70]	0.61 [0.43;0.75]
Potassium	0.50 [0.24;0.69]	0.61 [0.39;0.77]	0.41 [0.13;0.62]	0.58 [0.32;0.69]	0.62 [0.44;0.75]	0.58 [0.39;0.72]

[a] Energy-adjusted according to the residuals method. [b] Nutrients were log-transformed because of skewed distributions.

Ranking according to nutrient intake

Most correlation coefficients between nutrient intake assessed by FFQ and food records were between 0.40 and 0.70 (table 6.2). Crude correlation coefficients ranged from 0.26 for PUFA in females to 0.66 for carbohydrates in men. After deattenuation, correlation coefficients increased and ranged from 0.35 for PUFA in females to 0.73 for carbohydrates in men. After both deattenuation and energy-adjustment, most correlation coefficients decreased (table 6.2), and ranged from 0.10 for MUFA in men to 0.62 for protein in women. In general, the lowest correlation coefficients were observed for fatty acids and dietary fibre, and the highest correlation coefficients for carbohydrates.

Correlation coefficients between the FFQ and their respective biomarkers were much lower. In men, correlations were 0.24 for protein, 0.26 for potassium and 0.33 for PUFA (table 6.3). For women these correlation coefficients were 0.08 for protein, -0.05 for potassium and 0.27 for PUFA. Correlation coefficients between biomarkers and food records were generally considerably higher than between biomarkers and FFQ. Partial correlations for the total group adjusted for sex were between correlations in men and those in women (table 6.3). Validity coefficients of the FFQ with true intake as estimated by the method of triads for the total group were 0.50 for protein, 0.36 for potassium and 0.50 for PUFA.

Table 6.3 Correlation coefficients between protein, potassium and polyunsaturated fatty acids (PUFA) intake by FFQ, food records and biomarkers, and their validity coefficients estimated by the method of triads

Nutrient sex (*n*)	$r_{FFQ\text{-}Biomarker}$ [CI]	$r_{FFQ\text{-}FoodRecord}$ [CI]	$r_{Biomarker\text{-}FoodRecord}$ [CI]	$r_{FFQ\text{-}True}$ [CI]
Protein men (*n* = 44)	0.24 [-0.06; 0.50]	0.40 [0.11; 0.62]	0.20 [-0.10; 0.47]	0.70 [0.20; 1.8]
Protein women (*n* = 61)	0.08 [-0.18; 0.33]	0.37 [0.13; 0.57]	0.17 [-0.09; 0.40]	0.41 [0.15; 1.5]
Protein total group (*n* = 105)	0.15 [-0.04; 0.33]	0.38 [0.20; 0.53]	0.18 [-0.01; 0.36]	0.56 [0.26; 0.72]
Potassium men (*n* = 44)	0.26 [-0.04; 0.51]	0.48 [0.21; 0.68]	0.43 [0.15; 0.64]	0.53 [0.15; 0.76]
Potassium women (*n* = 61)	-0.05 [-0.30; 0.20]	0.53 [0.32; 0.69]	0.32 [0.07; 0.53]	NA
Potassium total group (*n* = 105)	0.09 [-0.10; 0.28]	0.51 [0.35; 0.64]	0.38 [0.20; 0.53]	0.35 [0.11; 0.55]
PUFA men (*n* = 46)	0.33 [0.04; 0.57]	0.49 [0.23; 0.68]	0.44 [0.17; 0.65]	0.61 [0.10; 0.88]
PUFA women (*n* = 63)	0.27 [0.02; 0.49]	0.31 [0.07; 0.52]	0.37 [0.13; 0.57]	0.47 [0.18; 0.75]
PUFA total group (*N* = 109)	0.30 [0.11; 0.46]	0.38 [0.20; 0.53]	0.39 [0.21; 0.54]	0.54 [0.23; 0.72]

Abbreviations: FFQ, food-frequency questionnaire; $r_{FFQ\text{-}Biomarker}$, correlation coefficient between FFQ and biomarker; $r_{FFQ\text{-}FoodRecord}$, correlation coefficient between FFQ and food record; $r_{Biomarker\text{-}FoodRecord}$, correlation between biomarker and food record; $r_{FFQ\text{-}True}$, validity coefficient of the questionnaire; 95% CI, 95% confidence interval.

Absolute intakes

Estimates of mean nutrient intakes by FFQ compared to biomarkers were according to expectations. Based on computations by the *Dutch FFQ-TOOL™*, the FFQ was expected to cover 70% of total protein intake in the population. Estimates of protein intake by FFQ were on average 64% of biomarker excretion in men and 76% in women (table 6.4). For potassium we expected the FFQ to cover 84% of intake. Estimates of mean potassium intake by FFQ were 74% of potassium excretion for men and 83% for women.

Participants with a BMI >25 kg/m^2 underreported 29 g protein and 1144 mg potassium per day, more than the 21 g protein and 892 mg potassium per day for participants with BMI ≤25 kg/m^2.

Evaluation of the FFQ

Ready-to-eat meals could not be reported by the FFQ, because they were not selected from the Dutch National Food Consumption Survey 1998. However, in the evaluation questionnaire which was filled out by all participants, 40 of the 109 participants reported to have consumed these meals during the past month. Furthermore 20 participants indicated to have used vitamin and mineral supplements. Forty-two respondents reported to have missed other food items in the food list. These foods included crackers and toast [12], Dutch spiced cake [8], vegetarian meat substitutes [7], and specific types of fish [4], vegetables [5], salads [5], grains [5]; soy foods [4] and alcoholic drinks [4].

Table 6.4 Correlation coefficients between protein, potassium and polyunsaturated fatty acids (PUFA) intake by FFQ, food records and biomarkers, and their validity coefficients estimated by the method of triads

	FFQ (g)	FFQ relative to biomarker (%)	Food record (g)	Food record relative to biomarker (%)	Biomarker (g)
Men					
Protein[a]	65.7 ± 16.1	64	96.7 ± 17.6	94	102.7 ± 25.0
Potassium[b]	3.0 ± 0.65	74	4.19 ± 0.98	102	4.09 ± 1.33
Women					
Protein[a]	59.9 ± 17.5	76	79.7 ± 16.5	101	78.6 ± 19.1
Potassium[b]	2.88 ± 0.72	83	3.65 ± 0.83	105	3.47 ± 0.97

[a] Food items in the FFQ together contributed to 70% of total population intake of protein
[b] Food items in the FFQ together contributed to 84% of total population intake of potassium

DISCUSSION

We validated an FFQ generated by a prototype of the *Dutch FFQ-TOOL™*, a new computer system which generates and processes FFQs, for its value in ranking participants according to their nutrient intake. Correlation coefficients between nutrient intakes by FFQ and food records were moderate to high, although correlations between nutrient intake and biomarkers were lower. The true validity is expected to lie between the rather low correlation coefficients with biomarkers, and the moderate to high validity coefficients obtained by the method of triads estimating the relationship with true intake.

Evaluation of the FFQ

The food list of our FFQ explained at least 87% of explained variance of all nutrients of interest and the questionnaire performed better to assess energy and carbohydrates than most FFQs [22]. Energy-adjustments according to the residuals method [17] decreased correlation coefficients between intake according to FFQ and food records, as also observed in other studies [23-25]. The validity of the tested FFQ was lower than FFQs previously developed at the division of Human Nutrition of Wageningen University, The Netherlands [26,27]. An important difference between the present and previous FFQs is that the older FFQs included a larger number of food items which together contributed to 90-95% of total population intake. In addition, two other FFQs that performed very well as compared to urinary nitrogen contained 350 food items [28] or 178 items that covered at least 90% of contribution to population intake of each nutrient of interest [29]. This would suggest that the validity of the generated FFQ might be improved by including more food items that explain a higher percentage of variance.

Protein

Validity coefficients between protein intake by FFQ and true protein intake were moderate to high and similar to those obtained in other validation studies [21,30,31]. However, the confidence intervals were wide due to a small study population. The upper limit of the confidence interval, which was even above one, may be explained by random sampling fluctuations in the bootstrap samples and is known as a Heywood case [21].

The low correlation coefficients with biomarkers may be explained by not taking into account day-to-day variation in the biomarker. We collected only one urine sample per participant, whereas protein intake is known to have a large day-to-day variation in intake [3]. This may be accounted for by collecting multiple 24-hour urines. Repeated urine collections showed higher correlation coefficients in some studies [28,29,32], but were only slightly higher in other studies [33-35]. In addition, after deattenuation of correlation coefficients between FFQ and biomarker, using data from another study, our estimates improved at maximum by 0.05. Thus, day-to-day variation in urine biomarkers is probably not the only explanation for the low correlation coefficients observed in our study.

Another explanation may be a suboptimal estimation of protein intake by our FFQ. The consumption of some protein rich foods, such as meat, is difficult to report in an FFQ (36). Also, respondents indicated that they had missed some sources of protein in the FFQ such as some types of fish and meat substitutes. On the other hand, estimates of the mean group level of protein intake by FFQ were as expected. This discrepancy indicates that especially the between-subject variation in intake was not captured sufficiently by our FFQ. Another possible reason is that food habits may have altered over the ten years since the Dutch Food Consumption Survey was performed that was used as a basis to generate the questionnaire.

Potassium

Validity coefficients for potassium intake for the FFQ according to triad analyses were high in men, but we were unable to estimate the validity coefficient for women due to a negative correlation coefficient between intake by FFQ and biomarker. This negative correlation coefficient is also a Heywood case that is most likely explained by random sampling fluctuations [21]. A negative correlation indicates that urinary potassium is a poor indicator of dietary potassium, possibly because of considerable within person variance in potassium excretion [37]. Correlation coefficients between potassium estimated by FFQ and biomarker were weak for men, but similar to previous studies [35,38].

PUFA

The validity coefficients between the FFQ and true intake of PUFA according to triad analyses were moderate to high. An important difference with the other biomarkers is that for PUFA both FFQ and biomarker covered a longer reference period that was about the same for both methods. Correlation coefficients of FFQ intake and polyunsaturated fatty acids in serum were comparable to those reported in other validation studies [26,39].

Generalisation

Motivated individuals participated in this study. They had on average an intermediate to high educational level, although also lower educated individuals participated. Although one might expect that lower educated respondents underreport more frequently as filling out an FFQ is a complex cognitive task [24], we did not observe any differences in the level of underreporting by FFQ compared to food records between the different educational levels. However, participants with a higher BMI underreported more than those with a higher BMI, which was according to expectations [40-43].

Possible improvement of the Dutch FFQ-TOOL™

Based on this study it may be expected that the FFQs developed by the *Dutch FFQ-TOOL™*, may improve by including more recent food consumption data which are collected over more than two days. However, this kind of data is not always available. We used the Dutch Food Consumption Survey of 1998 because it was the only dataset of this size available to us at the time of the study. In future new Dutch food consumption databases will become available and used to update the *Dutch FFQ-TOOL™*. Therefore, it may be recommended to improve food lists developed with older datasets by adding newer foods, which are continuously introduced at the market, when similar food items or food groups are selected by the *Dutch FFQ-TOOL™*.

Conclusion

The performance of the present FFQ which was developed using a prototype of the *Dutch FFQ-TOOL™*, was comparable with other FFQs. The performance may be improved by including food items that help to explain a higher level of variance and by updating food lists with foods that are new at the market by using more recent databases.

ACKNOWLEDGEMENTS

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

The authors thank Karin Borgonjen, Frederike Brouwer and Marieke Heinst for practical assistance during the validation study.

REFERENCE LIST

1. Kristal AR, Feng Z, Coates RJ, Oberman A, George V. Associations of race/ethnicity, education, and dietary intervention with the validity and reliability of a food frequency questionnaire: The women's health trial feasibility study in minority populations. Am J Epidemiol 1997;146:856-69.
2. Subar AF. Developing dietary assessment tools. J Am Diet Assoc 2004;104:769-70.
3. Willett W. Nutritional Epidemiology. second edition ed: Oxford university press, 1998.
4. Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol 1986;124:453-69.
5. Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. Am J Epidemiol 1996;143:514-21.
6. The Dutch Nutrition Centre. Zo eet Nederland: Resultaten van de Voedselconsumptiepeiling 1997-1998. Results of the Dutch National Food Consumption Survey 1997-1998; in Dutch Voedingscentrum, Den Haag. Den Haag: Voedingscentrum, 1998.
7. NEVO. Nederlandse Voedingsmiddelentabel; in Dutch; Dutch Food Composition Table. ed, 1996.
8. Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95:781-8.
9. NEVO. Nederlands Voedingsmiddelentabel; in Dutch; Dutch Food Composition Table. ed, 2006.
10. Bingham S, Cummings JH. The use of 4-aminobenzoic acid as a marker to validate the completeness of 24 h urine collections in man. Clin Sci (Lond) 1983;64:629-35.
11. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. J Nutr 2003;133 Suppl 3:921s-4s.
12. Jakobsen J, Pedersen AN, Ovesen L. Para-aminobenzoic acid (PABA) used as a marker for completeness of 24-hour urine: effects of age and dosage scheduling. Eur J Clin Nutr 2003;57:138-42.
13. Johansson G, Bingham S, Vahter M. A method to compensate for incomplete 24-hour urine collections in nutritional epidemiology studies. Public Health Nutr 1999;2:587-91.
14. Bingham SA, Cummings JH. Urine nitrogen as an independent validatory measure of dietary intake: a study of nitrogen balance in individuals consuming their normal diet. Am J Clin Nutr 1985;42:1276-89.
15. Tasevska N, Runswick SA, Bingham SA. Urinary potassium is as reliable as urinary nitrogen for use as a recovery biomarker in dietary studies of free living individuals. J Nutr 2006;136:1334-40.
16. Katan MB, Deslypere JP, van Birgelen AP, Penders M, Zegwaard M. Kinetics of the incorporation of dietary fatty acids into serum cholesteryl esters, erythrocyte membranes, and adipose tissue: an 18-month controlled study. J Lipid Res 1997;38:2012-22.
17. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997;65:1220s-8s.
18. Kleinbaum. Applied regression analysis and multivariable methods. 3 ed. Pacific Grove, CA: Brooks/Cole Publishing Company, 1998.
19. Kaaks RJ. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. Am J Clin Nutr 1997;65:1232s-9s.
20. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr 2002;5:915-23.

21. Ocke MC, Kaaks RJ. Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. Am J Clin Nutr 1997;65:1240s-5s.
22. Molag ML, de Vries JH, Ocke MC, Dagnelie PC, van den Brandt PA, Jansen MC, et al. Design characteristics of food frequency questionnaires in relation to their validity. Am J Epidemiol 2007;166:1468-78.
23. Goldbohm RA, van den Brandt PA, Brants HA, van't Veer P, Al M, Sturmans F, et al. Validation of a dietary questionnaire used in a large-scale prospective cohort study on diet and cancer. Eur J Clin Nutr 1994;48:253-65.
24. Crispim SP, Ribeiro RC, Silva MM, Rosado LE, Rosado GP. The influence of education in the validation process of a food frequency questionnaire for adults in Vicosa, Minas Gerais, Brazil. Eur J Clin Nutr 2006;60:1311-6.
25. Quandt SA, Vitolins MZ, Smith SL, Tooze JA, Bell RA, Davis CC, et al. Comparative validation of standard, picture-sort and meal-based food-frequency questionnaires adapted for an elderly population of low socio-economic status. Public Health Nutr 2007;10:524-32.
26. Feunekes GI, Van Staveren WA, De Vries JH, Burema J, Hautvast JG. Relative and biomarker-based validity of a food-frequency questionnaire estimating intake of fats and cholesterol. Am J Clin Nutr 1993;58:489-96.
27. Verkleij-Hagoort AC, de Vries JHM, Stegers MPG, Lindemans J, Ursem NTC, Steegers-Theunissen RPM. Validation of the assessment of folate and vitamin B12 intake in women of reproductive age: the method of triads. Eur J Clin Nutr 2007;61(5):610-5.
28. Riboli E, Elmstahl S, Saracci R, Gullberg B, Lindgarde F. The Malmo Food Study: validity of two dietary assessment methods for measuring nutrient intake. Int J Epidemiol 1997;26 Suppl 1:S161-73.
29. Ocke MC, Bueno de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.
30. Daures JP, Gerber M, Scali J, Astre C, Bonifacj C, Kaaks R. Validation of a food-frequency questionnaire using multiple-day records and biochemical markers: application of the triads method. J Epidemiol Biostat 2000;5:109-15.
31. Shai I, Rosner BA, Shahar DR, Vardi H, Azrad AB, Kanfi A, et al. Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR study. J Nutr 2005;135:573-9.
32. Kroke A, Klipstein Grobusch K, Voss S, Moseneder J, Thielecke F, Noack R, et al. Validation of a self-administered food-frequency questionnaire administered in the European Prospective Investigation into Cancer and Nutrition (EPIC) Study: comparison of energy, protein, and macronutrient intakes estimated with the doubly labeled water, urinary nitrogen, and repeated 24-h dietary recall methods. Am J Clin Nutr 1999;70:439-47.
33. Subar AF, Kipnis V, Troiano RP, Midthune D, Schoeller DA, Bingham S, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: the OPEN study. Am J Epidemiol 2003;158:1-13.
34. Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, et al. Validation of weighed records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. Br J Nutr 1995;73:531-50.

35. McKeown NM, Day NE, Welch AA, Runswick SA, Luben RN, Mulligan AA, et al. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. Am J Clin Nutr 2001;74:188-96.

36. Matt GE, Rock CL, Johnson Kozlow M. Using recall cues to improve measurement of dietary intakes with a food frequency questionnaire in an ethnically diverse population: an exploratory study. J Am Diet Assoc 2006;106:1209-17.

37. Dyer A, Elliott P, Chee D, Stamler J. Urinary biochemical markers of dietary intake in the INTERSALT study. Am J Clin Nutr 1997;65:1246S-53S.

38. Mente A, Irvine EJ, Honey RJ, Logan AG. Urinary potassium is a clinically useful test to detect a poor quality diet. J Nutr 2009;139:743-9.

39. Ma J, Folsom AR, Shahar E, Eckfeldt JH. Plasma fatty acid composition as an indicator of habitual dietary fat intake in middle-aged adults. The Atherosclerosis Risk in Communities (ARIC) Study Investigators. Am J Clin Nutr 1995;62:564-71.

40. Lissner L, Troiano RP, Midthune D, Heitmann BL, Kipnis V, Subar AF, et al. OPEN about obesity: recovery biomarkers, dietary reporting errors and BMI. Int J Obes (Lond) 2007;31:956-61.

41. Johansson L, Solvoll K, Bjorneboe GE, Drevon CA. Under- and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. Am J Clin Nutr 1998;68:266-74.

42. Heerstrass DW, Ocke MC, Bueno de Mesquita HB, Peeters PH, Seidell JC. Underreporting of energy, protein and potassium intake in relation to body mass index. Int J Epidemiol 1998;27:186-93.

43. Olafsdottir AS, Thorsdottir I, Gunnarsdottir I, Thorgeirsdottir H, Steingrimsdottir L. Comparison of women's diet assessed by FFQs and 24-hour recalls with and without underreporters: associations with biomarkers. Ann Nutr Metab 2006;50:450-60.

7

Discussion

INTRODUCTION

The aim of this thesis was to provide the scientific basis for development of a flexible data-based system, the *Dutch FFQ-TOOL™*, to generate and process tailored food frequency questionnaires using transparent, valid, and reproducible procedures for the assessment of dietary intake in the Netherlands. In this chapter, main findings of this thesis are summarized and discussed and recommendations for future research are given.

MAIN RESULTS

First, we studied design characteristics of FFQs in a meta-analysis of FFQ validation studies. We observed that FFQs with more items in the food list were better able to rank individuals according to their dietary intake. Inclusion of portion size questions did not consistently affect the ranking of different nutrients. Therefore, the choice between including questions about portion sizes or using standard portions must depend on the aim of the assessment by FFQ and is left to the investigators using the *Dutch FFQ-TOOL™*. Thus, the system includes the possibility to add portion size questions, but also to compute nutrient intake based on standard portion sizes.

After this in depth inventory of FFQ design, we needed a feasible approach for selecting food items in an automated way which could be incorporated in a computer system. As a source of information for selection of food items, food consumption databases of the Dutch population were used. Because FFQs are used to study the relationship between intake and health outcomes, food items explaining a large variation in intake between persons would be most informative. Regression analysis is very suitable to select food items for this purpose, but overloads the computer system. Therefore we validated a simple procedure that selected food items based on variance in intake, which tested only one combination of food items, without taking other co-varying items into account. This procedure, called *MOM2*, appeared suitable to select food items based on the highest variance in nutrient intake, if the explained cumulative variance by all selected foods was required to be at minimum 80%. Thus, this procedure was incorporated in the *Dutch FFQ-TOOL™*.

We then held a series of focus group interviews to develop clear questions for use in the *Dutch FFQ-TOOL™*. Respondents experienced problems in comprehension of example questions from FFQs. They tended to restrict their reports of consumption to the given examples in the questions, resulting in selective answers and less valid results. They also had difficulties in identifying foods, because categorization of foods was not logical to them, in aggregating frequencies of consumption of single foods into one food category, and in assessing relative frequencies of food intake. Using all the information from the focus group discussions, standard questions and answering categories were developed for use in the *Dutch FFQ-TOOL™*.

Using a prototype of the *Dutch FFQ-TOOL™*, we developed a first FFQ. This FFQ was validated for its ability to rank adults according to their intake of energy, macronutrients, fatty acids, potassium, and dietary fibre by comparison with biochemical markers of protein, potassium and fatty acid intake and a 3-day estimated food record, respectively. The validity of this FFQ was comparable to other FFQs, and may be improved by including more food items to achieve a higher additive level of explained variance, and by updating food lists with foods that are new at the market by using more recent information.

In the remainder of the discussion of this thesis, we will discuss the development of the *Dutch FFQ-TOOL™* based on the scientific requirements, requirements from a respondent's perspective and technical requirements.

SCIENTIFIC REQUIREMENTS

Food consumption data

We used the Dutch National Food Consumption Survey 1998 as a food intake database for the selection of food items in the *Dutch FFQ-TOOL™*. This is the most extensive and detailed database that covers all age groups, that was available to us. This database consists of estimated food records filled out on two consecutive days in 1997/1998 for 6250 individuals aged between 1 and 97 years (y). From this database, a subpopulation may be selected, defined by age and sex with the limitation that enough individuals remain to study between-person variation in food consumption in the selection. The reported food intake of the specified population is used as a source for the selection of food items for the FFQ. It is recommended to use at least 1000 to 2000 individuals for the selection of food items to develop a representative food list [1].

A major limitation of the Dutch National Food Consumption Survey 1998 is that it contains data that are over ten years old. Therefore food items that entered the market since 1998 will not be included in new FFQs, and must be added manually to the food list. Another limitation of the dataset used in our study was that only two subsequent days of intake were collected. This makes it difficult to estimate the true between-person variation in intake, especially of foods that are consumed infrequently such as fish. Thus, between person variance in this dataset was artificially high since this also contained part of day-to-day variation within persons [2]. With more recording days, the dataset would have better reflected the usual dietary pattern of individuals and would have been more suitable to select informative food items for an FFQ by regression analysis. Therefore, it is of utmost importance that more recent food consumption data, collected over multiple days will be used. Recently, a (semi-) continuous food consumption surveillance has started, pertaining to the general population aged from 4 to 69 y, consisting of two non-consecutive 24-hour dietary recalls in combination with a self-administered questionnaire. Thus the above described problems will become less serious in future [3].

Selection of food items

Because FFQs are used to study the relationship between dietary intake and health outcomes, food items that are able to rank respondents according to their intake between persons would be most informative. Regression analysis is very suitable to select food items for this purpose, because it evaluates many different combinations of food items and their estimated regression coefficients. Regression analysis is very demanding and if automated it may easily overload the computer system. Regression analysis tests many possible combinations in search of the most optimal combination of food items and large databases are used.

An alternative method, called *MOM2*, simply selects food items based on variance in nutrient intake [4] (chapter 3). Since this method does not take covariance in nutrient intake of different food items and their estimated regression coefficients into account, and tests only one combination of food items, it does not overload the computer system. Although this procedure, selected a few more food items than regression analysis, it appeared sufficient to select food items based on the highest variance in nutrient intake. In a next step, the selected food items are evaluated in the *Dutch FFQ-TOOL™* by computing the percentage of explained variance or percentage contribution of the selected food items in the food consumption data.

The above procedures assume that regression or *MOM2* are suitable procedures to explain variation in intake. However, a limitation of the procedures must be taken into account. Measurement errors in the variance of the food items add up in the total variance of a nutrient. In this respect it is especially problematic that multiple people from the same household were included in the Dutch National Food Consumption Survey. Food items that were consumed by the whole family will have an artificially high between-person variance and result in false entries. Thus food items with a large between-person error in estimation of frequency or portion size will result in false entries, whereas other foods may be omitted.

In addition, we explored the effects of energy-adjustments before selecting food items. We used energy-adjustments to remove variation due to differences in energy-intake. We hypothesized that only the most discriminating food items would be included in a food list using a-priori energy-adjustments. In order to study the suitability of energy-adjustment, we selected food items to explain crude and energy-adjusted nutrients using regression analyses in reported intake of 3524 adults aged 25-65 y from the Dutch National Food Consumption Survey 1998, as described above. Various nutrients representing different aspects of the food pattern were incorporated. We focused on the macronutrients carbohydrates and total fat; dietary fibre for vegetables, fruit, and cereals; and the micronutrients vitamin C, as a representative nutrient for vegetables, fruit, and vitamins; and calcium, for dairy foods and minerals. Energy-adjusted nutrient intakes were computed using the residual method 5. To classify individuals based on the smaller residual-energy-independent variation in nutrient intake, required much more food items for macronutrients than based on crude nutrient intakes. For example, more than 100 detailed food items explained only 26% of variance for carbohydrates. Thus, an important part of variance in macronutrient intake is explained by variation in energy intake. For energy-adjusted macronutrients, it was nearly impossible to explain more than 50 percent of variance in nutrient intake once variability in energy intake was accounted for. However, for dietary fibre and the micronutrients it did not affect the number of selected food items, because of a weak association of the tested micronutrients with energy intake [6]. In conclusion these results show that it is not very useful to adjust for energy intake before selecting food items.

Portion size questions

For assessment of portion sizes, a choice must be made between the inclusion of portion size questions or the use of standard portions. In the meta-analysis, described in chapter 2, the effects of including portion size questions on the validity of FFQs were still unclear. A reason to include portion size questions, is that estimates of the level of intake improved according to previous studies [7,8]. Ranking appeared to be worse for protein and vitamin C when portion-size questions were included in the FFQ instead of standard portions, but it improved for alcohol. Even when portion sizes were assessed in a user-friendly way, using pictures [9-11], or fitted from 24-hour dietary recalls [12], ranking of respondents according to their intake hardly improved compared to the use of standard portion sizes. Reasons may be that respondents have difficulties in estimating consumed portion sizes [13,14], vary a lot in their consumed portion sizes [15-17], or the listed portion sizes do not reflect their usually consumed portion sizes, for example as listed portion sizes are too small [18] or too large. The latter argument is not very likely as portion sizes in the *Dutch FFQ-TOOL™*, were based on common Dutch portion sizes [19].

In the above discussion it is important to realize that FFQs developed with the *Dutch FFQ-TOOL™* differ from 'Willett type' FFQs. These type of FFQs focus on frequency only and included questions ask how frequently a standard portion size was consumed. This type of questions is cognitively very difficult for respondents especially of a low-educated population. Willett validated his FFQs among well-educated nurses and male health professionals, which may explain that the problem did not manifest [20-22]. However, we aimed to develop FFQs for the general population and therefore our FFQs assess on how many days a food was consumed. Amount is addressed separately in terms of consumption per day and portion sizes or units. Thus, for us, both frequency and portion size are important. Also, the selection of food items in the *Dutch FFQ-TOOL™* is based on consumed amounts (i.e. frequency*portion). This implies that it is not possible to separate variance in intake in a frequency-based and portion size-based part. Therefore, we think it is better to add portion size questions to FFQs and we included this possibility in the *Dutch FFQ-TOOL™*. However, it is also possible to compute nutrient intakes based on standard portion sizes in the system.

Processing of FFQs

In addition, we needed valid and reproducible procedures for processing completed FFQs. Nutrient intakes must be computed based on the composition of the selected food items. The nutrient composition of each food item is derived from the weighted mean composition of the single foods of which the food item consisted, as weighted by the reported amounts in the food consumption database. Thus a limitation of these computations is that relative weights are based on reported amounts in the Dutch National Food Consumption Survey 1998. It is difficult to add new foods to existing food items, because this requires an estimate of the contribution of the new food to the existing item.

Validation of FFQs

From a scientific perspective, it is important to strive for valid FFQs. Because FFQs cannot measure dietary intake without errors, they are usually evaluated in a validation study. Evaluation of FFQs against another dietary assessment method, provides only limited information, as the validity is probably overestimated, because of correlated errors in intake estimates [23,24]. Preferably, the validity of FFQs is studied using biochemical markers of intake, because errors in estimations of biochemical markers are largely independent of errors in the FFQ [1,25]. However, biochemical markers also have their limitations. Suitable markers are available for a limited number of nutrients only, and correlations between intake by FFQ and biochemical markers are generally only modest, even if the dietary assessment is very accurate.

Recovery biomarkers can be used to evaluate both ranking of respondents as well as the level of intake provided by the FFQ. Even recovery biomarkers are influenced by within-person variation in intake, which result in fluctuations in biochemical indicator levels. The available recovery markers reflect intake of the very recent past, e.g. the previous day. Examples are doubly labelled water as a marker of 24-hour energy expenditure and 24-hour urinary nitrogen as a marker of protein intake. For practical reasons, it is often not possible to sample more than two to four replicates of biological samples as a standard for long-term dietary intake.

In addition, concentration markers may be analysed to evaluate ranking of respondents.

Besides day to day variation in intake, concentration markers are influenced by additional sources of variability. These include variations in metabolic processes between persons such as absorption and metabolism of most nutrients, variations in physiological mechanisms such as binding proteins, differences in bioavailability from different foods and technical error associated with laboratory measurements [1]. Thus, it is unclear whether the biomarker truly represents intake because so many intermediate factors between nutrient intake and biomarker level are involved. This implies that correlations between intake by FFQ and concentration markers tend to be low.

Unfortunately, for many nutrients no suitable biochemical markers of intake are available. Therefore, it is worth exploring possible new biomarkers. In our study, we tested the feasibility of using sucrose and fructose determined in 24-hour urine, as biomarkers of total sugars and carbohydrate intake in a large scale validation study. It has been reported that small amounts of sucrose cross the small intestine unchanged and are excreted in urine [26]. Recently, a few studies were conducted in a residential volunteer suite, where dietary intake was controlled and all blood and urine samples were collected [27]. In one study, participants were fed three different diets with constant levels of sugars, and in the second study participants were fed their usual diets for 30 days under controlled conditions in the volunteer suite. Both studies showed that the combined measure of urinary sucrose and fructose was highly correlated with both sucrose and total sucrose and fructose intake [28]. A third study confirmed these findings, and showed that BMI did not affect the validity of sucrose and fructose excretions in 24-hour urine used as biomarkers to estimate total sugars consumption [29]. Furthermore, these biomarkers were assessed in spot urines in an epidemiologic study, showing that odds ratios for a BMI>30 kg/m^2 were significantly increased across quintiles of sucrose excretion, whereas no association with dietary intake of sucrose was found [30]. Thus, urinary fructose and sucrose are promising biomarkers that can be used to estimate dietary intake of total sugars and sucrose.

To evaluate whether this biomarker can be used in a real-life setting, we assessed urinary fructose and sucrose in the 24-hour urines of 109 participants from the validation study described in chapter 6. For determination of the concentrations of sucrose and fructose we used similar enzymatic test kits as used by Tasevska et al. [28]. However, about fifty percent of the sucrose samples and forty percent of the fructose concentrations were below the detection limit of <5 mg per litre. In addition, we observed high coefficients of variation in laboratory assessment of these biomarkers. Detecting very low concentrations is possible by using mass spectrometry. Unfortunately, this was not possible in our study as we had used boric acid as a preservative, which interfered with the detection of sucrose and fructose in mass spectrometry. Because of the problems with samples below the detection limit and high coefficients of variation, the results of sucrose and fructose were not reported in chapter 6. For future large-scale studies, sucrose and fructose are also not suitable as biomarkers of dietary intake because low concentrations may be expected and mass spectrometry is very expensive.

For practical and financial reasons it is impossible to validate each new FFQ. In addition, as mentioned before, validation studies have their limitations as for most nutrients no biomarkers are available. Therefore, it would be very helpful to describe the construct validity of FFQs based on the available information. Construct validity is defined as the degree to which the data collected reflect or measure the variable of interest. Applied to an FFQ, we can frame this question as the magnitude of the variance shared between nutrient intake estimated from the FFQ and true nutrient intake [31]. Thus, in the *Dutch FFQ-TOOL™*, first estimates of the construct validity of FFQs are provided by explained variance and level of intake covered by the selected food items [1].

Energy adjustments

Furthermore, it is generally assumed that energy adjustment improves results of studies, controlling for confounding, because intakes are estimated from the same foods [24,32,33]. However, in the validation study presented in chapter 6, correlation coefficients between intake by FFQ and food record decreased after energy-adjustments. This may result from attenuation of the true correlation and contamination from the effect of an adjusting covariate [25]. Beaton et al already cautioned that differential biases in the reporting of intakes may exist, such as underreporting of unhealthy foods, not alleviating this problem by energy-adjustments [34,35]. Recently, underreporting of food intake was studied among low energy reporters that were identified by comparison of reported energy intake by FFQ with energy expenditure as estimated by doubly labelled water. It was confirmed that especially the frequency of consumption of unhealthy foods was underreported in the FFQ [36]. Thus these arguments imply that control for confounding by energy-adjustment, and disentangling the effects of different macronutrients, may be seriously limited by the quality of the information obtained from the respondents.

RESPONDENT'S PERSPECTIVE

For respondents, food items must be familiar and their consumption must be assessed in comprehensive questions with clear answering categories.

Food items

It is very important that respondents are motivated to fill out an FFQ, therefore they must be able to identify the food items in FFQs developed with the system. We improved recognition of food items in the *Dutch FFQ-TOOL™* by developing standardized hierarchical levels of food groups for use as items in the FFQ, based on expertise of experienced dieticians at Wageningen University and TNO, the Netherlands. Related food codes from the food composition table were combined into food items at different levels of detail, as required for a study. Grouping into food items was based on similarity in eating occasions, portion sizes and nutrient contents. For this, the dieticians took into account whether respondents could possibly answer questions about usual consumption of each food item. Another way to improve recognition of food items is by the addition of pictures of food items [37], this may be done in future. Another problem is that respondents may find it annoying that their favourite food is not listed in the FFQ. Therefore, an open space is sometimes provided to write down any missing food items [37]. However, these additional food items added little to ranking of respondents or to total nutrient intake of respondents [38]. This possibility was not included in the *Dutch FFQ-TOOL™*.

Reference period

In addition, the reference period of an FFQ is important. The researcher developing the FFQ decides which reference period is most suitable depending on the aim of the study and the nutrients of interest. A reference period of one month is suitable for assessing macronutrient intake, whereas one year is more suitable for assessing seasonally consumed food items and micronutrients [39]. It is often assumed that intake of the past month is representative of usual macronutrient intake. Moreover, it is even possible to include a few questions with a different reference period, for example to assess seasonally consumed food items. In this case it is important to carefully describe which reference period must be used for computations of seasonal intake.

Focus group discussions

Focus group discussions, described in chapter 4, showed that too many examples may confuse respondents or result in differential underreporting. Therefore, we recommended carefully considering whether including examples, as a clarification, is really necessary. Questions may be clarified by including cues that stimulate memory of respondents or remind them of consumed food items [37]. An example of a cue may be "think about consumption for breakfast, lunch, dinner and as in-between snack". During the focus group discussions most respondents were enthusiastic about cues. In the *Dutch FFQ-TOOL™*, we did not include standard cues for specific food items. For now, cues have to be entered manually and the researcher developing the FFQ decides for which questions cues are appropriate, for example if a food may be consumed at multiple meals.

Frequency questions

To describe frequency of consumption, respondents need clear questions, because they improve accuracy of FFQs [40]. We used findings from previous cognitive research and the focus group discussions described in chapter 4 to develop clear standard questions for the system [37,39]. Questions were formatted according to the nested approach described by Subar et al. [39]. In this approach, frequency of intake is first assessed followed by more specific questions about number of portions, portion sizes and preparation methods. Sometimes, summary questions are added to correct answers to individual questions. Summary questions have been shown to improve estimates of the level of intake, but resulted in larger errors if the interest was ranking of respondents according to their intake [41]. Therefore, we included the possibility to add summary questions in the *Dutch FFQ-TOOL™*. Furthermore, open-ended response formats are often used to allow respondents to report frequency of consumption in a way that suited their cognitive preferences, but they often failed to fill it out correctly [39]. Therefore, the *Dutch FFQ-TOOL™* includes only closed questions.

Portion size questions

Consumed amounts must be indicated in the FFQ using portion size questions that are easily understood and filled out by respondents. Respondents often skip portion size questions [39], probably because many respondents find it difficult to estimate portion sizes. In the literature, several methods to assess portion sizes have been described such as: describing if usual intake is smaller or larger than a specified medium portion size [42], describing intake of small, medium or large portion sizes [43,44], describing portion sizes in household measures such as spoons [45,46] or by including pictures of different portion sizes [9-11]. We did not include pictures for the estimation of portion sizes in the system. Usually only one component of a meal is shown on a picture, and respondents tend to choose the medium portion size. Therefore, ranking of respondents hardly improved when portion sizes were estimated using pictures compared to standard portions [7,10,11]. Cognitive strategies to report portions have been studied [39,45]. For reporting portions, respondents preferred to define ranges of portions in household measures, such as one cup or two cups [39]. Thus, we recommend adding portion size questions described in household measures.

TECHNICAL REQUIREMENTS

Standardizing procedures

In the *Dutch FFQ-TOOL™*, FFQ development must include the automated selection of food items and generating standard questions for the selected food items. In addition, the system must be able to compute nutrient intake of completed FFQs based on selected food items.

Selection of food items

A major challenge during the development of the *Dutch FFQ-TOOL™*, was the automated selection of food items in the system. As described above (see scientific requirements), it was not possible to select food items for ranking respondents according to their intake using regression analysis. Theoretically it would have been better to select food items using regression analysis. However, the gains of more precise classification should be balanced against other factors that influence the validity of FFQs. For example, respondents may underestimate their food intake if fewer food items are included in the FFQ [47], whereas increasing the number of items may lead to overreporting [41]. The order in which food items are listed in the FFQ also influences responses [44], for example putting specific items in the FFQ prior to general items was shown to increase reported intake relative to vice versa [48]. Moreover, it is important that a food list is comprehensible for respondents, and it may be desirable to add extra food items if this increases the face validity of FFQs. Modifications that increased comprehensiveness, such as clear questions, improved validity of estimates of nutrient intake [40]. All these factors support using a simpler method such as *MOM2*, because the precision gained by using forward regression is limited compared to the impact of above mentioned factors.

Level of aggregation

Food items may be enquired at different aggregation levels depending on the level of detail required for the aim of a study. Therefore, we divided foods into several subgroups at different levels of aggregation. Foods in the Dutch food composition table were categorized into 24 food groups such as 'bread', 'fruit' and 'vegetables'. For use as items in FFQs, these food groups were further subdivided into smaller food groups at three hierarchical levels of aggregation of food items. To compose a food list for an FFQ, a choice must be made on the most appropriate level of aggregation for each food item. To assess the level of intake for example for monitoring purposes highly aggregated food items are more suitable. The reason is that the level of intake is often overestimated when many detailed food items are included [41,47]. However, in epidemiology we are more interested in ranking individuals accurately according to their intake. In this case, food items at a detailed level are more suitable, because they better capture the between-person variation in intake. However, a disadvantage of selecting food items at a detailed level is that it results in a longer food list. We hypothesized that optimization processes such as linear programming could help to decide on the most informative level of aggregation [49,50]. However, this was not possible, because differences in nutrient contents of foods in a food item and frequency of consumption could not be taken into account. Thus other optimization techniques might be needed.

This optimization should also take into account that the dietician usually creates new food groups during the development process to differentiate between, for example, oranges that are very important to explain vitamin C intake, and 'other types of citrus fruit' that are important when grouped together. Because we were not able to automate the selection of food items, we created a screen in the *Dutch FFQ-TOOL™* that shows values of variance in intake or percentage contribution for food items at different levels of aggregation. The dietician can use this information to decide which level of aggregation is most appropriate. In a next step the selected food items are evaluated in the system by computing the percentage of explained variance or percentage contribution of the selected food items in the food consumption data. In future research, we aim to further standardize decisions about different levels of aggregation.

Composing a food list for multiple nutrients

Another difficulty, in composing a food list, is that an FFQ must be composed to assess multiple nutrients of interest. This is one of the reasons that developing a food list is a very time-consuming process for a dietician developing an FFQ. Foods added to explain further nutrients, also add to explained variance for the earlier selected nutrients. Therefore, there must be an optimum, explaining intake of all nutrients of interest with a minimum number of food items, as too long lists bore respondents.

We explored possibilities of linear programming to optimize the solution to this problem [51]. It was possible to further optimize selections for two nutrients of interest using the modelling program AIMMS® (of Paragon Decision Technology). This program uses two-dimensional graphs to decide which foods must be included. However, FFQs are usually developed for much more than two nutrients of interest. The use of two-dimensional graphs limits possibilities to study optimal solutions for FFQs that must assess more than two nutrients of interest. This would require three, four, five or even multidimensional graphs. Thus, further research must use more complex models to solve this optimization problem for multiple nutrients of interest.

Therefore, we needed a practical solution and studied whether the order of selections for different nutrients influenced the final food list for different combinations of nutrients (chapter 3). In the first approach, the order was from the nutrient of which 80% variance was explained by the lowest number of food items, to the nutrient explained by the highest number. In the alternative approach the reverse order was used. Analyses were done for the following sets of nutrients: vitamin C and carbohydrates, fat and carbohydrates, dietary fibre and carbohydrates. Selections made first for nutrients of interest explained by the lowest number of food items, followed by those explained by a high number of food items, led to the lowest total number of food items at the highest level of explained variance. Thus, combining a food list for multiple nutrients of interest should from the nutrient explained by the lowest number of foods to the highest, taking priorities in the ultimate study into account.

Processing of FFQs

Processing of completed FFQs must be automated in the *Dutch FFQ-TOOL™*, based on the selected food items. We thoroughly tested automated nutrient computations in the *Dutch FFQ-TOOL™*, and compared results with nutrients computations using standard procedures. In addition, minima and maxima of consumed amounts and missing values of completed FFQs are shown to the researcher who evaluates these values and adapts the data were needed. Thus, procedures for quality control were also semi-automated and included in the *Dutch FFQ-TOOL™*.

GENERALIZATION

Developing FFQs for different populations

In our objective, we aimed to develop tailored FFQs for different target populations with the *Dutch FFQ-TOOL™*. A first step for developing FFQs for these populations, is to import the appropriate food consumption data. For the development phase of the *Dutch FFQ-TOOL™* we have used the Dutch Food Consumption Survey 1997/1998. However, food consumption surveys are held every few years and can therefore be used to update food consumption data in the *Dutch FFQ-TOOL™*. Food consumption surveys of small children and young adults (18-30 y) are also available and will be imported [52-54]. However, smaller populations are included in these surveys making them less representative of the target population. Besides that, we must consider if the FFQ is the most appropriate dietary assessment method for the target population. For example in the elderly, dietary reporting may be inaccurate due to cognitive problems [55]. It is recommended to evaluate this before deciding to apply an FFQ. Also, FFQs developed for children must be adapted for them. For example, questions, or a selection of questions, must address their parents and FFQs must include smaller portion sizes than FFQs for adults. Thus, developing FFQs for different target populations than adults is possible if appropriate food consumption data is imported and adaptations are made to standard questions and answering categories.

Developing FFQs for individuals with different cultural backgrounds is more difficult. Culture-specific food intake data needs to be collected first as a source for the selection of foods for the FFQ [56,57]. In addition, completeness of food intake data must be evaluated. Respondents may have left out some foods because they believed that the interviewer was unfamiliar with some culture-specific foods or did not approve of certain foods [58]. In addition, focus group discussions are necessary to define culture-specific food groups and to define portion sizes. Also, a culture-specific nutrient database must be developed [58]. Thus, to develop FFQs for individuals with different cultural backgrounds with the *Dutch FFQ-TOOL™*, is still labour-intensive.

Developing FFQs for different nutrients of interest

In the *Dutch FFQ-TOOL^{TM}*, we aim to develop FFQs for assessing energy and nutrients of interest. In the different chapters of this thesis, we had to focus on a few nutrients. In chapter 2, we included energy and all energy-providing macronutrients (i.e. protein, carbohydrates, fat, and alcohol). We added vitamin C and dietary fibre to represent fruit and vegetables and some other food components, for example, calcium, to represent other specific foods. In chapter 3 we selected food items for carbohydrates, fat, vitamin C, dietary fibre and calcium. Thus, we used a selection of the nutrients used in chapter 2 for the evaluation of the selection procedures. In chapter 6, we developed an FFQ with the system that focused on nutrients for which biochemical markers were available i.e.: protein, potassium, mono- and disaccharides and fatty acids and some of the nutrients included in the previous studies: energy, carbohydrates, total fat, and dietary fibre. We assume that the *Dutch FFQ-TOOL^{TM}* is also suitable to develop FFQs for other nutrients. In future, we would like to include the possibility to select food groups.

FUTURE RESEARCH

In future research, we would like to focus on further development of the *Dutch FFQ-TOOL^{TM}*. To increase transparency, selection of food items at different hierarchical levels of aggregation and for multiple nutrients of interest needs to be further optimized. This may be done by exploring other possibilities of decision science than linear programming.

A challenge is to develop FFQs that are up to date regarding trends such as consumption of mixed dishes, functional foods and the use of supplements. There will always be a delay between collection of food consumption data and incorporation of this database in the *Dutch FFQ-TOOL^{TM}*. Therefore, it is necessary to add new food items that were introduced since the sampling of food consumption data used for the selection. Researchers must use their professional expertise to decide on which new food items must be added. Also, procedures must be included to add food items that increase the face validity of FFQs.

We also recommend developing web-based FFQs because these are more convenient for respondents and might help to reduce misreporting. In the Netherlands in 2008, 91% of all individuals had access to the internet. The percentage of individuals with access to the internet decreases with increasing age: 98% for 12 to 25 y old, 96% for 25 to 45 y old, 91% for adults between 45 and 65 y and 57% for adults between 65 and 75 y in 2008 [59]. This indicates that web-based FFQs are especially suitable to use in populations aged 65 y and younger.

Important advantages of web-based FFQs over paper versions include increased comprehension of web-based FFQs by including tailored pop-up. Also, web-based

FFQs may include pictures of unfamiliar foods to improve identification of foods (and estimation of portion sizes) [9,42,60]. Assessing frequencies of consumption may be improved by including more frequency categories and by providing information about the relative frequency categories. We recommend including portion size questions as both frequency and portion size are important for explaining variance in intake. These portion sizes must be described in household measures as these are most convenient for respondents. Also, reporting will be more accurate, as respondents can be compelled to fill out questions, and skip questions that do not apply to them. Thus, self-reports by web-based FFQs will reduce the amount of data cleaning, but require the respondents to have access to a computer.

CONCLUSION

From a scientific perspective it is of utmost importance that more recent food consumption data collected over multiple days will become available for development of new FFQs. Ideally, food items would have been selected using regression analysis in the *Dutch FFQ-TOOL™* for ranking respondents according to intake, however, procedures overloaded the system because large databases were used. Because of limitations in the food consumption database, and because variances dominate the selection process over covariance, it appeared suitable to select food items based on the highest variance in nutrient intake. We recommend including portion size questions as both frequency and portion size are important for explaining variance in intake. These portion sizes should be described in household measures as these are most convenient for respondents. In addition, the *Dutch FFQ-TOOL™* provides first estimates of the construct validity of FFQs by explained variance and level of intake covered by the selected food items. And the validity of a first FFQ in a validation study was comparable to other FFQs, and may be improved by including more food items to achieve a higher additive level of explained variance and by updating food lists with foods that are new at the market by using more recent information.

From a respondent's perspective it is important that we developed a system that generates FFQs that are clear to respondents by improved recognition of food items in the *Dutch FFQ-TOOL™*, by standardized hierarchical levels of food groups for use as items in the FFQ, and by clear questions to assess frequency of consumption. In future, web-based FFQs will become more convenient for respondents because questions that do not apply to them can automatically be skipped.

Because of technical limitations, it was impossible to automate selection of food items completely, but standard questions were automatically generated and procedures for processing FFQs were also automated. Although we were not able to automate all procedures in this system, new FFQs are developed considerably faster using this system than by a dietician that uses standard procedures. Time-consuming elements include identifying important foods for all nutrients of interest, aggregating these foods to food items, composing a food list from these food items, transposing the food list into standard questions with answering categories and adding questions about portion sizes and preparations, quality control of completed FFQs, composing a nutrient database for processing of filled out FFQs and finally quality control of processed FFQs. All have become easier or faster with this new approach. In addition automating procedures increased transparency of FFQ development, for example because food lists of FFQs are saved in the *Dutch FFQ-TOOL*™. These food lists show exactly which single foods were used to compose a food item.

Thus, we have developed a comprehensive data-based system in which researchers in the Netherlands may develop and process food frequency questionnaires (FFQs) in a standardized way to assess individual intakes for different research purposes and populations.

REFERENCE LIST

1. Willett W. Nutritional Epidemiology. second edition ed. New York, NY: Oxford University Press, 1998.
2. Lambe J, Kearney J, Leclercq C, Zunft HF, De Henauw S, Lamberg Allardt CJ, Dunne A, Gibney MJ. The influence of survey duration on estimates of food intakes and its relevance for public health nutrition and food safety issues. Eur J Clin Nutr 2000;54(2):166-73.
3. Ocké M, Hulshof KF. Naar een nieuw Nederlands voedingspeilingsysteem; in Dutch; Towards a new system of dietary monitoring in the Netherlands. Voeding Nu 2005;12:26-8.
4. Mark SD, Thomas DG, Decarli A. Measurement of exposure to nutrients: an approach to the selection of informative foods. Am J Epidemiol 1996;143(5):514-21.
5. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr 1997;65(4 Suppl):1220s-8s.
6. Mackerras D. Energy adjustment: the concepts underlying the debate. J Clin Epidemiol 1996;49(9):957-62.
7. Ocke MC, Bueno de Mesquita HB, Pols MA, Smit HA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. II. Relative validity and reproducibility for nutrients. Int J Epidemiol 1997;26 Suppl 1:S49-58.
8. Nothlings U, Hoofman K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. J Nutr 2007;137:2781-6.
9. Ocke MC, Bueno de Mesquita HB, Goddijn HE, Jansen A, Pols MA, van Staveren WA, Kromhout D. The Dutch EPIC food frequency questionnaire. I. Description of the questionnaire, and relative validity and reproducibility for food groups. Int J Epidemiol 1997;26 Suppl 1:S37-48.
10. Haraldsdottir J, Tjonneland A, Overvad K. Validity of individual portion size estimates in a food frequency questionnaire. Int J Epidemiol 1994;23(4):786-96.
11. Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semiquantitative food frequency questionnaire. Int J Epidemiol 1992;21(4):770-7.
12. Nothlings U, Hoffmann K, Bergmann MM, Boeing H. Fitting portion sizes in a self-administered food frequency questionnaire. J Nutr 2007;137(12):2781-6.
13. Huizinga MM, Carlisle AJ, Cavanaugh KL, Davis DL, Gregory RP, Schlundt DG, Rothman RL. Literacy, Numeracy, and Portion-Size Estimation Skills. Am J Prev Med 2009;36(4):324-8.
14. Faggiano F, Vineis P, Cravanzola D, Pisani P, Xompero G, Riboli E, Kaaks R. Validation of a method for the estimation of food portion size. Epidemiology 1992;3(4):379-82.
15. Tsubono Y, Kobayashi M, Takahashi T, Iwase Y, Iitoi Y, Akabane M, Tsugane S. Within- and between-person variations in portion sizes of foods consumed by the Japanese population. Nutr Cancer 1997;29(2):140-5.
16. Hunter DJ, Sampson L, Stampfer MJ, Colditz GA, Rosner B, Willett WC. Variability in portion sizes of commonly consumed foods among a population of women in the United States. Am J Epidemiol 1988;127(6):1240-9.
17. Young LR, Nestle MS. Portion sizes in dietary assessment: issues and policy implications. Nutr Rev 1995;53(6):149-58.
18. Young LR, Nestle M. Expanding portion sizes in the US marketplace: implications for nutrition counseling. J Am Diet Assoc 2003;103(2):231-4.

19. Donders-Engelen M, Van Der Heyden L, Hulshof K. Maten en Gewichten en Codenummers 1992; In Dutch; Food portion sizes and coding instructions. In: Donders-Engelen, Heyden VD, Hulshof, eds. Wageningen: Wageningen University and TNO Nutrition, Zeist, 2003.
20. Willett WC, Sampson L, Browne ML, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE. The use of a self-administered questionnaire to assess diet four years in the past. Am J Epidemiol 1988;127(1):188-99.
21. Willett WC, Sampson L, Stampfer MJ, Rosner B, Bain C, Witschi J, Hennekens CH, Speizer FE. Reproducibility and validity of a semiquantitative food frequency questionnaire. Am J Epidemiol 1985;122(1):51-65.
22. Rimm EB, Giovannucci EL, Stampfer MJ, Colditz GA, Litin LB, Willett WC. Reproducibility and validity of an expanded self-administered semiquantitative food frequency questionnaire among male health professionals. Am J Epidemiol 1992;135(10):1114-26.
23. Kipnis V, Midthune D, Freedman L, Bingham S, Day NE, Riboli E, Ferrari P, Carroll RJ. Bias in dietary-report instruments and its implications for nutritional epidemiology. Public Health Nutr 2002;5(6a):915-23.
24. Willett W, Kipnis V, Subar AF, Schatzkin A, Midthune D, Troiano RP, Schoeller DA, Bingham S, Freedman LS. Invited commentary: OPEN questions. Am J Epidemiol 2003;158(1):22-6.
25. Kipnis V, Midthune D, Freedman LS, Bingham S, Schatzkin A, Subar A, Carroll RJ. Empirical evidence of correlated biases in dietary assessment instruments and its implications. Am J Epidemiol 2001;153(4):394-403.
26. Menzies JS. Absorption of intact oligosaccharide in health and disease. Biochemical Society Transactions 1974;2(5):1042-7.
27. Bingham SA. Biomarkers in nutritional epidemiology. Public Health Nutr 2002;5(6a):821-7.
28. Tasevska N, Runswick SA, McTaggart A, Bingham SA. Urinary sucrose and fructose as biomarkers for sugar consumption. Cancer Epidemiol Biomarkers Prev 2005;14(5):1287-94.
29. Joosen AM, Kuhnle GG, Runswick SA, Bingham SA. Urinary sucrose and fructose as biomarkers of sugar consumption: comparison of normal weight and obese volunteers. Int J Obes (Lond) 2008;32(11):1736-40.
30. Bingham S, Luben R, Welch A, Tasevska N, Wareham N, Khaw KT. Epidemiologic assessment of sugars consumption using biomarkers: comparisons of obese and nonobese individuals in the European prospective investigation of cancer Norfolk. Cancer Epidemiol Biomarkers Prev 2007;16(8):1651-4.
31. Kristal AR, Potter JD. Not the time to abandon the food frequency questionnaire: counterpoint. Cancer Epidemiol Biomarkers Prev 2006;15(10):1759-60.
32. McKeown NM, Day NE, Welch AA, Runswick SA, Luben RN, Mulligan AA, McTaggart A, Bingham SA. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. Am J Clin Nutr 2001;74(2):188-96.
33. Bingham SA, Cassidy A, Cole TJ, Welch A, Runswick SA, Black AE, Thurnham D, Bates C, Khaw KT, Key TJ, et al. Validation of weihted records and other methods of dietary assessment using the 24 h urine nitrogen technique and other biological markers. Br J Nutr 1995;73(4):531-50.
34. Gibson RS. Principles of nutritional assessment. Second Edition. Vol. 2nd. Oxford, New York: Oxford University Press, Inc., 2005.
35. Beaton GH, Burema J, Ritenbaugh C. Errors in the interpretation of dietary assessments. Am J Clin Nutr 1997;65(4 Suppl):1100S-7S.

36. Millen AE, Tooze JA, Subar AF, Kahle LL, Schatzkin A, Krebs-Smith SM. Differences between food group reports of low-energy reporters and non-low-energy reporters on a food frequency questionnaire. J Am Diet Assoc 2009;109(7):1194-203.

37. Matt GE, Rock CL, Johnson Kozlow M. Using recall cues to improve measurement of dietary intakes with a food frequency questionnaire in an ethnically diverse population: an exploratory study. J Am Diet Assoc 2006;106(8):1209-17.

38. Block G, Mandel R, Gold E. On food frequency questionnaires: the contribution of open-ended questions and questions on ethnic foods. Epidemiology 2004;15(2):216-21.

39. Subar AF, Thompson FE, Smith AF, Jobe JB, Ziegler RG, Potischman N, Schatzkin A, Hartman A, Swanson C, Kruse L, et al. Improving food frequency questionnaires: a qualitative approach using cognitive interviewing. J Am Diet Assoc 1995;95(7):781-8.

40. Thompson FE, Subar AF, Brown CC, Smith AF, Sharbaugh CO, Jobe JB, Mittl B, Gibson JT, Ziegler RG. Cognitive research enhances accuracy of food frequency questionnaire reports: results of an experimental validation study. J Am Diet Assoc 2002;102(2):212-25.

41. Bogers RP, Dagnelie PC, Westerterp KR, Kester AD, van Klaveren JD, Bast A, van den Brandt PA. Using a correction factor to correct for overreporting in a food-frequency questionnaire does not improve biomarker-assessed validity of estimates for fruit and vegetable consumption. J Nutr 2003;133(4):1213-9.

42. Kumanyika SK, Tell GS, Shemanski L, Martel J, Chinchilli VM. Dietary assessment using a picture-sort approach. Am J Clin Nutr 1997;65(4 Suppl):1123s-9s.

43. Schlundt DG, Buchowski MS, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Separate estimates of portion size were not essential for energy and nutrient estimation: results from the Southern Community Cohort food-frequency questionnaire pilot study. Public Health Nutr 2007;10(3):245-51.

44. Kuskowska Wolk A, Holte S, Ohlander EM, Bruce A, Holmberg L, Adami HO, Bergstrom R. Effects of different designs and extension of a food frequency questionnaire on response rate, completeness of data and food frequency responses. Int J Epidemiol 1992;21(6):1144-50.

45. Chambers Et, Godwin SL, Vecchio FA. Cognitive strategies for reporting portion sizes using dietary recall procedures. J Am Diet Assoc 2000;100(8):891-7.

46. Welten DC, Carpenter RA, McPherson RS, Brodney S, Douglass D, Kampert JB, Blair SN. Comparison of a dietary record using reported portion size versus standard portion size for assessing nutrient intake. Public Health Nutr 2000;3(2):151-8.

47. Serdula M, Byers T, Coates R, Mokdad A, Simoes EJ, Eldridge L. Assessing consumption of high-fat foods: the effect of grouping foods into single questions. Epidemiology 1992;3(6):503-8.

48. Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires - a review. Public Health Nutr 2002;5(4):567-87.

49. Alidaee B, Kochenberger G, Lewis K, Lewis M, Wang H. A new approach for modeling and solving set packing problems. Eur J Oper Res 2008;186(2):504-12.

50. Delorme X, Gandibleux X, Rodriguez J. GRASP for set packing problems. Eur J Oper Res 2004;153(3 SPEC. ISS.):564-80.

51. Claassen GDH, Hendriks THB, Hendrix EMT. Decision Science: theory and applications. Wageningen: Wageningen Academic Publishers, 2007.

52. van Rossum CTM, Brants HA, Buurma-Rethans E, Fransen H, Niekerk M, Ocké M. Zo eten peuters en kleuters in Nederland. Resultaten van de Voedselconsumptiepeiling onder jonge kinderen. Voeding NU 2008;10(4):13-6.

53. Ocké M HK. Zo eten jongvolwassenen in Nederland. Voeding Nu 2004;6(10):16-9.

54. Ocké MC, Hulshof KFAM, van Rossum CTM. The Dutch national food consumption survey 2003. Methodological issues. Archives of Public Health 2005;63(5):227-41.

55. Pope SK, Kritchevsky SB, Morris MC, Block G, Tylavsky FA, Lee JS, Stewart S, Harris T, Rubin SM, Simonsick EM. Cognitive Ability is Associated with Suspected Reporting Errors on Food Frequency Questionnaires. J Nutr Health Aging 2007;11(1):55-8.

56. Buchowski MS, Schlundt DG, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Development of a culturally sensitive food frequency questionnaire for use in the Southern Community Cohort Study. Cell Mol Biol (Noisy le grand) 2003;49(8):1295-304.

57. Shahar D, Shai I, Vardi H, Brener Azrad A, Fraser D. Development of a semi-quantitative Food Frequency Questionnaire (FFQ) to assess dietary intake of multiethnic populations. Eur J Epidemiol 2003;18(9):855-61.

58. Teufel NI. Development of culturally competent food-frequency questionnaires. Am J Clin Nutr 1997;65(4 Suppl):1173s-8s.

59. CBS. CBS Statline databank. ICT gebruik van personen naar persoonskenmerken Vol. 2009. Den Haag/Heerlen: Centraal Bureau voor de Statistiek, 2009;http://statline.cbs.nl/StatWeb/publication/?DM=SLNL&PA=71098ned&D1=33-34&D2=0,3-13&D3=a&HDR=G1&STB=T,G2&VW=T.

60. Wengreen HJ, Munger RG, Wong SS, West NA, Cutler R. Comparison of a picture-sort food-frequency questionnaire with 24-hour dietary recalls in an elderly Utah population. Public Health Nutr 2001;4(5):961-70.

Summary/Samenvatting

SUMMARY

Food frequency questionnaires (FFQs) are widely used to assess dietary intake of large populations. The popularity of FFQs stems from their ease of administration, ability to assess dietary intake over an extended period of time, and low costs. They are therefore often used in epidemiologic studies to investigate the relation between diet and disease. A well-designed FFQ aims at assessing habitual intake of foods or nutrients of interest in a specific population. This implies that the food list of an FFQ has to be adapted to food consumption habits of the target population and updated, with new foods at the market, when re-used some time after initial development. Unfortunately, FFQs are often re-used without adaptations due to lack of expertise, time or finances. In order to make the development of new FFQs easier and more standardized, we developed a computer system to generate, apply and process FFQs, the so called '*Dutch FFQ-TOOL*™'. The overall aim of this thesis is to provide the scientific basis for development of a flexible data-based computer system to generate, apply and process tailored FFQs for multiple nutrients of interest and target groups and to evaluate this system.

In order to develop a scientifically sound system, we conducted a meta-analysis of 40 validation studies (chapter 2) and investigated the role of FFQ design, including length and use of portion-size questions, in ranking individuals according to their nutrient intake. In addition, we studied the ability of the FFQ to detect differences in energy intake between subgroups and to assess energy and protein intake. We observed that FFQs with longer food lists (>200 items) were better than shorter FFQs at ranking individuals for intake of most nutrients; results were statistically significant for protein, energy-adjusted total fat, and energy-adjusted vitamin C. We found that FFQs that included standard portions had higher correlation coefficients for energy-adjusted vitamin C (0.80 vs. 0.60, $p < 0.0001$) and protein (0.69 vs. 0.61, $p = 0.03$) than FFQs with portion-size questions. However, it remained difficult from this review to analyze the effects of using portion-size questions. Ranking was worse for protein and vitamin C determined by FFQs that used portion-size questions instead of standard portions, and ranking improved for alcohol when FFQs used portion-size questions. FFQs slightly underestimated gender differences in energy intake. We concluded that FFQs with more items are better able to rank people according to their intake than shorter FFQs. In addition, FFQs are able to distinguish between subpopulations, even though they underestimated the magnitude of these differences. Because of these findings, researchers have the possibility to add portion size questions in the *Dutch FFQ-TOOL*™ , but can also compute nutrient intake based on standard portions.

After this in depth exploration of FFQ design, we needed a feasible approach to automatically select food items in the computer system. For the selection of food items, several methods are available. Because FFQs are used to study the relationship between food intake and

health outcomes, food items explaining a large variation in intake between persons would be most informative. Regression analyses are not suitable in this case, because the procedure overloaded the databases in the system given the complexity of the analyses. Therefore, we used a simpler procedure, called *MOM2*, that selected food items explaining most of the variance in nutrient intake, but without taking the covariance with other items into account as regression analysis does. We compared food lists made by *MOM2*, with food lists made by forward regression (chapter 3). Food items were selected for the nutrients of interest from two days of recorded intake in 3524 adults aged 25-65 y. Food lists composed to explain 80% of variance by *MOM2* were compared to those by 80% explained variance for regression. They were compared based on differences between the number and type of food items, and were evaluated on 1) the percentage of explained variance by regression analysis and 2) percentage contribution to population intake computed for the selected items on the food list. *MOM2* selected the same food items for calcium, a few more for total fat and vitamin C, and a few less for carbohydrates and dietary fiber than regression analysis. Food lists by *MOM2* based on 80% explained variance in intake covered 75% to 87% of explained variance by regression analysis and contributed 53% to 75% to the total level of population intake for the nutrients of interest. We concluded that selections by variance in nutrient intake or *MOM2* is a practical procedure for developing food lists for FFQs and selections by *MOM2* were comparable to those by regression analysis. Thus, this procedure was incorporated in the *Dutch FFQ-TOOL™*.

After that, we held a series of focus group interviews to develop clear questions to be applied in FFQs generated by the *Dutch FFQ-TOOL™*. Filling out an FFQ is a complex task, because a variety of cognitive processes is involved in filling out an FFQ. Therefore problems experienced by Dutch respondents when filling out an FFQ were studied to improve questions for FFQs (chapter 4). Based on think-aloud interviews and literature, categories of problems were identified and discussed in seven focus group meetings with 40 respondents aged 25-40 y or 40-65 y. Example questions from FFQs used in the meetings had been improved according to principles of cognitive science. Results of discussions were transcribed and analyzed by means of thematic coding. Respondents experienced problems in comprehension of questions resulting in selective answers because they tended to restrict their reports to the given examples. They also had difficulties in identifying foods, because categorization of foods was not logical to them; in aggregating frequencies of consumption of single foods into one food category; and in assessing relative frequencies of foods. This study confirmed that filling out questions of an FFQ is a complex task for respondents. Questions from an FFQ that was improved by principles of cognitive science, may still cause problems relating to question comprehension, food identification and reporting of frequencies. FFQs may be further improved by removing or changing examples to the questions, and by including cues that explain for example how to answer questions about relative frequencies. Based on these findings, we did not add any examples to questions in the *Dutch FFQ-TOOL™*. Using all the information from the focus group discussions, standard questions and answering categories were developed for use in the *Dutch FFQ-TOOL™*.

A prototype of the *Dutch FFQ-TOOL™* was developed based on the information from the above described studies (chapter 5), and used to develop a first FFQ. This FFQ was validated for the aim to rank participants according to their intake of energy and selected nutrients (chapter 6). We generated an FFQ of 118 food items including food items that explained at least 80% of variance in intake of energy, macronutrients, fatty acids, potassium and dietary fiber of a representative sample of the Dutch population according to data from the Dutch National Food Consumption Survey 1998. Intakes according to FFQs completed by 46 men and 63 women of 25-65 y were compared with those by 3-d estimated food records, by nitrogen and potassium excretions in 24-h urine checked with PABA, and by concentrations of polyunsaturated fatty acids (PUFA) determined in serum cholesterylesters. Deattenuated correlation coefficients between nutrient intake by FFQ and food records ranged from 0.35 for PUFA to 0.73 for carbohydrates. Lower limits of the validity coefficient, estimated by correlation coefficients between protein, potassium, and PUFA intakes by FFQ with nitrogen and potassium in urine and PUFA in serum, were 0.24, 0.26 and 0.33 for men, and 0.08, -0.05 and 0.27 for women, respectively. We applied the methods of triads to estimate upper limits of the validity coefficients between FFQ and true intake for protein, potassium and PUFA by combining correlation coefficients between FFQ, food records and their respective biomarkers. These were 0.50, 0.36 and 0.50 for protein, potassium and PUFA, respectively. We conclude that this first FFQ developed by the *Dutch FFQ-TOOL™* performed rather well, although improvement of generated FFQs is possible by selecting food items explaining a higher level of variance and including new foods.

To further improve the *Dutch FFQ-TOOL™* it is of utmost importance that more recent food consumption data collected over multiple days will become available as input for development of new FFQs (chapter 7). In future, the *Dutch FFQ-TOOL™* must incorporate the possibility to develop web-based FFQs that are more convenient for respondents. An important advantage of web-based FFQs over paper versions is increased comprehension of web-based FFQs if tailored pop-ups and pictures of unfamiliar foods are included to improve identification of foods. Assessing frequencies of consumption may be improved by including more frequency categories and by providing information about the relative frequency categories. We recommend to include portion size questions as both frequency and amount are important for explaining variance in intake. These portion sizes must be described in household measures as that is most convenient for respondents. Also, reporting will be more accurate, as respondents can be compelled to fill out questions, and skip questions that do not apply to them. Thus, although self-reports by web-based FFQs require the respondents to have access to a computer, they will reduce the amount of data cleaning and are likely to improve quality of the obtained information.

Because of technical constraints during the project it was impossible to automate selection of food items completely, but standard questions are automatically generated and procedures for processing FFQs are also automated. Although we were not able to automate all procedures in this system, new FFQs are developed considerably faster using this system than by a dietician that uses best-practice standard procedures. Time-consuming procedures that have become easier or faster include the following: identification of important foods for all nutrients of interest, aggregation of these foods to food items, composition of a food list from these food items, transposition of the food list into standard questions with answering categories and addition of questions about portion sizes and preparations, quality control of completed FFQs, composition of a nutrient database for processing of filled out FFQs based on the Dutch food composition table and finally quality control of processed FFQs. In addition, automating procedures increased transparency of FFQ development. Thus, we have developed a comprehensive data-based system that can be used to efficiently develop and process food frequency questionnaires (FFQs) in a standardized way to assess individual intakes for different research purposes and populations.

SAMENVATTING

Voedselvragenlijsten (FFQs) worden veel gebruikt om de voedingsinname in te schatten van grote populaties. De populariteit van voedselvragenlijsten wordt verklaard door het gemak waarmee ze verspreid worden, de mogelijkheid om voedselinname gedurende langere tijd in te schatten en de relatief lage kosten. Ze worden daarom veel gebruikt in epidemiologisch onderzoek om de relatie tussen voeding en gezondheid te bestuderen. Een goed ontworpen FFQ heeft als doelstelling om de gebruikelijke inname van bepaalde voedingsmiddelen of nutriënten te meten in een specifieke populatie. Dit betekent dat de voedingsmiddelenlijst van een FFQ aangepast moet worden aan de voedingsgewoonten van de populatie en dat nieuwe voedingsmiddelen moeten worden toegevoegd, wanneer de FFQ na enige tijd opnieuw gebruikt wordt. Helaas worden FFQs vaak opnieuw gebruikt zonder aanpassingen vanwege gebrek aan expertise, tijd of geld om dit te doen. Om het aanpassen, ontwikkelen en verwerken van FFQs te vergemakkelijken en standaardiseren, hebben wij een computersysteem ontwikkeld, de zogenaamde '*Dutch FFQ-TOOL™*'. Het algemene doel van dit proefschrift is dan ook om de wetenschappelijke basis te leggen voor de ontwikkeling van een flexibel op data gebaseerd computersysteem voor het genereren, toepassen en verwerken van FFQs voor meerdere nutriënten en doelgroepen en om dit systeem te evalueren.

Om een wetenschappelijk onderbouwd systeem te ontwikkelen, hebben we een meta-analyse van 40 validatiestudies uitgevoerd (hoofdstuk 2) om de rol van het FFQ-design, inclusief de lengte en het gebruik van portiegroottevragen te bestuderen, voor het rangschikken van individuen naar hun nutriëntinname. Verder bestudeerden we de mate waarin FFQs in staat waren om onderscheid te maken tussen subgroepen met betrekking tot verschillen in energie en eiwitinname. We zagen dat FFQs met langere voedingsmiddelenlijsten (>200 items) beter waren dan kortere FFQs in het rangschikken van individuen naar hun inname voor de meeste nutriënten; de resultaten waren statistisch significant voor eiwit, energie-gecorrigeerd totaal vet en energie-gecorrigeerde vitamine C. We vonden dat FFQs die gebruik maakten van standaardporties hogere correlatiecoëfficiënten hadden voor energie-gecorrigeerde vitamine C (0.80 vs. 0.60, $p < 0.0001$) en eiwit (0.69 vs. 0.61, $p = 0.03$) dan FFQs met portiegroottevragen. Het bleef echter lastig om op basis van deze review de effecten van het gebruik van portiegroottevragen te analyseren. Het werd lastiger om individuen naar hun vitamine C- en eiwitinname te rangschikken wanneer portiegroottevragen gebruikt werden op de FFQ, terwijl portiegroottevragen dit juist verbeterden voor alcoholinname. Verder werden de geslachtsspecifieke verschillen in energie-inname licht onderschat met FFQs. We concludeerden dat FFQs met meer items beter in staat zijn om individuen te rangschikken naar hun inname dan kortere FFQs. Verder zijn FFQs in staat om onderscheid te maken tussen subgroepen ook al worden deze verschillen onderschat. Op basis van deze bevindingen hebben onderzoekers de mogelijkheid om portiegroottevragen toe te voegen in de '*Dutch FFQ-TOOL™*', maar is het ook mogelijk om de nutriëntinname te berekenen op basis van standaardporties.

Na deze grondige bestudering van het design van FFQs, hadden we een geschikte methode nodig om voedingsmiddelen te selecteren binnen het system. Voor de selectie van voedingsmiddelen zijn verschillende methodes beschikbaar. Omdat FFQs gebruikt worden om de relatie tussen voedselinname en gezondheidsuitkomsten te bestuderen, zouden voedingsmiddelen die een groot deel van de variantie in inname tussen personen verklaren het meest informatief zijn. Regressieanalyses zijn niet geschikt in dit geval, omdat de procedure de databases in het systeem doet overlopen vanwege de complexiteit van de analyses. Daarom gebruikten we een simpelere procedure die *MOM2* wordt genoemd. Deze procedure selecteert voedingsmiddelen die het grootste deel van de variantie in inname verklaren zonder rekening te houden met de covariantie van andere items zoals een regressieanalyse doet. We vergeleken een voedingsmiddelenlijst gemaakt met behulp van *MOM2*, met lijsten die met regressieanalyse waren gemaakt (hoofdstuk 3). Voedingsmiddelen werden geselecteerd voor de doelwitnutriënten uit tweedaagse dagboekjes bijgehouden door 3524 volwassenen van 25-65 jaar. Voedingsmiddelenlijsten op basis van 80% variantie met behulp van *MOM2* werden vergeleken met voedingsmiddelen die 80% van de variantie verklaarden met behulp van regressie. Ze werden vergeleken op basis van het aantal en type voedingsmiddelen en werden geëvalueerd op 1) percentage verklaarde variantie met behulp van regressieanalyse en 2) percentage bijdrage aan het niveau van inname in de populatie berekend voor de geselecteerde items op de voedingsmiddelenlijst. *MOM2* selecteerde dezelfde voedingsmiddelen voor calcium, een paar meer voor totaal vet en vitamine C, en een paar minder voor koolhydraten en voedingsvezel dan de regressieanalyse. Voedingsmiddelenlijsten op basis van 80% *MOM2* bereikten 75% tot 87% verklaarde variantie volgens regressieanalyse en droegen bij aan 53% tot 75% van het niveau van inname in de populatie voor de verschillende nutriënten. We concludeerden dat selecteren op basis van variantie in inname, oftewel *MOM2*, een praktische procedure is voor het ontwikkelen van voedingsmiddelenlijsten voor FFQs en de resultaten waren vergelijkbaar voor *MOM2* en regressieanalyses. Deze procedure werd daarom ingebouwd in de '*Dutch FFQ-TOOL™*'.

Daarna hielden we een aantal focusgroepinterviews om duidelijke vragen te ontwikkelen voor gebruik in FFQs gegenereerd door het computersysteem. Het invullen van een FFQ is een lastige taak, omdat een aantal verschillende cognitieve processen betrokken is bij het invullen van een FFQ. Daarom bestudeerden we problemen die Nederlandse respondenten ondervonden bij het invullen van een FFQ om vragen voor FFQs te verbeteren (hoofdstuk 4). Op basis van 'think aloud' interviews en de literatuur werden categorieën van problemen geïdentificeerd en bediscussieerd in zeven focusgroepbijeenkomsten met 40 respondenten van 25-40 jaar of 40-65 jaar. Voorbeeldvragen die gebruikt werden tijdens de bijeenkomsten kwamen uit FFQs die al verbeterd waren op cognitief gebied. Resultaten van deze discussies werden uitgeschreven en geanalyseerd met behulp van thematisch coderen. Respondenten begrepen sommige vragen niet goed waardoor ze hun antwoorden beperkten tot één of enkele van de gegeven voorbeelden. Ze hadden ook problemen met de identificatie van voedingsmiddelen omdat de indeling van voedingsmiddelen niet logisch voor hen was; met het schatten van de frequentie van consumptie wanneer dit gevraagd werd voor een aantal voedingsmiddelen tegelijk en met het schatten van relatieve frequenties waarmee bepaalde voedingsmiddelen werden gegeten. Dit onderzoek bevestigde dus dat het invullen van een FFQ een lastige taak is voor respondenten. FFQs zouden verder verbeterd kunnen worden door het verwijderen of veranderen van voorbeelden bij de vragen en door aanwijzingen toe te voegen die aangeven hoe vragen over bijvoorbeeld relatieve frequenties kunnen worden ingevuld. Op basis van deze bevindingen besloten we om geen voorbeelden toe te voegen aan de standaardvragen in de '*Dutch FFQ-TOOL*™'. Met behulp van alle informatie uit de focusgroepinterviews ontwikkelden we standaardvragen en antwoordcategorieën voor gebruik in de '*Dutch FFQ-TOOL*™'.

Een prototype van het systeem werd ontwikkeld op basis van de informatie uit de hierboven beschreven onderzoeken (hoofdstuk 5) en hiermee werd een eerste FFQ gemaakt. Deze FFQ werd gevalideerd op de doelstelling om volwassenen te rangschikken naar hun inname van energie en geselecteerde nutriënten (hoofdstuk 6). We genereerden een FFQ van 118 items door voedingsmiddelen op te nemen die minstens 80% van de variantie in energie, macronutriënten, vetzuren, kalium en voedingsvezel verklaarden in een representatieve groep uit de Nederlandse bevolking met behulp van data van de Nederlandse Nationale Voedselconsumptiepeiling 1998. Inname volgens FFQs ingevuld door 46 mannen en 63 vrouwen van 25-65 jaar werden vergeleken met hun inname volgens een 3-daags voedseldagboekje en stikstof- en kaliumuitscheiding in 24-uurs urine, gecontroleerd op compleetheid met behulp van PABA-tabletten en met concentraties van meervoudig onverzadigde vetzuren (PUFA) bepaald in serum cholesterylesters. Gedeattenueerde correlatiecoëfficiënten tussen nutriëntinname volgens FFQ en volgens het voedseldagboekje lagen tussen de 0.35 voor PUFA en 0.73 voor koolhydraten. Ondergrenzen van de validiteitscoëfficiënt, geschat door de correlatie tussen eiwit, kalium en PUFA-inname volgens FFQ met stikstof en kalium in urine en PUFA in serum, waren respectievelijk 0.24, 0.26 en 0.33 voor mannen en 0.08, -0.05 en 0.27 voor vrouwen. We gebruikten de triadmethode om de bovengrens van de validiteitscoëfficiënt te schatten tussen FFQ en de werkelijke inname voor eiwit, kalium en PUFA door de correlaties tussen FFQ, voedseldagboekje en de respectievelijke biomerkers te combineren. Deze waren respectievelijk 0.50, 0.36 en 0.50 voor eiwit, kalium en PUFA. We concludeerden dat de validiteit van de eerste FFQ ontwikkeld met de '*Dutch FFQ-TOOL*™' redelijk goed was, hoewel verbetering van de gegenereerde FFQs mogelijk is door het opnemen van voedingsmiddelen die een hoger percentage variantie verklaren en door voedingsmiddelen op te nemen die nieuw op de markt zijn gekomen.

Om de '*Dutch FFQ-TOOL™*' nog verder te verbeteren is het uiterst belangrijk dat recentere voedselconsumptiegegevens, verzameld gedurende meerdere dagen, beschikbaar worden als bronbestand voor de ontwikkeling van nieuwe FFQs (hoofdstuk 7). In de toekomst moet de '*Dutch FFQ-TOOL™*' de mogelijkheid krijgen om online vragenlijsten te ontwikkelen die gebruiksvriendelijker zijn voor respondenten. Een belangrijk voordeel van online vragenlijsten ten opzichte van papieren versies is een groter begrip doordat pop-ups en foto's van onbekende voedingsmiddelen kunnen worden bekeken om zo de identificatie van voedingsmiddelen te verbeteren. Het inschatten van consumptiefrequenties kan verbeterd worden door meer frequentiecategorieën op te nemen en door uitleg te geven bij de relatieve frequentiecategorieën. We raden aan om portiegroottevragen op te nemen, omdat zowel frequentie als hoeveelheid belangrijk zijn voor het verklaren van de inname. Deze porties moeten omschreven worden in huishoudelijke maten, aangezien dat het makkelijkste is voor respondenten. Ook wordt rapporteren nauwkeuriger als respondenten gedwongen worden om vragen in te vullen en zij vragen kunnen overslaan die niet relevant zijn. Kortom, ook al vereisen online vragenlijsten dat respondenten toegang hebben tot een computer, ze zullen ervoor zorgen dat er minder data hoeft te worden opgeschoond en daardoor de kwaliteit van de verzamelde data verbeteren.

Vanwege technische beperkingen tijdens het project was het niet mogelijk om de selectie van voedingsmiddelen volledig te automatiseren, maar standaardvragen worden automatisch gegenereerd en procedures voor het verwerken van FFQs zijn ook geautomatiseerd. Hoewel we niet in staat waren om alle procedures te automatiseren in het systeem, worden nieuwe FFQs veel sneller ontwikkeld met behulp van dit systeem dan door een diëtiste die de beste standaardprocedures volgt. Tijdrovende elementen die sneller of gemakkelijker zijn geworden zijn de volgende: identificatie van belangrijke voedingsmiddelen voor alle gekozen nutriënten, aggregatie van deze voedingsmiddelen tot items, samenstelling van een voedingsmiddelenlijst van deze items, het omzetten van deze lijst in standaardvragen met antwoordcategorieën en toevoeging van vragen naar portiegroottes en bereidingswijze, kwaliteitscontrole van ingevulde FFQs, samenstelling van een nutriëntentabel voor het verwerken van de ingevulde vragenlijsten gebaseerd op de Nederlandse voedingsmiddelentabel en uiteindelijke kwaliteitscontrole van de verwerkte FFQs. Daarnaast heeft het automatiseren van deze procedures ervoor gezorgd dat het ontwikkeltraject van FFQs transparanter werd. We hebben dus een eenvoudig op data gebaseerd systeem ontwikkeld dat gebruikt kan worden om efficiënt FFQs te ontwikkelen en verwerken op een gestandaardiseerde manier om individuele inname te meten voor verschillende onderzoeksvragen en populaties.

Dankwoord

Dit onderzoek heb ik niet alleen uitgevoerd, daarom wil ik iedereen die me daarbij geholpen heeft hartelijk bedanken.

Beste Jeanne, ik waardeer het enorm dat je me de kans hebt gegeven om als AIO met dit project aan de slag te gaan. Tijdens het onderzoek was je altijd erg betrokken, dat heb ik zeer gewaardeerd. Deze betrokkenheid bleek onder andere uit het feit dat zelfs je zoon, Niels, die wiskunde studeert, werd ingezet om ervoor te zorgen dat de ICT-mannen ook begrepen hoe onze formules in het computersysteem gezet moesten worden. Goed voedselconsumptieonderzoek ligt je erg aan het hart, en het frustreert je dan ook dat je regelmatig bestookt wordt met telefoontjes van onderzoekers die wel 'even' de invloed van voeding willen onderzoeken. Hopelijk worden ze zich bewust van de complexiteit van voedselconsumptieonderzoek, met de mooie tool die we ontwikkeld hebben.

Beste Pieter, ik waardeer je kritische en creatieve bijdrage aan mijn onderzoek. Soms kostte het enige moeite om een gaatje te vinden in je agenda, maar daarna konden we heerlijk discussiëren en daagde je me uit om nog wat diepgaander na te denken. Beste Piet, als tweede promotor was jij vanaf een afstand (Maastricht) betrokken bij het onderzoek, ik heb aan het eind veel gehad aan je kritische inbreng, mijn dank hiervoor.

Bovendien was mijn AIO-project een samenwerkingsverband tussen RIVM, TNO, Universiteit Maastricht en Wageningen Universiteit. Marga Ocké, Pieter Dagnelie, Henny Brants, Sandra Bausch-Goldbohm en Margje Jansen, ik wil jullie bedanken voor jullie tijd en de ontelbare reisjes naar Wageningen. Jullie zorgden voor een breed draagvlak voor de FFQ-tool en ik hoop dat jullie er veel profijt van gaan hebben. Daarnaast waren jullie kritische co-auteurs, die hielpen om mijn artikelen naar een hoger niveau te tillen, mijn dank hiervoor.

Saskia, we hebben veel samengewerkt aan de ontwikkeling van het systeem en je was een zeer fijne collega. Jouw ervaring met voedselvragenlijsten was zeer belangrijk voor het systeem. Vele tabellen hebben we gevuld met onder andere productgroepen, standaardvragen, antwoordcategorieën en portiegroottes. Hierbij konden we regelmatig ook rekenen op de hulp van Henny Brants. Daarnaast hebben we vele berekeningen moeten testen. Groot was de frustratie toen bleek dat we niet op tijd een vragenlijst konden genereren voor de validatiestudie. Maar gelukkig kwamen jij en Jeanne me gewoon helpen op Goede Vrijdag. Fijn dat jij de coördinatie van het systeem nu van me hebt overgenomen.

Karin, ik heb ook erg veel gehad aan jouw ondersteuning gedurende de focusgroepinterviews en de validatiestudie. Tijdens deze onderzoeken hielp je me vooral mijn ideeën te vertalen in hele concrete planningen en afspraken, dat heb ik erg gewaardeerd. Ook ben je druk bezig geweest met dagboekjes en vragenlijsten analyseren, waarbij ook mijn studenten moesten worden aangespoord. En ook achter de schermen heb je geholpen bij de ontwikkeling van het systeem. Daarnaast wil ik ook de andere diëtistes bedanken die hun steentje hebben bijgedragen aan de vulling van tabellen voor het systeem.

Els, Jeanne, Saskia, Anouk and Sandra, thanks for the nice time in Washington during the International Conference on Dietary Assessment and Activity Methods.

Joost en Sven en ook Herbert Voordes en alle andere medewerkers van I-design wil ik bedanken voor het daadwerkelijke bouwen van het computersysteem. Het duurde even voordat we dezelfde taal spraken en volledig begrepen wat de ander bedoelde met een bepaald begrip. Maar dat was altijd 'in basis geen probleem' volgens Joost. Het navragen van gegeten voedingsmiddelen bleek een stuk complexer dan jullie hadden gedacht, maar dankzij voortschrijdend inzicht ligt er nu een erg mooi systeem.

Niels Duif, je vormde een belangrijke schakel tussen de ICT-mannen en de dames die zich met het voedselconsumptieonderzoek bezig hielden. Verder wil ik ook Jack Thissen bedanken voor het meedenken over het geautomatiseerd selecteren van voedingsmiddelen voor voedselvragenlijsten.

Joke van Lemmen, jij vond mijn problemen erg interessant en ze bleken geschikt om met "operations research" aan te pakken. We wilden de maximale hoeveelheid informatie verzamelen met een minimale hoeveelheid voedingsmiddelen. Ellen Slegers ging hiermee aan de slag en werd zelfs beloond met een scriptieprijs voor haar werk. Ik wil jullie allebei erg bedanken voor jullie inspanningen, helaas moeten we deze problemen voorlopig nog steeds oplossen met de 'expert opinion'.

Wija, in eerste instantie begeleidde je mij als promotor. Maar al dat vergaderen na je pensionering werd je toch wat teveel en je besloot je rol aan Pieter over te dragen. Graag zou ik het systeem nog eens aan je demonstreren.

Reint-Jan Renes, voor hulp op het gebied van het uitvoeren van kwalitatief onderzoek kwamen we bij jou terecht. Je dwong me steeds om mijn onderzoeksvragen concreter te definiëren en dat is maar goed ook, want het beantwoorden bleek vrij gecompliceerd.

Paul Hulshof, je hebt me geholpen bij de labanalyses, vooral het meten van de suikers in urine was een grote uitdaging. Daarnaast wil ik graag Betty, Nhien, Tineke en alle andere labmedewerkers bedanken voor het uitvoeren van analyses. Verder wil ik Jantien Takens en Marga van der Steen bedanken voor hun ondersteuning. Lucy Ockma en de andere prikdames wil ik bedanken voor het bloedprikken tijdens de validatiestudie.

Verder wil ik de deelnemers van de FOCUS-studie bedanken voor het discussiëren over het invullen van een voedselvragenlijst. De deelnemers van de VALNED-studie wil ik bedanken voor het invullen van vragenlijsten en dagboekjes over voeding en daarnaast ook voor het afnemen van een beetje bloed en het verzamelen van 24-uurs urine. Een hele klus!

Karen Zweers en Gabriëlle wil ik bedanken voor hun secretariële ondersteuning. Het was altijd een onderneming om weer een vergadering in te plannen.

Tijdens mijn AIO-tijd heb ik verschillende studenten mogen begeleiden bij hun afstudeervak:
Marieke van Bakel, Tobias van den Berg, Geja Borsch, Ellen Slegers, Miranda Ruigers, Frederike Brouwer en Marieke Heinst. Ik heb veel geleerd tijdens de periode dat ik jullie mocht begeleiden en ik wil jullie bedanken voor de praktische hulp en jullie enthousiasme.

Daarnaast heb ik een erg plezierige tijd gehad in Wageningen dankzij mijn collega's: Gertrude: fijne en sportieve collega vanaf dag 1, Suzanne Kleikers mijn 1^{e} kamergenoot, Akke Botma mijn 2^{e} kamergenoot, Gerda Pot: veel gezelligheid, Renate Winkels: altijd goedgehumeurd, Anneleen en Olga, het was fijn van jullie ervaring als postdocs te mogen profiteren, en ook Pleunie, Sanne, Adrienne, Linda, Janette, Simone, Mariëlle, Cora, Nicolien, Marieke en Mariken bedankt voor de fijne tijd. Tussen het werken door heb ik vele rondjes gerend met Ondine, Gertrude, Pascalle, Elise U., Esmée en Rianne. Het was fijn al dat computeren af te wisselen met rennen door het Wageningse bos. Een hele bijzondere ervaring was de PhD Tour 2005 naar Engeland, Schotland en Ierland: reisgenoten bedankt voor de gezelligheid.

Ondine, ik waardeer het zeer dat je mijn paranimf wilt zijn. We waren allebei gek op hardlopen en we hebben dan ook vele wedstrijdjes gerend samen met als hoogtepunten onze eerste halve marathon in Utrecht en de de City-Pier-City in Den Haag. Ook vond ik het erg gezellig dat we samen op congres waren in Kopenhagen.

Sandra, I really appreciate that you are my other paranimf. You are a very nice colleague, very good at cocktail shaking, and I really enjoyed all our discussions about our validation studies and many papers. A highlight was our visit to the 7th ICDAM conference in Washington. You are the most international person I have ever met and I wish you all the best for finalizing your PhD project.

Verder wil ik ook mijn vrienden bedanken voor hun interesse in mijn onderzoek, hopelijk heb ik nu weer meer tijd voor jullie.

Hydrofielers, Anne, Bas en Joyce, Hilde, Linda en Dimitri, Luc en Hanneke, Heleen, Joppe en Ingrid, Edwin, Paula, Elske, Robert en Ellen, zwemmen doen we niet meer, maar ik hoop dat er nog vele avondjes in Jos, etentjes en Ardennenweekenden zullen volgen.

Fons, ik vind het geweldig dat je me hebt geholpen met de vormgeving van mijn proefschrift! Marijn, Marlieke en Monique, ik hoop dat er nog vele gezellige MMM-dates gaan volgen. Djamilla, Thijs, Jurgen en Fiona, genieten van lekker eten is ook belangrijk!

Ook mijn familie was altijd zeer geïnteresseerd. Mijn ouders, Trienke en Menso, stonden altijd klaar met een luisterend oor en goede adviezen. Helaas Arjen, ik zal de eerste Dr. Molag zijn, gelukkig heb je je plek nu weer gevonden. En Kevin bedankt voor je steun en dat je me altijd bijtijds meeneemt op vakantie.

Marja Molag


About the author

Curriculum vitae

List of publications

Overview of completed training activities

Colofon

CURRICULUM VITAE

Marja Molag was born on the 5th of March 1981 in Groningen the Netherlands. She moved to Apeldoorn with her parents and her brother when she was six years old. After completing secondary school at the 'Koninklijke scholengemeenschap Apeldoorn' in 1999, she studied Biomedical Health Sciences at the Radboud University, Nijmegen. First she focused on toxicology, and during her first MSc thesis she investigated the influence of parenteral lipid emulsions on the oxygen-radical production of the human neutrophil. In an internship at TNO she did a literature study on the influence of toxic smoke from fires in tunnels on the evacuation of people from tunnels. Afterwards, she decided to focus more on epidemiology and she wrote her final thesis at the institute of public health and the environment (RIVM) in Bilthoven, the Netherlands. She estimated the HIV prevalence in the Netherlands and studied determinants of sexual risk behavior among migrants in the Netherlands. She obtained her Master's degree in 2004 and she started as a PhD fellow at the Division of Human Nutrition of Wageningen University, the Netherlands. She started a project that aimed to develop a computer system that could generate, apply and process food frequency questionnaires for dietary assessment in the Netherlands. This project was a collaboration between experts in the field of dietary assessment from Wageningen University, RIVM, TNO and Maastricht University. She is currently working as an advisor at the 'Orde van Medisch Specialisten' in Utrecht where she will develop evidence-based guidelines.

PUBLICATIONS

Full papers

Molag ML, de Vries JHM, Ocké MC, Dagnelie PC, van den Brandt PA, Jansen MCJF, van Staveren WA, van `t Veer P. Design characteristics of Food Frequency Questionnaires. American Journal of Epidemiology 2007;166(12):1468-78

Ngo J, Engelen A, Molag M, Roesle J, Serra-Majem L. A review of the use of information and communication technologies for dietary assessment. British Journal of Nutrition 2009;101(Suppl 2):S102-S113.

Molag ML, de Vries JHM, Duif N, Ocké MC, Dagnelie PC, Goldbohm RA, van 't Veer P. Selecting informative food items to compose food frequency questionnaires: Comparison of procedures. [submitted]

Molag ML, Renes RJ, Meijboom S, Brants H, Dagnelie PC, de Vries JHM. Options for improvement of FFQs using cognitive interviewing. [submitted]

Molag ML, de Vries JHM, Meijboom S, Ocké MC, Brants HAM, Goldbohm RA, Dagnelie PC, van den Brandt PA, van 't Veer P. The Dutch FFQ-TOOLTM: development and use of a computer system to generate and process FFQs. [submitted]

Molag ML, van 't Veer P, Ocké MC, Meijboom S, Hulshof PJM, Goldbohm RA, Dagnelie PC, de Vries JHM. Validation of a food frequency questionnaire developed with a new tool for automated development and processing of FFQs against biomarkers and 3-d food records. [submitted]

Abstracts

Molag ML, van 't Veer P, Ocké MC, Meijboom S, Goldbohm RA, Dagnelie PC, de Vries JHM. Validation of a first FFQ developed using a computer system against independent biomarkers and three day food diaries. European Journal of Clinical Nutrition 2009;63 Suppl;S6. ISSN 0954-3007.

Molag ML, de Vries JHM, van 't Veer P, van den Brandt PA, Dagnelie PC, Ocké MC, Jansen MCJF, van Staveren WA . The influence of design characteristics of food frequency questionnaires on their validity to assess energy intake. The IEA-EEF European Congress of Epidemiology 2006, Epidemiology and Health care practice. European Journal of Epidemiology 2006;21 Suppl:91. ISBN 0393-2990.

OVERVIEW OF COMPLETED TRAINING ACTIVITIES

Discipline specific activities

- Seventh International Conference on Diet and Activity Measurements, ICDAM, Washington, US, 4-7th of June 2009
- Meeting Werkgroep Voedingsgewoonten (WEVO), Utrecht 2008
- Sixth International Conference on Dietary Assessment Methods, ICDAM, Copenhagen, Denmark, 27-29th of April 2006
- Symposium Nutrition and ageing, WUR/ NZO, Wageningen 2006
- Nutrition and lifestyle epidemiology, VLAG, Wageningen 2005
- Annual meeting of the Netherlands Epidemiology Society (WEON), Groningen 2008, Maastricht 2007, Utrecht 2006, Wageningen 2005
- Annual meeting NWO Nutrition, Deurne 2008, 2007 and 2006, Arnhem 2005

General courses

- Wageningen Nutritional Sciences Forum, Division of Human Nutrition, Wageningen University, Arnhem 2009
- Masterclass Nutrition Communication: new approaches towards effective health promotion, VLAG, Wageningen 2009
- Training in Food consumption research, Dietetics, Wageningen 2008
- Talent day 'Networking' and 'Career perspectives', NWO, Utrecht 2007
- Scientific Writing, CENTA, Wageningen 2007
- Masterclass Diet and Cancer, VLAG, Wageningen 2007
- Talent Class 'Negotiating', NWO, The Hague 2006
- PhD introduction course, VLAG, Bilthoven 2005

Optional courses and activities

- Concept and methods in epidemiology, Wageningen 2008 - 2009
- Literature group 'Oldsmobiles', Wageningen 2004 - 2009
- Research presentations Division of Human Nutrition, Wageningen 2004 - 2009
- Epidemiology research meetings, Wageningen 2004 - 2009
- Literature group 'Journal Club' , Wageningen 2004 - 2007
- PhD study tour to universities and research institutes in England, Scotland and Ireland, 2005
- Exposure assessment in Nutrition and Health Research, Wageningen 2005
- Preparation research proposal, Wageningen 2005

The project described in this PhD thesis was a cooperation between Wageningen University, Wageningen; National Institute for Public Health and the Environment (RIVM), Bilthoven; TNO Quality of Life, Leiden; and Department of Epidemiology, Maastricht University.

This study was supported by the Netherlands Organization for Health Research and Development (ZonMw, grant 40-00506-98-04-005).

The author gratefully acknowledges financial support for the printing of this thesis from Wageningen University.

Cover page, photographs and title pages of chapters: Fons van Bindsbergen
Photo editing: Heleen Visser
Layout: Renate Siebes, Proefschrift.nu
Printed by: Labor Grafimedia BV, Utrecht