Measurement errors in dietary assessment using duplicate portions as reference method

Laura Trijsburg

Thesis committee

Promotor
Prof. Dr P. van 't Veer
Professor of Nutrition and Epidemiology
Wageningen University

Co-promotors
Dr A. Geelen
Assistant professor, Division of Human Nutrition
Wageningen University

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

Other members
Prof. Dr R.F. Witkamp, Wageningen University
Dr H. van der Voet, Wageningen University
Dr M.C. Ocké, National Institute for Public Health and the Environment (RIVM), Bilthoven
Dr E. Corpeleijn, University of Groningen

Measurement errors in dietary assessment using duplicate portions as reference method

Laura Trijsburg

To Martin

Abstract

Background: As Food Frequency Questionnaires (FFQs) are subject to measurement error, associations between self-reported intake by FFQ and outcome measures should be corrected for measurement error with data from a reference method. Whether the correction is adequate depends on the characteristics of the reference method used in the validation study. The duplicate portion method (DP), compared to the often used 24h recall (24hR), seems a promising reference method as correlated errors between FFQ and DP, such as memory bias, errors in portion size estimations and food composition databases, are not expected.

Aim: This thesis aimed to determine the validity of the DP compared to the 24hR as a reference method for FFQ validation. The second aim was to explore the validity of nutrient densities for DP, 24hR and FFQ. The third aim was to determine the factors associated with misreporting of energy, protein and potassium as estimated by DP, 24hR and FFQ.

Methods: Within the DuPLO-study, a Dutch validation study which is part of the NQplus study, two DPs, two FFQs, two blood and urinary biomarkers and one to fifteen 24hRs (web-based and/or telephone-based) were collected in 198 subjects, within 1.5 years. Also, one or two doubly labelled water measurements were available for 69 participants. Multivariate measurement error models were used to assess proportional scaling bias, error correlations with the FFQ, validity coefficients and attenuation factors. Furthermore linear regression analysis was used to determine the association between misreporting and various factors.

Results: The DP was less influenced by proportional scaling bias, had lower correlated errors with the FFQ and showed higher attenuation factors than the 24hR for potassium, sodium and protein. Also, the DP seemed a better reference method than the 24hR for the assessment of validity coefficients for the FFQ for various fatty acids. The attenuation factors for the FFQ, using either the DP or 24hR as reference method, agreed reasonably well. Furthermore, the DP showed, when using plasma fatty acids as reference, slightly better ranking of participants according to their intake of n-3 fatty acids (0.33) and the n-3/LA ratio (0.34) than the 24hR (0.22 and 0.24, respectively). Less group level bias was observed for protein and sodium densities compared to their absolute intakes for FFQ, 24hR and DP, but not for potassium. Overall the validity coefficients and attenuation factors for DP, 24hR and FFQ did not improve for nutrient densities compared to absolute intakes, except for the attenuation factor for sodium

density. Lastly, BMI proved to be the most consistent determinant associated with misreporting (group level bias) of energy, protein and potassium for DP, 24hR and FFQ. Men tended to underreport protein by the DP, FFQ and 24hR and persons of older age underreported potassium but only by the 24hR and FFQ. Other explorative determinants did not show a consistent association with misreporting of energy or nutrients by the different dietary assessment methods.

Conclusion: With respect to error correlations and attenuation factors the DP performed slightly better than the 24hR as a reference method for validating FFQs in epidemiological research. Furthermore, the use of nutrient densities does not necessarily improve the validity of the dietary intake estimates from DP, 24hR and FFQ. Moreover, it was shown that BMI is an important determinant of misreporting of energy, protein and potassium for these three assessment methods.

Table of Contents

Chapter 1

General Introduction

Background

In nutritional epidemiology the associations between diet and disease are studied, whereby dietary intake is often assessed by a food frequency questionnaire (FFQ). The FFQ is relatively cheap, it assesses usual intake, it can be used to assess the past diet and is easy to administer (1). Evidence shows that FFQs are subject to considerable measurement error (2-6) thereby attenuating diet-disease associations. Understanding the impact of measurement errors is of importance to the interpretation and correction of nutritional epidemiological outcomes The research described in this thesis explores the nature and extent of measurement error in a FFQ and the potential of the duplicate portion method (DP) as a reference method for FFQ validation.

Types of measurement error affecting dietary assessment

Measurement errors are generally classified as random errors and systematic errors. Random errors decrease the precision of the measurement while systematic errors decrease the accuracy of the method (1). Random errors are due to for example day-to-day variation, and they do not influence the mean intake of a population but attenuate (weaken) the strength of the association between diet and disease. Systematic errors result from an under- or overestimation of the mean intake. Proportional scaling bias is a form of systematic error on the group level and indicates the extent of under- or overestimation related to the true intake. Proportional scaling bias for example takes place when persons with high intakes tend to overestimate their intake more than persons with low intakes. Another systematic error on the individual level (but the error is random on the group level), person specific bias, is the difference between an individual's reported intake, and the persons true usual intake, after taking constant bias and proportional scaling bias into account (7). It differs between persons and person characteristics possibly influence the extent of person specific bias. To illustrate the difference between proportional scaling bias and person specific bias an example is given: it is known that obese people tend to underestimate their energy intake to a larger extent than non-obese people (8). As obese people have a higher intake and a higher underestimation than people with a normal weight there is a group level of underestimation by obese people which is indicated by proportional scaling bias, in addition every obese individual also deviates from his/her true usual intake, and this is indicated by the person specific bias.

Awareness of the presence of measurement errors prompted researchers to incorporate validation studies in epidemiological studies. Of special importance in such a study is the correlated error between a reference method and the

method to be validated. This error occurs when the two dietary assessment methods share errors, for example because they make use of the same food composition table with imperfect data or of the same erroneously estimated portion sizes, or they suffer from similar memory difficulties when food intake is reported (1). Also, due to short term fluctuations of diet over time, if measurements of the methods are taken close in time, random variation of the two measurements may add to the degree of correlated error.

Correction and identification of measurement error

Measurement errors can be (partly) corrected for by statistical methods, and certain methods can be used to quantify the different measurement errors. Some evidence exists that using the nutrient density approach (5, 6, 9) decreases the error in the dietary intake variable for FFQs. Furthermore adjusting for person characteristics could diminish the measurement error in the dietary intake variable (10).

Statistical methods to correct and identify measurement errors

When only random errors are expected in dietary intake data, taking large numbers of repeated measurements or taking few replicated and subsequently correcting for random error will suffice. Various methods such as the National Cancer institute (NCI) method (11) and Iowa State University (ISU) method (12, 13) have been developed to correct for random error.

In addition, systematic errors can be dealt with by obtaining extra information from a validation study in which a superior reference method is used to assess dietary intake. Regression calibration (14), where the reference method is regressed on the main method, can be used to correct dietary intake data assessed by the main method. With the method proposed by Rosner (15), an attenuation factor can be derived, which is estimated as the slope of true on observed dietary intake. An attenuation factor can be used to correct the association between diet and disease and provides information about the extent to which diet-disease associations are attenuated (weakened) by measurement error. If no proportional scaling bias is expected, the validity coefficient can be used as a correction factor instead of the attenuation factor. The validity coefficient is also used to measure the loss of statistical power to detect a diet-disease association, and indicates how well a method is able to rank participants according to their unknown true dietary intake. Among others, the method of triads (16) can be used to obtain a validity coefficient, although a second reference method with uncorrelated errors with the other two methods is necessary.

Dietary measurement error models (10, 17, 18) can be used to calculate attenuation factors and validity coefficients. Various models are known (19) and in this thesis we used a model where the magnitude of the different measurement errors (as described in the previous section) can be quantified and also the correlation between errors of different dietary assessment methods can be calculated from the measurement error model. The set-up of the measurement error models used in this thesis is described below. An identifiable measurement error model should include one reference method Z which is assumed to 1) be unbiased and 2) have errors that are not correlated with the errors from the main method (17, 20), in our case the main method is the FFQ.

$$Zij = T + \varepsilon_{Zij}$$

Where i denotes the person and j indicates the occasion, T indicates the true unknown intake and ε_{Zij} is the random error with mean zero and constant variance.
The basic model for the (error prone) FFQ, can be specified as follows:

$$Xij = \alpha_X + \beta_X T + w_{xi} + \varepsilon_{Xij}$$

Where α_X expresses the constant bias for the FFQ and β_X is the proportional scaling bias, the person specific bias of the method is given by w_{xi} and ε_{Xij} is the random error with mean zero and constant variance.
To allow for the quantification of correlated error between the FFQ and a third dietary assessment method, a third equation should be included in the measurement error model:

$$Yij = \alpha_Y + \beta_Y T + w_{Yi} + \varepsilon_{Yij}$$

Where α_Y expresses the constant bias for method Y and β_Y is the proportional scaling bias, the person specific bias of the method is given by w_{Yi} and ε_{Yij} is the random error with mean zero and constant variance. Correlated error between method X and Y can be calculated from the measurement error model estimates.

Reference methods for validation studies
In nutritional epidemiology often multiple 24h recalls (24hRs) are used as the reference method for FFQ validation. The 24hRs are probably less affected by under- or overreporting than FFQs (21) however correlated errors between FFQ

and 24hR have been demonstrated for protein and energy intake (3, 4) and the 24hR also suffers from bias (2-6).

The duplicate portion method (DP) is not memory based, portion sizes do not need to be estimated and data from food composition databases is not necessary. This eliminates major error sources expected to result in less correlated error with the FFQ. For a DP a participant collects an equal portion of everything he/she eats and drinks over one or more days and the nutrient content will be chemically analysed. In a Swedish study where the DP was sampled on the same day as urine, the DPs were found to accurately measure sodium and potassium intake (22). The DP method also seemed to be a favourable method to measure contaminants (23), trace elements (24-26) and fatty acid composition (27, 28) since valid data about these components is in some cases lacking in food composition databases. On the other hand, collections of DPs may lead to reactivity bias demonstrating a change in the respondents' intake on the collection day, mostly resulting in underestimation of intake (29-32).

Table 1: Characteristics of different dietary assessment methods

Characteristic	FFQ	24hR*	DP	Biomarker
Food composition table needed	yes	yes	no	no
Memory based	yes	yes	To a lesser degree	To a lesser degree
Portion size estimation needed	yes	yes	no	no
Use of frequencies	yes	no	no	no
Altering of intake due to monitoring	no	no	yes	no
Laboratory measurements	no	no	yes	yes
Questions	closed	open	NA	NA
Assesses	usual intake	actual intake	actual intake	Timeframe depends on nutrient/ biological fluid

*Small differences between a web-based 24hR and telephone-based 24hR exist, e.g. questions for the web-based 24hR are semi-open, where for a telephone-based 24hR the questions are completely open.

Like DPs, biomarkers are not memory based, and do not require portion size estimation and food composition databases. Errors in biomarkers are mostly due to physiological issues and are therefore expected to be independent from errors in a FFQ (33). Biomarkers can be defined as biochemical indicators of dietary intake or nutritional status from e.g. urine or blood. However, objective biomarkers that directly reflect nutrient intake, called recovery biomarkers, are only known for a limited number of nutrients. Recovery biomarkers, measured in

urine, have a well-known relationship between intake and excretion and are known for energy (34), protein (35, 36), sodium (1, 37) and potassium (38). The second class of biomarkers often measured in serum or plasma, concentration biomarkers, for e.g. fatty acids are suitable to evaluate the ranking of participants according to their intake but this biomarker group is not suitable to estimate the absolute level of dietary intake (39). Table 1 lists the characteristics of the reference methods mentioned above, for comparability the FFQ is also included in the table.

In the past decades, research about the impact and structure of measurement errors for the nutrients for which a recovery biomarker is available has been performed (2-6, 40, 41). To extend research about the measurement error structure to other nutrients, suitable reference methods have to be identified. The DP seems a promising reference method because of its assumed independent errors with the FFQ. However research about the suitability of the DP as a reference method for FFQ validation is lacking.

Nutrient densities and correction for measurement error

A common approach in nutritional epidemiology is to focus on dietary composition (relative contribution of a specific nutrient to total energy intake) by the use of nutrient densities. Data from a large American pooling project indicated that protein, potassium and sodium densities showed higher validity compared to the absolute nutrient intakes for the FFQ but this was not so notable for the 24hR (5, 6). However, there is also evidence that showed the opposite (42), which could be due to a large amount of measurement error in the estimation of energy intake (43). Furthermore it has been observed that proportional scaling bias increased for nutrient densities for protein (3). Eating habits differ per country and population, therefore a FFQ cannot just be applied in another population group or a different study (44). To what extent nutrient densities influence measurement errors for a FFQ specifically designed for the Dutch population has never been assessed and would add to our understanding of measurement errors in our FFQ but also for FFQs in general.

Determinants of misreporting and correction for measurement error

Identifying determinants associated with misreporting of dietary intake is essential and could help to improve the assessment of dietary intake and correction methods for errors in dietary intake (45). Strong evidence exists for the association between BMI and misreporting, a higher BMI is associated with a higher degree of underreporting (46-51). Furthermore misreporting has also been associated with several other factors including: gender, age, level of

education, physical activity level, restrained eating, dieting behaviour and social desirability (46, 47). Which determinants are associated with misreporting in our study population could help in our understanding and ways to correct for measurement errors and is therefore of interest as a research topic in this thesis.

Aim and outline of this thesis

The first aim of this thesis was to determine the validity of the DP compared to the 24hR as a reference method for FFQ validation for energy and several nutrients. Our second aim was to explore the validity of nutrient densities for DP, 24hR and FFQ. Thirdly, we wanted to determine the factors associated with misreporting of energy and various nutrients for DP, 24hR and FFQ.

In chapter 2, the validity of the DP as a reference method for FFQ validation is assessed for nutrients with a known recovery biomarker: protein, potassium and sodium. Since the 24hR is a commonly used reference method in nutritional epidemiology, a comparison between 24hR and DP as reference methods for FFQ validation was made. Chapter 3 compares the use of the DP and 24hR as reference methods for validation of the intake of saturated fatty acids, monounsaturated fatty acids, n-3 fatty acids, linoleic acid and the n-3/linoleic acid ratio estimated by FFQ. Additionally, biomarker (plasma fatty acids) data were used to objectively assess the ability of the DP and 24hR to rank individuals according to their intake of n-3 fatty acids, linoleic acid and the n-3/linoleic acid ratio. In chapter 4, a comparison was made between the validity of the intake of protein, potassium and sodium and their densities for the DP, 24hR and FFQ. Recovery biomarker measurements were used as the reference method to assess the validity of the absolute and nutrient density intakes. The association of BMI, gender, age and level of education with misreporting of energy, protein and potassium intake from DP, 24hR and FFQ, was evaluated in chapter 5. Additionally, the association between BMI-related and other determinants and misreporting was explored. The recovery biomarkers for the respective nutrients were used as reference method to assess the degree of misreporting. The final chapter 6 summarizes the main outcomes, places them in a larger context, discusses methodological issues, gives implications of our work and proposes further research topics.

References

1. Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.
2. Ferrari P, Roddam A, Fahey MT, et al. A bivariate measurement error model for nitrogen and potassium intakes to evaluate the performance of regression calibration in the European Prospective Investigation into Cancer and Nutrition study. *European Journal of Clinical Nutrition* 2009;63(SUPPL. 4):S179-S87.
3. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* 2003;158(1):14-21.
4. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology* 2011;174(5):591-603.
5. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014;180(2):172-88.
6. Freedman LS, Commins JM, Moler JE, et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Potassium and Sodium Intake. *American Journal of Epidemiology* 2015;181(7):473-87.
7. National Institutes of Health NCI. Measurement error webinar series. (http://appliedresearch.cancer.gov/measurementerror/). (Accessed December 2012).
8. Johansson L, Solvoll K, Bjørneboe G-E, et al. Under-and overreporting of energy intake related to weight status and lifestyle in a nationwide sample. *The American journal of clinical nutrition* 1998;68(2):266-74.
9. Willett W. Commentary: Dietary diaries versus food frequency questionnaires-a case of undigestible data. *International Journal of Epidemiology* 2001;30(2):317-9.
10. Kipnis V, Midthune D, Freedman L, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public health nutrition* 2002;5(6A):915-24.
11. Tooze JA, Kipnis V, Buckman DW, et al. A mixed-effects model approach for estimating the distribution of usual intake of nutrients: The NCI method. *Statistics in medicine* 2010;29(27):2857-68.
12. Guenther PM, Kott PS, Carriquiry AL. Development of an approach for estimating usual nutrient intake distributions at the population level. *Journal of Nutrition* 1997;127(6):1106-12.
13. Nusser SM, Carriquiry AL, Dodd KW, et al. A semiparametric transformation approach to estimating usual daily intake distributions. *Journal of the American Statistical Association* 1996;91(436):1440-9.
14. Carroll RJ, Ruppert D, Stefanski LA, et al. *Measurement error in nonlinear models: a modern perspective*. CRC press; 2006.
15. Rosner B, Spiegelman D, Willett WC. Correction of logistic regression relative risk estimates and confidence intervals for measurement error: The case of multiple covariates measured with error. *American Journal of Epidemiology* 1990;132(4):734-45.
16. Kaaks R. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *American journal of Clinical Nutrition* 1997;65(4):1232S-9S.
17. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *American Journal of Epidemiology* 2001;153(4):394-403.
18. Kaaks R, Ferrari P, Ciampi A, et al. Part H. Advances in the statistical evaluations and interpretation of dietary data: Uses and limitations of statistical accounting for random

error correlations, in the validation of dietary questionnaire assessments. *Public health nutrition* 2002;5(6 A):969-76.

19. Kipnis V, Carroll RJ, Freedman LS, et al. Implications of a new dietary measurement error model for estimation of relative risk: Application to four calibration studies. *American Journal of Epidemiology* 1999;150(6):642-51.

20. Kaaks R, Riboli E, Esteve J, et al. Estimating the Accuracy of Dietary Questionnaire Assessments - Validation in Terms of Structural Equation Models. *Statistics in Medicine* 1994;13(2):127-42.

21. Kirkpatrick SI, Reedy J, Butler EN, et al. Dietary Assessment in Food Environment Research: A Systematic Review. *American Journal of Preventive Medicine* 2014;46(1):94-102.

22. Abdulla M, Aly KO, Andersson I, et al. Nutrient intake and health status of lactovegetarians: chemical analyses of diets using the duplicate portion sampling technique. *American journal of Clinical Nutrition* 1984;40(2):325-38.

23. Wittsiepe J, Schnell K, Hilbig A, et al. Dietary intake of nickel and zinc by young children-results from food duplicate portion measurements in comparison to data calculated from dietary records and available data on levels in food groups. *Journal of Trace Elements in Medicine and Biology* 2009;23(3):183-94.

24. Malde MK, Dahl L, Zerihun L, et al. Intake of iron, zinc and iodine in 28 Ethiopian children living in Wonji Shoa Sugar Estate, assessed by duplicate portion technique. *Scientific Research and Essays* 2010;5(8):730-6.

25. Yoshida M, Ôgi N, Iwashita Y. Estimation of mineral and trace element intake in vegans living in Japan by chemical analysis of duplicate diets. *Health* 2011;3(11):672-6.

26. Lightowler HJ, Davies GJ. Iodine intake and iodine deficiency in vegans as assessed by the duplicate-portion technique and urinary iodine excretion. *British Journal of Nutrition* 1998;80:529-35.

27. Brady L, Lesauvage S, Saini N, et al. Comparison of dietary fat and fatty acid intake estimated by the duplicate diet collection technique and estimated dietary records. *Journal of Human Nutrition and Dietetics* 2003;16(6):393-401.

28. Roshanai F, Sanders T. Assessment of fatty acid intakes in vegans and omnivores. *Human nutrition Applied nutrition* 1984;38(5):345.

29. Isaksson B. A critical evaluation of the duplicate-portion technique in dietary surveys. *European journal of clinical nutrition* 1993;47(7):457-60.

30. Johansson G, Åkesson A, Berglund M, et al. Validation with biological markers for food intake of a dietary assessment method used by Swedish women with three different with dietary preferences. *Public health nutrition* 1998;1(03):199-206.

31. Kim WW, Mertz W, Judd JT, et al. Effect of making duplicate food collections on nutrient intakes calculated from diet records. *American journal of Clinical Nutrition* 1984;40(6):1333-7.

32. Stockley L. Changes in habitual food intake during weighed inventory surveys and duplication diet collections. A short review. *Ecology of Food and Nutrition* 1985;17(3):263-9.

33. Freedman LS, Kipnis V, Schatzkin A, et al. Can we use biomarkers in combination with self-reports to strengthen the analysis of nutritional epidemiologic studies? *Epidemiologic Perspectives and Innovations* 2010;7(1).

34. Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *Journal of Nutrition* 1988;118(11):1278-89.

35. Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *Journal of Nutrition* 2003;133(3 SUPPL.):921S-4S.

36. 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. *American journal of Clinical Nutrition* 1985;42(6):1276-89.

37. Holbrook JT, Patterson KY, Bodner JE. Sodium and potassium intake and balance in adults consuming self-selected diets. *American Journal of Clinical Nutrition* 1984;40(4):786-93.

38. 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. *Journal of Nutrition* 2006;136(5):1334-40.

39. Jenab M, Slimani N, Bictash M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human genetics* 2009;125(5):507-25.

40. Neuhouser ML, Tinker L, Shaw PA, et al. Use of recovery biomarkers to calibrate nutrient consumption self-reports in the Women's Health Initiative. *American Journal of Epidemiology* 2008;167(10):1247-59.

41. Bingham SA, Gill C, Welch A, et al. 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. *International Journal of Epidemiology* 1997;26(SUPPL. 1):S137-S51.

42. Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *American journal of Clinical Nutrition* 1997;65(4 Suppl):1130S-7S.

43. Freedman LS, Kipnis V, Brown CC, et al. Comments on" Adjustment for total energy intake in epidemiologic studies". *American journal of Clinical Nutrition* 1997;65(4):1229S-31S.

44. Cade JE, Burley VJ, Warm DL, et al. Food-frequency questionnaires: A review of their design, validation and utilisation. *Nutrition Research Reviews* 2004;17(1):5-22.

45. Krebs-Smith S, Graubard BI, Kahle LL, et al. Low energy reporters vs others: A comparison of reported food intakes. *European Journal of Clinical Nutrition* 2000;54(4):281-7.

46. Livingstone MB, Black AE. Markers of the validity of reported energy intake. *Journal of Nutrition* 2003;133 Suppl 3:895s-920s.

47. Maurer J, Taren DL, Teixeira PJ, et al. The psychosocial and behavioral characteristics related to energy misreporting. *Nutrition reviews* 2006;64(2):53-66.

48. Briefel RR, Sempos CT, McDowell MA, et al. Dietary methods research in the third National Health and Nutrition Examination Survey: Underreporting of energy intake. *American Journal of Clinical Nutrition* 1997;65(4 SUPPL.):1203S-9S.

49. Price GM, Paul AA, Cole TJ, et al. Characteristics of the low-energy reporters in a longitudinal national dietary survey. *British Journal of Nutrition* 1997;77(6):833-51.

50. Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. *American Journal of Clinical Nutrition* 2004;79(5):795-804.

51. Voss S, Kroke A, Klipstein-Grobusch K, et al. Is macronutrient composition of dietary intake data affected by underreporting? Results from the EPIC-Potsdam study. *European Journal of Clinical Nutrition* 1998;52(2):119-26.

Chapter 2

Comparison of duplicate portion and 24h recall as reference methods for validating a food frequency questionnaire using urinary markers as the estimate of true intake

Laura Trijsburg
Jeanne H.M. de Vries
Hendriek C. Boshuizen
Paul J.M. Hulshof
Peter C.H. Hollman
Pieter van ´t Veer
Anouk Geelen

Abstract

As Food Frequency Questionnaires (FFQs) are subject to measurement error, associations between self-reported intake by FFQ and outcome measures should be adjusted by correction factors obtained from a validation study. Whether the correction is adequate depends on the characteristics of the reference method used in the validation study. Preferably reference methods should 1) be unbiased and 2) have uncorrelated errors with those in the FFQ. The aim of the present study was to assess the validity of the duplicate portion (DP) technique as a reference method and compare its validity with that of a commonly used reference method, the 24 hour recall (24hR), for protein, potassium and sodium using urinary markers as the unbiased reference method. For 198 subjects, two DPs, two FFQs, two urinary biomarkers and between one and fifteen 24hRs (web-based and/or telephone-based) were collected within 1.5 years. Multivariate measurement error models were used to estimate bias, error correlations between FFQ and DP or 24hR and attenuation factors of these methods. The DP was less influenced by proportional scaling bias (0.58 for protein, 0.72 for K and 0.52 for Na) and correlated errors between DP and FFQ were lowest (protein 0.28, K 0.17 and Na 0.19) compared to the 24hRs. Attenuation factors (protein 0.74, K 0.54 and Na 0.43) also indicated that the DP performed better than the 24hRs. Therefore the DP is probably the best available reference method for FFQ validation for nutrients that currently have no generally accepted recovery biomarker.

Introduction

FFQs are often used to determine diet-disease relationships in epidemiological research because they are inexpensive and pose a low burden on participants compared with other dietary assessment methods. However, the association between disease and dietary exposure, assessed by an FFQ, is biased because of measurement errors in the FFQ (1). Therefore, a validation study should be performed to assess the amount of measurement error in order to correct the observed associations. However, whether the correction is adequate depends among others on the characteristics of the reference method used in the validation study. A reference method should 1) be unbiased and 2) have uncorrelated errors with the errors in the method to be validated (2). Recovery biomarkers are assumed to meet these requirements but are only available for energy and for a few nutrients such as K, Na and protein (3, 4). Therefore, other dietary assessment methods such as replicate 24 hour recalls (24hR) and food records have been used as reference methods. However, previous research showed that these methods do not entirely correct for measurement errors (1, 5-7) because they are biased and have correlated errors with the FFQ.

Bias is present when dietary intake is over- or under-estimated because of, for example, incorrect portion size estimation, inaccuracies in food composition databases (FCDs) or a lack of detail to identify foods consumed. The second criterion for a valid reference method, that is, uncorrelated errors between the reference method and FFQ, is violated when, for example, both methods make use of data from the same FCD, rely on memory or estimate portion sizes by using the same household measures (4). The duplicate portion (DP) technique partially overcomes these limitations as it does not depend on FCD data, is not memory based, and does not use standardized portion sizes. For a DP, participants collect a second equal portion of each food and drink they consume over one or more days. Afterwards, the dietary composition of the DP is determined by chemical analysis. Because of this, the magnitude of correlated errors of this method with an FFQ is expected to be lower than for a 24hR for which correlated errors are a known limitation (1). On the other hand, collections of DPs may lead to reactivity bias, demonstrating a change in the respondents' intake on the collection day, mostly resulting in underestimation of intake (8-11).

Our study aimed to evaluate the suitability of the DP technique as a reference method for an FFQ to assess protein, K and Na, using multivariate measurement error models. As the 24hR is often used as a reference method in evaluation studies, our secondary aim was to compare the validity of the 24hR and DP as reference methods for an FFQ. To this end, recovery biomarkers for protein, K

and Na were determined and assumed to be unbiased with independent measurement error.

Methods

Subjects and design

In this study, the DuPLO study, a random subsample of 200 Dutch adults (92 men and 108 women) from the NQplus study was included. The NQplus study is a longitudinal study designed for multiple aims: to validate a newly developed FFQ, to start a reference database for nutrition research and to study the association between diet and intermediate health outcomes. Participants for the NQplus study were recruited by sending invitations to randomly selected persons aged 20-70 years, living in Wageningen, Ede, Renkum and Arnhem. Subjects participating in the NQplus study at that time (N≈630) received an email invitation to join the DuPLO study. Recruitment for DuPLO started in November 2011 until April 2013. After reaching the intended sample size for DuPLO (N=200), recruitment for the NQplus study was still ongoing.

Baseline measurements consisted of, amongst others, a physical examination, including weight and height, and general and lifestyle questionnaires (including questions about education, health and smoking habits). Within a timeframe of 3 years each participant collected two DPs (~ 5 months apart), and two urine samples (~1 year apart). In addition, two self-reports by FFQ (~ 7 months apart) were handed in. The 24hR was administrated in two ways; by means of a telephone interview by a trained dietitian (telephone-based 24h recall collection (24hRT)) (0 to 8 replicates, ~ 4 months apart) or filled in by the participant in a web-based program (web-based 24h recall collection (24hRW)) (0 to 9 replicates, ~ 3 months apart). An overview of the timeframe and sample size of the data collection is presented in Appendix I. The large variety in replicates for the 24hRT and 24hRW is mainly due to the fact that participants were difficult to reach by telephone or people felt the burden of participation was too much and therefore cancelled invitations for the 24hR. The Dutch FCD of 2011 (12) was used to calculate nutrient intake for the 24hRT, 24hRW and FFQ. Participants with missing data for one or more of the methods were included in the analysis because they provided information for the other dietary assessment methods. In total 198 participants were included for analysis; 92 males and 106 females. Two participants became pregnant during the study. As it was expected that they deviated from their habitual dietary intake, they were excluded from analysis. This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the medical ethical committee of Wageningen University. Written informed consent was obtained from all subjects.

Dietary assessment

Duplicate Portion collection: Participants got verbal and written instructions to collect a second identical edible portion of all foods and drinks consumed over a 24h period. Foods and drinks were collected in separate baskets in a cool box (5°C). Participants received a monetary reimbursement for the products collected for the DP. The collected and consumed portions were measured using the same household measures. The collection cool boxes were brought to the participant's home one day before collection and picked up the day after collection. In the laboratory, collected DPs were weighed, homogenized in a blender (Waring Commercial model 34BL22) and 2,5 mL 0.02% tert-butylhydrochinon (BHQ) in ethanol was added per kg of DP as antioxidant during blending. The homogenized samples were stored within 1 h at -20°C until further analysis. A part of the sample was freeze dried before analysis.

Food Frequency Questionnaire: Participants completed an online self-administered 180 items FFQ using the online open-source survey tool Limesurvey^{TM} (LimeSurvey Project Team/Carsten Schmitz. Hamburg, Germany, 2012). Portion sizes were assessed by commonly used household measures and the reference period for reporting intake was the past month. The performance of the FFQ had been evaluated for energy (ρ=0.65 as compared with three 24hRs), fats (ρ ranged between 0.29-0.75 as compared with three 24hRs), selected vitamins (ρ ranged between 0.46-0.86 as compared with three 24hRs) and dietary fibre intake (ρ=0.82 as compared with three 24hRs) (13). The estimated mean energy intake by the FFQ appeared to be accurate (14) and in comparison with a replicate 24hR the FFQ showed an acceptable to good ranking ability for most nutrients (13).

Web-based 24-hour recall collection: participants received an unannounced email invitation, which was valid for 24 hours, to self-administer a recall over the previous day in the web-based program Compl-eat. This program is based on the five-step multiple-pass method (15), which enables participants on a step-by-step basis to accurately report the foods and drinks consumed the previous day. If participants did not fill in the 24hRW, a new invitation was randomly sent within 3-10 days. Portion sizes of foods or recipes were reported by using household measures, standard portion sizes, weight in grams, or volume in litres (16). The 24hRW were checked for completeness and unusual or missing values and if necessary, adjustments were made using standard portion sizes (16) and recipes following a standard internal protocol.

Telephone-based 24-hour recall collection: trained dietitians of the Division of Human Nutrition of Wageningen University made an unannounced phone call to the participant. The dietitian asked about foods and drinks consumed the previous day according to a standardised protocol based on the five-step multiple-pass method (15). The 24hRTs were coded using Compl-eat. For various components (energy, nutrients and foods) the highest and lowest ten values were checked for errors, such as errors in coding numbers or in the amounts (e.g. 150 cups instead of 150 gram of milk).

Urine collection: Participants received verbal and written instructions for 24-hour urine collections. The urine collection started after discarding the first voiding on the morning of the collection day and finished after the first voiding on the morning of the next day. The preservative lithium dihydrogenphosphate (25 g) was added to the collection containers. Subjects were instructed to ingest a tablet containing 80 mg para-aminobenzoic acid (PABA) (PABA check, Elsie Widdowson Laboratory) during breakfast, lunch and dinner on the day of collection to check for completeness of urines. Participants were also instructed to register possible deviations from the protocol (e.g. missing urine). At the study centre, the urine collections were mixed, weighted and aliquoted and stored at -20°C until further analyses.

Laboratory analysis

Protein analysis: Total nitrogen in the urine and in the DP was analysed by automated Kjeldahl method (17) using a Foss Kjeltec^{TM} 2300 Analyzer (Foss Tecator AB). The amount of protein was calculated using a nitrogen to protein conversion factor of 6.25 (18). Protein intake was calculated from nitrogen excretion assuming an average ratio of urinary to dietary nitrogen of 0.81 (19). For the DP, the within-run CV was <1% and between-run CV was <1%. For the urine analysis, the within-run CV was 1.6% and between-run CV was 1.3%.

Potassium and sodium analysis: K and Na in urine were determined with an ion-selective electrode (Roche 917 analyser; Roche). K and Na intake assessed by urinary excretion was calculated taking into account 19% K (20) and 14% Na (21) extra-renal and faecal losses. Participation in the External Quality Assessment Scheme of the Dutch Foundation for Quality Assessment in Medical Laboratories showed bias of -1.6% and +1.1% and analytical variation was 1.6% and 1.2% for K and Na, respectively. A within-run CV of <1% and a between-run CV of <1% for K and a within-run CV of <1% and a between-run CV of <1% for Na were observed. K and Na in the DP were determined after digestion of the samples in PTFE tubes using a MarsXpress microwave digestor (CEM), with inductively coupled plasma

atomic emission spectroscopy (ICP-AES, Varian Australia Pty LtdISO, 2010) at the Chemical Biological Soil Laboratory of Wageningen University with a within-run CV of <1% and a between-run CV of <1% for K and a within-run CV of 1.1% and a between-run CV of 1.7% for Na.

PABA analysis: PABA was measured by means of HPLC after alkaline hydrolysis of the urine samples to convert PABA metabolites into PABA (22). Using a minimum of 78% PABA recovery as a cut-off point for complete urine collection, which is proposed if PABA is analysed by HPLC (22), 16.7% of the urine samples were judged incomplete. The total CV for the PABA analysis was 9%. The within-run CV for PABA was 1.9% and the between-run CV for PABA was 1.3%.

Measurement error model

We assumed protein, Na and K intake assessed by urinary excretion to be unbiased in assessing usual intake (3), which we assumed not to vary within the 3 years of study. All our measurement error models assumed a linear relationship between DP, 24hRT, 24hRW, FFQ, biomarker and the true unknown intake *T*. In our measurement error model i is the person and j indicates the occasion. Furthermore, α_X expresses the constant bias for reference method X (X being DP for the DP method, 24hRT for the telephone-based 24hR, and 24hRW for the web-based 24hR) and β_X is the proportional scaling bias where α_Q and β_Q are similar respective parameters for the FFQ. The person-specific bias of the reference method is given by w_{xi} and for the FFQ by v_i. Finally, ε_{Xij} is the random error with mean zero and constant variance for the reference method, whereas ε_{Qij} is the random error for the FFQ.

Reference method X:	$Xij = \alpha_X + \beta_X T + w_{xi} + \varepsilon_{Xij}$	(1)
Food Frequency Questionnaire:	$Qij = \alpha_Q + \beta_Q T + v_i + \varepsilon_{Qij}$	(2)
Biomarker:	$Mij = T + \varepsilon_{Mij}$	(3)

Statistical analysis

Descriptive statistics were presented in percentages and as means with their standard deviation. Presence of bias between the mean of the recovery biomarker and the mean of the available replicates of FFQ, DP, 24hRW and 24hRT was tested by performing a Student's paired *t* test. The significance level was set at a two-sided *P* value of 0.05.
A Bayesian approach (23), Markov Chain Monte Carlo, the PROC MCMC procedure in SAS, was used to estimate the parameters of our measurement error models for which uninformative priors were set to make the model data

driven (syntax can be found in Appendix II). The sensitivity of our measurement error model was tested by using different distributions for the parameters and changing the prior estimates. As little variation in model outcomes was observed, we assumed the model to be robust. Sex-specific models for Na did not converge (because of the low variance of the person specific biases compared with within- and between-person variances) and are therefore not reported. To assess whether the reference method adequately corrects for measurement error it should be unbiased, which is indicated by the absence of proportional scaling bias (a β_x equal to one in equation 1 of the measurement error model indicates that there is no proportional scaling bias present). Furthermore, the reference method should have uncorrelated errors with the errors in the FFQ that is, the error correlation should be 0. The error correlation (ρ_{XQ}) is calculated according to formula 4 specified below from the measurement error model outcomes. From the model outcomes we also calculated the attenuation factor (λ_X) for each reference method according to formula 5 as specified below. Note that this is not the attenuation factor for the FFQ using the reference method, but the attenuation factor for the reference method using the biomarker as reference.

$$\rho_{XQ} = \frac{cov_{wivi}}{\sqrt{(var\ \varepsilon_{Xij} + varw_{xi}) * (var\ \varepsilon_{Qij} + varv_i)}} \qquad (4)$$

$$\lambda_X = \frac{\beta_X * varT}{\beta_X^2 * varT + \frac{var\ \varepsilon_{Xij}}{k} + varw_{xi}} \qquad (5)$$

where cov$_{wivi}$ is the covariance between the error in the FFQ and the error in the reference method X, varε_{Xij} is the variance of the random error of the reference method X, varw$_{xi}$ indicates the variance of the person-specific bias of method X, varv$_i$ is the variance of the person-specific bias of the FFQ and varε_{Qij} is the variance of the random error of the FFQ and β_X is the proportional scaling bias of method X. To obtain the estimates of the attenuation factor for multiple DPs and 24hRs, the variance of the random error of the method (varε_{Xij}) was divided by the number of measurements (k) of the method. All statistical tests were performed in SAS version 9.3 (SAS Institute Inc. Cary, NC, USA, 2012).

A sensitivity analysis was performed, comparing the model outcomes from the complete urine dataset with the model outcomes after exclusion of the urine samples with <78% PABA recovery (22). Measurement error model outcomes did not differ substantially when no urine samples were excluded compared with excluding urines with PABA <78%. This points in the same direction as the finding of Subar *et al.* (24), who observed a modest effect on correction factors when urines were excluded based on PABA recovery compared with not excluding

urines in the OPEN study (24). We therefore report the results based on the complete urine set in this article.

Results

At baseline, participants were on average 55.7 (SD 10.2) years of age and women were slightly younger than men (53.8 vs 58.0 y, Table 1). The average body mass index (BMI) was 25.1 (SD 3.7) kg/m^2 and a higher percentage of women (64%) had a healthy BMI (18.5-25.0 kg/m^2) compared with men (46%). Furthermore, 58% of the men and 48% of the women were classified as highly educated (university or college).

The percentage of the number of 24hRT and 24hRW varied between 18% and 29% over the seasons (Table 2). The variation in the number of urine collections per season was larger, and varied between 4% collected in spring and 51% in summer. Most DPs (34%) were collected in spring (Table 2). For the FFQ, 39% was collected in autumn and 12% in winter. The DP, 24hRT, 24hRW, and urine collections were evenly distributed between week (range Mon-Fri 63-76%) and weekend days (range Sat-Sun 24-37%).

Table 1: Baseline characteristics of the study population
(mean values and standard deviations; percentages)

	Total (N=198)	Women (N=106)	Men (N=92)
Age y (mean±SD)	55.7±10.2	53.8±10.6	58.0±9.3
BMI[a] (mean±SD)	25.1±3.7	24.6±3.8	25.8±3.5
BMI[a] (%)			
<25	55.6	64.1	45.6
25-30	33.8	25.5	43.5
≥30	10.6	10.4	10.9
Education (%)			
Low[b]	18.7	21.7	15.2
Intermediate[c]	28.8	30.2	27.2
High[d]	52.5	48.1	57.6

[a] Body Mass Index in kg/m^2
[b] primary or lower education
[c] secondary or higher vocational education
[d] university or college

Table 2: Percentage of the number of collection days distributed over the seasons and weekend vs week days

	Biomarker (N=197)	DP (N=198)	FFQ (N=194)	24hRT (N=155)	24hRW (N=193)
Season					
Winter	13.5	25.0	12.0	25.3	18.6
Spring	3.8	33.5	21.9	17.5	28.7
Summer	50.9	18.8	27.5	28.4	23.4
Autumn	31.8	22.8	38.5	28.2	29.3
Weekend[a]	36.7	23.8	23.8	33.0	31.6
Week days	63.3	76.2	76.2	67.0	68.5

DP=duplicate portion, FFQ=food frequency questionnaire, 24hRT=telephone based 24 hour recall, 24hRW= web based 24 hour recall
[a] Weekend days are Saturdays and Sundays

The DP underestimated protein by 20.9%, K by 6.8% and Na by 33.5% (Table 3). For all nutrients, underestimation was smallest using the 24hRT (protein 12.7%, K 4.7% and Na 28.7%). The FFQ, the method to be validated, underestimated protein (22.6%) and Na (41.6%) to the largest extent. A similar pattern was observed for men and women. Overall, women tended to underestimate to a lesser extent than men for all dietary assessment methods and nutrients.

A proportional scaling bias, as indicated with β_x in Table 4, closer to 1 means less bias. In general, the estimates for the DP were closest to one, 0.58 for protein, 0.72 for K and 0.52 for Na, compared with those for 24hRT and 24hRW (Table 4). For the sex-specific models, the proportional scaling bias was closest to one for the DP for K for women (0.77) and for protein for men (0.72). However the 24hRT performed better for protein for women (0.62) and for K for men (0.93).

In the total population, the correlated errors between the DP and FFQ were the lowest for the two micronutrients, Na (0.19) and K (0.17) (Table 4). For protein, the error correlations with the FFQ were comparable between the three reference methods (0.28 for the DP and 24hRT, and 0.27 for the 24hRW). The range of correlated errors was comparable for men (0.12-0.28) and women (0.08-0.29).

Table 3: Mean intake and bias for the intake of protein, K and Na, compared to the urinary excretion marker (mean values and standard deviations)

	Total			Women[a]		Men[a]	
	Protein(g)	K(mg)	Na(mg)	Protein(g)	K(mg)	Protein(g)	K(mg)
Urinary biomarker (reference) (N=197)							
Mean±SD	100.0±26.6	3865±1045	3983±1264	85.8±16.0	3517±846	116.5±27.0	4272±1110
Duplicate portion (N=198)							
Mean±SD	76.8±19.0	3484±867	2505±807	69.7±14.8	3270±739	84.9±20.2	3732±938
Bias (%)	-20.9[a]	-6.8[a]	-33.5[a]	-16.7[a]	-4.3[a]	-25.8[a]	-9.7[a]
24hR telephone based (N=155)							
Mean±SD	81.3±18.9	3477±821	2568±743	75.4±15.3	3268±643	90.8±20.4	3819±961
Bias (%)	-12.7[a]	-4.7[a]	-28.7[a]	-9.9[a]	-3.4[a]	-17.3[a]	-7.0[a]
24hR web based (N=193)							
Mean±SD	78.3±23.6	3280±850	2519±859	70.7±16.5	3074±792	86.7±27.3	3510±857
Bias (%)	-19.4[a]	-12.2[a]	-31.7[a]	-15.4[a]	-9.6[a]	-23.9[a]	-15.2[a]
Food Frequency Questionnaire (N=194)							
Mean±SD	74.4±21.1	3407±871	2137±708	67.6±16.3	3194±732	82.4±23.3	3657±955
Bias (%)	-22.6[a]	-8.2[a]	-41.6[a]	-18.9[a]	-5.9[a]	-26.9[a]	-10.9[a]

N=number of participants

[a] Values were significantly different from the biomarker P<0.01

* For Na, sex-specific models did not converge because of the low variance of the person specific biases compared with within- and between-person variances and are therefore not reported

Table 4: Proportional scaling bias and correlated error with the FFQ for the intake of protein, potassium and sodium (mean values and standard deviations)

	Total[*]			Women[†,‡]		Men[†,‡]	
	Protein	K	Na	Protein	K	Protein	K
Proportional scaling bias (β_x)							
Duplicate portion	0.58±0.06	0.72±0.10	0.52±0.11	0.52±0.14	0.77±0.15	0.72±0.10	0.89±0.19
24hR telephone based	0.53±0.08	0.68±0.09	0.32±0.13	0.62±0.15	0.66±0.18	0.64±0.14	0.93±0.18
24hR web based	0.48±0.08	0.58±0.09	0.35±0.13	0.53±0.15	0.65±0.19	0.56±0.13	0.69±0.16
Correlated error with FFQ (ρ_{xq})							
Duplicate portion	0.28±0.06	0.17±0.07	0.19±0.07	0.29±0.07	0.08±0.08	0.22±0.10	0.14±0.12
24hR telephone based	0.28±0.05	0.27±0.06	0.29±0.05	0.22±0.06	0.25±0.09	0.27±0.10	0.12±0.12
24hR web based	0.27±0.04	0.23±0.05	0.32±0.04	0.22±0.05	0.14±0.07	0.28±0.07	0.27±0.09

[*] Adjusted for BMI and sex
[†] Adjusted for BMI
[‡] For Na, sex-specific models did not converge because of the low variance of the person specific biases compared with within and between person variances and are therefore not reported

Table 5: Attenuation factors for the reference methods for the intake of protein, potassium and sodium (mean values and standard deviations)

	Total*			Women[†,‡]		Men[†,‡]	
	Protein	K	Na	Protein	K	Protein	K
Duplicate portion							
1 DP	0.74±0.07	0.54±0.06	0.43±0.08	0.30±0.07	0.41±0.07	0.73±0.08	0.55±0.08
2 DPs	0.93±0.08	0.71±0.07	0.65±0.10	0.41±0.10	0.59±0.09	0.90±0.10	0.68±0.10
24hR telephone based							
1 24hRT	0.45±0.06	0.47±0.05	0.19±0.06	0.20±0.05	0.34±0.06	0.42±0.08	0.48±0.07
2 24hRTs	0.63±0.07	0.63±0.07	0.30±0.10	0.31±0.07	0.49±0.09	0.57±0.11	0.62±0.09
3 24hRTs	0.73±0.08	0.72±0.07	0.38±0.12	0.38±0.08	0.57±0.10	0.64±0.12	0.69±0.10
24hR web based							
1 24hRW	0.30±0.05	0.31±0.05	0.18±0.05	0.14±0.04	0.20±0.05	0.26±0.06	0.38±0.07
2 24hRWs	0.45±0.07	0.46±0.06	0.28±0.08	0.22±0.06	0.30±0.07	0.37±0.08	0.54±0.09
3 24hRWs	0.54±0.08	0.54±0.07	0.36±0.10	0.28±0.07	0.36±0.08	0.44±0.09	0.63±0.09

DP, duplicate portion; 24hRT, telephone based 24 hour recall; 24hRW, web based 24 hour recall
* Adjusted for BMI and Gender
[†] Adjusted for BMI
[‡] For Na, sex-specific models did not converge because of the low variance of the person specific biases compared with within and between person variances and are therefore not reported

An attenuation factor close to one indicates an overall better estimation of the nutrient intake. In the total population, looking at estimates for single measurements, attenuation factors for the DP were highest for all three nutrients (0.74 for protein, 0.54 for K and 0.43 for Na), whereas for the 24hRW attenuation factors tended to be the lowest for all nutrients (0.30 for protein, 0.31 for K and 0.18 for Na) (Table 5). The same trend was seen for women and men separately. Attenuation factors increased when the number of replicates was expanded. For protein, the attenuation factor for one measurement of the DP was 0.74, whereas for the 24hRT three measurements gave a similar attenuation factor (0.73). In general, attenuation factors for all dietary assessment methods tended to be higher for men than for women.

Discussion

In this Dutch validation study, we found that all dietary assessment methods underestimated the intake of protein, K and Na compared with the biomarker measurements where the 24hRT showed the smallest underestimation. Furthermore, all dietary assessment methods were biased (affected by proportional scaling bias) and showed correlated errors with the FFQ for protein, K and Na. However, dietary intake measures from the DP were less affected by proportional scaling bias compared with the 24hRT and 24hRW. Furthermore, error correlations between the DP and FFQ were lowest. Attenuation factors also indicated that the DP had the best performance (attenuation factors were closer to one).

To our knowledge, this is the first study assessing error correlations between the FFQ and DP, proportional scaling bias for DP and estimating attenuation factors for the DP. Research on 24hRs has among others been performed in a pooled analysis of five American validation studies comparing protein intakes assessed by the FFQ and 24hR with urinary nitrogen excretion (25). Freedman *et al* (25) found wide ranges of study-specific attenuation factors (0.14-0.54) for the 24hR. This is comparable with our results, but we found estimates at the higher end of this range. One of the possible explanations is that our study population was highly motivated; they were willing to collect, in addition to filling out multiple 24hRs and various food and lifestyle questionnaires, two urine and two DP samples. Above that, a high percentage of our participants were highly educated. Furthermore, cultural differences in dietary patterns and the design of the FFQ and 24hR could also explain our findings to be in the upper part of the range.

Proportional scaling bias for the 24hRT for protein was similar to that found in the OPEN study, a large American study from Montgomery County, Maryland, for women (0.62 for DuPLO vs 0.60 for OPEN), but our estimate was slightly

lower for men (0.64 for DuPLO vs 0.70 for OPEN) (1). Error correlations between the 24hR and FFQ were slightly higher in our study compared to the EPIC study, a large European multi-centre study, showing 0.21 for K and 0.21 for protein (5) and the OPEN-study (showing 0.24 for protein for women and 0.18 for men) (1). Prentice & Huang. (26) found slightly higher error correlations between their FFQ and 24hR for protein (0.33) (26). Differences between error correlations of the 24hRs with the FFQ in studies are expected because of different sets of covariates included, different modes of administration (web-based and interviewer administered) and numbers of replicates of a 24hR, varying ways of portion size estimations and differences between the study populations (ethnic groups, social economic status, age).

The attenuation factor for Na intake for the DP (0.43) was remarkably higher than for both 24hR administrations (0.19 for the 24hRT and 0.18 for the 24hRW) and taking a second replicate for the DP increased the attenuation factor to 0.65. The DP for Na was also less affected by proportional scaling bias (β_{DP}=0.52) and demonstrated a lower error correlation with the FFQ (0.19) compared with the 24hRT and 24hRW. Accurately assessing Na intake is challenging because of the high variability of Na content of foods (27), which is not always accurately reflected in FCD. In addition, it is difficult to accurately report the amount of salt added during cooking or at the table. In the 24hR and FFQ in this study, there is no question included about added salt during cooking or at the table. The accuracy of dietary intake estimates of Na from 24hR and FFQ is therefore expected to be limited. This is supported by other research about Na estimation from 24hRs, FFQs and dietary records (27). The higher attenuation factor and proportional scaling bias for the DP could be explained by the fact that salt added during cooking was included as a sample of the cooked meal was collected and the DPs were chemically analysed and estimates did not depend on information in FCD. However, attenuation factors for Na for the DP were still notably lower than those for protein and K intake.

Correlated errors between the FFQ and reference methods for protein intake tended to have the same order of magnitude for all methods, while for K and Na intake, the DP showed lower error correlations than the 24hRT and 24hRW. Thus, there must be a source of error equally influencing the estimation of protein in all four methods apart from the correlated errors that are expected between the FFQ and 24hR (use of the same FCD to calculate nutrients, estimation of portion sizes and memory based). A similar error source for all four methods (FFQ, DP, 24hRT and 24hRW) could be response errors, meaning that people tended to forget (for FFQ and 24hRs) or not collect (for DP), either on purpose or not, protein-rich products.

A weakness of this study is the unequal spread of biomarker measurements over the seasons (summer was over-represented and spring under-represented), while they were assumed unbiased in our measurement error model. This assumption was based on evidence from the literature that does not indicate seasonal variation of nutrient intake in western populations (28, 29). Furthermore, the different methods did not exactly cover the same time period. However, we were interested in a person's usual intake and not in the dietary intake on a specific day. We assumed that energy and nutrient intake of a person would be fairly stable over a longer time period. Thus, although intake data measured by the different dietary assessment methods did not cover the same time period, they could be all considered to represent a person's usual energy and nutrient intake. Therefore, comparisons between methods can be made.

We reported the results based on all urines collected, independent of the PABA results. This was based on a sensitivity analysis to exclude urine samples based on PABA focussing on the main outcomes; attenuation factor and correlated error. These main outcomes did not differ substantially between inclusion of all urine samples and inclusion of only the complete urine samples (based on PABA recovery). Furthermore, not excluding urine samples provided a larger sample size. However, results for bias (i.e. difference between levels of intake) must be regarded rather carefully as they differed significantly for protein and K when incomplete urines were excluded.

Taking into account that in general the DP showed lesser proportional scaling bias, the highest attenuation factors and the lowest error correlations with the FFQ, this method appeared more promising as a reference method than did the 24hR. Important considerations in the collection of DPs are that it is burdensome for participants, requires a lot of time from the researcher, is expensive to perform and reactivity bias, mostly causing underestimation of habitual intake, is expected. We carefully instructed our participants not to deviate from their habitual intake and provided them with written instructions, including tips to remind the participant to include everything in the collection baskets. Nevertheless, the DP showed substantial underestimation for protein, K and Na.

Attenuation factors calculated for FFQs using the 24hR as a reference method are affected by correlated errors between the two methods (30). Better estimates of attenuation factors will be obtained if these correlated errors between the FFQ and 24hR are taken into account. The error correlations between the 24hR and FFQ found in this study could be considered in the calculation of attenuation factors, however, generalizing results from one study population to another should always be done conservatively, taking into account the characteristics of both study populations and the study setup.

Conclusion

We conclude that the DP violated the requirements to be used as unbiased reference method for validating an FFQ, however, to a lesser extent than a telephone-based 24hR and, even more, a web-based 24hR. As the proportional scaling bias was less for the DP, the DP-FFQ error correlations were lowest, and the attenuation factors were highest, we propose that the DP is probably the best available reference method for FFQ validation for nutrients that currently have no generally accepted recovery biomarker.

Acknowledgements

References

1. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* 2003;158(1):14-21.
2. Kaaks R, Riboli E, Esteve J, et al. Estimating the Accuracy of Dietary Questionnaire Assessments - Validation in Terms of Structural Equation Models. *Statistics in Medicine* 1994;13(2):127-42.
3. Jenab M, Slimani N, Bictash M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human genetics* 2009;125(5):507-25.
4. Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.
5. Ferrari P, Roddam A, Fahey MT, et al. A bivariate measurement error model for nitrogen and potassium intakes to evaluate the performance of regression calibration in the European Prospective Investigation into Cancer and Nutrition study. *European Journal of Clinical Nutrition* 2009;63(SUPPL. 4):S179-S87.
6. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology* 2011;174(5):591-603.
7. Geelen A, Souverein OW, Busstra MC, et al. Comparison of approaches to correct intake-health associations for FFQ measurement error using a duplicate recovery biomarker and a duplicate 24 h dietary recall as reference method. *Public health nutrition* 2014;18(2):1-8.
8. Isaksson B. A critical evaluation of the duplicate-portion technique in dietary surveys. *European journal of clinical nutrition* 1993;47(7):457-60.
9. Johansson G, Åkesson A, Berglund M, et al. Validation with biological markers for food intake of a dietary assessment method used by Swedish women with three different with dietary preferences. *Public health nutrition* 1998;1(03):199-206.
10. Kim WW, Mertz W, Judd JT, et al. Effect of making duplicate food collections on nutrient intakes calculated from diet records. *American journal of Clinical Nutrition* 1984;40(6):1333-7.
11. Stockley L. Changes in habitual food intake during weighed inventory surveys and duplication diet collections. A short review. *Ecology of Food and Nutrition* 1985;17(3):263-9.
12. Nevo. Dutch Food Composition Database The Hague: Stichting Nevo, 2011.
13. Streppel MT, De Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. *Nutrition Journal* 2013;12(1).
14. Siebelink E, Geelen A, De Vries JHM. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *British Journal of Nutrition* 2011;106(2):274-81.
15. Conway JM, Ingwersen LA, Vinyard BT, et al. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *American Journal of Clinical Nutrition* 2003;77(5):1171-8.
16. Donders-Engelen M, van der Heijden L. *Maten, gewichten en codenummers 2003*. Wageningen: Wageningen UR, Vakgroep Humane Voeding; 2003.
17. Hambleton LG, Noel RJ. Protein analysis of feed, using a block digestion. *Journal of the AOAC* 1975;58:143-5.
18. Jones DB. *Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins*. US Department of Agriculture Washington, DC; 1941.

19.   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. *American journal of Clinical Nutrition* 1985;42(6):1276-89.

20.   Freisling H, Van Bakel MME, Biessy C, et al. Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *British journal of nutrition* 2012;107(6):910-20.

21.   Holbrook JT, Patterson KY, Bodner JE. Sodium and potassium intake and balance in adults consuming self-selected diets. *American Journal of Clinical Nutrition* 1984;40(4):786-93.

22.   Jakobsen J, Ovesen L, Fagt S, et al. Para-aminobenzoic acid used as a marker for completeness of 24 hour urine: Assessment of control limits for a specific HPLC method. *European Journal of Clinical Nutrition* 1997;51(8):514-9.

23.   Richardson S, Gilks WR. A Bayesian approach to measurement error problems in epidemiology using conditional independence models. *American Journal of Epidemiology* 1993;138(6):430-42.

24.   Subar AF, Midthune D, Tasevska N, et al. Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *European journal of clinical nutrition* 2013;67(8):863-7.

25.   Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014;180(2):172-88.

26.   Prentice RL, Huang Y. Measurement error modeling and nutritional epidemiology association analyses. *Canadian Journal of Statistics* 2011;39(3):498-509.

27.   Champagne CM, Cash KC. Assessment of salt intake: how accurate is it? *The Proceedings of the Nutrition Society* 2013;72(3):342-7.

28.   Ma Y, Olendzki BC, Li W, et al. Seasonal variation in food intake, physical activity, and body weight in a predominantly overweight population. *European journal of clinical nutrition* 2006;60(4):519-28.

29.   van Staveren WA, Deurenberg P, Bureman J, et al. Seasonal variation in food intake, pattern of physical activity and change in body weight in a group of young adult Dutch women consuming self-selected diets. *International Journal of Obesity* 1986;10(2):133-45.

30.   Wong MY, Day NE, Bashir SA, et al. Measurement error in epidemiology: The design of validation studies I: Univariate situation. *Statistics in medicine* 1999;18(21):2815-29.

Appendix I: Overview of timeframe and sample size of the data collection

	N
FFQ 1	194
FFQ 2	179
DP 1	198
DP 2	198
urine 1	197
urine 2	184
24hRT	779
24hRT 1	155
24hRT 2	147
24hRT 3	145
24hRT 4	140
24hRT 5	120
24hRT 6	55
24hRT 7	13
24hRT 8	4
24hRW	973
24hRW 1	193
24hRW 2	184
24hRW 3	175
24hRW 4	158
24hRW 5	119
24hRW 6	87
24hRW 7	42
24hRW 8	14
24hRW 9	1

Timeline columns (2011–2014): JUL, AUG, SEPT, OCT, NOV, DEC, JAN, FEB, MRT, APR, MAY, JUN, JUL ...

Appendix II: Syntax for the MCMC procedure

We provide the SAS syntax for our measurement error model using proc MCMC. We used generic labels for the variables as can be found in the model statement. The dataset is called mydata. The array statement is used for identifying the (latent) person-specific biases and the error covariance. In the parms statements, the starting values are given for each model parameter, and in the prior statement a distribution is given. The estimates for the correlated errors and attenuation factors are calculated at the bottom of the syntax.

```
ods graphics on;
Proc MCMC data=mydata seed=20000 nmc=300000 thin=20 NBI=50000 Maxtune=50
MONITOR=(_PARMS_ corrDPQ rhoDPT lambdaDP rho2DPT lambda2DP rhoQT lambdaQ)
outpost=postdata;
ARRAY WIVI[2] VI wi;
array wivi_0[2] (0,0) ;
ARRAY varwivi[2,2];
array s[2,2] (1 0 0 1);
parms bDP 0.5 aDP 76 sdEDP 10;
parms aQ 71 bQ 0.5 sdEQ 10;
parms muT 100 sdT 25 sdEM 15;
parms bBMI1 0 bBMI2 0;
parms bG1 0 bG2 0;
parms varwivi {1 0 0 1};
parms wiscale 15;
parms viscale 25;
        prior aQ ~ normal (0, var=10000);
        prior bQ ~ normal (0, var=10000);
        prior aDP ~ normal (0, var=10000);
        prior bDP ~ normal (0, var=10000);
        prior bBMI1 ~ normal (0, var=10000);
        prior bBMI2 ~ normal (0, var=10000);
        prior bG1 ~ normal (0, var=10000);
        prior bG2 ~ normal (0, var=10000);
        prior muT ~ normal (0, var=1000000);
        prior sdEQ ~ uniform(0, 50);
        prior sdEDP ~ uniform(0, 50);
        prior sdEM ~ uniform(0, 50);
        prior sdT ~ uniform(0, 50);
        prior viscale ~ uniform(0, 50);
        prior wiscale ~ uniform(0, 50);
        prior varwivi ~ iwish(3,s);
```

```
varEQ = sdEQ**2; varEDP = sdEDP**2; varEM = sdEM**2; varT = sdT**2;
Random T ~ normal (muT, var=varT) subject = _OBS_;
Random wivi ~ mvn (wivi_0, varwivi) subject = _OBS_;
muQ=aQ + bQ*(T-100) + bG1*gender + bBMI1*(BMI-25) + viscale*vi;
muDP=aDP + bDP*(T-100) + bG2*gender + bBMI2*(BMI-25) + wiscale*wi;
muM = T;
        model FFQ_1_prot ~ normal (muQ, var=varEQ);
        model FFQ_2_prot ~ normal (muQ, var=varEQ);
        model DP1_prot_gr ~ normal (muDP, var=varEDP);
        model DP2_prot_gr ~ normal (muDP, var=varEDP);
        model T0_urine_eiwit ~ normal (muM, var=varEM);
        model T1_urine_eiwit ~ normal (muM, var=varEM);
corrDPQ=wiscale*viscale*varwivi[1,2]/sqrt(((wiscale**2)*varwivi[2,2]+varEDP)*((viscale
**2)*varwivi[1,1]+varEQ));
lambdaDP = (1/(1+(varEDP+(wiscale**2)*varwivi[2,2])/(bDP**2*varT)))/bDP;
lambda2DP = (1/(1+((varEDP/2)+(wiscale**2)*varwivi[2,2])/(bDP**2*varT)))/bDP;
        run;
        ods graphics off;
```

Chapter 3

The duplicate portion as a reference method for validating fatty acid intake as estimated by a food frequency questionnaire

Laura Trijsburg
Jeanne H.M. de Vries
Peter C.H. Hollman
Paul J.M. Hulshof
Pieter van ´t Veer
Hendriek C. Boshuizen
Anouk Geelen

Abstract

Measurement errors in fatty acid intake estimates obtained with Food Frequency Questionnaires (FFQs) are expected to underlie inconclusive results about their association with disease risks. We assumed the duplicate portion (DP) to be a better reference method for validation of a FFQ than the often used 24 hour recall (24hR), and compared their performance. Plasma fatty acids were used to objectively compare ranking of individuals based on fatty acid intakes from DP and 24hR. Intakes of specific fatty acids were estimated for 198 Dutch subjects by chemical analysis of two DPs and two plasma samples, and by on average five 24hRs and two FFQs. Multivariate measurement error models were used to estimate validity coefficients and attenuation factors. Validity coefficients for fatty acid estimates by the FFQ were lower or similar when using the DP as reference method than when the 24hR was used. Attenuation factors for the FFQ, using the DP as reference method, tended to be slightly higher for mono-unsaturated fatty acids (0.34 vs. 0.21), and similar for the other fatty acids compared to those when using the 24hR as reference method. Furthermore, when using plasma fatty acids as reference, the DP showed comparable to slightly better ranking of participants according to their intake of n-3 fatty acids, linoleic acid and their ratio than the 24hR. Altogether, the use of the 24hR as reference method gives slightly different results compared to the DP, which seems a promising reference method for FFQ validation of fatty acid intake.

Introduction

Inconclusive results about the risks of intake of total fat and various fatty acids on diseases such as breast cancer (1, 2) and coronary diseases (3, 4) plague epidemiological research. This inconclusiveness may originate from limitations and errors in food composition databases and dietary assessment methods to assess total fat and fatty acid intake. Food frequency questionnaires (FFQs) are often used in epidemiological studies, since they are relatively cheap and pose a low burden on the participants. However, they are suspected to be affected by systematic and random errors that together obscure the true variation in fat intake between subjects. The observed association between fat intake and disease can be adjusted for these measurement errors by an attenuation factor derived from a validation study. The reference method used in the validation study should generate unbiased dietary intake data (i.e. no proportional scaling bias should be present) and have uncorrelated errors with the FFQ (5, 6). However for most nutrients, including fatty acids, only imperfect reference methods are available, e.g. 24-hour recalls (24hRs) or concentration biomarkers. 24hRs are able to assess the intake of a wide array of fatty acids, but are biased and showed correlated errors with FFQs for energy and protein (7, 8). Concentration biomarkers are less susceptible to have correlated errors with the FFQ but are only informative on ranking of individuals according to their intakes and not on their absolute levels of intake. Furthermore, use of plasma fatty acids as biomarkers of intake is limited to fatty acids that are not endogenously produced (i.e. n-3 and n-6 fatty acids) (9). Previous research concluded that the duplicate portion method (DP) is a suitable reference method and might be preferable over a 24hR for FFQ validation for nutrients for which no recovery biomarker is available (10). The DP does not depend on the availability and quality of the nutrient values in food composition databases, and also biases related to memory and estimation of portion sizes are less of a problem as compared to methods like 24hR and FFQ. Altogether, the DP proved to be less affected by proportional scaling bias and had a lower degree of correlated errors with the FFQ than the 24hR for protein, potassium and sodium (10). In the present paper, we therefore assumed the DP to be a better reference method for FFQ validation for fatty acids than the often used reference method, the 24 hour recall (24hR), and compared their performance.

Our overall aim was to compare the validity of the intake assessed by FFQ for saturated fatty acids (SFA), monounsaturated fatty acids (MUFA), n-3 fatty acids, linolenic acid (LA) and the n-3/LA ratio using the DP or the 24hR as the reference method. We additionally assessed the ability of DP and 24hR to rank individuals according to their intake of n-3 fatty acids, LA and the n-3/LA ratio using an objective biomarker (plasma fatty acids) as reference method.

Subjects and Methods

Subjects and design

In this Dutch validation study called DuPLO, which is part of the NQplus study, 200 Dutch adults (92 men, 108 women) were enrolled. The recruitment and study procedures are described elsewhere (10). Briefly, between July 2011 and July 2014 each participant collected two DPs (~ 5 months apart), and two blood samples (~13 months apart). Also two FFQs (~ 7 months apart) were filled out. An average of five 24hRs per subject was administrated by a telephone interview by a dietician (~ 4 months apart). A varying number of 24hRs per person (between 0 and 8 measurements) was collected because participants were enrolled in different sub-studies of the NQplus study. Participants with missing data for one or more of the methods were included in the analysis because they provided information for the other dietary assessment methods. This validation study was approved by the medical ethical committee of Wageningen University.

24-hour recalls and FFQ

The 24hR administration followed a standardized protocol based on the 5-step multiple pass method (11). Participants got an unannounced phone call from a trained dietician. Portion sizes of foods or recipes were reported using household measures, standard portion sizes, weight in grams, or volume in liters (12).
The 180 item FFQ (13, 14) was administered via the web using the online open-source survey tool Limesurvey^{TM}. The reference period for the FFQ was one month and frequencies of intake were combined with standard portion sizes and household measures to assess amounts of intake (12). Self-reported dietary intake data from 24hR and FFQ were converted into nutrient data using the Dutch food composition database (FCD) of 2011 (15).

Duplicate portion collection and analytical methods

Participants got verbal and written instructions preceding the collection of the DP. Participants collected all edible foods and drinks consumed over a 24-hour period in collection baskets and stored them in a cool box (5°C). At the study center, DPs were weighed, homogenized in a blender (Waring Commercial model 34BL22) and 2.5 mL 0.02% tert-butylhydrochinon (BHQ) in ethanol was added per kg of DP as antioxidant. For each DP, an aliquot of the homogenized sample was stored within 1 hour at -20°C, until further analysis. Total fat was measured gravimetrically by acid hydrolysis (AOAC method 14.019) (16).

Blood sampling and fatty acid assessment

Blood samples were collected from the participants in a fasting state. EDTA plasma was stored at -80°C until further analysis. Cholesteryl esters from plasma were isolated using solid phase extraction silica columns and fatty acid profiles of the plasma cholesteryl esters were analyzed by gas chromatography as previously described (17).

Statistical analysis and measurement error models

In total 198 participants were included for analysis, 92 males and 106 females. Two participants got pregnant during the study. As it was expected that they had altered their habitual dietary intake they were excluded from analysis. Means and 95% confidence intervals were estimated for SFA, MUFA, n-3 fatty acids, and LA in grams and as a percentage of the total amount of fatty acids for DP, 24hR and FFQ. Today's western diets are high in n-6 and low in n-3 fatty acids. This is suggested to promote the pathogenesis of many diseases, including inflammatory, cancer and cardiovascular diseases (18). An n-3/LA ratio (LA is an n-6 fatty acid) closer to one indicates a healthier distribution and this ratio is therefore included as an additional outcome measure in this research. Because of their skewed distribution, a log transformation was used for all variables.

Our measurement error models assumed a linear relationship between the log(intake) according to DP, 24hR, FFQ or biomarker and the true unknown intake T, with intakes of the specific fatty acids expressed as percentages of the total fatty acid intake. Measurement error models were adjusted for BMI and gender. In our measurement error models i indicates the person and j the occasion. Furthermore, in all measurement error models α expresses the constant bias and β the proportional scaling bias. The person specific bias for the method is given by w_{Xi} and the random error by ε_{Xij} with mean zero and constant variance.

To evaluate the comparability of the 24hR and the DP as reference methods for the FFQ, model 1 (with equations 1 and 2) is defined as below. In this model the assumptions of negligible error correlation between reference method and FFQ and between replicates of the reference method, and absence of proportional scaling bias in the reference method ($\beta_X = 1$) were made to enable estimation of the model parameters.

Reference method X (24hR or DP): $Xij = T + \varepsilon_{Xij}$ (1)

Food Frequency Questionnaire: $Qij = \alpha_Q + \beta_Q T + w_{Qi} + \varepsilon_{Qij}$ (2)

Validity coefficients (ρ_{XT}, formula 3) were estimated to assess the ability of the dietary assessment method to rank participants according to their intake:

$$\rho_{XT} = \sqrt{\frac{\beta_X{}^2 \; varT}{\beta_X{}^2 \; varT + \frac{var\,\varepsilon_{Xij}}{k} + varw_{Xi}}} \tag{3}$$

Where varT is the variance of the true nutrient intake; $var\varepsilon_{Xij}$ the variance of the random error of method X and $varw_{Xi}$ the variance of the person specific bias for method X.

The attenuation factor (λx, formula 4) provides information about the extent to which diet-health associations are affected by measurement error:

$$\lambda_X = \frac{\rho_{XT}{}^2}{\beta_X} \tag{4}$$

As an additional check of the performance of the two reference methods, we used the biomarker to objectively compare the ranking based on individual fatty acid intakes when using the DP and the 24hR. Since the biomarker is only valid for n-3 and n-6 fatty acids (9) this was only done for the n-3 fatty acids, LA and the n-3/LA ratio. Therefore we specified measurement error model 2 (with equations 5 and 6) as given below. In this model the assumptions of negligible error correlation between biomarker and DP or 24hR and between replicates of the biomarker and absence of proportional scaling bias for the biomarker ($\beta_M = 1$) were made to enable estimation of the model parameters.

Biomarker: $Mij = T + \varepsilon_{Mij}$ (5)

Method X (24hR or DP): $Xij = \alpha_X + \beta_X T + w_{Xi} + \varepsilon_{Xij}$ (6)

All statistical tests were performed in SAS version 9.3 (SAS Institute Inc. Cary, NC, USA, 2012).

Results

Baseline characteristics of the study population

At baseline, mean age of the study population was 55.7 (SD 10.2) years and mean BMI was 25.1 (SD 3.7) kg/m^2. 52.5 percent completed a high level (university or college) and 18.7 percent a low level of education (primary or lower education).

Table 1: Mean intake of SFA, MUFA, n-3 fatty acids, LA, and n-3/LA ratio in grams and as a percentage of total fatty acids for the DP, 24hR and FFQ

	SFA		MUFA		n-3		LA		n-3/LA ratio	
	Mean	CI	Mean	CI	Mean	CI	Mean	CI	Mean	CI
Intake in grams										
DP	31.2	29.9-32.6	32.3	31.0-33.7	2.49	2.26-2.71	14.3	13.5-15.2	0.18	0.17-0.20
24hR	30.1	28.7-31.5	27.9	26.6-29.2	2.02	1.89-2.15	13.5	12.7-14.2	0.17	0.16-0.18
FFQ	26.9	25.6-28.3	28.7	27.4-30.0	2.25	2.14-2.35	14.6	13.9-15.4	0.16	0.16-0.17
Intake in percentage of total FA										
DP	37.4	36.6-38.3	38.4	37.7-39.0	2.98	2.76-3.20	17.2	16.5-18.0	0.18	0.17-0.20
24hR	40.2	39.4-41.1	36.8	36.1-37.4	2.83	2.66-3.01	18.0	17.3-18.7	0.17	0.16-0.18
FFQ	35.5	34.7-36.2	37.8	37.4-38.1	3.04	2.93-3.14	19.2	18.7-19.7	0.16	0.16-0.17

24hR= 24hour recall, CI=confidence interval, DP=duplicate portion, FA=fatty acids, FFQ=food frequency questionnaire, LA=linoleic acid, MUFA= mono-unsaturated fatty acids, n-3=n-3 fatty acids, SFA=saturated fatty acids

Table 2: Validity coefficients and attenuation factors of the FFQ for fatty acids (expressed as % of total fatty acids) with DP or 24hR as reference methods

Ref method	SFA	CI	MUFA	CI	n-3	CI	LA	CI	n-3/LA ratio	CI
Validity coefficient										
DP	0.76	0.63-0.89	0.37	0.19-0.54	0.47	0.32-0.62	0.64	0.48-0.79	0.33	0.17-0.48
24hR	0.82	0.77-0.86	0.65	0.56-0.74	0.62	0.48-0.76	0.80	0.75-0.85	0.76	0.70-0.82
Attenuation factor										
DP	0.57	0.46-0.68	0.34	0.17-0.50	0.63	0.41-0.85	0.60	0.45-0.76	0.49	0.25-0.73
24hR	0.46	0.38-0.53	0.21	0.15-0.27	0.56	0.41-0.71	0.55	0.44-0.66	0.45	0.32-0.58

24hR= 24hour recall, CI=confidence interval, DP=duplicate portion, LA=linoleic acid, MUFA= mono-unsaturated fatty acids, n-3=n-3 fatty acids, SFA=saturated fatty acids

Models were adjusted for BMI and gender

Mean intakes of fatty acids

Mean intakes of the specific fatty acids in grams and expressed as percentages of the total amount of fatty acids are shown in Table 1. SFA intake by the FFQ was lower (26.9 g) than by the DP (31.2 g) and 24hR (30.1 g). MUFA intake was highest when assessed by the DP (32.3 g) and n-3 fatty acid intake was highest when assessed by DP (2.49 g) and lowest by the 24hR (2.02 g). For LA, DP, 24hR and FFQ intake assessments were rather similar as was the case for the n-3/LA ratio. SFA intake as percentage of total fatty acids was highest when assessed by the 24hR (40.2%), followed by the DP (37.4%) and FFQ (35.5%). The LA intake percentage was highest when assessed by the FFQ (19.2%). For MUFA intake percentages were similar for the three dietary assessment methods as was the case for n-3 fatty acids and the n-3/LA ratio.

DP and 24hR as reference methods for FFQ validation

The highest validity coefficient for the FFQ was seen for SFA and was comparable regardless of whether the DP or 24hR was used as the reference method (0.76 for DP, 0.82 for 24hR, Table 2). For the other fatty acids, validity coefficients for the FFQ were lower when the DP was used as reference method than when the 24hR was used as reference method. This was especially true for MUFA (0.37 for DP, 0.65 for 24hR) and the n-3/LA ratio (0.33 for DP, 0.76 for 24hR, Table 2). Attenuation factors for the FFQ were rather similar when the DP was used as the reference method compared to the 24hR (Table 2), except for MUFA for which the attenuation factor was higher when the DP was used (0.34) as the reference method than when the 24hR was used (0.21). The attenuation factor for the FFQ for MUFA was lowest, irrespective of whether the DP or 24hR was used as the reference method, as compared to the other fatty acids.

Table 3: Validity coefficients of the DP and 24hR for n-3, LA and n-3/LA ratio where the mean of two plasma fatty acid values (expressed as % of total fatty acids) were used as reference method

		n-3		LA		n-3/LA ratio	
	k		CI		CI		CI
DP	1	0.33	0.20-0.45	0.18	0.07-0.30	0.34	0.22-0.47
	2	0.39	0.25-0.54	0.22	0.09-0.36	0.41	0.26-0.56
24hR	1	0.22	0.11-0.32	0.21	0.12-0.29	0.24	0.15-0.34
	2	0.28	0.15-0.41	0.27	0.16-0.39	0.32	0.20-0.45

24hR= 24hour recall, CI=confidence interval, DP=duplicate portion, LA=Linoleic acid, k = number of measurements, n-3=n-3 fatty acids
Models were adjusted for BMI and gender
Estimates were obtained using model 2 (equation 5 and 6) and formula 3

Ranking ability of DP and 24hR

To additionally compare the performance of the DP and 24hR, biomarker measurements were used as reference method. Validity coefficients were used to assess the ability of both methods to rank individuals according to their fatty acid intake. Validity coefficients for the ranking based on a single DP for the n-3 fatty acids (0.33) and n-3/LA ratio (0.34) were slightly higher than those for a single 24hR (0.22 and 0.24 respectively, Table 3). For LA, validity coefficients were similar (0.18 for DP and 0.21 for 24hR). A similar pattern was observed for validity coefficients based on two DP and two 24hR measurements.

Discussion

Fatty acid intakes expressed in grams were (slightly) higher when assessed by the DP as compared to the 24hR. For the fatty acid intakes expressed as percentages of total fatty acids, differences between the dietary assessment methods did not show a clear pattern. Validity coefficients for fatty acid estimates by the FFQ were lower or comparable when the DP was used as reference method than when the 24hR was used. For attenuation factors however, the DP as reference method showed a slightly higher value for MUFA, and similar results for the other fatty acids as compared to the 24hR. Using plasma fatty acids as reference method showed that the DP was able to rank participants according to their intake of n-3 fatty acids, LA and the n-3/LA ratio to a similar degree or slightly better than the 24hR.

Intakes of fatty acids in our study population were comparable with those of the general Dutch population based on the 2007-2010 Dutch National Food Consumption Survey (DNFCS) (19). The DNFCS intake data are based on two telephone-based 24hRs and the same FCD (2011) as we used to calculate nutrient intakes. Assessment of nutrient intake is among others limited by the availability and quality of the data in the FCD. Fatty acid composition of foods may change over time and vary amongst different brands. However, a study comparing calculated and analysed test diets for controlled dietary interventions found a reasonable agreement between the two for SFA and MUFA (20) indicating the Dutch FCD performs reasonably well for these fatty acids.
Published data on validity coefficients for FFQs for fatty acids intake estimates are scarce. One study, using the method of triads with the biomarker and weighed food records as reference method, found a validity coefficient of 0.50 for n-3 fatty acids assessed by FFQ (21), which is comparable to our results. A study by Kabagambe *et al*, also using the method of triads, found validity coefficients for the FFQ for LA between 0.77 and 0.89 (22), using the biomarker

and 24hR as reference methods. This is in line with our findings for LA when using the 24hR as reference method. Although differences in the statistical method to assess validity coefficients, adjustment for different covariates, study population, validity of the FCD and characteristics of the FFQ may hamper comparability of studies, our findings were in the same order of magnitude as the results previously published.

To be able to estimate the model parameters, we had to make assumptions, which in practise probably do not hold. These assumptions are universally made when the 24hR is used as reference method and are not specifically related to the use of measurement error models. In our first model we made the assumption of negligible error correlation between FFQ and DP or 24hR and between replicates of the reference methods, and the absence of proportional scaling bias for the DP and 24hR. Previous research showed that correlated errors between FFQ and 24hR and also between FFQ and DP were present and so was proportional scaling bias for the DP and 24hR for energy, protein, potassium and sodium intake (7, 8, 10). It would thus be likely that correlated errors and proportional scaling bias are also present when assessing fatty acid intake. The presence of correlated errors between FFQ and reference method will lead to an overestimation of validity coefficients and attenuation factors for the FFQ when using DP or 24hR as reference method (23). We previously showed that less correlated errors were present between DP and FFQ than between 24hR and FFQ (10). This would imply that the validity coefficients of the FFQ obtained with the DP as the reference method would show less overestimation. We indeed observed lower validity coefficients for fatty acid estimates by the FFQ when the DP was used as reference method than when the 24hR was used. However, correlation of errors between replicates are also to be expected and would cause the validity coefficient to be underestimated (23). In practice it is unclear which of these two biases, under- or overestimation of validity coefficients, will predominate. For attenuation factors the influence of the proportional scaling bias also needs to be taken into account. Assuming this bias is mostly smaller than one (7, 10, 24), the attenuation factor will be overestimated. However, it is unclear what the net effect of all of this will be on the attenuation factors.

In our second model we assumed negligible error correlation between biomarker and DP or 24hR and between replicates of the biomarker. In addition, absence of proportional scaling bias for the biomarker was assumed, however if this assumption is not met this does not affect the comparability of validity coefficients for DP and 24hR. The assumption of uncorrelated errors between biomarker and DP or 24hR is likely to hold since the errors in the biomarker measurement are assumed to be mostly physiological where the errors in DP and 24hR are due to the reporting of dietary intake. However, an individual's

digestion, absorption and metabolism are likely to influence concentration biomarker measurements (25), causing error correlations between replicates of the biomarker. Due to this error correlation, validity coefficients for the DP and 24hR will be underestimated which limits their interpretation as the calculated values should be interpreted as lower limit of the range of potential validity coefficient estimates. However, errors in the biomarker estimates are assumed to influence the validity coefficients for DP and 24hR equally, thus comparison of the two methods is possible. The finding that the DP had comparable or slightly better ranking abilities than the 24hR is therefore sound.

Using DP or 24hR as reference methods for FFQ validation enables to assess the validity of a wide range of fatty acids, while plasma fatty acids can only be used to evaluate ranking based on intakes of fatty acids that are not endogenously produced. Furthermore, DPs and 24hRs can be used to assess the validity of absolute FFQ fatty acid intakes, while the plasma fatty acids can only be expressed as percentage of total fatty acids. DPs are not affected by errors originating from the FCD, which is the case for 24hRs, while also portion size estimation bias and the influence of memory are expected to be smaller for DP. Moreover, if a combination of 24hRs and FFQ is used as dietary assessment method in future studies, 24hRs can no longer be used as independent reference method to evaluate the performance of the combined method while DP could very well fulfil that role.

In conclusion, taking into account that the assumptions made in our models do not allow us to draw firm conclusions, validity of assessment of fatty acid intake by FFQ differs slightly when the DP is used as reference method as compared to the conventionally used 24hR. The DP seems to perform slightly better than the 24hR when used to obtain validity coefficients for the FFQ, where for attenuation factors for the FFQ the use of DP or 24hR as reference method seem comparable. Therefore, the DP seems a promising reference method for FFQ validation of fatty acid intake.

References

1. Vera-Ramirez L, Ramirez-Tortosa MC, Sanchez-Rovira P, et al. Impact of Diet on Breast Cancer Risk: A Review of Experimental and Observational Studies. *Critical Reviews in Food Science and Nutrition* 2013;53(1):49-75.
2. Khodarahmi M, Azadbakht L. The association between different kinds of fat intake and breast cancer risk in women. *International Journal of Preventive Medicine* 2014;5(1):6-15.
3. Chowdhury R, Warnakula S, Kunutsor S, et al. Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Annals of internal medicine* 2014;160(6):398-406.
4. Kromhout D, Geleijnse JM, Menotti A, et al. The confusion about dietary fatty acids recommendations for CHD prevention. *British Journal of Nutrition* 2011;106(5):627-32.
5. Kaaks R, Riboli E, Esteve J, et al. Estimating the Accuracy of Dietary Questionnaire Assessments - Validation in Terms of Structural Equation Models. *Statistics in Medicine* 1994;13(2):127-42.
6. Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *American Journal of Epidemiology* 2001;153(4):394-403.
7. Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* 2003;158(1):14-21.
8. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology* 2011;174(5):591-603.
9. Arab L. Biomarkers of fat and fatty acid intake. *Journal of Nutrition* 2003;133(3 SUPPL.):925S-32S.
10. Trijsburg L, de Vries JH, Boshuizen HC, et al. Comparison of duplicate portion and 24 h recall as reference methods for validating a FFQ using urinary markers as the estimate of true intake. *British Journal of Nutrition* 2015;114(8):1304-12.
11. Conway JM, Ingwersen LA, Vinyard BT, et al. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *American Journal of Clinical Nutrition* 2003;77(5):1171-8.
12. Donders-Engelen M, van der Heijden L. *Maten, gewichten en codenummers 2003*. Wageningen: Wageningen UR, Vakgroep Humane Voeding; 2003.
13. Siebelink E, Geelen A, De Vries JHM. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *British Journal of Nutrition* 2011;106(2):274-81.
14. Streppel MT, De Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. *Nutrition Journal* 2013;12(1).
15. Nevo. Dutch Food Composition Database The Hague: Stichting Nevo, 2011.
16. Horwitz W. *Official methods of analysis of the Association of Official Analytical Chemists*. Washington D.C.: Association of Official Analytical Chemists; 1975.
17. de Goede J, Verschuren WMM, Boer JMA, et al. N-6 and N-3 Fatty Acid Cholesteryl Esters in Relation to Fatal CHD in a Dutch Adult Population: A Nested Case-Control Study and Meta-Analysis. *PloS one* 2013;8(5).
18. Simopoulos AP. The importance of the omega-6/omega-3 fatty acid ratio in cardiovascular disease and other chronic diseases. *Experimental biology and medicine (Maywood, NJ)* 2008;233(6):674-88.

19.   van Rossum CTM, Fransen HP, Verkaik-Kloosterman J, et al. Dutch National Food Consumption Survey 2007-2010 : Diet of children and adults aged 7 to 69 years RIVM, 2011.

20.   Siebelink E, de Vries JHM, Trijsburg L, et al. Evaluation of calculated energy and macronutrient contents of diets provided in controlled dietary intervention trials by chemical analysis of duplicate portions. *Journal of Food Composition and Analysis* 2015;43:68-74.

21.   McNaughton SA, Hughes MC, Marks GC. Validation of a FFQ to estimate the intake of PUFA using plasma phospholipid fatty acids and weighed foods records. *British Journal of Nutrition* 2007;97(3):561-8.

22.   Kabagambe EK, Baylin A, Allan DA, et al. Application of the method of triads to evaluate the performance of food frequency questionnaires and biomarkers as indicators of long-term dietary intake. *American Journal of Epidemiology* 2001;154(12):1126-35.

23.   Kaaks R, Ferrari P, Ciampi A, et al. Part H. Advances in the statistical evaluations and interpretation of dietary data: Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public health nutrition* 2002;5(6 A):969-76.

24.   Freedman LS, Midthune D, Dodd KW, et al. A statistical model for measurement error that incorporates variation over time in the target measure, with application to nutritional epidemiology. *Statistics in Medicine* 2015.

25.   Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.

Chapter 4

Validity of absolute intake and nutrient density of protein, potassium and sodium assessed by various dietary assessment methods

Laura Trijsburg
Anouk Geelen
Paul J.M. Hulshof
Pieter van ´t Veer
Hendriek C. Boshuizen
Peter C.H. Hollman
Gertjan van Dijk
Edith J.M. Feskens
Jeanne H.M. de Vries

Abstract

It is suggested that nutrient densities are less affected by measurement errors than absolute intake estimates of dietary exposure. We compared the validity of absolute intakes and densities of protein (kJ from protein/ total energy (kJ)), potassium and sodium (potassium or sodium (in mg)/ total energy (kJ)) assessed by different dietary assessment methods. For 69 Dutch subjects, two duplicate portions (DPs), five to fifteen 24-hour recalls (24hRs, telephone-based and web-based) and two food frequency questionnaires (FFQs) were collected and compared to duplicate urinary biomarkers and one or two doubly labelled water measurements. Multivariate measurement error models were used to estimate validity coefficients (VCs) and attenuation factors (AFs). This research showed that, group bias diminished for protein and sodium densities for all assessment methods as compared to the respective absolute intakes, but not for potassium. However, for the four methods and nutrients considered, the VCs and AFs for the nutrient densities compared to absolute intakes, did not improve; except for the AF for sodium density (0.71) of the FFQ which was better than that of the absolute sodium intake (0.51). Thus, using nutrient densities rather than absolute intakes does not necessarily improve the performance of the DP, FFQ or 24hR.

Introduction

In nutritional epidemiology, it is common practice to focus on the variation in dietary composition, by using either energy adjustment or nutrient densities (1). These methods reduce between person variation due to extraneous factors (which are not confounders) such as differences in body composition (1). Moreover, these methods may reduce the impact of measurement errors on estimates of dietary exposure, and thus strengthen the observed diet-disease associations (2). In the OPEN-study, protein densities instead of absolute intakes estimated by 24-hour recalls (24hRs) and food frequency questionnaires (FFQs) were indeed less affected by measurement errors (3). On the other hand, there is also evidence that showed a weakening of the observed diet-disease association based on nutrient densities as compared to absolute intakes (4). This could be ascribed to substantial measurement error in the estimated energy intake (5). To objectively evaluate energy intakes, the doubly labeled water (DLW) technique should be used, under the assumption of a stable body weight of the subjects (6).

A common method to express dietary composition is by nutrient densities, where the energy intake derived from the nutrient, or the absolute amounts consumed (for non-energy bearing nutrients) is divided by total energy intake. Nutrient densities can be calculated directly from the data on the individual level (1). A pooling of 5 American validation studies, including the before mentioned OPEN study, showed that protein, potassium and sodium densities were less affected by measurement error compared to the absolute nutrient intakes estimated by the FFQs but this was not so pronounced for 24hRs (7, 8). In the present study, we aimed to compare the validity of nutrient densities and absolute intakes of protein, potassium and sodium estimated by four dietary assessment methods: FFQ, telephone-based 24hR (24hRT), web-based 24hR (24hRW) and duplicate portion (DP). As reference methods we used the respective recovery biomarkers of these nutrients and DLW for the intake of energy.

Methods

Study participants and design

The study set-up has previously been described (9). In short: 200 participants of DuPLO, a Dutch validation study which is part of the NQplus study, were invited by email to have their energy expenditure assessed by DLW. Recruitment stopped after the targeted sample size of 70 participants was reached. Furthermore, 30 of these subjects completed a second energy expenditure measurement by DLW (~ 5 months later). The participants, aged 20-70 years,

lived in Wageningen and surroundings, the Netherlands. Baseline measurements included a physical examination, and general and lifestyle questionnaires (including questions about health and education). Within a timeframe of 1.5 years, participants collected two DPs (~ 5 months apart), two 24-hour urines (~ 1 year apart), completed zero to nine (average 6) 24hRW (~ 3 months apart) and zero to eight (average 5) 24hRT (~ 4 months apart), and filled out two FFQs (~ 7 months apart). The number of 24hRT and 24hRW per subject varied because part of the participants was enrolled in a sub-study of the NQplus study in which larger numbers of 24hRs were aimed at. In the DuPLO study we used all available 24hR data from the NQplus study. As for all methods the time between replicates varied per person, we report the average time between the replicates. Participants with DLW data but missing data for one or more of the other methods were included in the data analysis. Written informed consent was obtained from every participant. This study was approved by the medical ethical committee of Wageningen University.

Dietary assessment by 24-hour recall and food frequency questionnaire

Both the telephone and web-based 24hR assessments followed a standardized protocol according to the 5-step multiple pass method (10). For the 24hRW, participants received an unannounced email invitation to fill out the 24hRW in the web-based program Compl-eat, to report intake of the day before. If participants did not complete the 24hRW within 24 hours a new invitation was sent within three to ten days. For the 24hRT, participants got an unannounced phone call from a trained dietician. Portion sizes for 24hRW and 24hRT were reported using household measures, standard portion sizes, weight in grams, or volume in liters (11).
A previously validated 180 item FFQ (12, 13) was administered via the web using the online open-source survey tool Limesurvey^{TM}. The reference period for the FFQ was one month and standardized household measures were used to assess portion sizes (11).
Self-reported dietary intake data from 24hRW, 24hRT and FFQ were converted into energy and nutrient data using the Dutch food composition database of 2011 (14).

Duplicate portion and 24-hour urine collection

Participants received verbal and written instructions preceding the collection of the DP and 24-hour urine. For the DP the participants collected an identical portion of all edible foods and drinks consumed over a 24-hour period in collection baskets and stored them in a cool box (5°C). At the study center, DPs were weighed, homogenized in a blender (Waring Commercial model 34BL22)

and 2.5 ml 0.02% tert-butylhydrochinon in ethanol was added per kg of DP as antioxidant. Samples were stored within 1 hour at -20°C until further analysis. Part of the sample was freeze dried before analysis.

The 24-hour urine collection started after discarding the first voiding on the morning of the collection day and included the first voiding on the morning of the next day. The preservative lithium dihydrogenphosphate (25 g) was added to the collection containers. Subjects were instructed to ingest in total three 80 mg para-aminobenzoic acid (PABA) tablets (PABA check, Elsie Widdowson Laboratory, Cambridge, UK) during breakfast, lunch and dinner on the day of collection to check for completeness of the urine collection. At the study center, the urine collections were mixed, weighted, aliquoted and stored at -20°C until further analyses.

Laboratory analysis

Potassium and sodium from urine were analyzed by ion selective electrode (Roche 917 analyzer, Indianapolis, USA) and the intake was calculated taking into account extra-renal and fecal losses of 19% for potassium (15) and 14% for sodium (16). Potassium and sodium of the DP were assessed by inductively coupled plasma atomic emission spectroscopy (ICP-AES, Varian Australia Pty Ltd., Mulgrave, Australia; ISO, 2010). Nitrogen was assessed with the Kjeldahl technique (17) in both DP and urine. The amount of protein was calculated using a nitrogen to protein conversion factor of 6.25 (18), and an average ratio of urinary nitrogen excretion to dietary nitrogen of 0.81 (19) was assumed. The fat content of the DP was assessed by the acid hydrolysis method (20), ash by heating the freeze dried food in a muffle furnace at 550°C (21), alcohol by gas chromatography (22), and the moisture content was assessed by drying in a vacuum oven (21). PABA in urine was analyzed by HPLC (23). See Appendix I for quality control measures.

We assumed water, ash, fat, protein, alcohol and total carbohydrates (including dietary fiber) to sum up to 100% of the total weight of the DPs (24). Total carbohydrates were calculated by difference (25). Energy content of the DPs was subsequently calculated from the total amount of protein, fat, total carbohydrates and alcohol using the general Atwater factors for these nutrients: 17, 37, 17 and 29 kJ per gram respectively.

Energy expenditure measured by doubly labeled water

Total energy expenditure for each participant covering an eleven day period was assessed by DLW method using the two-point protocol (26). In the morning on the first day of the DLW period, weight and height of each subject were

measured to the nearest 0.1kg and 0.1cm, respectively. Next, baseline urine and saliva samples were collected followed by ingestion of a dose of DLW. Saliva samples were collected as back up samples if urine samples would not be sufficient or generated invalid results. Subjects received a mixture of 1.8 g 10% enriched $H_2^{18}O$ (Centre for Molecular Research Ltd, Moscow, Russia) and 0.12 g 99.8% enriched ^2H_2O (Cambridge Isotope Laboratories, Inc, Andover, MA, USA) per kg body water. It was assumed that body weight of males and females comprised 55% and 50% body water respectively [27]. Additional urine and saliva samples were collected three and four hours post dose. Eleven days after the dosing, subjects revisited the study center at the same time as the three hours post dose collection on day 1. Body weight was re-measured and two urine and saliva samples were collected with an interval of one hour. Isotopic enrichment of the samples and diluted doses were analyzed at the Center for Isotope Research, Groningen, The Netherlands as described elsewhere [28]. Enrichments expressed as delta units were converted into parts per million excess [26, 29]. ^2H and ^{18}O dilution spaces were calculated from the plateau enrichments at three and four hours post dose. Total body water was calculated as the average of the ^2H dilution space divided by 1.041 and ^{18}O dilution space divided by 1.01 to account for non-aqueous isotope exchange [30]. The rate of carbon dioxide production was calculated by the equation proposed by Schoeller et al [31]. Total energy expenditure was calculated using the modified Weir equation [32] with a respiratory quotient of 0.85. See Appendix I for quality control measures.

Measurement error model

In the measurement error model it was assumed that protein, sodium and potassium intake assessed by urinary excretion and energy expenditure assessed by DLW were unbiased estimates of true intake [33]. A linear relationship between dietary intake assessed by DP, 24hRT, 24hRW, FFQ, or biomarker with the true (unknown) intake T was assumed. In the measurement error model, i indicates the person and j the occasion; α_X the constant bias; β_X proportional scaling bias; w_{xi} person specific bias (psb); ε_{xij} the random within person error with mean zero and constant variance for method X and ε_{Mij} similarly for the biomarker. Replicates contributed to the estimation of within person random error. Method X is either: DP, 24hRT, 24hRW or FFQ.

Biomarker:	$Mij = T + \varepsilon_{Mij}$	(1)
Method X:	$Xij = \alpha_X + \beta_X T + w_{Xi} + \varepsilon_{Xij}$	(2)

Statistical analysis

Data of one participant were excluded because of physiologically implausible body water changes between repeated measurements while body weight remained stable. Thus data of 69 participants (37 men, 32 women) were included for analysis, 29 of them had duplicate measurements (16 men, 13 women). Descriptive statistics are presented as mean±SD or percentages. Protein densities were calculated by dividing the energy provided by protein (1 g protein = 17 kJ) by total energy (in kJ). For potassium and sodium we used the ratio of the total amount of the nutrient (in mg) to total energy (in kJ). For the denominator of the biomarker densities we used the average energy expenditure per person from DLW if two measurements were available; otherwise the single DLW estimate was used (40 subjects). Using the average of two DLW measurements caused unwanted correlation between densities at the two time points. A sensitivity analysis, where densities for the urinary biomarkers at baseline were calculated with the first DLW measurement and for those participants with a second DLW measurement at year one with the second DLW measurement, did not substantially affect the model outcomes. We therefore report the data using the average of two DLW measurements. Visual inspection of QQ-plots of the data did not show evidence of non-normality. The validity coefficient (VC, ρ_{XT}, formula 3) was used to assess the loss of statistical power to detect a diet-disease association and the ability to rank participants according to their intake, whereas the attenuation factor (AF, λ_x, formula 4) provides information about the extent to which diet-disease associations are affected by measurement error.

$$\rho_{XT} = \sqrt{\frac{\beta_X^2 \ varT}{\beta_X^2 \ varT + \frac{var \ \varepsilon_{Xij}}{k} + varw_{Xi}}} \tag{3}$$

$$\lambda_X = \frac{\rho_{XT}^2}{\beta_X} \tag{4}$$

Where varT is the variance of the true nutrient intake, $var\varepsilon_{Xij}$ the variance of the random within person error, $varw_{Xi}$ the variance of the psb and k the number of replicates of the method. We assessed the theoretical case of obtaining an infinite (∞) number of measurements, in which within person variation cancels out from the equation.

To understand observed differences in VCs and AFs between methods the size of the different variances of psb and random within person error and the proportional scaling bias are relevant. To facilitate these comparisons between

methods and to enable comparison between different nutrients, we expressed the variances of the errors for all nutrients relative to the estimate of the variance of the true intake (3):

Random error variance ratio:
$$\frac{var\ \varepsilon_{Xij}}{varT} \quad (5)$$

Person specific bias variance ratio:
$$\frac{varw_{Xi}}{varT} \quad (6)$$

Combined error variance ratio:
$$\frac{var\ \varepsilon_{Xij} + varw_{Xi}}{varT} \quad (7)$$

Using the proposed minimum 78% PABA recovery as a cut-off point for complete urine collection (23), N=19 (14%) of the urines were judged incomplete. A sensitivity analysis in which urines with <78% PABA recovery were excluded did not substantially change the model outcomes. This was in line with findings of Subar *et a.* who observed a modest effect on correction factors when urines were excluded based on PABA recovery (34). We therefore report the results based on the complete urine set. For all statistical tests SAS version 9.3 (SAS Institute Inc. Cary, NC, USA, 2012) was used.

Results

At baseline, participants (N=69) were on average 57.3 (SD 9.1) years old, had a mean body mass index of 25.5 (SD 3.6) kg/m^2 and 20.3% was classified as low educated (primary or lower education) while 47.8% was classified as high educated (university or college degree).

On average, all four dietary assessment methods underestimated energy and nutrient intakes as compared to the biomarkers. Energy intake was underestimated by on average 20% (between methods range 16.1-21.8%) as compared to the energy expenditure measured by DLW (Table 1). Compared to their respective urinary biomarkers protein intake was underestimated to a comparable extent (between methods range 10.4-23.0%) while sodium was seriously underestimated (between methods range 27.3-41.3%). The percentage bias for protein and sodium densities was smaller than those for their respective absolute intakes. In contrast, bias in percent for potassium densities was larger than for absolute potassium intake.

For energy, the VC for a single FFQ measurement (0.63) was comparable with that based on three measurements for the DP (0.59, Table 2). The VC for energy based on three 24hRT was only 0.14 while it was 0.48 for three 24hRW. The AF for energy roughly followed a similar pattern: for a single FFQ measurement it

Table 1: Mean Intake and Bias for Energy, Protein, Potassium, Sodium and their Nutrient Densities for the Biomarker, DP, 24hRT, 24hRW and FFQ

	Energy (MJ)	Protein (g)	Protein density[a]	Potassium (mg)	Potassium density[b]	Sodium (mg)	Sodium density[b]
Biomarker							
Mean±SD	11.4±2.5	101.5±27.1	0.15±0.03	3888±886	0.35±0.09	4069±1263	0.36±0.12
DP							
Mean±SD	8.9±2.0	77.3±19.5	0.15±0.03	3554±868	0.41±0.08	2580±853	0.29±0.09
Bias[c,d] (%)	-19.7±17.2	-21.4±19.3	0.87±24.8	-6.4±22.3	21.7±34.3	-34.3±18.8	-16.0±25.5
24hRT							
Mean±SD	9.0±1.7	83.2±19.4	0.16±0.02	3506±691	0.40±0.08	2646±674	0.30±0.06
Bias[c,d] (%)	-16.1±20.5	-10.4±20.3	9.7±22.0	-4.9±19.2	18.9±25.7	-27.3±27.7	-10.4±34.0
24hRW							
Mean±SD	8.8±2.4	82.2±25.1	0.16±0.03	3429±1005	0.40±0.08	2663±832	0.31±0.07
Bias[c,d] (%)	-20.6±21.2	-16.5±23.0	7.8±21.3	-10.4±20.3	20.1±30.4	-30.1±25.6	-7.5±32.3
FFQ							
Mean±SD	8.8±2.6	76.4±25.6	0.15±0.02	3466±1018	0.40±0.06	2242±833	0.25±0.04
Bias[c,d] (%)	-21.8±19.0	-23.0±22.5	-0.25±20.8	-9.5±21.7	19.2±27.5	-41.3±23.8	-25.1±23.1

[a] kJ from protein divided by total energy in kJ
[b] mg potassium or sodium divided by total energy in kJ
[c] % bias was calculated on the individual level using the biomarker as the true intake and then averaged
[d] Mean±SD

Table 2: Validity Coefficients (VC) for Energy, Protein, Protein, Potassium, Sodium and their Nutrient Densities for the DP, 24hRT, 24hRW and FFQ with the Biomarker as the Reference Method

	k^a	Energy	Protein	Protein density	Potassium	Potassium density	Sodium	Sodium density
DP	1	0.49±0.12	0.70±0.07	0.28±0.11	0.66±0.10	0.31±0.12	0.53±0.09	0.48±0.09
	2	0.56±0.14	0.77±0.07	0.32±0.13	0.76±0.11	0.36±0.14	0.66±0.09	0.61±0.10
	3	0.59±0.14	0.80±0.07	0.34±0.13	0.81±0.12	0.39±0.15	0.73±0.08	0.69±0.09
	∞	0.68±0.16	0.87±0.07	0.39±0.15	0.93±0.14	0.46±0.18	1.00[b]	1.00[b]
24hRT	1	0.10±0.15	0.45±0.12	0.34±0.08	0.56±0.09	0.37±0.09	0.18±0.12	0.15±0.08
	2	0.13±0.18	0.54±0.13	0.43±0.09	0.68±0.10	0.45±0.11	0.23±0.16	0.20±0.11
	3	0.14±0.20	0.58±0.13	0.48±0.10	0.74±0.11	0.50±0.12	0.26±0.18	0.24±0.14
	∞	0.18±0.25	0.71±0.14	0.71±0.14	0.94±0.13	0.65±0.15	0.42±0.28	0.66±0.41
24hRW	1	0.34±0.09	0.44±0.08	0.31±0.07	0.57±0.08	0.21±0.09	0.49±0.11	0.22±0.09
	2	0.43±0.12	0.54±0.09	0.40±0.09	0.70±0.09	0.27±0.12	0.61±0.14	0.29±0.12
	3	0.48±0.12	0.59±0.10	0.45±0.10	0.76±0.10	0.30±0.13	0.67±0.14	0.34±0.13
	∞	0.68±0.15	0.75±0.10	0.63±0.13	0.97±0.12	0.41±0.17	0.92±0.17	0.63±0.24
FFQ	1	0.63±0.11	0.70±0.08	0.37±0.11	0.73±0.11	0.41±0.13	0.46±0.16	0.38±0.15
	2	0.65±0.11	0.72±0.08	0.40±0.12	0.76±0.11	0.44±0.14	0.48±0.17	0.42±0.16
	3	0.65±0.11	0.73±0.08	0.42±0.12	0.76±0.11	0.45±0.14	0.48±0.17	0.43±0.16
	∞	0.67±0.11	0.74±0.08	0.45±0.13	0.78±0.12	0.47±0.14	0.50±0.17	0.47±0.17

All measurement error models were adjusted for BMI and gender
[a] k=number of measurements, [b] No person specific bias was observed

was 0.51, but it was higher for three replicates of the DP (0.69, Table 3), and considerably lower for three 24hRT (0.17) and 24hRW (0.40).

Below VCs and AFs were compared between nutrient densities and absolute nutrient intakes for single measurements. Increasing the number of measurements (up to infinite), showed a comparable pattern as described for single measurements. VCs for protein densities were lower than for absolute intakes for the DP (0.28 vs 0.70, Table 2) and FFQ (0.37 vs 0.70) whereas for the 24hRT and 24hRW VCs for protein densities and absolute protein intake were comparable. For potassium, VCs were lower for the densities than for absolute potassium intake for all four methods. For sodium, VCs for densities and absolute intake were comparable for all methods except for the 24hRW, for which lower estimates for sodium density than for absolute sodium intake were observed (0.22 and 0.49 respectively).

Comparing AFs for the same method between protein densities and absolute protein intakes showed comparable estimates, except for the DP for which lower estimates for protein densities than for absolute protein intake were observed (0.30 and 0.78 respectively, Table 3). AFs for potassium densities for the DP, 24hRT and 24hRW were lower than for absolute potassium intakes while they were comparable for the FFQ. AFs for sodium density and absolute sodium intake were comparable for both DP and 24hRT, whereas for the 24hRW, a lower AF for sodium density (0.15) than for absolute intake (0.35) was seen. In contrast, for the FFQ, a higher AF for sodium density (0.71) than for absolute intake (0.51) was observed.

Table 4 shows the variances of the error components relative to the variance of the true intake. A lower ratio means the estimated intakes were less affected by random within person error, psb or combined error (the sum of the variances of random within person error and psb). Consistent with the concept of the methods, the combined error variance ratio for energy intake was highest for the 24hRW (2.60, Table 4) due to the high random error variance ratio (2.19). The FFQ had the highest psb variance ratio (0.76).

For DP and 24hRT, combined error variance ratios were higher for the nutrient densities than for the absolute intakes. For the FFQ, combined error variance ratios were lower for the nutrient densities than for the absolute intakes: for protein (0.44 vs 0.58, Table 4), potassium (0.70 vs 1.05) and sodium (0.25 vs 0.65), which can be largely attributed to the lower psb variance ratios for the nutrient densities. For the 24hRW, the combined error variance ratio for sodium density was higher, while for protein density it was slightly lower and for potassium it was lower (2.48 vs 3.03) than for the absolute intakes. The latter was due to the diminished random error variance ratio (from 2.92 to 1.87).

Table 3: Attenuation Factors (AF) for Energy, Protein, Protein, Potassium, Sodium and their Nutrient Densities for the DP, 24hRT, 24hRW and FFQ with the Biomarker as the Reference Method

	k^a	Energy	Protein	Protein density	Potassium	Potassium density	Sodium	Sodium density
DP	1	0.46±0.12	0.78±0.10	0.30±0.12	0.45±0.08	0.25±0.10	0.50±0.11	0.39±0.10
	2	0.61±0.15	0.96±0.11	0.39±0.16	0.60±0.10	0.34±0.14	0.78±0.14	0.63±0.13
	3	0.69±0.16	1.03±0.12	0.44±0.18	0.67±0.12	0.39±0.16	0.96±0.15	0.79±0.14
	∞	0.90±0.23	1.23±0.17	0.57±0.24	0.89±0.19	0.56±0.24	1.80±0.41	1.67±0.44
24hRT	1	0.09±0.13	0.42±0.11	0.27±0.07	0.40±0.08	0.23±0.07	0.15±0.10	0.11±0.06
	2	0.14±0.20	0.60±0.14	0.44±0.11	0.58±0.10	0.35±0.09	0.25±0.17	0.20±0.11
	3	0.17±0.24	0.70±0.15	0.56±0.13	0.69±0.11	0.43±0.11	0.33±0.22	0.29±0.16
	∞	0.29±0.40	1.04±0.21	1.21±0.33	1.11±0.21	0.73±0.20	0.84±0.58	2.09±1.96
24hRW	1	0.20±0.06	0.31±0.07	0.24±0.06	0.27±0.05	0.13±0.06	0.35±0.08	0.15±0.06
	2	0.32±0.08	0.46±0.08	0.39±0.10	0.40±0.06	0.21±0.09	0.54±0.12	0.27±0.11
	3	0.40±0.10	0.55±0.09	0.49±0.12	0.48±0.07	0.26±0.11	0.67±0.14	0.36±0.14
	∞	0.78±0.19	0.88±0.15	0.99±0.25	0.77±0.13	0.47±0.21	1.25±0.26	1.23±0.56
FFQ	1	0.51±0.09	0.65±0.09	0.52±0.17	0.49±0.09	0.45±0.15	0.51±0.18	0.71±0.28
	2	0.54±0.10	0.69±0.09	0.62±0.19	0.52±0.09	0.50±0.17	0.54±0.20	0.85±0.33
	3	0.55±0.10	0.70±0.09	0.66±0.21	0.53±0.09	0.53±0.17	0.56±0.20	0.90±0.35
	∞	0.57±0.10	0.72±0.10	0.75±0.24	0.55±0.10	0.57±0.19	0.59±0.21	1.04±0.41

All measurement error models were adjusted for BMI and gender
[a] k=number of measurements

Table 4: Error Variance Ratios and Proportional Scaling Biases for Energy, Protein, Potassium, Sodium and their Nutrient Densities for the DP, 24hRT, 24hRW and FFQ with the Biomarker as the Reference Method (Mean±SE)

	Energy	Protein	Protein density	Potassium	Potassium density	Sodium	Sodium density
DP							
Combined ratio[a]	0.84±0.23	0.41±0.10	0.81±0.20	1.21±0.40	1.40±0.43	0.81±0.27	1.19±0.42
Random ratio[b]	0.53±0.16	0.29±0.08	0.43±0.12	1.06±0.39	0.86±0.28	0.81±0.27	1.19±0.42
Psb ratio[c]	0.30±0.14	0.12±0.07	0.39±0.14	0.15±0.27	0.54±0.24	_[d]	_[d]
Prop scaling bias	0.51±0.16	0.62±0.10	0.26±0.11	0.97±0.25	0.38±0.17	0.56±0.13	0.60±0.16
24hRT							
Combined ratio[a]	1.21±0.31	0.91±0.20	1.39±0.32	1.38±0.43	2.15±0.63	1.39±0.51	1.91±0.71
Random ratio[b]	0.83±0.21	0.68±0.15	1.22±0.28	1.30±0.43	1.68±0.50	1.19±0.44	1.86±0.70
Psb ratio[c]	0.38±0.14	0.23±0.09	0.17±0.09	0.08±0.17	0.46±0.22	0.21±0.12	0.06±0.09
Prop scaling bias	0.11±0.16	0.48±0.15	0.42±0.10	0.80±0.20	0.58±0.17	0.21±0.15	0.21±0.13
24hRW							
Combined ratio[a]	2.60±0.64	1.63±0.35	1.51±0.34	3.03±0.92	2.48±0.72	1.50±0.52	2.05±0.76
Random ratio[b]	2.19±0.55	1.32±0.28	1.26±0.28	2.92±0.97	1.87±0.55	1.41±0.53	1.89±0.71
Psb ratio[c]	0.41±0.18	0.32±0.13	0.25±0.10	0.11±0.34	0.61±0.24	0.08±0.16	0.16±0.11
Prop scaling bias	0.59±0.18	0.63±0.14	0.41±0.10	1.21±0.29	0.35±0.16	0.68±0.23	0.32±0.15
FFQ							
Combined ratio[a]	0.92±0.27	0.58±0.16	0.44±0.11	1.05±0.40	0.70±0.22	0.65±0.25	0.25±0.09
Random ratio[b]	0.16±0.05	0.11±0.03	0.15±0.04	0.26±0.10	0.18±0.06	0.11±0.04	0.09±0.04
Psb ratio[c]	0.76±0.24	0.47±0.14	0.28±0.09	0.79±0.36	0.52±0.18	0.54±0.22	0.15±0.07
Prop scaling bias	0.78±0.19	0.75±0.14	0.26±0.09	1.11±0.29	0.38±0.14	0.42±0.19	0.21±0.09

All measurement error models were adjusted for BMI and gender

Prop scaling bias=proportional scaling bias, β_x in equation 2

[a] Combined error variance ration; calculated according to equation 7, [b] Random error variance ratio; calculated according to equation 5, [c] Psb ratio = Person specific bias variance ratio, calculated according to equation 6, [d] No person specific bias was observed

Proportional scaling bias, indicated by β_X in equation 2, is less if its value is closer to 1. Proportional scaling bias influenced the energy intake assessed by the 24hRT to a major extent (0.11, Table 4), whereas the FFQ was least influenced (0.78). Overall, proportional scaling biases affected the nutrient densities for all dietary assessment methods to a larger extent than the absolute nutrient intakes. However, absolute sodium intake and sodium density had comparable proportional scaling biases for DP (0.56 and 0.60) and 24hRT (both 0.21).

Discussion

In our study, the DP, both 24hRs and FFQ showed comparable patterns for group bias: bias for protein and sodium densities was less than that of the absolute intakes, whereas bias for potassium density was larger than that for absolute potassium intake. The VCs and AFs for DP, both 24hRs and FFQ did not improve for nutrient densities compared to absolute intakes of protein, sodium, and potassium, except for the AF of the FFQ for sodium. For potassium, densities performed less than absolute intakes, but also for protein and sodium this was seen for some of the VCs and AFs. Proportional scaling bias, random within person error and psb, all affected protein, potassium and sodium density estimates to a larger extent than their absolute nutrient intakes. Exceptions to this observation were seen for the FFQ, where the psb was smaller for all nutrient densities than for the absolute intakes, and for the 24hRW, where the random error for potassium density was smaller than for absolute potassium intake.

VCs and AFs of energy intake were highest for the FFQ and DP, followed by the 24hRW and least for the 24hRT. The poor validity for the 24hRT is consistent with findings from an American pooling project (7). It appeared partly due to a large proportional scaling bias (0.11). Although an explanation for the latter is lacking, it is clear that the errors in estimated energy intake carry forward to the estimated densities, most seriously for the 24hRT followed by the 24hRW and least for DP and FFQ.

Comparing our findings with an American pooling project showed that our finding of the higher group level bias of potassium density compared to potassium was similar. However they did not observe the consistent improvement of group level bias for protein and sodium densities as compared to absolute intakes. When we compare the VCs and AFs of the nutrient densities and absolute nutrient intakes our results were not in line with their findings either. They observed that VCs of FFQs improved for nutrient densities compared to absolute nutrient intakes, where for the 24hR, the VCs were rather comparable (7, 8). In our study the VCs did not show an improvement for any of the dietary assessment methods for the nutrient densities, and especially for

potassium density, they worsened compared to the absolute intakes. In the Pooling project the AFs improved for both the FFQ and 24hR for the nutrient densities compared to the absolute intakes (7, 8). We only observed an improvement of the AF for sodium density for the FFQ and especially for potassium density, AFs worsened compared to the absolute intakes. Our VCs and AFs for the absolute intakes generally tended to be higher than those observed in the Pooling project, where our VCs and AFs of the nutrient densities were of similar magnitude. Since our absolute nutrient intake already had a relatively high validity there was not much room for improvement of validity when using nutrient densities. Differences in validity were to be expected as the dietary assessment methods were not exactly the same, and also the dietary pattern of our population differed from that in the pooling project. Unfortunately, inference on such issues is limited by the precision of our estimates, because of the sample size of our study.

We observed that for all absolute nutrient intakes for the FFQ the psb variance ratios were larger than for the DP and 24hRs. This might be due to the specific methodological characteristics of the FFQ: grouping of foods into a limited number of food items limits the freedom to report specific foods which increases the person specific bias variance, especially when comparing to open ended dietary assessment methods that allow much more specificity at the food level (DP and 24hRs).

The combined error variance ratios for our FFQ for all nutrient densities were smaller than for the absolute intakes. Michels *et al* observed that error correlations between nutrients and energy from a FFQ were larger than those from a food diary (35). As error correlations between nutrient and energy intake partially cancel out when using densities (2, 36), this might explain the smaller combined error variance ratios for our FFQs for the nutrient densities compared to the absolute intakes. However this did not improve the VCs and AFs for the nutrient densities, as the proportional scaling bias was larger for the nutrient densities than the absolute nutrient intakes.

The different methods did not exactly cover the same time period. However, our interest was to evaluate the validity of a person's usual dietary intake not the dietary intake on a specific day. We assumed energy and nutrient intakes of a person to be fairly stable over the 1.5 year in which the person's measurements were taken. Thus although intake data measured by the different dietary assessment methods did not cover the same time period, the estimates could be considered to represent a person's usual energy and nutrient intakes.

We found that accounting for energy by means of energy densities does not necessarily diminish the impact of measurement errors on estimates of dietary exposure. These results serve to highlight the obvious, that validation studies

should be incorporated in the study design, irrespective of whether absolute dietary intake or nutrient densities are the measure of interest of dietary exposure in nutritional epidemiology.

From this study it can be concluded that in this rather small, highly educated Dutch population, expressing diet in terms of nutrient densities rather than absolute intakes did not improve the performance of the assessment methods for protein, potassium and sodium.

References

1.   Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. *American Journal of Clinical Nutrition* 1997;65(4 Suppl):1220S-8S; discussion 9S-31S.

2.   Willett W. Commentary: Dietary diaries versus food frequency questionnaires-a case of undigestible data. *International Journal of Epidemiology* 2001;30(2):317-9.

3.   Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* 2003;158(1):14-21.

4.   Bingham SA, Day NE. Using biochemical markers to assess the validity of prospective dietary assessment methods and the effect of energy adjustment. *American journal of Clinical Nutrition* 1997;65(4 Suppl):1130S-7S.

5.   Freedman LS, Kipnis V, Brown CC, et al. Comments on" Adjustment for total energy intake in epidemiologic studies". *American journal of Clinical Nutrition* 1997;65(4):1229S-31S.

6.   Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.

7.   Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014;180(2):172-88.

8.   Freedman LS, Commins JM, Moler JE, et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Potassium and Sodium Intake. *American Journal of Epidemiology* 2015;181(7):473-87.

9.   Trijsburg L, de Vries JH, Boshuizen HC, et al. Comparison of duplicate portion and 24 h recall as reference methods for validating a FFQ using urinary markers as the estimate of true intake. *British Journal of Nutrition* 2015;114(8):1304-12.

10.   Conway JM, Ingwersen LA, Vinyard BT, et al. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *American Journal of Clinical Nutrition* 2003;77(5):1171-8.

11.   Donders-Engelen M, van der Heijden L. *Maten, gewichten en codenummers 2003*. Wageningen: Wageningen UR, Vakgroep Humane Voeding; 2003.

12.   Siebelink E, Geelen A, De Vries JHM. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *British Journal of Nutrition* 2011;106(2):274-81.

13.   Streppel MT, De Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. *Nutrition Journal* 2013;12(1).

14.   Nevo. Dutch Food Composition Database The Hague: Stichting Nevo, 2011.

15.   Freisling H, Van Bakel MME, Biessy C, et al. Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *British journal of nutrition* 2012;107(6):910-20.

16.   Holbrook JT, Patterson KY, Bodner JE. Sodium and potassium intake and balance in adults consuming self-selected diets. *American Journal of Clinical Nutrition* 1984;40(4):786-93.

17.   Hambleton LG, Noel RJ. Protein Analysis of Feeds, Using a Block Digestion. *Journal of the AOAC* 1975;58:143-5.

18.   Jones DB. *Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins*. US Department of Agriculture Washington, DC; 1941.

19.   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. *American journal of Clinical Nutrition* 1985;42(6):1276-89.

20.   Horwitz W. *Official methods of analysis of the Association of Official Analytical Chemists*. Washington D.C.: Association of Official Analytical Chemists; 1975.

21.   Osborne D, Voogt P. *The analysis of nutrients in foods*. London: Academic Press Inc. Ltd., 24/28 Oval Road, London NW1 7DX.; 1978.

22.   Klee M, Meng C. Ambient Headspace Analysis with the Agilent 7683 Automatic Liquid Sampler Agilent Technologies. 1998.

23.   Jakobsen J, Ovesen L, Fagt S, et al. Para-aminobenzoic acid used as a marker for completeness of 24 hour urine: Assessment of control limits for a specific HPLC method. *European Journal of Clinical Nutrition* 1997;51(8):514-9.

24.   Food and Agriculture Organization of the United Nations. Food Energy-methods of analysis and conversion factors. Rome: FAO, 2003.

25.   Siebelink E, de Vries JHM, Trijsburg L, et al. Evaluation of calculated energy and macronutrient contents of diets provided in controlled dietary intervention trials by chemical analysis of duplicate portions. *Journal of Food Composition and Analysis* 2015;43:68-74.

26.   IAEA. *Assessment of Body Composition and Total Energy Expenditure in Humans Using Stable Isotope Techniques*. Vienna: International Atomic Energy Agency 2009.

27.   Chumlea WC, Guo SS, Kuczmarski RJ, et al. Body composition estimates from NHANES III bioelectrical impedance data. *International Journal of Obesity* 2002;26(12):1596-609.

28.   Guidotti S, Jansen HG, Aerts-Bijma AT, et al. Doubly Labelled Water analysis: Preparation, memory correction, calibration and quality assurance for δ2H and δ18O measurements over four orders of magnitudes. *Rapid communications in mass spectrometry* 2013;27(9):1055-66.

29.   Slater C, Preston T, Weaver LT. Stable isotopes and the international system of units. *Rapid Communications in Mass Spectrometry* 2001;15(15):1270-3.

30.   Racette SB, Schoeller DA, Luke AH, et al. Relative dilution spaces of 2H- and 18O-labeled water in humans. *American journal of physiology: endocrinology and metabolism* 1994;267(4 30-4):E585-E90.

31.   Schoeller DA, Ravussin E, Schutz Y, et al. Energy expenditure by doubly labeled water: Validation in humans and proposed calculation. *American journal of physiology Regulatory, integrative and comparative physiology* 1986;250(5).

32.   Weir JBD. New methods for calculating metabolic rate with special reference to protein metabolism. *Journal of Physiology-London* 1949;109(1-2):1-9.

33.   Jenab M, Slimani N, Bictash M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human genetics* 2009;125(5):507-25.

34.   Subar AF, Midthune D, Tasevska N, et al. Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *European journal of clinical nutrition* 2013;67(8):863-7.

35.   Michels KB, Bingham SA, Luben R, et al. The effect of correlated measurement error in multivariate models of diet. *American Journal of Epidemiology* 2004;160(1):59-67.

36.   Day N, Wong M, Bingham S, et al. Correlated measurement error—implications for nutritional epidemiology. *International Journal of Epidemiology* 2004;33(6):1373-81.

Appendix I: Laboratory quality control measures

Participation in the External Quality Assessment Scheme of the Dutch Foundation for Quality Assessment in Medical Laboratories showed bias of -1.6% and +1.1% and analytical variation was 1.6% and 1.2% for urinary K and Na respectively. Within run coefficients of variation (CVw) and between run coefficients of variation (CVb) in urine were: CVw<1% and CVb<1% for both K and Na, for protein CVw=1.6% and CVb=1.3% and for PABA CVw=1.9% and CVb=1.3%. For the DPs quality control measures were as follows: protein CVw<1% and CVb<1%, K CVw<1% and CVb<1%, Na CVw=1.1% and CVb=1.7%, fat CVw = 0.9% and CVb=4.0%, ash CVw = 0.7% and CVb = 1.1% and for alcohol CVw = 4.3% and CVb=10.8%.

For the DLW analysis of reference waters (biomedical enriched waters gravimetrically prepared from Vienna Standard Mean Ocean Water) showed analytical variations <0.5% for both isotopes and accuracy defined as deviation from the certified values were <1% for δ^2H and <0.3% for δ^{18}O. Isotope enrichment of ^{2}H and ^{18}O at three and four hour post dose differed on average 1.1% (range 0.0-4.5%) and 0.2% (range 0.0-1.1%) respectively. The ratio of deuterium dilution space to ^{18}O dilution space was on average 1.031 (range 1.000-1.073). Urine enrichments on the final day (day 11) were on average 44 ppm (range 24-65 ppm) above baseline for ^{2}H and 55ppm (range 27-90 ppm) for ^{18}O. Baseline values for ^{2}H and ^{18}O were 152 ppm (range 148-155 ppm) and 1994 ppm (range 1990-1998 ppm) respectively.

Chapter 5

BMI is the most consistent determinant related to misreporting of energy, protein and potassium intake measured by different dietary assessment methods

Laura Trijsburg
Anouk Geelen
Peter C.H. Hollman
Paul J.M. Hulshof
Edith J.M. Feskens
Pieter van ´t Veer
Hendriek C. Boshuizen
Jeanne H.M. de Vries

Abstract

Misreporting, mostly underreporting, of dietary intake is a generally known problem in nutritional research and is consistently shown to be associated with a high body mass index (BMI). The associations of basic determinants (BMI, gender, age and level of education) with misreporting of energy, protein and potassium intake from the duplicate portion method (DP), 24-h recall (24hR) and food frequency questionnaire (FFQ), were evaluated. Additionally, the association between BMI-related and other determinants, and misreporting was explored. Of 197 subjects, two DPs, two FFQs, two 24 hour urinary biomarkers and two 24hRs were collected within 1.5 years. Also of 69 subjects one or two doubly labelled water measurements were obtained. We assessed the association between the extent of misreporting by DP, 24hR and FFQ with the determinants using linear regression analysis. Higher BMI was associated with underreporting of dietary intake assessed by the different dietary assessment methods for energy, protein and potassium, except for potassium by the DP. Men tended to underreport protein by the DP, FFQ and 24hR and persons of older age underreported potassium but only by the 24hR and FFQ. If corrected for the basic determinants, the BMI-related and other determinants did not show a consistent association with misreporting of energy or nutrients by the different dietary assessment methods. As BMI was the most consistent determinant associated with misreporting, we conclude that BMI should always be taken into account when assessing and correcting dietary intake.

Introduction

Misreporting, mostly underreporting, of dietary intake is a generally known problem in nutritional research and is shown to affect self-report diet assessment methods, including food frequency questionnaires (FFQs), 24h recalls (24hRs), food records and the duplicate portion method (DP) (1-5). Identifying the determinants associated with misreporting of dietary intake may help to facilitate the adjustments of dietary assessment methods or development of correction methods.

A large body of evidence demonstrated that various determinants are associated with underreporting of energy intake. The identified determinants depended on the research question and analysis, population and the availability of different sets of determinants in the studies (6, 7). To assess the degree of misreporting of energy intake, energy expenditure measured by the doubly labelled water method (DLW), a recovery biomarker, is the preferred reference method (8). The method assumes that participants are in energy balance. For a limited number of nutrients other recovery biomarkers are available, including protein, potassium and sodium (8, 9). However, relatively few studies looked at the association of determinants with misreporting of these nutrients (10-12). Having a high body mass index (BMI) was consistently associated with underreporting of energy and nutrients for different dietary assessment methods (10, 11, 13-16). Results for gender do not always point in the same direction: underreporting of energy intake was found to be more prone in women than men for 24hRs (13, 17), and underreporting of energy assessed by FFQ was higher in men (18). Having a lower level of education (15, 18) and being of older age (13, 18) were also associated with underreporting of energy intake for both FFQ and 24hR. Although studies investigating misreporting include different sets of determinants, usually the determinants, BMI, gender, age and level of education (or another indicator of social economic status) are included; we will in this article refer to these determinants as the basic determinants. Other determinants reported to be associated with energy misreporting by 24hRs include, but are not limited to, body fatness (17, 19), smoking status and physical activity level (13). Our research aimed to assess the associations of the basic determinants BMI, gender, age and education level, with misreporting of energy, protein and potassium for three dietary assessment methods (FFQ, 24hR and DP). Our secondary aim was to additionally assess the associations of explorative determinants with misreporting of the same nutrients for these dietary assessment methods. We included a set of explorative determinants distinguishing between BMI-related determinants, as BMI is strongly associated with misreporting, and other determinants including personal characteristics (Table 1). The recovery biomarkers for energy, protein and potassium were used

to assess the degree of misreporting for energy and the respective nutrients. To our knowledge, determinants associated with misreporting have not yet been studied for three nutrients and three conceptually different dietary assessment methods.

Subjects and methods

Subjects and design

The recruitment and the DuPLO-study procedures, conducted between July 2011 and July 2014, are described elsewhere (20). Briefly, a subsample of 200 Dutch adults (92 men, 108 women) from the NQplus study, aged 20-70 years and living in the surroundings of Wageningen were recruited. Baseline measurements consisted of, amongst others, a physical examination, including weight and height, general questionnaires (including questions about education, health and smoking habits), and lifestyle and psychosocial questionnaires. In a timeframe of 1.5 years, each participant collected two DPs (~ 5 months apart), and two urine samples (~1 year apart). Also two self-reports by FFQ (~ 7 months apart) were handed in and telephone based 24hRs (~ 7 months apart) were performed of which the first two were used in this analysis. For 70 participants (37 men, 33 women) energy expenditure was assessed by doubly labelled water (DLW) and 30 of these participants completed a second DLW measurement. Energy expenditure by DLW was assessed between September 2012 and September 2013. This study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects were approved by the medical ethical committee of Wageningen University. Written informed consent was obtained from all subjects.

Dietary intake assessment

The 24hR was administered by trained dieticians following a standard protocol based on the five step multiple-pass method (21). The 180-item FFQ (22, 23) was self- administered using the open-source online survey tool Limesurvey^{TM} (LimeSurvey Project Team/Carsten Schmitz. Hamburg, Germany 2012). The Dutch food composition database of 2011 (24) was used to calculate energy, protein and potassium intake for the 24hR and FFQ.

For the DP all foods and drinks consumed over a 24-hour time period were collected by the participant and weighed, homogenized and stored until further analysis by the researcher. On the day of the 24-hour urine collections, participants were instructed to ingest three 80 mg para-aminobenzoid (PABA) tablets to check for urine completeness. At the research centre urines were mixed, weighted and aliquoted and stored at -20°C until further analyses.

Laboratory measures

Nitrogen was assessed with the Kjeldahl technique (25) in both DP and urine. The amount of protein was calculated using a nitrogen to protein conversion factor of 6.25 (26), and an average ratio of urinary nitrogen excretion to dietary nitrogen of 0.81 (27) was assumed. Potassium in urine was determined with an ion-selective electrode and intake was calculated taking into account 19% potassium (10) extra-renal and faecal losses. PABA in urine was assessed by HPLC method and based on the cut-off value of 78% PABA recovery (28), 70% of the urines was considered complete. Potassium in the DP was determined, after digestion of the samples in PTFE tubes using a MarsXpress microwave digestor (CEM, Matthews, NC, USA), with inductively coupled plasma atomic emission spectroscopy (ICP-AES, Varian Australia Pty Ltd., Mulgrave, Australia; ISO, 2010). The fat content of the DP was assessed by the acid hydrolysis method (29), ash by heating the freeze dried food in a muffle furnace at 550°C (30), alcohol by gas chromatography (31), and the moisture content was assessed by drying in a vacuum oven (30). We assumed water, ash, fat, protein, alcohol and total carbohydrates (including dietary fiber) summed up to 100% of the total weight of the DPs (32). Total carbohydrates were calculated by difference (33). Energy content of the DPs was subsequently calculated from the total amount of protein, fat, total carbohydrates (including dietary fiber) and alcohol using the general Atwater factors for these nutrients: 17, 37, 17 and 29 kJ per gram respectively.

Total energy expenditure for each participant covering an eleven day period, was assessed by doubly labeled water (DLW) method using the two-point protocol (34). Total energy expenditure was calculated using the modified Weir equation (35) where the respiratory quotient was assumed to be 0.85. A detailed description of the DLW protocol can be found elsewhere (Chapter 4).

Determinants

Physical measurements
Physical measurements were done at baseline by trained research assistants following a standardized protocol. Height was measured to the nearest 0.1 centimetre without shoes with a stadiometer (SECA, Germany). Weight was measured with empty pockets and without shoes and sweaters to the nearest 0.1 kg on a digital scale (SECA, Germany). BMI was calculated by dividing weight (kg) by the square of the body height in meters. Body fat percentage was measured by a Dual-energy X-ray Absorptiometry (DXA) scan (Lunar prodigy, GE healthcare). DXA quality-assurance measurements were performed daily to ensure scanner reliability. In case the participant's body did not fit the outline of

the scanner (N=1), only the right side of the body was scanned and results were doubled. In a subsample (n=27), fat percentage was measured using the Tanita body composition analyser (BC418MA, Tanita Corporation) instead.

General questionnaire
Participants were contacted by email and asked to complete questionnaires online using the open-source survey tool Limesurvey[TM]. The general questionnaire administered at baseline included questions about age, gender, household composition, current and previous smoking habits, dieting habits, opinion about body weight, highest and lowest weight in the past five years and education level. Never smokers were those who had not smoked in the past month and never smoked for a full year. Subjects who smoked in the past month or ever smoked for a full year and did not stop smoking were classified as current smokers. Subjects who ever smoked for a full year, but had not smoked in the past month and stopped smoking were classified as former smokers. Subjects with no education or primary or lower vocational education as highest completed education were classified as having a low education level. Subjects who completed lower secondary or intermediate vocational education were classified as having an intermediate education level and subjects with a high education level were those who completed higher secondary education, higher vocational education or university.

Physical activity
Physical activity was assessed by accelerometer, the triaxial GT3X or triaxial GT3X+ (Actigraph, Pensacola, Florida), and expressed in Metabolic Equivalents (METs) per day. Participants wore the accelerometer for seven continuous days on their hip and kept a record of daily activities. ActiLife version 6.7.3 (Actigraph) was used to assess daily MET scores with the equation developed by Swartz et al (2000) (36).

Perceived Stress Scale
The Perceived Stress Scale (PSS) measures the degree to which situations are considered as stressful. The PSS asks about feelings and thoughts in the last two weeks (37). Within this study, the 4-item version of the PSS, i.e. the PSS4, was used. A total PSS4 score was derived by reversing the scores of the two positively stated items and then summing across all 4 items (range 0 to 16). The internal consistency of the PSS4 was acceptable (alpha= 0.72) and the test-retest reliability was fair (0.55) (37).

Dutch Eating Behaviour Questionnaire

The Dutch Eating Behaviour Questionnaire (DEBQ) ranks participants on a scale of 1 to 5 on three eating styles; restrained eating i.e. conscious restriction of food intake, emotional eating i.e. eating resulting from negative moods, and external eating i.e. eating as a response on smell or sight of food (38). The questionnaire comprises of 33 statements to be rated on a 5-point scale. The mean of the total score for each eating style was taken and used for the analysis. The DEBQ was found to successfully identify the three dimensions of eating style in clinical and non-clinical groups (39).

Nutrition behaviour questionnaire

Based on the theory of planned behaviour and the trans-theoretical theory a questionnaire was developed to assess self-identity, knowledge and perceived barriers for healthy eating in general as previously described (40). In short, for self-identity the mean score of three statements about ones identification to be a healthy eater, with answering scales ranging from 1 to 7, was used. Knowledge was assessed by two types of questions, the first set consisted of statements about the Dutch dietary guidelines for a healthy diet and the second set asked participants to select the healthier choice from pairs of foods. For each correct answer the participants received one point and the sum score ranged from 0 to 17 points. Perceived barriers were assessed by 13 statements related to barriers for healthy eating on a 7-point scale. Mean scores ranging from 1 to 7 were used in this analysis.

Statistical analysis

Two participants got pregnant during the study. As it was expected that they changed their habitual dietary intake they were excluded from analysis. One participant did not collect urine samples and was therefore also excluded from analysis. In total 197 participants were included for analysis of protein and potassium misreporting, 91 males and 106 females. Furthermore, DLW energy expenditure data of one participant were excluded because of physiologically implausible body water changes between repeated measurements while body weight remained stable. Thus data of 37 men and 32 women were included for the analysis of energy misreporting. Participants with missing data for one or more of the methods were included in the analysis because they provided information for the other dietary assessment methods. In each table N represents the number of participants included for the analysis for the specific dietary assessment method. Misreporting (group level bias) was calculated as the intake assessed by a single measurement of FFQ, 24hR or DP minus the

Table 1: Classification of the determinants

Classification	Determinant	Method used to assess
Basic determinants	BMI (kg/m^2)	Physical measurements
	Age (years)	General questionnaire
	Gender	General questionnaire
	Education level (high-intermediate-low)	General questionnaire
Explorative BMI-related determinants	Opinion BW - too high (yes-no)	General questionnaire
	Dieting - sometimes (yes-no)	General questionnaire
	Total dieting attempts	General questionnaire
	5-yr weight difference (kg)	General questionnaire
	Physical activity (METs per day)	Accelerometer
	Body fat (in %)	DXA scan or Tanita body composition analyser
Explorative other determinants	Stress level (score 1-5)	Perceived Stress Scale (PSS4)
	Restrained eater (score 1-5)	Dutch Eating Behaviour Questionnaire
	Emotional eater (score 1-5)	Dutch Eating Behaviour Questionnaire
	External eater (score 1-5)	Dutch Eating Behaviour Questionnaire
	Knowledge about healthy eating (score 0-17)	Nutrition Behaviour Questionnaire
	Self-identity with healthy eating (score 1-7)	Nutrition Behaviour Questionnaire
	Perceived barriers for healthy eating (score 1-7)	Nutrition Behaviour Questionnaire
	Smoking (Never-yes-former)	General questionnaire
	Living with partner (yes-no)	General questionnaire
	Living with children (yes-no)	General questionnaire

BMI=body mass index, BW=body weight

mean of two measurements of the recovery biomarker intake for protein, potassium, or energy (where for DLW for N=40 participants only one measurement was available). This was done for both measurements of FFQ, 24hR and DP separately. The mean of the two biases was reported. The percentage bias was calculated by taking the mean of bias percentages at the individual level. A Student's paired *t* test between mean of the recovery biomarkers and the mean of the two replicates of FFQ, DP or 24hR was performed to test for statistical significance of misreporting. Descriptive statistics were presented in percentages and as means with their standard deviation. Multiple imputations were used to impute missing determinant data. Linear regression analysis was performed to relate the basic determinants (BMI, gender, age and education level), all in one model, to the difference between reported intake by FFQ, 24hR or DP and estimated intake based on the biomarker for energy, protein and potassium intake. For education two dummies were included in the model. Recovery of PABA (complete yes or no) was included in all models for protein and potassium as it is a methodological factor related to the urine collection. Next, the explorative determinants were added to this multivariate linear regression model, one at a time. For smoking, two dummies were included simultaneously in the model. All statistical tests were performed using SAS version 9.3 (SAS Institute Inc. Cary, NC, USA, 2012).

Results

Baseline characteristics of the study population

The participants had a mean BMI of 25.1 (SD 3.7) kg/m^2 and a mean age of 55.8 (SD 10.1) years (Table 2). Slightly more women were enrolled (53.8%) in the study than men and 52.8% of the participants completed a high level of education (university or college) while 18.8% finished a low level of education (primary or lower education).

Baseline values of the explorative BMI-related determinants showed that 54.3% found their body weight to be too high, 31.8% was sometimes dieting, and on average 4 dieting attempts were done. Furthermore, a median self-reported maximum weight difference within a 5-year period of 6 kg was observed. The mean physical activity level was 1.78 METs per day and participants had on average 28% body fat. For the explorative other determinants; mean stress scores and the scores on the DEBQ (restrained, emotional or external eater) were fairly average (around 2.5) and participants on average answered 15 out of 17 questions correct regarding knowledge about healthy eating. Furthermore, 7.8% was a current smoker, most participants were living with a partner (80.2%) and 21.2% lived with children.

Table 2: Baseline characteristics of the study participants (N=197)

	Mean (SD)
BMI (kg/m^2)	25.1±3.7
Age (y)	55.8±10.1
Total dieting attempts	4±10
5 year weight difference (kg)	6*
Physical activity (METs per day)	1.78±0.19
Body fat (%)	28.3±9.26
Stress level (1-5)	2.55±2.90
Restrained eater (1-5)	2.97±0.69
Emotional eater (1-5)	2.03±0.70
External eater (1-5)	2.49±0.50
Knowledge (0-17)	14.7±1.8
Self-identity (1-7)	5.51±0.87
Perceived barriers (1-7)	2.28±0.79
	Percentage
Gender - men	46.2
Gender - women	53.8
EDU-low	18.8
EDU-intermediate	28.4
EDU-high	52.8
Opinion BW-too high	54.3
Dieting-sometimes	31.8
Smoking-never	40.8
Smoking-current	7.8
Smoking-former	51.4
Living with partner	80.2
Living with children	21.8

*As this was a highly skewed variable the median is reported
A higher stress score means the perceived stress is higher
A higher score on restrained, emotional or external eating, means the eating behaviour of the person is more leaning to the specific eating pattern
A higher score on knowledge means the person has a higher knowledge about healthy eating
A higher score on self-identity means the person identifies him-/herself with eating healthy
A higher score on perceived barriers means the perceived barriers for eating healthy are higher

Misreporting of energy, protein and potassium intake

All dietary assessment methods significantly underestimated the intake of energy, protein and potassium compared to the biomarker (Table 3). Energy and protein were underestimated by approximately 20%, except for protein assessed by the 24hR, which was underestimated by 12% (data not shown). Potassium

intake was underestimated to a smaller extent; 8.2%, 6.8% and 3.9% by FFQ, DP and 24hR, respectively.

Basic determinants associated with misreporting

A higher BMI was associated with underreporting of energy and protein to a similar degree for all methods (Table 3). An increase in BMI of 1 kg/m2 led to an increase in underreporting of energy of 279 kJ for the DP, 204 kJ for the 24hR and 272 kJ for the FFQ and of protein of 1.3 g, 1.3 g and 2.0 g for the respective dietary assessment methods. BMI was also associated with underreporting of potassium assessed by 24hR and FFQ, but not by DP. Being of older age was associated with misreporting of potassium intake, also only by 24hR and FFQ. For a 1 year increase in age, potassium was underreported by an additional 17 mg by the 24hR and 16 mg by the FFQ. Men showed higher underreporting than women for protein assessed by all three dietary assessment methods but not for energy. No significant association was observed between misreporting and level of education.

Explorative determinants associated with misreporting

The explorative determinants were added to the model, one at a time, in addition to all basic determinants and the PABA recovery variable. For most determinants no significant associations were observed, below only statistically significant results ($p \leq 0.05$) are described. With respect to the BMI-related determinants participants who were of the opinion that they were too heavy, on average overreported protein assessed by the FFQ by 9.8 g as compared to those who were of the opinion they had a right body weight (Table 4). In contrast, a 1 kg larger weight difference within 5 years was associated with protein underreporting of 0.52 g but only when assessed by the DP. Also, a 1 METs higher physical activity level was associated with a higher level of underreporting of energy assessed by the 24hR and FFQ (3916 KJ and 3799 KJ respectively) and protein assessed by the 24hR (22.7 g). However, a higher percentage of body fat was associated with protein and potassium overreporting, just for the FFQ (0.91 g and 43.1 mg respectively).

Further exploration of the other determinants indicated a higher perception of barriers to eat healthy to be associated with overreporting of energy assessed by the 24hR (984 KJ). However, for those living in a household with children, energy underreporting, only by the DP (1495 kJ), was observed compared to those living without children. Also, being a current smoker was associated with underreporting, though just for protein (16.7 g as compared to never smokers) and potassium (600 mg) intake as assessed by the FFQ.

Table 3: Association between basic determinants and misreporting of energy,

| | Duplicate portion | | | 24hR |
	Energy(kJ) (N=69)	Protein(g) (N=197)	K(mg) (N=197)	Energy(kJ) (N=58)
Mean intake	8917±1915	76.8±19.1	3484±869	8809±2123
Bias (absolute)	-2445±2416*	-23.2±20.1*	-381±878*	-2278±2308*
BMI	-279±78.9*	-1.28±0.36*	-24.5±17.1	-204±93.4*
Age	6.69±31.2	0.04±0.13	1.40±6.32	-30.1±36.6
Gender	-776±554	-13.6±2.72*	-242±129°	-300±621
EDU-low	297±713	-0.84±3.55	-19.3±169	139±767
EDU- intermediate	184±631	1.16±3.05	51.3±145	1016±710

*significant at p≤0.05, ^{o}significant at p≤0.10
All determinants were put in the model at the same time together with PABA completeness (yes=0, no=1)
Coding: gender: 0=women, 1=men; EDU: relative to highly educated;

Table 4: Association between explorative determinants and misreporting of analyses

| | Duplicate portion | | | 24hR |
	Energy(kJ) (N=69)	Protein(g) (N=197)	K(mg) (N=197)	Energy(kJ) (N=58)
BMI-related determinants				
Opinion BW-too high	-544±655	2.79±3.30	-41.6±157	76.8±707
Dieting-sometimes	-307±635	-4.37±2.98	-154±146	-812±702
Total dieting attempts	-31.8±29.3	-0.08±0.14	-0.84±6.70	-49.2±41.0
5 year weight difference	11.4±91.7	-0.52±0.26*	-16.4±12.2	90.2±106
Physical activity	-2290±1600	-3.85±7.39	8.42±364	-3916±1824*
Body fat (%)	-55.8±62.4	0.29±0.28	3.50±13.2	61.7±73.4
Other determinants				
Stress level	-93.9±99.9	0.65±0.45	15.4±21.4	16.5±108
Restrained eater	-402±401	-1.23±1.94	25.8±92.5	-516±440
Emotional eater	-594±450	-0.42±2.06	-72.5±97.9	366±495
External eater	197±604	2.40±2.74	-27.0±131	897±649
Knowledge	-5.09±142	-0.15±0.80	-6.50±39.2	130±141
Self-identity	-291±316	-1.09±1.66	45.9±78.6	-528±332
Perceived barriers	542±405	-3.26±1.98	-109±95.6	984±444*
Smoking-current	1763±1144	-7.51±5.10	-116±254	1559±1122
Smoking-former	183±611	1.16±2.74	173±135	-11.9±710
Living with partner	59.4±741	-3.28±3.33	-215±158	743±873
Living with children	-1495±730*	-2.55±3.60	-157±168	-402±827

Determinants were entered in a multivariate linear regression model in addition to the basic determinants (BMI, gender, age and education level) and PABA completeness, one at a time
*significant at p≤0.05, ^{o}significant at p≤0.10

protein and potassium by DP, 24hR and FFQ according to regression analyses

FFQ				
Protein(g) (N=155)	K(mg) (N=155)	Energy(kJ) (N=69)	Protein(g) (N=193)	K(mg) (N=193)
81.8±22.0	3499±910	8504±2336	74.3±21.1	3400±868
-13.8±23.6*	-282±930*	-2592±2460*	-25.3±26.3*	-465±1067*
-1.33±0.48*	-42.2±19.4*	-272±81.7*	-2.04±0.48*	-61.6±20.3*
-0.25±0.19	-17.3±7.49*	-38.4±32.4	-0.24±0.18	-16.0±7.70*
-8.76±3.88*	28.2±155	-255±575	-11.6±3.64*	-132±155
2.38±4.90	-14.3±196	-293±739	0.63±4.73	79.7±201
4.22±4.19	260±168	-556±655	-1.65±4.06	96.4±172

energy, protein and potassium by DP, 24hR and FFQ according to regression

FFQ				
Protein(g) (N=155)	K(mg) (N=155)	Energy(kJ) (N=69)	Protein(g) (N=193)	K(mg) (N=193)
4.23±4.54	209±181	265±682	9.80±4.37*	245±187
-6.49±4.22	1.10±177	-806±641	-4.26±4.10	-200±175
0.19±0.20	1.90±7.91	-40.6±30.1	-0.04±0.19	-9.52±7.97
-0.35±0.35	-10.6±13.3	-26.0±95.2	-0.47±0.35	-15.9±14.7
-22.7±10.7*	-692±422	-3799±1714*	-11.9±9.98	-254±449
0.53±0.41	19.0±16.6	43.1±64.9	0.91±0.36*	43.1±15.4*
0.45±0.58	-9.92±23.4	-5.98±104	0.83±0.60	7.82±25.5
-3.45±2.71	-15.3±109	-372±417	-3.77±2.57	-93.3±110
4.25±2.79	30.4±113	297±471	3.22±2.75	37.0±118
4.20±3.73	174±149	596±623	2.63±3.66	109±156
-0.83±1.09	-7.69±45.9	102±141	-0.04±1.09	-29.5±46.6
-1.92±2.16	103±86.5	-318±323	-2.09±2.24	-11.7±94.2
-0.35±2.80	61.2±116	341±413	-1.67±2.69	-78.0±115
-0.18±6.93	220±275	-192±1163	-16.7±6.89*	-600±288*
-0.20±3.94	99.1±157	-428±645	-2.52±3.65	-24.7±157
2.49±4.56	-127±182	-285±767	2.64±4.45	31.4±189
1.92±5.07	-49.8±208	-1269±764°	-3.98±4.73	28.2±204

Discussion

As expected, this research showed that BMI is an important determinant associated with misreporting of energy, protein and potassium intake for the different dietary assessment methods, except for potassium assessed by the DP. Persons of older age underreported potassium to a larger extent by the 24hR and FFQ and men tended to underreport protein to a larger extent than women on the DP, 24hR and FFQ. Adding the explorative determinants, BMI-related and other, to the basic model did not show associations with misreporting for most of the determinants. For the explorative determinants which did show a significant association with misreporting, no consistent pattern over the three nutrients or dietary assessment methods was observed.

Our findings that a higher BMI was associated with underreporting were consistent with those of other studies (10, 13-16). The reason for the association between higher misreporting and a higher BMI might be found in the selective underreporting of certain foods (such as unhealthy snacks) by people with a higher BMI (41). Selective underreporting of foods may result in differential reporting of nutrients. Selective misreporting of nutrients was proposed by Subar *et al* (42) who observed larger underreporting of energy than of protein, suggesting a bias toward more underreporting of fat, carbohydrates or alcohol. Social desirability has also been suggested to underlie the underreporting of people with a higher BMI, however Taren *et al* (43) demonstrated that social desirability and BMI influence misreporting independently. In our study no variables about number of meals and snacks were included as this is not traceable for the DP. For macronutrients such as carbohydrates and fats no recovery biomarkers are presently known (8) so the degree of misreporting for such nutrients can only be estimated using imperfect reference methods. Furthermore, no data on social desirability were available. It may be worth investigating the relation between diet characteristics and social desirability in future research.

The other basic determinants which showed an association with misreporting were gender and age. Our results indicated that men showed higher underreporting than women for protein on the DP, 24hR and FFQ. In our study standard portion sizes were used for 24hR and FFQ (44), as men tend to eat more than women, this might have caused underestimation of intake for men. For example; the major source of protein intake for men is meat (29%) (45) thus if the meat consumption of men is underestimated this will lead to an underestimation of protein intake. However, this does not explain why men also underestimated protein intake by the DP, for the DP it might be possible that men considered it a waste to hand in meat for the DP and thus diminished their

meat consumption on the day of collection. Older age was significantly associated with higher under reporting of potassium intake in the models, but only for FFQ and 24hR. In our population there was little variation in age, thus drawing hard conclusions based on these findings cannot be done.

When looking at the significant explorative BMI related determinants, our findings are consistent with those of another study that also reported associations between the extent of misreporting of energy intake and a higher physical activity level (13). For body fatness, body image (which is related to opinion about one's body weight) and weight differences also associations with misreporting of energy were found (17). For the other explorative determinants, studies found an association between smoking status and misreporting of energy intake (7, 13). To our knowledge no other studies found associations with misreporting and perception of barriers or living with children. Direct comparison of these findings with our results is difficult as in these studies only misreporting of energy intake was investigated, and it should be taken into account that not all of these studies used DLW to assess the level of misreporting. Furthermore due to the high costs of DLW, the sample size for energy was small in our study. Also, physical activity level can be measured in different ways (e.g. with accelerometers (as in our study) or questionnaires), populations differ, and different questionnaires were used in the different studies. Percentage body fat is different for women and men. In our study women had a mean body fat percentage of 34% and men of 22%. However in none of the models the interaction term for body fat and gender was significant. Still this research was intended to explore the association of potential determinants with misreporting and the determinants showing a significant association reported here could be considered in future research.

It is worthwhile to discuss some of the determinants that did not show a significant association with misreporting, such as education, knowledge and restrained eating. For the first two variables, the reason for not observing an association may be the limited variation. Over 50% of our study population was highly educated and only 19% was classified as low educated. Next, underreporters could be expected to have a higher knowledge on healthy foods and thus (selectively) not report unhealthy foods. Our knowledge questionnaire might not have had a high discriminatory power (the average score was 15 out of 17) for our participants, as subjects were highly motivated (they were willing to fill in multiple questionnaires, collect two DPs, fill in two FFQs, administer multiple 24hR and 70 participants joined the DLW study) and could therefore also have a higher knowledge about healthy eating. Although we did not observe an association between restrained eating and misreporting, restrained eating has been associated with misreporting of energy intake (46, 47). However, there

are restrained eaters who report adequately and non-restrained eaters who report poorly (6, 7, 15) thus hard conclusions cannot be drawn.

Misreporting of dietary intake is often associated with weight status and specific food groups, thus true associations between diet and diseases may be distorted by this bias (48). Various methods have been proposed to adjust energy intake for misreporting, most of these methods are based on the exclusion of the group of implausible energy reporters (49, 50). The problem arises that this might cause selection bias (51), the people identified as being implausible reporters of energy intake might also be the group with specific characteristics, e.g. a higher BMI than the plausible reporters.

This study investigated the association between determinants and misreporting. Identifying determinants associated with misreporting could help in our understanding of the how, why and what processes are ongoing during dietary reporting. Eventually such information could be used to improve the assessment of dietary intake. An attempt has been made to develop a dietary assessment instrument for use in obese individuals, however after several validation studies in different obese target groups it did not prove to consistently show valid results (41).

Furthermore, information on these variables could be collected in nutritional epidemiologic studies and inclusion in calibration models could improve the calibration model. However deciding which covariables to include in a calibration model is complex, especially BMI is difficult, because of its multiple possible roles (52, 53). For example; BMI could be a reflection of energy consumption and physical activity level, adding up to an indicator of energy balance (53). Further research is needed as to how best to choose the covariates in a calibration model without introducing other sources of bias.

We conclude that in this Dutch study population, concerning the investigated determinants; BMI was most consistently associated with misreporting of energy or nutrients by the different dietary assessment methods. Thus BMI should always be taken into account when assessing and correcting dietary intake.

References

1. Crispim SP, Geelen A, De Vries JHM, et al. Bias in protein and potassium intake collected with 24-h recalls (EPIC-Soft) is rather comparable across European populations. *European Journal of Nutrition* 2012;51(8):997-1010.

2. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014;180(2):172-88.

3. Trabulsi J, Schoeller DA. Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. *American Journal of Physiology-Endocrinology And Metabolism* 2001;281(5):E891-E9.

4. Isaksson B. A critical evaluation of the duplicate-portion technique in dietary surveys. *European journal of clinical nutrition* 1993;47(7):457-60.

5. Stockley L. Changes in habitual food intake during weighed inventory surveys and duplication diet collections. A short review. *Ecology of Food and Nutrition* 1985;17(3):263-9.

6. Maurer J, Taren DL, Teixeira PJ, et al. The psychosocial and behavioral characteristics related to energy misreporting. *Nutrition reviews* 2006;64(2):53-66.

7. Livingstone MB, Black AE. Markers of the validity of reported energy intake. *Journal of Nutrition* 2003;133 Suppl 3:895s-920s.

8. Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.

9. Jenab M, Slimani N, Bictash M, et al. Biomarkers in nutritional epidemiology: applications, needs and new horizons. *Human genetics* 2009;125(5):507-25.

10. Freisling H, Van Bakel MME, Biessy C, et al. Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *British journal of nutrition* 2012;107(6):910-20.

11. Mossavar-Rahmani Y, Shaw PA, Wong WW, et al. Applying Recovery Biomarkers to Calibrate Self-Report Measures of Energy and Protein in the Hispanic Community Health Study/Study of Latinos. *American Journal of Epidemiology* 2015;181(12):996-1007.

12. Mossavar-Rahmani Y, Tinker LF, Huang Y, et al. Factors relating to eating style, social desirability, body image and eating meals at home increase the precision of calibration equations correcting self-report measures of diet using recovery biomarkers: Findings from the Women's Health Initiative. *Nutrition Journal* 2013;12(1).

13. Briefel RR, Sempos CT, McDowell MA, et al. Dietary methods research in the third National Health and Nutrition Examination Survey: Underreporting of energy intake. *American Journal of Clinical Nutrition* 1997;65(4 SUPPL.):1203S-9S.

14. Voss S, Kroke A, Klipstein-Grobusch K, et al. Is macronutrient composition of dietary intake data affected by underreporting? Results from the EPIC-Potsdam study. *European Journal of Clinical Nutrition* 1998;52(2):119-26.

15. Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. *American Journal of Clinical Nutrition* 2004;79(5):795-804.

16. Price GM, Paul AA, Cole TJ, et al. Characteristics of the low-energy reporters in a longitudinal national dietary survey. *British Journal of Nutrition* 1997;77(6):833-51.

17. Novotny JA, Rumpler WV, Riddick H, et al. Personality characteristics as predictors of underreporting of energy intake on 24-hour dietary recall interviews. *Journal of the American Dietetic Association* 2003;103(9):1146-51.

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

19.   Johnson RK, Soultanakis RP, Matthews DE. Literacy and body fatness are associated with underreporting of energy intake in US low-income women using the multiple-pass 24-hour recall: A doubly labeled water study. *Journal of the American Dietetic Association* 1998;98(10):1136-40.

20.   Trijsburg L, de Vries JH, Boshuizen HC, et al. Comparison of duplicate portion and 24 h recall as reference methods for validating a FFQ using urinary markers as the estimate of true intake. *British Journal of Nutrition* 2015;114(8):1304-12.

21.   Conway JM, Ingwersen LA, Vinyard BT, et al. Effectiveness of the US Department of Agriculture 5-step multiple-pass method in assessing food intake in obese and nonobese women. *American Journal of Clinical Nutrition* 2003;77(5):1171-8.

22.   Siebelink E, Geelen A, De Vries JHM. Self-reported energy intake by FFQ compared with actual energy intake to maintain body weight in 516 adults. *British Journal of Nutrition* 2011;106(2):274-81.

23.   Streppel MT, De Vries JH, Meijboom S, et al. Relative validity of the food frequency questionnaire used to assess dietary intake in the Leiden Longevity Study. *Nutrition Journal* 2013;12(1).

24.   Nevo. Dutch Food Composition Database The Hague: Stichting Nevo, 2011.

25.   Hambleton LG, Noel RJ. Protein Analysis of Feeds, Using a Block Digestion. *Journal of the AOAC* 1975;58:143-5.

26.   Jones DB. *Factors for converting percentages of nitrogen in foods and feeds into percentages of proteins*. US Department of Agriculture Washington, DC; 1941.

27.   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. *American journal of Clinical Nutrition* 1985;42(6):1276-89.

28.   Jakobsen J, Ovesen L, Fagt S, et al. Para-aminobenzoic acid used as a marker for completeness of 24 hour urine: Assessment of control limits for a specific HPLC method. *European Journal of Clinical Nutrition* 1997;51(8):514-9.

29.   Horwitz W. *Official methods of analysis of the Association of Official Analytical Chemists*. Washington D.C.: Association of Official Analytical Chemists; 1975.

30.   Osborne D, Voogt P. *The analysis of nutrients in foods*. London: Academic Press Inc. Ltd., 24/28 Oval Road, London NW1 7DX.; 1978.

31.   Klee M, Meng C. Ambient Headspace Analysis with the Agilent 7683 Automatic Liquid Sampler Agilent Technologies. 1998.

32.   Food and Agriculture Organization of the United Nations. Food Energy-methods of analysis and conversion factors. Rome: FAO, 2003.

33.   Siebelink E, de Vries JHM, Trijsburg L, et al. Evaluation of calculated energy and macronutrient contents of diets provided in controlled dietary intervention trials by chemical analysis of duplicate portions. *Journal of Food Composition and Analysis* 2015;43:68-74.

34.   IAEA. *Assessment of Body Composition and Total Energy Expenditure in Humans Using Stable Isotope Techniques*. Vienna: International Atomic Energy Agency 2009.

35.   Weir JBD. New methods for calculating metabolic rate with special reference to protein metabolism. *Journal of Physiology-London* 1949;109(1-2):1-9.

36.   Swartz AM, Strath SJ, Bassett D.R, Jr., et al. Estimation of energy expenditure using CSA accelerometers at hip and wrist sites. *Medicine and Science in Sports and Exercise* 2000;32(9 SUPPL.):S450-S6.

37.   Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *Journal of health and social behavior* 1983:385-96.

38.   Van Strien T, Frijters JER, Bergers GPA, et al. The Dutch eating behavior questionnaire (DEBQ) for assessment of restrained, emotional, and external eating behavior. *International Journal of Eating Disorders* 1986;5(2):295-315.

39.   Wardle J. Eating style: A validation study of the Dutch eating behaviour questionnaire in normal subjects and women with eating disorders. *Journal of Psychosomatic Research* 1987;31(2):161-9.

40.   Hooft van Huysduynen EJC. *Towards healthy diets for parents: effectiveness of a counselling intervention* [Dissertation]. 2014.

41.   Lissner L. Measuring food intake in studies of obesity. *Public health nutrition* 2002;5(6a):889-92.

42.   Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The OPEN study. *American Journal of Epidemiology* 2003;158(1):1-13.

43.   Taren DL, Tobar M, Hill A, et al. The association of energy intake bias with psychological scores of women. *Eur J Clin Nutr* 1999;53(7):570-8.

44.   Donders-Engelen M, van der Heijden L. *Maten, gewichten en codenummers 2003*. Wageningen: Wageningen UR, Vakgroep Humane Voeding; 2003.

45.   van Rossum CTM, Fransen HP, Verkaik-Kloosterman J, et al. Dutch National Food Consumption Survey 2007-2010 : Diet of children and adults aged 7 to 69 years RIVM, 2011.

46.   Asbeck I, Mast M, Bierwag A, et al. Severe underreporting of energy intake in normal weight subjects: Use of an appropriate standard and relation to restrained eating. *Public Health Nutrition* 2002;5(5):683-90.

47.   Bathalon GP, Tucker KL, Hays NP, et al. Psychological measures of eating behavior and the accuracy of 3 common dietary assessment methods in healthy postmenopausal women. *American Journal of Clinical Nutrition* 2000;71(3):739-45.

48.   Mendez MA. Invited commentary: Dietary misreporting as a potential source of bias in diet-disease associations: future directions in nutritional epidemiology research. *American journal of epidemiology* 2015;181(4):234-6.

49.   Mendez MA, Popkin BM, Buckland G, et al. Alternative methods of accounting for underreporting and overreporting when measuring dietary intake-obesity relations. *American Journal of Epidemiology* 2011;173(4):448-58.

50.   Rhee JJ, Sampson L, Cho E, et al. Comparison of methods to account for implausible reporting of energy intake in epidemiologic studies. *American Journal of Epidemiology* 2015;181(4):225-33.

51.   Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology (Cambridge, Mass)* 2004;15(5):615-25.

52.   Prentice RL, Huang Y. Measurement error modeling and nutritional epidemiology association analyses. *Canadian Journal of Statistics* 2011;39(3):498-509.

53.   Zheng C, Beresford SA, Van Horn L, et al. Simultaneous association of total energy consumption and activity-related energy expenditure with risks of cardiovascular disease, cancer, and diabetes among postmenopausal women. *American Journal of Epidemiology* 2014;180(5):526-35.

Chapter 6

General Discussion

The aim of this thesis was to determine the validity of the duplicate portion method (DP) compared to the 24h recall (24hR) as a reference method for food frequency questionnaire (FFQ) validation. Our second aim was to explore the validity of nutrient densities for DP, 24hR and FFQ. The third aim was to determine the factors associated with misreporting of energy and protein and potassium estimated by DP, 24hR and FFQ. In this chapter we will summarize the main findings, discuss methodological issues and implications of our work, propose possible directions for future research and give an overall conclusion.

Main findings

In chapter 2 we observed that the DP was less influenced by proportional scaling bias, had lower correlated errors with the FFQ and showed higher attenuation factors than the 24hR for potassium, sodium and protein. We therefore concluded that the DP is probably the best available reference method for FFQ validation for nutrients that currently have no generally accepted recovery biomarker. Following this observation, we compared the validity of the FFQ for saturated fatty acids, monounsaturated fatty acids, n-3 fatty acids, linoleic acid (LA) and the n-3/LA ratio, using DP and 24hR as reference methods (chapter 3). The DP seemed a better reference method than the 24hR for the assessment of validity coefficients. The attenuation factors for the FFQ, using either the DP or 24hR as reference method, agreed reasonably well. Furthermore, the DP showed, when using plasma fatty acids as reference, slightly better ranking of participants according to their intake of n-3 fatty acids (0.33) and the n-3/LA ratio (0.34) than the 24hR (0.22 and 0.24, respectively). In the next chapter, a comparison between the validity of the absolute intakes of protein, potassium and sodium and their densities showed that less group level bias was observed for protein and sodium densities compared to their absolute intakes for FFQ, 24hR and DP, but not for potassium (chapter 4). Overall the validity coefficients and attenuation factors for DP, 24hR and FFQ did not improve for nutrient densities compared to absolute intakes, except for the attenuation factor for sodium density (0.71) of the FFQ which was better than that of the absolute sodium intake (0.51). Lastly, BMI proved to be the most consistent determinant associated with misreporting (group level bias) of energy, protein and potassium for DP, 24hR and FFQ (chapter 5). Men tended to underreport protein by the DP, FFQ and 24hR and persons of older age underreported potassium but only by the 24hR and FFQ. Other explorative determinants did not show a consistent association with misreporting of energy or nutrients by the different dietary assessment methods.

Methodological considerations

The following paragraphs discuss methodological issues related to the research conducted. This includes the assessment of usual intake, the influence of timing of collections on usual intake estimates, the assumptions made in the different measurement error models as used in this thesis, and a discussion of the DP method.

Assessment of usual intake

Usual dietary intake is defined as an individual's long-term average daily intake of a food or nutrient (1). In our study we assessed an individual's usual intake over a timeframe of 1.5 years and assumed dietary intake over this time period would be stable. FFQs assess usual intake, where a single 24hR or DP assess actual intake, as does a single biomarker measurement for protein, potassium and sodium. Usual intake can be assessed with these methods if random day-to-day variance is accounted for by taking repeated measurements. Day-to-day variation in dietary intake could originate from e.g. differences in the eating pattern between week and weekend days (2) and seasonal variations in nutrient intake and food choice (3). The extent of variation depends on the nutrient of interest (4). Assuming uncorrelated errors in replicates, the measurement error models as used in the chapters 2, 3 and 4 allowed for the estimation of an individual's random day-to-day variance given that at least two replicates per individual were available for each method used. We estimated an individual's usual intake based on two days for the DP and biomarker. For both the telephone-based 24h recall and the web-based 24h recall, an average of 5 measurements was obtained. Validity coefficients and attenuation factors reported in chapters 2, 3 and 4 were corrected for random day-to-day variance in the methods. As correction for day-to-day variance was based on only two days for the DP, and biomarker (chapters 2 to 5), and the sample size was relatively small (chapter 4), our estimates might be imprecise estimates of usual dietary intake (5).

Timing of collections and usual intake

We assumed dietary intake over the 1.5 year to be fairly stable. However, intake could vary over time (6). It has recently been suggested that variation in time should be accounted for in the analysis (6, 7). Freedman *et al* showed a better fit of their time-varying usual intake model than of the fixed-time usual intake model that we used for our data analyses. Their findings indicated that for the time-varying usual intake model, estimates of validity coefficients and attenuation factors for FFQs were slightly larger but for 24hRs they were smaller

(6, 7). In the study by Freedman *et al* this could be explained by the timing of biomarker measurements related to the 24hR, as these were close in time. This could have caused high correlations using the fixed-time usual intake model for the 24hR where the time-varying usual intake model eliminates the influence of this proximity in time. The FFQ is probably less affected by the proximity in time with the biomarker since it assesses usual intake rather than actual intake. Therefore, the time-varying usual intake model does not much affect the estimates for this method. In our study we avoided taking biomarker measurements close to the 24hR or DP collections. Thus using the fixed-time model instead of the time-varying model is expected to be less of a problem with respect to overestimation of validity coefficients and attenuation factors for 24hR and DP.

Measurement error models and assumptions

The estimated model parameters depend highly on the model assumptions made, and wrong assumptions can seriously affect the calculated validity coefficients and attenuation factors (8, 9). Two characteristics of the reference method are crucial to the applicability of the measurement error models (10). To this end, the reference method should:

1) Have errors that are uncorrelated to errors in the method to be validated (i.e. the FFQ)
2) Be unbiased with respect to the true intake (no proportional scaling bias should be present)

Violation of these requirements will cause the correction to be flawed, and this could be in either direction (over- or under-correction). In which direction the incomplete correction will go, depends on how the different types of errors operate jointly and is an ongoing topic of research. In practice, estimates of attenuation factors obtained with an imperfect reference method have been shown to be inflated (results look too optimistic) (11).

Correlated errors between the reference method and the method to be validated (i.e. FFQ) will cause an overestimation of the true between person variance and thus the validity coefficient and attenuation factor will also be overestimated (12). However, this error correlation between methods is especially important if the true validity coefficients of the reference method and the main method are poor; if the methods perform reasonably well, the influence of correlated errors is moderate (13). Aside from the correlated errors between methods, correlated errors between replicates of the reference

methods may affect the validity of the estimated model parameters. Such errors may cause an underestimation of the validity coefficient of the main method (12, 13).

In this thesis different reference methods have been used in the different chapters. Below these different methods including recovery biomarker, concentration biomarker and the 24hR are described. Since an important aim of our study was to determine the validity of the DP as a reference method for FFQ validation, the DP is discussed separately.

Recovery biomarkers as reference method
In chapters 2, 4 and 5 recovery biomarkers were used as reference methods. The theory that errors from recovery biomarkers are uncorrelated with errors from self-report methods is based on the assumption that errors in biomarker measurements are mostly physiological. These errors are different from the errors in reported dietary intake; therefore the assumption of uncorrelated errors between FFQ, DP or 24hR and recovery biomarker seems reasonable. The assumption that 24-h urinary nitrogen yields unbiased estimates (no group level bias and not proportional scaling bias are present) of protein intake is well founded (14-16). Likewise evidence for unbiased estimates from doubly labelled water (DLW), as a measure for energy intake, is large (17-19), although in the calculation of energy expenditure assumptions have to be made about body water percentage and the respiratory quotient. However, DLW studies are very expensive, which limits the numbers of participants in validation studies.
Some controversy exists regarding urinary potassium and sodium. Varying conversion factors for the calculation of potassium from urinary potassium have been used (20) and potassium excretion differed between black and white participants in the DASH trial (21). Sodium excretion in the urine is shown to exhibit a rhythmic excretion pattern changing over the week, with a constant salt intake (22, 23) and the percentage of excreted sodium is different over the seasons, a lower urinary sodium excretion was found in summer probably due to a higher excretion of sodium with sweat (24). However, the correlation between urinary excretion and dietary intake of sodium and potassium is high (24). Furthermore, the deviations in sodium and potassium urinary excretion are small, thus urinary sodium and potassium excretion are considered to be reliable recovery biomarkers (25, 26).

Urinary PABA check
Complete urine collections are required in order to obtain valid protein, potassium and sodium estimates. To check for completeness of 24h urine

collection, para-aminobenzoic acid (PABA), taken as 80 mg tablets three times a day, is often used in nutritional studies. PABA recovery in urine in our study was assessed by the High Performance Liquid Chromatography (HPLC) method (where recoveries of 78% or higher are generally considered complete) (27). The HPLC method is more specific than the previously used spectrophotometric method in which use of drugs like paracetamol and sulphonamides, interfered. To our knowledge no validated adjustment method for PABA recoveries below the cut-off value of the HPLC method have been published. Based on PABA recovery, excluding incomplete urine samples would considerably diminish our sample size. Since we preferred not to reduce our sample size we checked whether excluding urines based on PABA would be necessary in our study by a sensitivity analysis, comparing the model outcomes (chapters 2 and 4) from the complete urine dataset with the model outcomes after exclusion of the urines with <78% PABA recovery. Estimated model parameters did not differ substantially, thus calculations were done based on the complete urine data set. This points in the same direction as the finding of Subar *et al* (28), who observed a modest effect on correction factors when urines were excluded based on PABA recovery compared with not excluding urines in the OPEN study. However, in chapter 5 our main outcome was the difference in level of intake between methods, and, as expected, the results differed significantly when incomplete urines were excluded. Therefore in this chapter, incomplete PABA recovery was included as a covariable in the model.

Concentration biomarkers as reference method
In chapter 3 the concentration biomarker for fatty acids was used as a reference method to assess validity coefficients. The first requirement for a reference method, uncorrelated errors with the method to be validated, is likely to hold since the errors in the biomarker measurement are assumed to be mostly physiological. However, variations in the concentration biomarker measurements are not solely determined by diet, but also depend on an individuals' characteristics like variations in digestion and absorption, metabolism and excretion, thus repeated measurements tend to be correlated (29). Furthermore, concentration biomarkers do not have a quantitative relationship with dietary intake (12, 25). As concentration biomarkers are based on the concept of the existence of a correlation between reported dietary intake and intake measured from the biological sample, they can at the most be used to rank participants according to their dietary intake. The magnitude of correlation is informative and enables comparison between the performances of different dietary assessment methods.

24hR as reference method

Research showed that error correlations between 24hR and FFQ exist for energy, protein and potassium (9, 11, 30, 31) and that the 24hR does not generate unbiased estimates (both group level bias and proportional scaling bias) of energy, protein, potassium and sodium intake (9, 11, 31-33). These finding were confirmed by this research as presented in chapters 2 and 4. Replicates of the 24hRs were taken apart in time (4 months for the telephone-based 24hR and 3 months for the web-based 24hR), but correlated errors between replicates cannot be ruled out.

In this research both web-based and telephone-based 24hR were used (chapters 2 and 4). The above mentioned methodological considerations are assumed to hold for both modes of administration. However, due to the different ways of administering, different error sources could influence the intake estimates obtained, and influence the magnitude of validity for both 24hRs differently. Our results indicated that the validity (on the group level and for attenuation factors) of the telephone-based 24hR was slightly better than that of the web-based 24hR, which is in line with findings from a large pooling project in the USA (32, 33). Validation and improvements of our web-based 24hR are currently ongoing.

The Duplicate portion method

Central in this research was the question whether the DP would be a good reference method for validation studies. Both chapters (2 and 3) addressing this question concluded the DP is a promising reference method for validation studies. Generally, the DP is not often used in nutritional research because the collection of DPs is burdensome for participants, logistics are extensive, it requires a lot of time from the researcher and the chemical analysis of the nutrient content of the DPs in the laboratory is time consuming resulting in high costs. Collections of DPs may lead to reactivity bias demonstrating a change in the respondents' intake on the collection day; thus a deviation from their habitual intake may be expected. We carefully instructed our participants not to deviate from their habitual intake and provided them with verbal and written instructions, including tips to remind the participant to include everything in the collection baskets and to prepare enough food including the extra portion for the collection of the DP. No specific reminders about salt added at the table were included, which might have influenced our sodium outcomes. Nevertheless, the DP showed substantial underestimation for energy, protein, potassium and sodium (Chapters 2, 3 and 5) and this is supported by results from other studies (34-37). To our knowledge no research about the magnitude of proportional scaling bias of the DP and correlated errors between DP and FFQ has been performed. In chapters 2 and 4 it was shown that proportional scaling

bias was present for energy, protein, potassium and sodium. Furthermore we found that correlated errors between DP and FFQ were present (chapter 2). Repeated measures of the DPs were taken apart in time (5 months) however, correlation of errors between replicates are expected.

Generalizability

Generalization should always be done conservatively taking into account different factors involved. This section discusses possibilities to generalize our findings to other populations, nutrients and dietary assessment methods.

Generalizing to other populations

Participants included in this study were a sub-sample of the NQplus study (38). Since it was a subsample, data could be generalized to the whole NQplus study; however it should be taken into account that a selective sample may have been included in this study, with highly motivated and health-conscious participants. Participants individually agreed to enrol in this study and were willing to collect two duplicate portions and 70 of them also joined the DLW study. A high number (over 50%) of the participants enrolled in this study were highly educated, where in the general Dutch population this is only 28% (39), indicating that this sample is not representative of the general Dutch population. Dietary underreporting is shown to be larger in lower educated individuals (40, 41), thus our results might show less underreporting than would be the case for the general Dutch population. Evidence from the EPIC-study on attenuation factors and validity coefficients for protein assessed by FFQ differed considerably between different countries. Similarly, estimates of the different error components varied for the different countries (9). Thus, as expected, patterns of measurement error differ for different countries. Generalizing our results to the general Dutch population or to other countries should be done conservatively.
The study population in this research consisted of adults aged 20-70 years. The assessment of dietary intake for children, adolescents and elderly faces different challenges than for adults (42). Thus the pattern of measurement error for diverse age groups is likely different and generalizing our results to other age groups than adults cannot be granted.

Generalizing to other nutrients and dietary assessment methods

Regarding error correlations, including the magnitude of proportional scaling bias and attenuation factors, the DP performed slightly better than the 24hR as a reference method for protein, potassium and sodium (chapter 2). These findings seem supported by our findings of fatty acid intake (chapter 3). As the findings

are consistent for nutrients with a recovery biomarker and nutrients with a concentration biomarker, it seems reasonable to extend our finding to other nutrients. Thus, this implies that for validation purposes in epidemiological research the DP seems the best reference method, for nutrients that have no generally accepted recovery biomarker.

Contrary to what we expected, correlated errors between DP and FFQ were of similar magnitude for protein, and less for potassium and sodium than between 24hR and FFQ (chapter 2). The DP is a prospective method, nutrients can be directly analysed in the lab and portion sizes are the same as those eaten whereas the 24hR is a retrospective method, depends on valid data from food composition databases and portion sizes are estimated (43). Correlated errors with the FFQ thus seem a more universal problem than we anticipated when starting this research.

Overall, using nutrient densities rather than absolute intakes of protein, potassium and sodium did not improve the validity of the dietary intake data from DP, FFQ and 24hR (chapter 4). These results were not in line with data from a large American Pooling project (32, 33). The sample size of our study was small (data from 69 participants were included), the methods were not exactly the same, and also the dietary pattern of our population differed from that in the Pooling project. Therefore, generalizing our findings to other methods and nutrients should be done cautiously.

In chapter 5, body mass index (BMI) was found the most consistent determinant associated with misreporting (group level bias) of energy, protein and potassium assessed by three conceptually different dietary assessment methods (FFQ, 24hR and DP). This is supported by findings from other research groups (20, 40, 44-46). Thus it seems reasonable to extend this finding to other nutrients and dietary assessment methods for group level bias.

Implications

As described in chapter 2 correlated errors between DP and FFQ were present. This was contrary to our expectations and thus the hypothesis that correlated errors mainly originate from memory problems, wrong estimations of portion sizes or incorrect data from food composition databases should be rejected.

Furthermore, this research showed that validity of nutrient densities is not necessarily better than absolute intakes of protein, potassium and sodium (chapter 4). Thus using nutrient densities for the reason of diminishing the

influence of measurement error on the dietary outcomes does not seem favourable.

Lastly, this research showed that BMI is an important determinant associated with misreporting of energy and different nutrients for DP, FFQ and 24hR (chapter 5). This is supported by evidence from literature (20, 40, 44-48) and suggests that the influence of BMI needs to be taken into account if the outcome of interest is group level bias.

Future research

We started this research with the hypothesis that error correlations between DP and FFQ would be non-existent but definitely less than those between 24hR and FFQ. The finding that correlated errors are notably present between DP and FFQ raises the question where these correlated errors come from. Is it maybe due to person characteristics that these error correlations exist and would correction for these determinants diminish error correlations? This would be an interesting continuation of our research.

Although statistical methods to correct for measurement errors have been improved in the last decades, these methods require the researcher to make assumptions which in practise do not hold. Searching for new or improved statistical methods, taking into account the limitations of the current dietary assessment methods, could facilitate a better correction for measurement errors. An example of this is the development of a varying-time usual intake model as mentioned before (6, 7). As the use of a varying-time usual intake model showed a better fit of dietary intake data in a large Pooling project with comparable objectives as our study (6, 7), exploring the influence of using such a model for our data would be worthwhile.

Research about the validity of the DP as an alternative reference method should continue for different nutrients but also for food groups. Nowadays, in nutrition research assessing dietary patterns using e.g. pre-existing diet quality indices (49, 50) is common practice. For this purpose also the correct assessment of foods and food groups is of importance. Research should focus on diminishing and correcting of the measurement error in food and food group intake assessment. Concentration biomarkers for the intake of e.g. fruit and vegetables (51, 52) and whole grain wheat and rye intake (53) exist, although thus far no golden standard for measuring food intake for free living individuals is known. The DP might be an alternative method to assess the intake of different food groups for e.g. dairy products, with C15 and C17 as a proxy for intake (54) and fish intake, using the n-3 fatty acids DHA and EPA as a proxy for intake (55). As

fish intake is often episodically, this will also add challenges for estimating usual fish intake.

This thesis focussed on correcting for measurement error using a reference method. Another approach is to improve the dietary assessment method as such. Promising results have been shown for optimising the food list of an FFQ for the specific research population and research question by means of linear programming (56). Measurement errors in the dietary intake estimate could also be diminished by combining different dietary assessment instruments (57). FFQ data can add information about the frequency of foods consumed and enhance usual intake estimates from 24hRs, and especially the assessment of episodically consumed foods could benefit from such an approach. Including images (pictures or video) could aid in the estimation of portion sizes. Furthermore allowing participants to make pictures of their diet could support participants in memorizing what they consumed. However, the validity of the use of images is currently limited (58). Furthermore new technologies to improve dietary assessment, such as mobile phones (which could also be used for images) provide a new way to facilitate dietary assessment (59) but need further exploration, especially for use in large epidemiological studies. In this light it should be kept in mind that similar errors as in the original dietary assessment method are expected, although their magnitude might differ.

In light of the few recovery biomarkers available, the search for more such biomarkers is worthwhile. Research into combining biomarker data with genetic and environmental data (60) could increase the validity of concentration biomarkers as reference methods. Furthermore promising research is ongoing about food metabolomics as novel dietary biomarkers to measure dietary exposure (61). This information could be used to validate self-reported dietary intakes.

This research about correction and estimation of measurement error was performed in the Netherlands and other research on this topic has been performed in the USA (11, 31). Nowadays, much dietary consumption studies are performed in low and middle income countries, where different challenges are faced by researchers: imprecise and missing food composition table data, a lack of reliable dietary intake assessment methods for the population and a poor infrastructure (62). Measurement error patterns might therefore be very different for such countries. Taking into account that low and middle income countries often deal with the double burden of disease, highlighting the importance of valid dietary intake data, there is a clear need for the focus on

measurement error research in these countries. The DP could be useful in this light although different challenges will influence the validity of the DP.

Overall Conclusion

The work described in this thesis showed that with respect to error correlations and attenuation factors the DP performed slightly better than the 24hR as a reference method for validating FFQs in epidemiological research. We also found that the use of nutrient densities does not necessarily improve the validity of the dietary intake estimates from DP, 24hR and FFQ. Moreover, it was shown that BMI is an important determinant of misreporting (i.e. group level bias) of energy, protein and potassium for these three assessment methods and should be taken into account when assessing and correcting dietary intake.

References

1.	National Institutes of Health NCI. Dietary Assessment Primer. (http://dietassessmentprimer.cancer.gov/). (Accessed May 2015).
2.	McCarthy S. Weekly patterns, diet quality and energy balance. *Physiology & Behavior* 2014;134(0):55-9.
3.	Shahar DR, Yerushalmi N, Lubin F, et al. Seasonal variations in dietary intake affect the consistency of dietary assessment. *European Journal of Epidemiology* 2001;17(2):129-33.
4.	Basiotis PP, Welsh SO, Cronin FJ, et al. Number of days of food intake records required to estimate individual and group nutrient intakes with defined confidence. *Journal of Nutrition* 1987;117(9):1638-41.
5.	Illner AK, Nothlings U, Wagner K, et al. The assessment of individual usual food intake in large-scale prospective studies. *Annals of nutrition & metabolism* 2010;56(2):99-105.
6.	Freedman LS, Midthune D, Carroll RJ, et al. Application of a New Statistical Model for Measurement Error to the Evaluation of Dietary Self-report Instruments. *Epidemiology (Cambridge, Mass)* 2015.
7.	Freedman LS, Midthune D, Dodd KW, et al. A statistical model for measurement error that incorporates variation over time in the target measure, with application to nutritional epidemiology. *Statistics in Medicine* 2015.
8.	Keogh RH, White IR, Rodwell SA. Using surrogate biomarkers to improve measurement error models in nutritional epidemiology. *Statistics in medicine* 2013;32(22):3838-61.
9.	Kipnis V, Midthune D, Freedman L, et al. Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public health nutrition* 2002;5(6A):915-24.
10.	Kaaks R, Riboli E, Esteve J, et al. Estimating the Accuracy of Dietary Questionnaire Assessments - Validation in Terms of Structural Equation Models. *Statistics in Medicine* 1994;13(2):127-42.
11.	Kipnis V, Subar AF, Midthune D, et al. Structure of dietary measurement error: Results of the OPEN biomarker study. *American Journal of Epidemiology* 2003;158(1):14-21.
12.	Kaaks R, Ferrari P, Ciampi A, et al. Part H. Advances in the statistical evaluations and interpretation of dietary data: Uses and limitations of statistical accounting for random error correlations, in the validation of dietary questionnaire assessments. *Public health nutrition* 2002;5(6 A):969-76.
13.	Wong MY, Day NE, Bashir SA, et al. Measurement error in epidemiology: The design of validation studies I: Univariate situation. *Statistics in medicine* 1999;18(21):2815-29.
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. *American journal of Clinical Nutrition* 1985;42(6):1276-89.
15.	Kipnis V, Midthune D, Freedman LS, et al. Empirical evidence of correlated biases in dietary assessment instruments and its implications. *American Journal of Epidemiology* 2001;153(4):394-403.
16.	Bingham SA. Urine nitrogen as a biomarker for the validation of dietary protein intake. *Journal of Nutrition* 2003;133(3 SUPPL.):921S-4S.
17.	Schoeller DA. Measurement of energy expenditure in free-living humans by using doubly labeled water. *Journal of Nutrition* 1988;118(11):1278-89.
18.	Schoeller DA, Hnilicka JM. Reliability of the doubly labeled water method for the measurement of total daily energy expenditure in free-living subjects. *Journal of Nutrition* 1996;126(1):348s-54s.
19.	Schoeller DA, Schoeller DA. Validation of habitual energy intake. *Public Health Nutrition* 2002;5(6a):883-8.

20. Freisling H, Van Bakel MME, Biessy C, et al. Dietary reporting errors on 24 h recalls and dietary questionnaires are associated with BMI across six European countries as evaluated with recovery biomarkers for protein and potassium intake. *British journal of nutrition* 2012;107(6):910-20.

21. Turban S, Thompson CB, Parekh RS, et al. Effects of sodium intake and diet on racial differences in urinary potassium excretion: Results from the dietary approaches to stop hypertension (DASH)-sodium trial. *American Journal of Kidney Diseases* 2013;61(1):88-95.

22. Ortiz-Melo D, Coffman TM. A trip to inner space: insights into salt balance from cosmonauts. *Cell metabolism* 2013;17(1):1-2.

23. Rakova N, Juttner K, Dahlmann A, et al. Long-term space flight simulation reveals infradian rhythmicity in human Na(+) balance. *Cell metabolism* 2013;17(1):125-31.

24. Holbrook JT, Patterson KY, Bodner JE. Sodium and potassium intake and balance in adults consuming self-selected diets. *American Journal of Clinical Nutrition* 1984;40(4):786-93.

25. Willett W. *Nutritional epidemiology*. Oxford: Oxford University Press; 2013.

26. 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. *Journal of Nutrition* 2006;136(5):1334-40.

27. Jakobsen J, Ovesen L, Fagt S, et al. Para-aminobenzoic acid used as a marker for completeness of 24 hour urine: Assessment of control limits for a specific HPLC method. *European Journal of Clinical Nutrition* 1997;51(8):514-9.

28. Subar AF, Midthune D, Tasevska N, et al. Checking for completeness of 24-h urine collection using para-amino benzoic acid not necessary in the Observing Protein and Energy Nutrition study. *European journal of clinical nutrition* 2013;67(8):863-7.

29. Kaaks R. Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *American journal of Clinical Nutrition* 1997;65(4):1232S-9S.

30. Ferrari P, Roddam A, Fahey MT, et al. A bivariate measurement error model for nitrogen and potassium intakes to evaluate the performance of regression calibration in the European Prospective Investigation into Cancer and Nutrition study. *European Journal of Clinical Nutrition* 2009;63(SUPPL. 4):S179-S87.

31. Prentice RL, Mossavar-Rahmani Y, Huang Y, et al. Evaluation and comparison of food records, recalls, and frequencies for energy and protein assessment by using recovery biomarkers. *American Journal of Epidemiology* 2011;174(5):591-603.

32. Freedman LS, Commins JM, Moler JE, et al. Pooled results from 5 validation studies of dietary self-report instruments using recovery biomarkers for energy and protein intake. *American Journal of Epidemiology* 2014;180(2):172-88.

33. Freedman LS, Commins JM, Moler JE, et al. Pooled Results From 5 Validation Studies of Dietary Self-Report Instruments Using Recovery Biomarkers for Potassium and Sodium Intake. *American Journal of Epidemiology* 2015;181(7):473-87.

34. Isaksson B. A critical evaluation of the duplicate-portion technique in dietary surveys. *European journal of clinical nutrition* 1993;47(7):457-60.

35. Johansson G, Åkesson A, Berglund M, et al. Validation with biological markers for food intake of a dietary assessment method used by Swedish women with three different with dietary preferences. *Public health nutrition* 1998;1(03):199-206.

36. Kim WW, Mertz W, Judd JT, et al. Effect of making duplicate food collections on nutrient intakes calculated from diet records. *American journal of Clinical Nutrition* 1984;40(6):1333-7.

37.   Stockley L. Changes in habitual food intake during weighed inventory surveys and duplication diet collections. A short review. *Ecology of Food and Nutrition* 1985;17(3):263-9.

38.   Wageningen University DoHN. NQplus study. 2011. (http://www.eetmeetweet.nl/nqplus/1214/5/0/60). (Accessed 15-9-2015).

39.   statistiek CBvd. Onderwijsniveau bevolking gestegen. 2013. (http://www.cbs.nl/nl-NL/menu/themas/onderwijs/publicaties/artikelen/archief/2013/2013-3905-wm.htm). (Accessed September 2015).

40.   Briefel RR, Sempos CT, McDowell MA, et al. Dietary methods research in the third National Health and Nutrition Examination Survey: Underreporting of energy intake. *American Journal of Clinical Nutrition* 1997;65(4 SUPPL.):1203S-9S.

41.   Klesges RC, Eck LH, Ray JW. Who underreports dietary intake in a dietary recall? Evidence from the Second National Health and Nutrition Examination Survey. *Journal of consulting and clinical psychology* 1995;63(3):438.

42.   Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. *British Journal of Nutrition* 2004;92 Suppl 2:S213-22.

43.   Garrow JS, James WPT, Ralph A. *Human nutrition and dietetics*. Edinburgh: Churchill Livingstone; 2000.

44.   Price GM, Paul AA, Cole TJ, et al. Characteristics of the low-energy reporters in a longitudinal national dietary survey. *British Journal of Nutrition* 1997;77(6):833-51.

45.   Tooze JA, Subar AF, Thompson FE, et al. Psychosocial predictors of energy underreporting in a large doubly labeled water study. *American Journal of Clinical Nutrition* 2004;79(5):795-804.

46.   Voss S, Kroke A, Klipstein-Grobusch K, et al. Is macronutrient composition of dietary intake data affected by underreporting? Results from the EPIC-Potsdam study. *European Journal of Clinical Nutrition* 1998;52(2):119-26.

47.   Subar AF, Kipnis V, Troiano RP, et al. Using intake biomarkers to evaluate the extent of dietary misreporting in a large sample of adults: The OPEN study. *American Journal of Epidemiology* 2003;158(1):1-13.

48.   Mendez MA. Invited commentary: Dietary misreporting as a potential source of bias in diet-disease associations: future directions in nutritional epidemiology research. *American journal of epidemiology* 2015;181(4):234-6.

49.   Guenther PM, Reedy J, Krebs-Smith SM. Development of the Healthy Eating Index-2005. *Journal of American Dietetic Association* 2008;108(11):1896-901.

50.   van Lee L, Geelen A, van Huysduynen EJ, et al. The Dutch Healthy Diet index (DHD-index): an instrument to measure adherence to the Dutch Guidelines for a Healthy Diet. *Nutrition Journal* 2012;11:49.

51.   McGrath AJ, Hamill LL, Cardwell CR, et al. Combining vitamin C and carotenoid biomarkers better predicts fruit and vegetable intake than individual biomarkers in dietary intervention studies. *European Journal of Nutrition* 2015.

52.   Van Kappel AL, Steghens JP, Zeleniuch-Jacquotte A, et al. Serum carotenoids as biomarkers of fruit and vegetable consumption in the New York Women's Health study. *Public Health Nutrition* 2001;4(3):829-35.

53.   Ross AB. Present status and perspectives on the use of alkylresorcinols as biomarkers of wholegrain wheat and rye intake. *Journal of Nutrition and Metabolism* 2012;2012.

54.   Brevik A, Veierød MB, Drevon CA, et al. Evaluation of the odd fatty acids 15:0 and 17:0 in serum and adipose tissue as markers of intake of milk and dairy fat. *European Journal of Clinical Nutrition* 2005;59(12):1417-22.

55. Zock PL, Mensink RP, Harryvan J, et al. Fatty acids in serum cholesteryl esters as quantitative biomarkers of dietary intake in humans. *American Journal of Epidemiology* 1997;145(12):1114-22.

56. Gerdessen JC, Souverein OW, van 't Veer P, et al. Optimising the selection of food items for FFQs using Mixed Integer Linear Programming. *Public Health Nutr* 2014:1-7.

57. Carroll RJ, Midthune D, Subar AF, et al. Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *American Journal of Epidemiology* 2012;175(4):340-7.

58. Gemming L, Utter J, Ni Mhurchu C. Image-assisted dietary assessment: a systematic review of the evidence. *Journal of the Academy of Nutrition and Dietetics* 2015;115(1):64-77.

59. Sharp DB, Allman-Farinelli M. Feasibility and validity of mobile phones to assess dietary intake. *Nutrition* 2014;30(11-12):1257-66.

60. Puiggròs F, Solà R, Bladé C, et al. Nutritional biomarkers and foodomic methodologies for qualitative and quantitative analysis of bioactive ingredients in dietary intervention studies. *Journal of Chromatography A* 2011;1218(42):7399-414.

61. Scalbert A, Brennan L, Manach C, et al. The food metabolome: A window over dietary exposure. *American Journal of Clinical Nutrition* 2014;99(6):1286-308.

62. Pisa PT, Landais E, Margetts B, et al. Inventory on the dietary assessment tools available and needed in Africa: a prerequisite for setting up a common methodological research infrastructure for nutritional surveillance, research and prevention of diet-related non-communicable diseases. *Critical Reviews in Food Science and Nutrition* 2014:0.

Summary

As Food Frequency Questionnaires (FFQs) are subject to measurement error, associations between self-reported intake by FFQ and outcome measures should be corrected for measurement error using data from a reference method. Whether the correction is adequate depends on the characteristics of the reference method used in the validation study. The duplicate portion method (DP) seemed a promising reference method as correlated errors between FFQ and DP, such as memory bias, errors in portion size estimations and food composition databases, are not expected. The influence of measurement errors on dietary intake measures is also suggested to be less when nutrient densities are used. Furthermore, identifying determinants associated with misreporting of dietary intake may help to facilitate adjustments of dietary assessment methods or the development of correction methods for measurement errors. The research described in this thesis addresses the above mentioned issues in the different chapters where data from the DuPLO-study, a subsample of the NQPlus-study, are used. In the DuPLO-study, a Dutch validation study, two DPs, two FFQs, two urinary biomarkers and between one and fifteen 24-hour Recalls (24hR, web-based and/or telephone-based) were collected in 198 subjects, within 1.5 years.

The research started with assessing the performance of the DP as compared to the 24hR as reference method for FFQ validation for nutrients with a known recovery biomarker namely: protein, potassium and sodium (**chapter 2**). Preferably a reference method should 1) not be affected by proportional scaling bias and 2) have uncorrelated errors with those in the FFQ. Multivariate measurement error models were used to estimate proportional scaling bias. Also error correlations between FFQ and DP or 24hR and attenuation factors for DP and 24hR were calculated from the model estimates. We observed that the DP proved to be less affected by proportional scaling bias, had lower correlated errors with the FFQ and showed higher attenuation factors than the 24hR for protein, potassium and sodium. We therefore concluded that the DP is probably the best available reference method for FFQ validation for nutrients that currently have no generally accepted recovery biomarker.

Following these findings, the DP was used as a reference method for validation of fatty acids assessed by the FFQ (**chapter 3**). For fatty acids no unbiased reference method is currently known. Therefore, we determined intakes of fatty acids by chemical analysis of two DPs and two blood plasma samples in the DuPLO-study. Both the DP and 24hR, a commonly used reference method for FFQ validation, were used as reference methods for validation of the intake of saturated fatty acids, monounsaturated fatty acids, n-3 fatty acids, linoleic acid (LA) and the n-3/LA ratio estimated by the FFQ. Biomarker data could only be

used to objectively compare ranking of individuals based on fatty acid intakes from DP and 24hR for n-3 fatty acids, LA and the n-3/LA acid ratio. Multivariate measurement error models were used to estimate validity coefficients and attenuation factors. The DP appeared a better reference method than the 24hR for the assessment of validity coefficients. Attenuation factors for the FFQ, using the DP and 24hR as reference methods, tended to be similar, but for mono-unsaturated fatty acids slightly higher (0.34 vs. 0.21). Furthermore, when using plasma fatty acids as reference, the DP showed comparable to slightly better ranking of participants according to their fatty acid intake. Altogether, the use of the 24hR as reference method gives slightly different results compared to the DP. The DP seems a promising reference method for FFQ validation of fatty acid intake.

In **chapter 4**, a comparison was made between the validity of the absolute intakes of protein, potassium and sodium and their densities for the DP, 24hR and FFQ. In a subsample of 69 subjects of the DuPLO-study an additional one or two doubly labelled water measurements were obtained for assessment of total energy expenditure. Recovery biomarker measurements were used as the reference method for each method. Multivariate measurement error models were used to estimate validity coefficients and attenuation factors. Group level bias was less for protein and sodium densities for all assessment methods as compared to the respective absolute intakes, but not for potassium. However, for all dietary assessment methods and nutrients considered, the validity coefficients and attenuation factors for the nutrient densities compared to absolute intakes did not improve; except for the attenuation factor for sodium density (0.71) of the FFQ which was better than that of the absolute sodium intake (0.51). Thus, we concluded that using nutrient densities rather than absolute intakes does not necessarily improve the performance of the DP, FFQ or 24hR.

In **chapter 5** we studied associations between determinants and misreporting (group level bias) for the basic determinants, body mass index (BMI), gender, age and level of education, which are often included in validation studies and epidemiological models in general. Additionally, the association between BMI-related and other determinants, and misreporting was explored. Misreporting of energy, protein and potassium intake by DP, FFQ and 24hR was assessed for 197 subjects of the DuPLO-study. The recovery biomarkers of the respective nutrients were used as reference method to assess the degree of misreporting. We assessed the association between the extent of misreporting by DP, 24hR and FFQ with the determinants using linear regression analysis. Higher BMI was associated with underreporting of dietary intake assessed by the different

dietary assessment methods for energy, protein and potassium, except for potassium by the DP. Men tended to underreport protein by the DP, FFQ and 24hR and persons of older age underreported potassium but only by the 24hR and FFQ. If corrected for the basic determinants, the BMI-related and other determinants did not show a consistent association with misreporting of energy or nutrients by the different dietary assessment methods. As BMI was the most consistent determinant associated with misreporting, we conclude that BMI should always be taken into account when assessing and correcting dietary intake.

The last chapter (**chapter 6**) describes the most important findings of this research and places them in a larger context. Methodological aspects related to the assessment of usual intake, the influence of timing of collections on usual intake estimates, the assumptions made in the different measurement error models as used in this thesis, and the application of the DP method are discussed. Furthermore, the possibilities of generalizing these results to other populations and nutrients and dietary assessment methods are described. The chapter ends with suggestions for future research.

From this research it can be concluded that, with respect to error correlations and attenuation factors, the DP performed slightly better than the 24hR as a reference method for validating FFQs in epidemiological research. We also found that the use of nutrient densities does not necessarily improve the validity of the dietary intake estimates from DP, 24hR and FFQ. Moreover, it was shown that BMI is an important determinant of misreporting (i.e. group level bias) of energy, protein and potassium for these three assessment methods and should be taken into account when assessing and correcting dietary intake.

Acknowledgements

Nu ik dit schrijf is mijn PhD reis bijna ten einde, maar niet zonder even stil te staan en over mijn schouder terug te kijken en alle mensen die (delen van) deze reis hebben meegelopen te bedanken!

First and foremost I want to thank my husband, **Martin**, you always know how to make me laugh and life is just more beautiful with you at my side. Nïwega Müno for being my shoulder to lean on, my voice of reason, the one to cover my back and my partner in crime.

Zonder mijn zes begeleiders had deze reis nooit geresulteerd in dit mooie eindproduct. Het was een uitdaging om met zo veel disciplines in 1 project de gouden middenweg te vinden en hierin ook mijn eigen stukje in te brengen. Ik heb niet alleen inhoudelijk maar ook persoonlijk enorm veel van geleerd, bedankt hiervoor! Ik wil mijn promotor **Pieter** bedanken voor de inspirerende brainstormsessies, mijn co-promotoren; **Jeanne** voor je luisterend oor als mijn hoofd overliep, ik liep altijd met een geordend hoofd je deur uit en **Anouk**, voor je snelle en to the point feedback, dank jullie wel! **Hendriek** ik ben je enorm dankbaar voor je hulp in de statistische jungle, alle problemen verdwenen als sneeuw voor de zon na een overleg. **Peter** dank je wel voor je aanwijzingen in het Engels en de structuur voor mijn artikelen en **Paul** dank voor je raad en daad voor de laboratorium kant van het verhaal.

Edith je keek van de zijlijn als co-auteur mee en gaf me dat extra steuntje in de rug waar nodig, dank je wel hiervoor! **Gertjan** dank je wel voor alle input voor de dubbel gemerkt water studie. **Fré** dank je wel voor je luisterend oor op de momenten dat het echt nodig was.

Mira, **Apostolos**, **Marije** en **Cecilia**, thank you for doing your thesis in my project, you might not realize it but I learned a lot from you all and I am grateful for all the (practical) work you helped me with.

Collega's **Linde**, **Eveline**, **Rianne**, **Elise**, **Janne**, **Sandra**, **Geerke**, **Maaike**, **Yfke**, **Moniek**, **Marije**, **Rieneke** en **Pim** dank jullie wel voor de gezellige koffie, lunch en thee pauzes of gewoon de praatjes tussendoor. De PhD tour commissie 2013, **Annemarie**, **Agnes**, **Anouk**, **Swetlana**, **Nikkie**, **Lieke** en **Inge**, dank jullie voor de gezellige vergaderingen voor en na onze reis down-under! En **Agnes**, ik voel me vereerd dat je mijn paranimf bent.

George, thank you for walking this path together and to my African sisters **Akwilina**, **Wanjiku**, **Phyllis** and **Sophie** thank you for the laughter and stories! **Anna Carla** I often think back about our conversations with a smile on my face.

Meiden van de middelbare school **Esther**, **Rosanne**, **Elena**, **Renée**, **Linda** onze spontane afspraken waren altijd hilarisch, ontspannen en door het geluidsniveau, misschien enigszins angstaanjagend voor onze partners. **Loes**, **Kim**, **Carla**, **Marianne** en **Susan** dank jullie wel voor de gezellige verjaardagen en vakantie in Oostenrijk! **Inge**, gezellig om onder het genot van een kopje thee bij te kletsen. **Jitske**, een wandelingetje en een drankje na, de perfecte setting om bij te kletsen. **Daphne**, in de zon de ups en downs van onze levens bespreken, ik hoop dat we dat nog vaak mogen doen! **Heleen**, via Wageningen naar Den Haag en weer terug in Wageningen, bijzonder en gezellig om je tijdens deze rondreis telkens weer tegen te komen.

Als laatste wil ik mijn familie bedanken; lieve **pap** en **mam** dank jullie wel voor de "retraite" weekenden in de Achterhoek waar ik alle PhD drukte even kon vergeten, **Jean – Paul** en **Jessica**, mijn PhD heeft met twee keer down under gebracht en jullie deur stond altijd voor me open, dank jullie wel! **Aebele** dank je wel voor het mooie omslag, het was een feest er samen aan te werken en **Anne** dank, voor je tot he point aanwijzingen tijdens dit proces. **Roseanne** het is altijd een feest om aan te komen en jouw enthousiaste knuffels te krijgen. **Opa** en **oma** ik ben zo blij met jullie als mijn grootouders! And last be not least I want to thank my **family in law** for making me feel at home in Kenya.

About the Author

Curriculum Vitae

Laura Trijsburg was born on March 23, 1984 in Laren, the Netherlands. After completing secondary school at Staring College in Lochem and a gap year at the Vrije Hogeschool in Driebergen she started her study Nutrition and Health at Wageningen University. Her minor thesis took place at the department of toxicology. Her major thesis was at the division of Human Nutrition in the field of international nutrition. For her internship she developed a tool for Doctors without borders Holland about hunger as a weapon of war. Shortly after obtaining her MSc degree in 2009, Laura worked as Product Specialist R&D at VitOrtho BV. Then she volunteered in a research project in Kenya on iron supplementation in pregnant women in a malaria endemic area and taught at Maseno University, Kenya. In 2011 she started her PhD project of which the results are described in this thesis. During her PhD she attended several conferences and courses and she was involved in teaching and supervising BSc and MSc students. She was a member of the organizing committee of the PhD study tour to Australia in 2013 and in 2015 she was selected for the European Nutrition Leadership Programme (ENLP). Currently, Laura works on a systematic literature review about adolescent nutrition.

List of Publications

Peer reviewed publications

1. **Trijsburg L**, de Vries JH, Boshuizen HC, Hulshof PJM, Hollman PC, van 't Veer P, Geelen A. Comparison of duplicate portion and 24 h recall as reference methods for validating a FFQ using urinary markers as the estimate of true intake. *Br J Nutr.* 2015;114(8):1304-12.
2. Mwangi MN, Roth JM, Smit MR, **Trijsburg L**, Mwangi AM, Demir AY, Wielders JP, Mens PF, Verweij JJ, Cox SE, Pentice AM, Brouwer ID, Savelkoul HF, Andang'o PE, Verhoef H. Effect of daily antenatal iron supplementation on Plasmodium infection in Kenyan women: a randomized clinical trial. *JAMA.* 2015; 314(10):1009-20
3. Siebelink E, de Vries JHM, **Trijsburg L**, Hulshof PJM. Evaluation of calculated energy and macronutrient contents of diets provided in controlled dietary intervention trials by chemical analysis of duplicate portions. *J Food Comp Anal.* 2015;43; 68-74
4. Mwangi MN, Maskey S, Andang o PE, Shinali NK, Roth JM, **Trijsburg L**, Mwangi AM, Zuilhof H, van Lagen B, Savelkoul HF, Demir AY, Verhoef H. Diagnostic utility of zinc protoporphyrin to detect iron deficiency in Kenyan pregnant women. *BMC Med.* 2014;12:229.
5. Rani V, **Trijsburg L**, Brouwer ID, Khetarpaul N. Dietary non-heme iron bioavailability among children (ages 5-8) in a rural, high-anemia-prevalent area in North India: comparison of algorithms. *Ecol Food Nutr.* 2010;49(4):262-78.

Other publications

1. **L. Trijsburg**. To change your coffee drinking habit or not, that's the question, *ENLPress,* Issue 11, September 2015

Abstracts and presentations

1. **Trijsburg L**, de Vries JHM, Boshuizen HC, Hulshof PJM, Hollman PCH, van´t Veer P, Geelen A. Comparison of duplicate portion and 24h recall as reference methods for validating a food frequency questionnaire using urinary markers as the estimate of true intake. 9th International Conference on Diet and Acitivity Methods. Sept 2015. Brisbane, Australia. Oral presentation
2. **Trijsburg L**, Geelen A, Hulshof PJM, van´t Veer P, Boshuizen HC, Hollman PCH, van Dijk G, Feskens EJM, de Vries JHM. Comparison of the validity of absolute intake and nutrient density of protein, potassium and sodium assessed by various dietary assessment methods. 9th International Conference on Diet and Acitivity Methods. Sept 2015. Brisbane, Australia. Poster presentation
3. **Trijsburg L**, Seves M, de Vries JHM, Hollman PCH, Hulshof PJM, van 't Veer P, Geelen A. Validity of protein intake from duplicate portions and 24-h recalls compared to 24-h urinary nitrogen within the duplo study. NWO (Dutch Nutrition Science Days). Deurne, The Netherlands. Oct, 2013. Oral presentation
4. **Trijsburg L**, de Vries JHM, Geelen A, Hollman PCH, Hulshof PJM, van´t Veer P. Correction for measurement errors in an FFQ by using duplicate portions, 24 hour recalls and biomarkers as reference methods. 8th International Conference on Diet and Acitivity Methods. April 2012. Rome, Italy. Poster presentation

Overview of completed training activities

Course	Organizer and place	Year
Discipline specific activities		
Measurement error webinar series	US-NCI, Washington(US)	2011
8th International congress on diet and activity methods (ICDAM 8)	ICDAM, Rome(IT)	2012
Measurement error workshop	ICDAM, Rome(IT)	2012
Exposure assessment in nutrition research	VLAG, Wageningen(NL)	2012
Masterclass longitudinal data analysis	VLAG, Wageningen(NL)	2013
Nutrition Science Days	NWO, Deurne(NL)	2013
Measurement errors symposium	VLAG, Wageningen(NL)	2014
9th International congress on diet and activity methods (ICDAM 9)	ICDAM, Brisbane(AUS)	2015
General courses		
Masterclass analysis in R	VLAG, Wageningen(NL)	2012
Statistics for the Life Sciences	WIAS, Wageningen(NL)	2012
VLAG PhD week	VLAG, Wageningen(NL)	2012
Acklas	ECS, Wageningen(NL)	2012-2013
Masterclass confounding	VLAG, Wageningen(NL)	2014
Scientific Writing	WGS, Wageningen(NL)	2014
Optional Courses		
Preparing research proposal	WUR, Wageningen(NL)	2011
Concepts and methods in epidemiology	HNE, Wageningen(NL)	2011-2012
Methodology club	HNE, Wageningen(NL)	2011-2015
Seminars and presentations	HNE, Wageningen(NL)	2011-2015
EPI research meetings	HNE, Wageningen(NL)	2011-2015
Measurement error club	HNE, Wageningen(NL)	2012-2015
PhD tour 2013 Australia	HNE, Wageningen(NL)	2013
Analytical epidemiology	HNE, Wageningen(NL)	2014

Notes

Notes

Notes