Validity of dietary data in young populations and implications of measurement errors

Dissertation

Zur Erlangung des Grades eines Doktors der Naturwissenschaften

– Dr. rer. nat. –

Claudia Börnhorst

November 2012

BIPS – Institut für Epidemiologie und Präventionsforschung GmbH

Universität Bremen, Fachbereich 03: Mathematik und Informatik

Erstgutachterin:	Prof. Dr. rer. nat. Iris Pigeot-Kübler
Zweitgutachterin:	Dr. Inge Huybrechts

Danksagung

Zu allererst möchte ich meiner Doktormutter Iris Pigeot danken. Trotz deines vollen Terminplans schaffst du es immer, dir für deine Doktoranden Zeit einzuräumen und zu jedem ein persönliches Verhältnis aufzubauen. Daher möchte ich dir nicht nur für die wertvollen inhaltlichen Hilfestellungen danken, sondern insbesondere für die unglaublich nette, persönliche Betreuung.
Weiter möchte ich meiner Zweitbetreuerin Inge Huybrechts danken. Du hattest immer ein offenes Ohr und hast mir trotz der Entfernung eine sehr intensive, freundschaftliche Betreuung gegeben. Ich habe sehr gerne mit dir zusammengearbeitet und hoffe, dass das auch mit der Abgabe dieser Arbeit nicht endet.
Zudem danke ich dem IDEFICS Konsortium für die Bereitstellung der Daten und für die Möglichkeit im Rahmen dieses Projekts zu promovieren. Insbesondere möchte ich mich bei allen Ko-Autoren bedanken, die mit mir an den Publikationen gearbeitet haben.
Natürlich geht auch ein großer Dank an meine Kollegen und Kolleginnen am BIPS, vor allem an die WiMis der Abteilung Biometrie und EDV und die Mitglieder des Doktorandenkolloquiums.
Insbesondere danke ich meinen lieben Eltern Antonius und Rita Börnhorst. Danke, dass ihr mir das Studium ermöglicht, mich immer gefördert und bei jeglichen Entscheidungen unterstützt habt. Zudem möchte ich euch dafür danken, dass ihr nie Druck ausgeübt habt, sondern mich "habt machen lassen" und einfach stolz ward. Weiter danke ich meinen Schwestern Karin, Sandra, Manuela und Marion sowie meinen Freunden für's "Kopf frei kriegen" durch das weitgehende Ausklammern des Themas "Doktorarbeit". Das hat sehr dazu beigetragen, zwischenzeitlich den nötigen Abstand zu gewinnen.
Zum Schluss möchte ich noch Alexander Hesmer dafür danken, dass er mich durchgehend unterstützt, aufgefangen und motiviert hat - selbst in Phasen, in denen ich mich selbst als "unausstehlich" bezeichnet hätte. Danke!

Abstract

A high number of relationships between dietary intakes and health outcomes has been suggested and investigated during the last decades (Ezzati & Riboli, 2012; Vargas & Thompson, 2012; Alinia et al., 2009; Howarth et al., 2005; Kushi, 1992). Diet is of special interest as it is a modifiable risk factor. However, to date little is known with certainty on the complex relations between diet and specific diseases as respective research requires accurate, quantitative information on dietary intakes. Hence, the description of dietary intakes is one of the main tasks of dietary monitoring surveys and epidemiological studies.

Assessment of dietary intakes is challenging due to changes in diet during life as well as due to the day-to-day variation that characterizes dietary intakes in general. In addition, the estimation of long-term consumption frequencies and amounts is difficult for most people as it relies on long-term memory and the capability of correct averaging. Strictly spoken, dietary intakes cannot be measured without error and will presumably never be. Researchers investigating associations between diet and specific diseases or distributions of dietary intakes need to account for various measurement errors to avoid drawing erroneous conclusions. The nature and magnitude of measurement errors in dietary data depend on the study population under investigation as well as on the assessment instrument.

To date, there are only few recommendations available for measuring dietary intakes among children. As young children do not have the cognitive ability to report their dietary intakes themselves, usually parents are asked to proxy-report their child's intakes. This means that additional problems emerge from meals that are not under parents' control like e.g. school meals leading to unintentional misreporting. Little is known about the validity of proxy-reported dietary data, potential determinants of misreporting and additional sources of measurement errors in young populations yet.

Therefore, this thesis aimed to investigate the extent and effects of measurement errors when assessing and modeling dietary data in young children. Special emphasis was put on differential measurement errors resulting from misreporting

where different methods to counteract attenuation or distortion of risk estimates were encountered and evaluated.
In summary, differences in the determinants of misreporting were found for proxy-reported dietary data compared to those previously reported for self-reported data where the problem of misreporting seemed to be even more severe in case of proxy-reports. Misreporting strongly affected effect estimates of associations between diet and overweight/obesity. Results strongly depended on the chosen statistical model where even reversed signs were observed when accounting or not-accounting for reporting errors. These findings suggest that studies on diet-disease associations based on proxy-reported dietary data are problematic as there is still no formal way to handle differential measurement errors caused by misreporting. In the absence of objective validation data, the true effects remain unknown. A large number of associations reported so far in epidemiological studies may be biased due to the application of models relying on the assumption of non-differential measurement errors only.

Main chapters of the thesis:
This thesis is mainly structured into two parts: Chapters 1 to 5 introduce different dietary assessment methods as well as approaches to validate dietary data and to evaluate the extent of misreporting in general. Furthermore, statistical concepts to handle measurement errors in dietary data are outlined in this part. The second part shortly introduces the IDEFICS and HELENA study and describes the results of different validation studies conducted within the framework of these projects as well as two studies on determinants of misreporting in proxy-reports and on the effect of misreporting on associations between diet and overweight/obesity. The thesis concludes with an overall discussion and future perspectives.

Zusammenfassung

In den letzten Jahren wurde eine Vielzahl an möglichen Zusammenhängen zwischen Ernährung und gesundheitlichen Folgen untersucht (Ezzati & Riboli, 2012; Vargas & Thompson, 2012; Alinia et al., 2009; Howarth et al., 2005; Kushi, 1992), da Ernährung als modifizierbarer Risikofaktor von besonderem Interesse ist. Bisher lassen sich allerdings nur wenig sichere Aussagen über die komplexen Zusammenhänge zwischen Ernährung und potenziellen Auswirkungen treffen, da entsprechende Forschung eine präzise, quantitative Erhebung der Nahrungsaufnahme erfordert.

In epidemiologischen Studien sowie in Studien zum Ernährungs-Monitoring besteht daher eine der Herausforderungen in der adäquaten Erfassung der Nahrungsaufnahme der vorliegenden Studienpopulation. Aufgrund von Änderungen in der Ernährung im Laufe des Lebens sowie aufgrund der typischerweise täglichen Variation im Verzehr ist die Erhebung von Ernährungsdaten schwierig. Darüber hinaus ist die Schätzung von Langzeitverzehrhäufigkeiten und -portionsgrößen für die meisten Probanden problematisch, da sie sich zum einen über lange Zeiträume erinnern und zudem in der Lage sein müssen, ihr "durchschnittliches" Essverhalten zu beschreiben und zu quantifizieren. Nahrungsaufnahmen sind, strikt genommen, nicht fehlerfrei messbar und werden es voraussichtlich auch nie sein. Somit müssen Wissenschaftler, die Zusammenhänge zwischen Ernährung und bestimmten Erkrankungen untersuchen, unterschiedliche Messfehler berücksichtigen, um keine falschen Schlussfolgerungen zu ziehen. Dabei hängen sowohl die Eigenschaften als auch das Ausmaß von Messfehlern in Ernährungsdaten von der untersuchten Studienpopulation als auch vom verwendeten Erhebungsinstrument ab. Insbesondere fehlt es bis heute an Empfehlungen zur Erfassung von Ernährungsdaten bei kleinen Kindern. Da kleine Kinder noch nicht über die kognitiven Fähigkeiten verfügen, ihren Verzehr adäquat anzugeben, werden gewöhnlich Proxies, meist die Eltern, gebeten den Verzehr ihrer Kinder zu berichten. Hieraus resultieren zusätzliche Probleme, wie z.B. unbeabsichtige Falschangaben durch Mahlzeiten, die von den Eltern nicht beobachtet werden. Bisher ist wenig über

die Validität von proxy-berichteten Ernährungsdaten, möglichen Determinanten für Falschangaben sowie zusätzlichen Quellen für Messfehler in jungen Studienpopulationen bekannt.
Das Ziel dieser Arbeit bestand daher darin, die Validität von proxy-berichteten Ernährungsdaten sowie mögliche Quellen und Determinanten von Messfehlern zu untersuchen. Ein besonderer Schwerpunkt wurde auf differentielle Messfehler gelegt, die aus Falschangaben resultieren können. Dazu wurden verschiedene Methoden betrachtet, um einer Abschwächung oder Verzerrung von Risikoschätzern als Folge von Messfehlern entgegenzuwirken.
Zusammenfassend wurden Unterschiede zwischen den für Selbstangaben bekannten und den hier in proxy-berichteten Ernährungsdaten beobachteten Determinanten für Falschangaben gefunden. Des Weiteren wirkten sich Falschangaben stark auf die Effektschätzer in Zusammenhangsanalysen zwischen Ernährung und Übergewicht/Adipositas aus. Dabei hingen die Ergebnisse sehr vom gewählten statistischen Modell ab. Sogar umgekehrte Vorzeichen von Effektschätzern wurden beim Vergleich von Modellen mit und ohne Berücksichtigung von Falschangaben beobachtet. Da keine objektiven Validierungsdaten vorlagen, konnte über den wahren Effekt keine Aussage getroffen werden. Insgesamt kann vermutet werden, dass eine große Anzahl an Effektschätzern, die bis heute in epidemiologischen Studien berichtet wurden, verzerrt sind. Die verwendeten statistischen Modelle basieren oft auf der Annahme, dass in Ernährungsdaten nur nichtdifferentieller Messfehler vorliegen. Diese Annahme spiegelt jedoch aufgrund von Falschangaben häufig nicht die tatsächliche Messfehlerstruktur wider.

Aufbau der Arbeit:
Diese Arbeit gliedert sich in zwei Teile: In Kapitel 1 bis 5 werden allgemein verschiedene Ernährungserhebungsmethoden eingeführt sowie Verfahren zur Validierung von Ernährungdaten und zur Bewertung des Ausmaßes an Falschangaben. Zudem werden in diesem Teil statistische Methoden zum Umgang mit Messfehlern in Ernährungsdaten beschrieben. Im zweiten Teil werden kurz die IDEFICS und die HELENA Studie vorgestellt und die Ergebnisse aus verschiedenen im Rahmen dieser Projekte durchgeführten Validierungsstudien zusammengefasst. Weiter werden die Ergebnisse einer Studie zu Determinaten von Falschangaben und einer Studie zum Effekt von Falschangaben auf Zusammenhangsanalysen zwischen Ernährung und Übergewicht/Adipositas präsentiert. Die Arbeit schließt mit einer übergreifenden Diskussion und einem Ausblick auf zukünftige Forschung.

Contents

List of Abbreviations

BMR	Basal metabolic rate
CEHQ	Children Eating Habits Questionnaire
CEHQ-FFQ	Children Eating Habits Questionnaire - Food Frequency Questionnaire
CL	Confidence limit
DHA	Docosahexaenoic acid
DLW	Doubly labeled water
EI	Energy intake
EFCOVAL	European Food Consumption Validation
EPA	Eicosapentaenoic acid
FAO	Food and Agriculture Organization of the United Nations
FFQ	Food frequency questionnaire
HBSC	Healthy Behaviour in School-Aged Children
HELENA	Heathy Lifestyle in Europe by Nutrition in Adolescents
HELENA-DIAT	HELENA - Dietary Assessment Tool
IDAMES	Innovation of Dietary Assessment Methods for Epidemiological Studies and Public Health
IDEFICS	Identification and Prevention of Dietary- and Lifestyle-Induced Health Effects in Children and Infants
MSM	Multiple Source Method
NCI	National Cancer Institute
OPEN	Observing Protein and Energy Nutrition
PAL	Physical activity level
PDA	Personal Digital Assistant
SACINA	Self-Administered Children and Infants Nutrition Assessment
SD	Standard deviation
TEE	Total energy expenditure
UNU	United Nations University
WHO	World Health Organization
YANA-C	Young Adolescents' Nutrition Assessment on Computer
24-HDR	24-hour dietary recall

List of Tables

List of Figures

Chapter 1

Introduction

A high number of relationships between dietary intakes and health outcomes has been suggested and investigated during the last decades (Ezzati & Riboli, 2012; Vargas & Thompson, 2012; Alinia et al., 2009; Howarth et al., 2005; Kushi, 1992). Diet is of special interest as it is a modifiable risk factor. However, little is known with certainty on the complex relations between diet and specific diseases as respective research requires accurate, quantitative information on dietary intakes. Hence, the description of dietary intakes is one of the main tasks of dietary monitoring surveys and epidemiological studies. Assessment of dietary intakes is challenging due to changes in diet during life as well as due to the day-to-day variation that characterizes dietary intakes in general. In addition, the estimation of long-term consumption frequencies and amounts is difficult for most people as it relies on long-term memory and the capability of correct averaging (Willett, 1995). Strictly spoken, dietary intakes cannot be measured without error and will presumably never be.

The nature and magnitude of measurement errors in dietary data depend on the study population under investigation as well as on the assessment instrument. In cohort studies, dietary intakes are typically measured by means of food frequency questionnaires (FFQ), which were shown to be particularly prone to memory errors (Thompson & Subar, 2012). Other commonly applied instruments are food records and 24-hour dietary recalls (24-HDR) which assess the intakes of single days. Here, problems especially emerge from day-to-day and seasonal variations in diet which are not captured when assessing only single days. Consequently, researchers investigating associations between diet and specific diseases or distributions of dietary intakes need to account for various measurement errors to avoid drawing erroneous conclusions (Carroll et al., 1998). Measurement errors strongly impact the ability to detect nutritional factors that affect health (Paeratakul et al., 1998) and can, depending on the type of error, attenuate, exaggerate

or even hide diet-disease associations. They further result in a loss of power such that even true associations may not be detectable.

Furthermore, researchers are typically interested in usual dietary intakes being defined as long-run-average intakes because intakes of single days do typically not lead to a disease with a later onset (Tooze et al., 2010). Serious health outcomes that are caused by single intakes like e.g. allergies or infections are not in the focus of this thesis. Also recommendations for nutrient adequacy are related to usual intakes. As nutrients can be stored in the body, it does not make sense to promote the adherence to intake targets on every single day. However, dietary intakes strongly vary from day to day such that the estimation of population distributions of usual intakes, which is needed e.g. to determine population groups under risk of mal- or overnutrition, poses another challenge in nutritional epidemiology (Dodd et al., 2006). The presence of random within-person variation in dietary data results in a misleading estimate of the percentage of the population at risk as the increased standard deviation broadens the observed intake distribution (see also Figure 3.1 on page 35). This problem applies in particular when assessing dietary data by means of short-term instruments. Next to the flattening of the distribution, the estimated intake distribution may be shifted to the left or right because of misreporting or other systematic errors.

Not accounting for the effects of random and systematic measurement errors when estimating population distributions of intake may result in wrong definitions of population groups under risk and consequently in wrong recommendations of dietary adequacy. Different procedures mainly based on short-term instruments like 24-HDR have been proposed to estimate usual dietary intake distributions (Souverein et al., 2011; Tooze et al., 2006, 2010; Hoffmann et al., 2002; Haubrock et al., 2011; Dodd et al., 2006). They primarily aim to correct for the inflated intra-individual variability due to additional day-to-day variation. Some recent methods additionally consider correlations between the probability of consumption of a certain food and the consumption amount, person-specific effects or include information from FFQs to enhance the precision of the estimates especially in case of non-daily consumed food items (Tooze et al., 2010; Haubrock et al., 2011). However, established methods were developed for adults and cannot correct for differential misreporting. Currently, little is known on whether direct application of 'adult methods' to data obtained from children results in adequate estimates. Doubt is justified as dietary intakes heavily vary during childhood, especially in periods of growth.

Two European projects funded within the 6th EU Framework Programme, namely

EFCOVAL[1] and IDAMES[2] (de Boer et al., 2011), recommended a combination of methods as e.g. two non-consecutive 24-HDR interviews together with an FFQ to measure dietary intakes for monitoring nutritional adequacy and food safety among adults. Also for epidemiological cohort studies on diet-related disease risks a combination of methods was recommended (Carroll et al., 2012).
To date, only few recommendations for measuring dietary intakes among young children are available (Andersen et al., 2011). Dietary validation studies in young children, especially those including objective biomarker information, are rare. The main problem lies in the increased complexity of conducting an epidemiological survey in young children which is due to the following reasons:

- Written informed consent cannot be obtained from young children such that parents as well as the children need to be convinced to participate in the study;
- Biological samples are difficult to obtain from young children for medical and ethical reasons, e.g. only small samples of blood can be drawn;
- Physical examinations are more complex, e.g. it is difficult to keep a child still when measuring blood pressure;
- Children may be too young to adequately report the required information.

In dietary assessment, mainly the latter item poses a challenge. As young children do not have the cognitive ability to report their dietary intakes by themselves, usually parents are asked to proxy-report their child's intakes (Livingstone & Robson, 2000). This means that additional problems emerge from meals that are not under parent's control like e.g. school meals leading to unintentional misreporting. Furthermore, the extent to which social desirability may affect the parental reporting accuracy is unknown. Little is known about the validity of proxy-reported dietary data, potential determinants of misreporting and additional sources of measurement errors in young populations yet. As misreporting may result in differential measurement errors that are difficult to correct and that may even reverse associations (Buzas et al., 2004), careful consideration of this problem would be desirable. Yet, only few methods to account for misreporting when analyzing diet-disease associations were suggested and applied in

[1] European Food Consumption Validation

[2] Innovation of Dietary Assessment Methods for Epidemiological Studies and Public Health

adult populations (Nielsen & Adair, 2007; Huang et al., 2005) where none of these methods is able to correct for differential misreporting. To the author's knowledge, no study to date addressed this topic in young children though the problem of misreporting may be even more pronounced in data relying on proxy-reports. The IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) study, which was funded by the 6th EU Framework Programme, is one of few large European multi-center studies aiming to investigate the causes of diet- and lifestyle-induced health effects in children and infants. Various lifestyle factors based on parental reports but also biomarker measurements were obtained from more then 16,200 children aged 2 to 9 years (Ahrens et al., 2011). Dietary information was assessed using repeated proxy-reported 24-HDRs as well as the so-called Children's Eating Habits Questionnaire (CEHQ) querying general information on eating behaviors, parents' attitudes towards diet and meal frequencies. The CEHQ further includes a food frequency questionnaire part. This study thus provides a valuable database to evaluate the validity of proxy-reported dietary information in young children.
The misreporting behavior of the parents is of special interest here, as unintentional misreporting due to meals out of parental control brings an additional source of error in proxy-reports. Reference information on energy expenditure measured by the doubly labeled water technique, various concentration biomarkers from blood and urine as well as anthropometric measurements enable an in depth analysis of the degree and determinants of misreporting in proxy-reports as well as on the potential effect of misreporting on diet-disease associations. This database further facilitates the exemplary application of different methods to correct for misreporting and an evaluation of their usefulness.

Summing up, there is a lack of information on how to adequately measure and model dietary data in young children accounting for the various measurement errors. Therefore, this thesis aims to investigate the validity and reliability of proxy-reported dietary data in young populations and to identify potential sources, determinants and effects of measurement errors. Special emphasis will be put on differential measurement errors resulting from misreporting. Methods to assess the extent of misreporting and to counteract attenuation of risk estimates caused by measurement errors will be encountered and evaluated based on the data assessed within the framework of the IDEFICS study (Ahrens et al., 2011). Shortly also results concerning the validity of dietary data assessed in adolescents will be presented based on the data of the HELENA study (Moreno et al., 2008). In

addition, concepts to handle broadened distributions due to artificially increased standard deviations when investigating dietary data in young populations will be outlined to provide starting points for future research.

Structure of the thesis

The thesis is structured into nine chapters, which can be subdivided into a theoretical part summarizing basic concepts in nutritional epidemiology (Chapters 2 to 5) and into a practical part describing the results of different validation studies conducted within the framework of the IDEFICS study (Chapters 6 to 8).

The general introduction (Chapter 1) is followed by a short review of the dietary assessment methodology commonly applied in epidemiological studies (Chapter 2). Subsequently, Chapter 3 introduces basic concepts of measurement error theory focussing on the effects of measurement errors in dietary data on measures of associations and on estimates of population intake distributions. In this context, also methods to correct for measurement errors will be outlined. Different options to check dietary data for their reliability and validity as well as how to evaluate the extent of misreporting are presented in Chapters 4 and 5.

The description of the theoretical background is followed by a general introduction to the IDEFICS and HELENA studies (Chapter 6) and a summary of the results of different dietary validation studies conducted in course of these projects (Chapter 7). An analysis of the prevalence and determinants of misreporting is then presented in Chapter 8. The presented results form the basis for the subsequent study on methods to account for misreporting in proxy-reported dietary data where different approaches are illustrated by exemplary application to the IDEFICS data. The studies outlined in Chapters 7 and 8, to which I contributed either as first author or as a co-author, are published or currently under review in epidemiological journals with peer-review. The entire articles are printed in Appendix B on page 119ff. Finally, the thesis concludes with an overall discussion of the results and a section on suggestions for future research and perspectives.

Chapter 2

Dietary assessment

In most instances, the purpose of measuring dietary intake or exposure is to obtain quantitative information on energy and nutrient intakes ingested by the study population under investigation (Rutishauser, 2005). However, dietary intake assessments are performed indirectly in terms of food intakes[3] but not in terms of energy and nutrient intakes, as people do usually not know their energy or nutrient intakes. For example, when assessing an individual's dietary intake of a single day, the data collected consist of several different food items where each has to be specified and quantified. These food intakes need to be linked with food composition data then to estimate nutrient and energy intakes. In the end, a measurement of dietary intake does not consist of one but many single measurements each being prone to measurement errors (Rutishauser, 2005).
This chapter is mainly based on Willett (1995) and Thompson & Subar (2012) and shortly summarizes dietary assessment methods that are frequently applied in epidemiological studies. In general, dietary assessment instruments can be classified into prospective and retrospective assessment instruments. Another common classification differentiates between short-term and long-term instruments. Short-term instruments intend to measure the actual intake on single days or other specified short time periods whereas long-term instruments intend to measure usual intakes, also referred to as long-term average intakes. The latter means that the respondent is asked to estimate his/her average daily intake over a specified period of time ranging in most cases from four weeks up to one year where both, consumption and non-consumption days, are taken into account.

[3] Here and in the following the term 'food' refers to all foods and beverages consumed by the oral route.

2.1 Assessment instruments

In large epidemiological studies, the food frequency questionnaire (FFQ) is usually the method of choice as it is most convenient and inexpensive to use. It is almost always designed for self-completion and was originally collected paper-pencil based, but currently also Personal Digital Assistant (PDA), mobile-phone, interactive computer- and web-based technologies become more common (Illner et al., 2012). In case of paper-pencil based assessments, often the design allows optical scanning to save time on data entries and data checks (Rutishauser, 2005). An FFQ queries retrospectively the consumption frequencies of selected food items from a closed food list ranging from few up to more than 200 items for a defined period of time. The food lists vary depending on the purpose of the study and depending on the study population under investigation. The 'frequency of consumption categories' also vary by questionnaire but usually include frequencies per day, per week, or per month. It is commonly distinguished between quantitative, semi-quantitative and qualitative (non-quantitative) FFQs. Qualitative FFQs only query consumption frequencies whereas quantitative and semi-quantitative FFQs additionally collect information on portion sizes. The respondent is then either asked to estimate his/her usual portion sizes as precisely as possible in grams or liters (quantitative) or based on predefined standard portions described in the questionnaire (semi-quantitative).
FFQs intend to measure long-term intake and allow distinguishing between consumers and non-consumers. The main disadvantage lies in the lack of details obtained about foods and in the closed food list. Furthermore, FFQs are prone to memory errors and errors in estimating usual consumption frequencies or amounts as they rely on the participants capability of correct averaging. In addition, information obtained from FFQs was reported to be affected by recent diet and also systematic bias at the individual level, i.e. under- and overreporting, has been observed. An exemplary extract of an FFQ that was applied in the IDEFICS study (cf. Section 6) is given in Figure 2.1 for illustration.

Commonly used short-term instruments include food records and 24-hour dietary recall (24-HDR) interviews. Such methods are more costly and work-intense for both, study personnel and participants, but were shown to introduce less bias compared to the FFQ (Kipnis et al., 2003). For this reason, replicates of short-term instruments assessed over a reference period are also commonly used to measure long-term intake.

	Never / less than once a week	1 - 3 times a week	4 - 6 times a week	1 time per day	2 times per day	3 times per day	4 or more times per day	I have no idea
Vegetables								
Cooked vegetables, potatoes and beans (also in mixed recipes)	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Legumes (only in Greece)	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Fried potatoes, potato croquettes	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Raw vegetables (mixed salad, carrot, fennel, cucumber, lettuce, tomato etc.)	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Fruits								
Fresh fruits (also freshly squeezed, fruit smoothie) ***without*** added sugar	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Fresh fruits (also as freshly squeezed, fruit smoothie) ***with*** added sugar	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Drinks								
Water	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8
Fruit juices (orange juice, apple juice, *<local examples>* etc.)	❍ 1	❍ 2	❍ 3	❍ 4	❍ 5	❍ 6	❍ 7	❍ 8

Figure 2.1: Extract from the food frequency questionnaire used in the IDEFICS study to query consumption frequencies of selected food items

Food records prospectively assess the food intakes at the time of consumption over a number of days where the assessment days are not necessarily sequential. Here the respondents are usually asked to enter their food intake information on hard copy forms, though tape-recording, bar-coding, or recently smartphone applications have also been used to collect the information. Portion sizes are either estimated by the study participant using household measures or based on portion size estimation aids like photographs of food portions. In case of weighted food records, the respondent is asked to weigh all foods and beverages consumed. Due to the prospective assessment, food records do not rely on memory and are hence less cognitively challenging. However, the intake may be affected by the recording as study subjects may change their diets when knowing to be under study.

The 24-HDR is a retrospective assessment instrument in which the respondent is asked to recall and describe all foods and beverages consumed in the preceding 24 hours or another defined time span which usually ranges from one to seven days (Willett, 1998). The interview may be conducted face-to-face, by telephone, via paper-pencil, computer-assisted or recently also web-based. Portion size estimation aids can assist the respondent to recall the portion sizes consumed (Thompson & Subar, 2012). Such estimation aids may be photos in a picture book or photos shown on the screen in case of computer-based assessments but also household measures, bread shapes or bean bags are typically used. Figure 2.2 presents an example of such pictures with increasing portion sizes obtained from a computer-assisted 24-HDR that was previously used in the IDEFICS study. Details on the dietary assessment methodology applied in IDEFICS are presented in Section 6.1.1.

Because the 24-HDR is open-ended, it is also suitable for large populations of different ethnicity. Furthermore, it provides much detail and puts relatively low burden on the respondents.

The main disadvantages of all short-term instruments are the high workload and costs of coding as well as the need of multiple assessments to be able to estimate long-term intakes. Various factors contribute to the day-to-day variation like the day of the week and the season. Therefore, a single recall or record is an imprecise measure of the true usual intake (Carroll et al., 1998). In particular, a single recall cannot reflect both, consumption and non-consumption days of typically episodically consumed foods like fish or olives. Although food recalls and food records are often assumed as an unbiased estimate for dietary intake, studies involving recovery biomarkers such as doubly labeled water (see also Section 5.1.1) or urinary nitrogen suggest a systematic bias on average towards underreporting

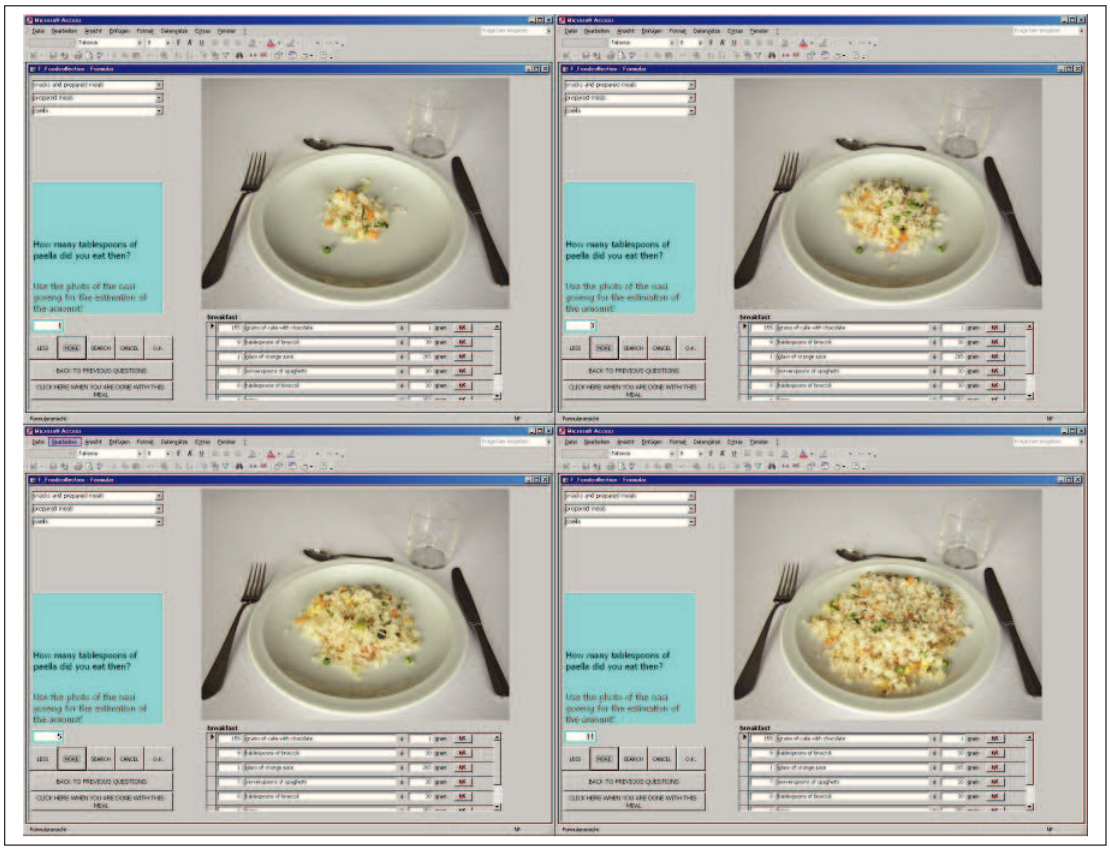


Figure 2.2: Screenshots showing pictures with increasing portion sizes of paella obtained from a computer-assisted 24-HDR that was used in the IDEFICS study

and further that individuals differ in their reporting accuracy (Kipnis et al., 2001).

Besides these methods, dietary intake can be assessed by means of direct observation. This method is either used to objectively assess an individual's intake, which is mainly applied in a clinical setting, or in order to validate another dietary assessment method. It is quite resource intense as it requires trained staff that directly observes and documents the study subject's food intakes and portion sizes. The main limitation of direct observation is 'reactivity' which means that the study subjects are likely to alter their diet if they know that they are under study. Preferably, the participant should be observed in his/her natural environment, but often this method is applied under laboratory conditions which may further increase the likelihood of alteration in diet.
Depending on the chosen assessment instrument, the subsequent conversion of the food intake information into nutrient intake information can be achieved by means of linkage of the reported food items to food composition tables, by laboratory analysis of samples of the foods consumed or by analysis of duplicate diets (Rutishauser, 2005). The latter two options can only be applied in case of prospective dietary assessments.
A more comprehensive description of dietary assessment instruments can be found in Willett (1995) and Thompson & Subar (2012).

2.2 Proxy-reports vs. self-reports

The assessment methods shortly reviewed in Section 2.1 are mostly based on self-completion which means that the study participant reports his/her own intake by him-/herself (referred to as 'self-report' in the following). However, young children lack the cognitive skills to adequately report their own intakes. In the literature, the age of 8 to 12 years is considered as transition period in which children develop the abilities for accurate self-assessments (Burrows et al., 2010). In children younger than eight years, the conceptualization of time, the ability to conceptualize frequencies, the capacity for remembering, the attention span as well as knowledge of names of single food items are not sufficiently pronounced for valid self-reports (Livingstone & Robson, 2000). Therefore, data in young children often rely on proxy-reporters, mainly the parents (referred to as 'proxy-report'). This may introduce additional error because children are not permanently under parental control and respectively parents do not observe all intakes of their children. Additional assessment of meals eaten at school or kindergarten by means

of direct observation through caretakers or teachers is one possibility to reduce this problem but still not all intakes may be captured due to e.g. intakes at other peoples' homes or snacking underhand.
Next to dietary assessments based on self- or proxy-reports, also interviewer-administered assessments are sometimes conducted meaning that an interviewer prompts the respondent to describe his/her dietary intakes and documents the given information. Such interviews can be done in-person, by telephone, paper-pencil based or computer-assisted. However, interviewer-administered assessments are not in the focus of this thesis and will not be further discussed in the following chapters.

2.3 Choice of an assessment instrument

The choice of the most adequate assessment instrument depends on the aim of the study as well as on the target population. Sometimes it may also be desirable to assess data in such a way that these are comparable to former studies or internationally comparable. Besides, there may be limitations in terms of time and costs.
In general, for group mean comparisons it is adequate to use either short-term or long-term instruments (Rutishauser, 2005). Here the choice mainly depends on available resources but also on the required level of precision. Measuring the diet of as many individuals as possible on a single day is usually considered the most efficient approach for group mean comparisons. This approach is also often applied in national surveys that target a high response proportion but also detailed quantitative information (Sempos et al., 1992).
If one is, however, interested in determining a population group 'under risk of dietary inadequacy or excess', a reliable estimate of the usual intake distribution is needed. For this purpose, at least two non-consecutive, repeated short-term measures are needed to be able to statistically adjust for the within-person variation caused by day-to-day variation. In large samples, it may be sufficient to obtain repeated measures in a representative subsample only (Carriquiry, 1999). Recently, the use of two complementary dietary assessment instruments, e.g. 24-HDR and FFQ, was shown to be beneficial to obtain valid estimates on usual intakes (Carroll et al., 2012; Haubrock et al., 2011). It is noteworthy, that statistical techniques that correct for the artificially increased standard deviation caused by day-to-day variation provide improved estimates of the population proportion below and above defined cut-offs, but do not provide information on the actual

individuals 'at risk'. In case the interest lies in the assessment of individual usual diets based on short-term instruments, dietary information over a period of at least one week is required where the number of days differs depending on the nutrient under investigation (Willett, 1995). Another option is the application of sophisticated statistical methods based on repeated 24-HDR and additional FFQ information to obtain estimates of individual usual intakes like e.g. proposed by Kipnis et al. (2009) and Haubrock et al. (2011).

For the estimation of associations between diet and health outcomes, the assessment instrument must be able to rank individuals according to their intake levels. For this purpose, mainly the FFQ was used in former studies but recently short-term instruments, especially repeated 24-HDRs, have been suggested to become the main instrument to adequately model diet-disease associations in future (Kipnis et al., 2009). In case of short-term assessments, the statistical power strongly depends on the number of replicates as well as on the total study size due to the large random errors that go along with short-term assessments (cf. Section 2.1). Therefore, if feasible, recent research suggests the combination of different self-report instruments with objective measures of diet like biomarkers to overcome the limitations of the single instruments and at the same time to optimally exploit their strengths (Illner et al., 2012; Freedman et al., 2010).

Chapter 3

Measurement error in dietary data

"There is not, and presumably never will be, a method that can estimate dietary intake without error."
(Beaton, 1994)

This chapter shortly reviews basic concepts and models of measurement errors, procedures to correct for measurement errors as well as consequences of ignoring measurement errors in the statistical analysis of dietary data. Potential effects of measurement errors on distributions of dietary intake, prevalence estimates and associations between diet and an health outcome are illustrated in Figure 3.1 and will be further described in the following.
This chapter is mainly based on information obtained from Carroll (1998), Buzas

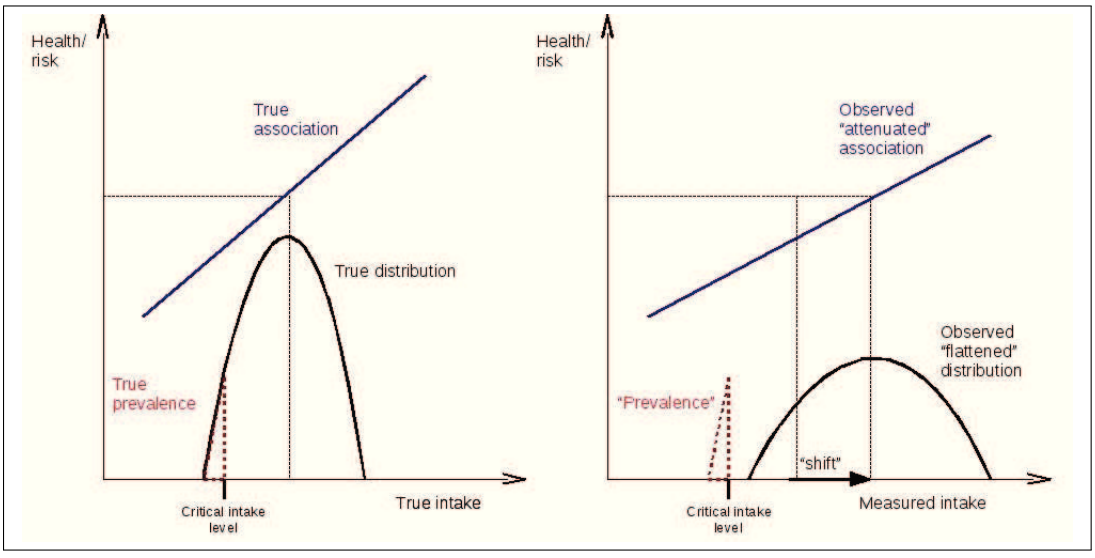


Figure 3.1: Effects of measurement errors on distributions of dietary intake (black), prevalence estimates (red) and associations between diet and health outcomes (blue).

et al. (2004) and Willett (1998) and focuses on the association between an outcome and an exposure in the presence of measurement error in the exposure

variable. In Section 3.3.4, the effects of measurement errors on population distributions of usual dietary intakes and prevalence estimates of dietary inadequacy will be outlined.
In nutritional epidemiology, one is typically interested in the association between the risk of a disease D and the true long-term usual intake T of a given nutrient, where associations are mainly modeled applying (multi-level) logistic or linear regression models. The true intake T is a latent variable, meaning that it cannot be measured directly. Therefore, dietary assessment instruments are used to assess surrogate measures Q for the dietary intake. As described in Section 2.1, such surrogate measures have several limitations resulting in measurement errors. Hence, simple substitution of the true intake T by the measurement Q in the diet-disease model may result in biased estimates and misleading inferences. To account for the measurement errors in statistical models, attention needs to be paid to the nature and type of the error.

3.1 Types of measurement errors and their effects

3.1.1 Random vs. systematic measurement error

There are two types of measurement error, namely random and systematic error (Rutishauser, 2005). Both can occur either within or between persons meaning that they can occur at the individual or at the group level.
Random errors are randomly distributed and can be attributed to chance such that they are unpredictable. The expected value of random errors is zero implying that these errors are scattered around the true value, i.e. estimates of group mean values in the presence of random error only are unbiased. Random errors reduce the precision of dietary intake estimates as they increase the variance. This poses a challenge when investigating diet-disease associations as random errors tend to attenuate correlations and regression coefficients towards the null and relative risk estimates towards the one and may therefore lead to false-negative conclusions (cf. Section 3.3). The effect of random errors can be reduced by increasing the number of observations because the average of many repeated values measured with random error approaches the true value due to the well-known 'law of large numbers'.
Random within-person variation caused by day-to-day variation is the major source of error in dietary data assessed through short-term instruments. The term 'error' may be misleading in this context as even though the intake on a

single study day may have been measured correctly, it is a poor estimate for the study subject's average (usual) intake due to the random day-to-day variation (Beaton, 1994). Here, the effect of random errors on measures of association can be reduced by either increasing the number of study participants or by increasing the number of days collected from each study participant. The degree of random within-person error can be measured by replicate assessments in a sample of subjects, i.e. in a so-called reliability study.

Non-randomly distributed errors in a group of subjects or in data from a given subject result in **systematic errors** where the effect of systematic error cannot be reduced by increasing the number of observations. Three types of systematic within-person errors can be distinguished:

(a) Additive error

In the presence of additive error, the instrument introduces a bias that is equal for all individuals as depicted in Figure 3.2. Additive error shifts regression lines or intake distributions by a constant factor β_0 as all reported intakes differ by β_0 from the true intake. If a regression line is shifted up or down without changing the slope, hypothesis tests as well as the statistical power to detect an association are not affected. However, shifting an intake distribution to the right or left, including the group mean, may result in wrong definitions of population groups under risk as shown in Figure 3.1 (cf. also Section 3.3.4).

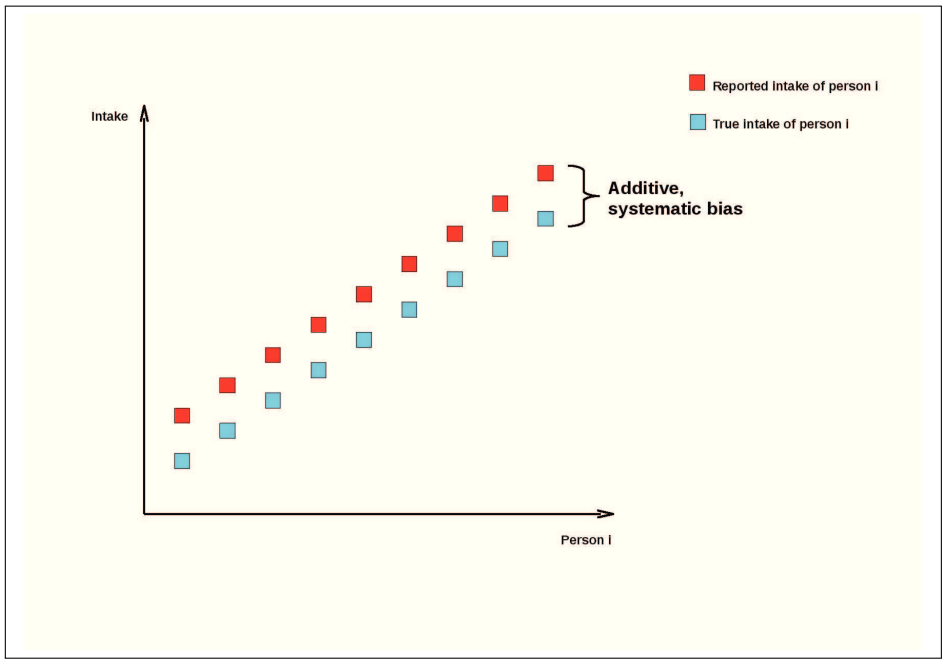


Figure 3.2: Systematic additive error

(b) Intake-related error (multiplicative error; scaling error)
This is a bias that is proportional by a factor β_1 to the true intake of the nutrient or food item as shown in Figure 3.3. In the presence of intake-related bias, odds ratio estimates are scaled by the factor $1/\beta_1$. If the proportional bias β_1 was known, it would be easy to correct for it based on its inverse. Again the scatter around regression line would remain unchanged such that hypothesis tests as well as the statistical power to detect an association were not affected. Unfortunately, β_1 is in general unknown such that the observed standard deviation will not reflect the truth and therefore affects hypothesis tests as well as the statistical power.

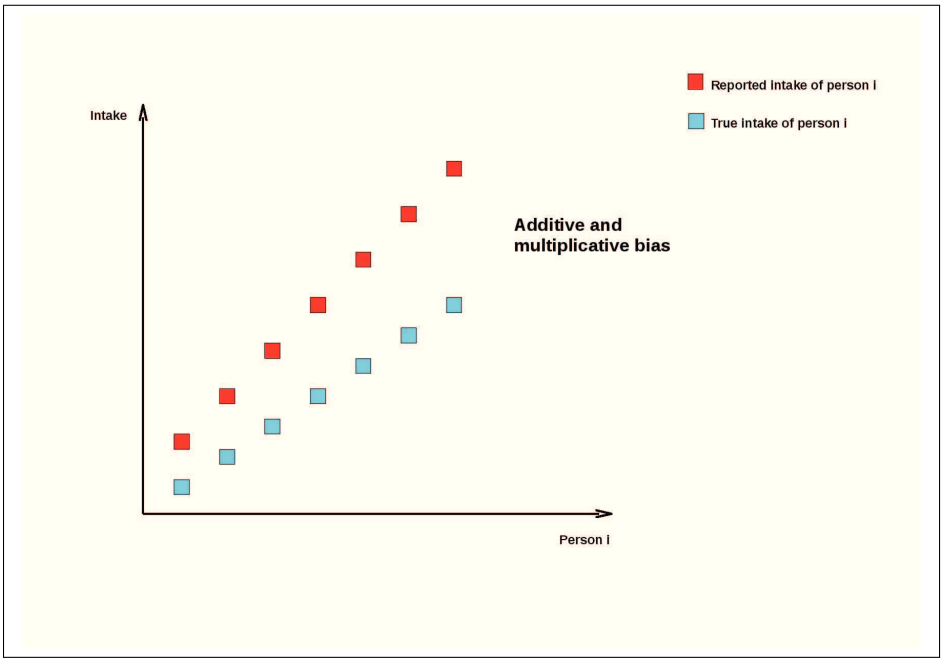


Figure 3.3: Systematic additive and multiplicative error

(c) Person-specific error
This kind of bias is related to personal characteristics of the study subjects, like e.g. age, sex or weight status and is actually a bias at the individual level. It is specific to an individual and can differ among individuals (see Figure 3.4). If it occurs randomly such that it cancels out on group level, it increases the scatter around the regression line. Then odds ratio estimates are attenuated, significance tests are less powerful and the study power is decreased. Unfortunately, the assumption of random person-specific bias may in most cases be violated due to e.g. differential misreporting (see also Chapter 5). Non-random person-specific errors introduce differential error

that is difficult to handle which is further discussed in Section 3.1.2 and 3.3.

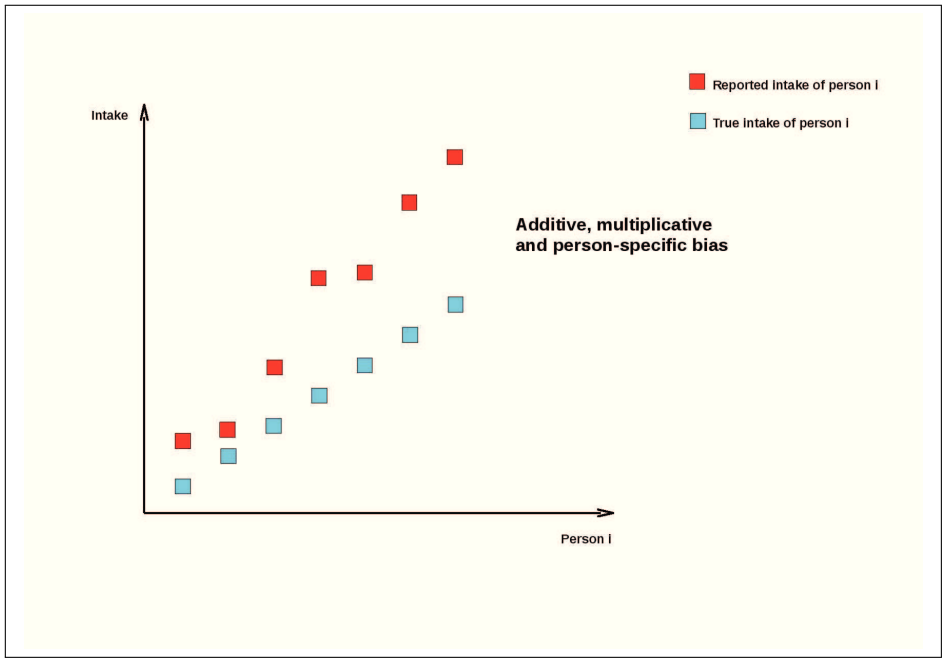


Figure 3.4: Intake-related bias (person-specific bias)

For example under- and overreporting of intake by some subjects introduces systematic within-person error. This may again result in systematic between-person error (Paeratakul et al., 1998) meaning that the mean value for the group is incorrect. Such bias could appear to be randomly distributed in the study population, i.e. some subjects under- and some overreport, but only until introducing an additional variable like e.g. the weight status in the analysis. The bias may co-vary with a variable of analytical interest e.g. in the presence of an opposite reporting bias in lean and obese subjects. If lean people overreported whereas at the same time obese ones underreported, associations would be masked or even reversed resulting in erroneous conclusions (Beaton, 1994).
Further common systematic errors in dietary data emerge from errors in food composition tables for certain food items. Such errors affect intake data differently for different study subjects as the consumption frequencies or amounts of the affected food items also differ by study subject.
In general, the different types of error occur simultaneously such that the magnitude of the single error types and their effects on measures of associations or estimates of dietary intake distributions are difficult to distinguish and hence to determine. To date, there is no suitable procedure how to deal with systematic errors. A second, superior measurement of the exposure variable is needed to

determine the magnitude of systematic errors, i.e. a validation or so-called calibration study (for details see Chapter 4). Ideally an internal validation study should be used in such cases. If such data are not available, corrections may be performed based on external validation data. However, the latter approach should only be applied if the assumption of similar error structures in both, the primary data and the external data, seems justified. For example one requirement might be that the datasets are based on similar study populations in terms of the characteristics of the study participants.
The ideal dietary assessment instrument would have only random error, i.e. it should be unbiased for true intake. In the OPEN (Observing Protein and Energy Nutrition) biomarker study, attenuation factors, which give an indication of the measurement error (see Section 3.3), and correlations between measured intakes and recovery biomarkers (see Section 4.2.1) were shown to be substantially better for repeated 24-HDRs compared to FFQs (Kipnis et al., 2003). Although also for 24-HDRs systematic biases were found (Westerterp & Goris, 2002), the structure of measurement errors in 24-HDRs reported so far makes it the best approximation of an unbiased instrument for the estimation of usual intake distributions and associations between diet and health outcomes. However, recently combinations of 24-HDRs and FFQs were shown to be superior, where the additional information obtained from the FFQ is especially useful when the interest lies in episodically consumed foods as the FFQ provides information on non-consumers (cf. also Section 2.3).

3.1.2 Differential vs. non-differential errors

Next to the differentiation between random and systematic errors, it is usually distinguished between differential and non-differential measurement errors. This is mainly done for conceptual reasons as several correction methods of epidemiological measures and statistical models require the assumption of errors being non-differential. This means that it is assumed that errors occur equally in groups of cases and non-cases (e.g. diseased and non-diseased subjects). This formally requires the distribution of the disease D given (T, Q) depending only on the true intake T. Errors that do not occur randomly in relation to disease, i.e. systematic differences in measurement errors between diseased and non-diseased subjects, are called 'differential' errors. These have serious consequences on measures of association and population distributions but can usually not be corrected in the absence of validation data (Buzas et al., 2004). Although the assumption of only

non-differential measurement errors is questionable in case of dietary data due to the known problem of differential misreporting (Black & Cole, 2001; Livingstone et al., 1992), it has become a statistical convention to assume e.g. 24-HDRs to be an unbiased estimate of true intake. Nevertheless, this should be considered rather as a pragmatic approach due to the lack of a formal procedure to handle differential errors in 24-HDR data.

3.2 Measurement error models

Improvement of exposure assessment to prevent measurement errors in an early stage is the first step to obtain good estimates of associations between diet and diseases. But at some stage it is not feasible to further reduce measurement errors. Then the next step is to determine the type and magnitude of measurement errors and to attempt to understand the effects of measurement errors on the investigated relationship.
Different measurement error models were suggested to represent the relation between the true intake T and its measured value Q aiming to obtain nearly unbiased effect estimates and valid inferences. Throughout all models it is assumed that the errors in the dietary assessment instrument are non-differential with respect to the disease D, i.e. that the surrogate measure Q adds no additional information about disease risk beyond that provided by the true intake T.
For a long period of time, the 'classical measurement error model' was the most commonly applied model when analyzing nutritional data. Many error structures can be transformed into the classical error model, which is the reason for its wide application (Buzas et al., 2004). It accounts for random errors in the surrogate measure Q and was subsequently extended to account also for the other types of measurement error mentioned in Section 3.1. The classical measurement error model assumes an independent, unbiased additive measurement error such that:

$$Q_{i,j} = T_i + \epsilon_{Q_{i,j}},$$

where $Q_{i,j}$ denotes the reported intake of individual i ($i = 1, ...M$) on day j ($j = 1, ...N$), T_i denotes the true long term usual intake of individual i and $\epsilon_{Q_{i,j}}$ denotes the random measurement error of individual i on day j with $\epsilon_{Q_{i,j}} \sim N(0, \sigma^2_{\epsilon_Q})$. The error $\epsilon_{Q_{i,j}}$ is assumed to be independent of the latent true variable T. In this model, the variance of Q can be expressed as sum of the variances of the true

intake σ_T^2 and of the random error $\sigma_{\epsilon_Q}^2$, i.e. $\sigma_Q^2 = \sigma_T^2 + \sigma_{\epsilon_Q}^2$. In addition, it is assumed that $Cov(Q,T) = Var(T) = \sigma_T^2$, where $Cov(Q,T)$ denotes the covariance between Q and T. Furthermore, it holds that $E(Q_{i,j}|T_i) = T_i$ meaning that the surrogate Q is an unbiased estimator of the true intake T.
This model has further been extended adding a parameter β_0 for an additive, systematic bias that is equal for all individuals as well as a parameter β_1 for a multiplicative, intake-related bias (see also Section 3.1.1):

$$Q_{i,j} = \beta_0 + \beta_1 T_i + \epsilon_{Q_{i,j}}.$$

In this model all reported intakes are assumed to differ by a constant amount β_0 from the true intake. In addition to the additive error, the intake is scaled up or down by a factor β_1.
Another model extension suggested amongst others by Kipnis et al. (2002) further introduces a random person-specific effect U, i.e.

$$Q_{i,j} = \beta_0 + \beta_1 T_i + U_i + \epsilon_{Q_{i,j}}$$

with $U_i \sim N(0, {\sigma_U}^2)$. The person-specific effect is specific to an individual but differs among individuals. It is actually a bias at the individual level and reflects the idea that two people that eat exactly the same may nonetheless report their intakes differently. In the above model the person-specific effect is assumed to be random such that it cancels out at the group level. This model will be further referred to as the 'Kipnis model'.

After selecting an appropriate measurement error model for the relation between the measurement Q and the true intake T, the magnitude and effects of measurement error on measures of association can be evaluated as further described in the following sections. However, the usefulness of these models strongly depends on the appropriateness of the underlying model assumptions.

3.3 Correction for the effects of measurement errors

In this section, different commonly applied statistical concepts to correct for measurement errors in nutritional epidemiology are shortly summarized. A lot of publications are based on the classical measurement error model assuming that the measurement error is simply additive. This model yields estimates of the magnitude of the error and allows to apply respective corrections.

Let us assume that the interest lies in the association between an outcome D and the true intake T which is assumed to be linear:

$$D = \alpha_0 + \alpha_T T + \epsilon_T. \tag{3.5}$$

As T cannot be observed it is commonly substituted by a measured value Q of the true exposure T assuming the classical measurement error model, i.e.

$$D = \tilde{\alpha}_0 + \tilde{\alpha}_T Q + \tilde{\epsilon}_T \tag{3.6}$$

with $\tilde{\alpha}_0$ and $\tilde{\alpha}_T$ denoting the regression coefficients obtained when using Q instead of T and assuming $Q = T + \epsilon_Q$. In this situation the relation between the true regression coefficient α_T and the regression coefficient $\tilde{\alpha}_T$ obtained based on Q can be described as

$$\tilde{\alpha}_T = \frac{Cov(D,Q)}{\sigma_Q^2} = \frac{Cov(D, T+\epsilon_Q)}{\sigma_{T+\epsilon_Q}^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{\epsilon_Q}^2} \frac{Cov(D,T)}{\sigma_T^2} = \lambda_{TQ}\alpha_T.$$

Here $\sigma_Q^2 = \sigma_{T+\epsilon_Q}^2 = \sigma_T^2 + \sigma_{\epsilon_Q}^2$, σ_T^2 and $\sigma_{\epsilon_Q}^2$ denote the variances of Q, T and ϵ_Q, respectively, $Cov(D,Q)$ denotes the covariance of D and Q, $Cov(D,T)$ denotes the covariance of D and T and

$$\lambda_{TQ} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{\epsilon_Q}^2}. \tag{3.7}$$

The factor λ_{TQ} is termed 'attenuation factor' and can be used as an indicator for the measurement error (Kipnis et al., 2002). It gives the degree to which a regression coefficient is biased towards the null where lower values indicate larger attenuation.

As the attenuation factor corresponds to the slope when regressing T on Q, i.e. $T = \lambda_0 + \lambda_{TQ} Q + \epsilon$, it can also be written as:

$$\lambda_{TQ} = \frac{Cov(Q,T)}{{\sigma_Q}^2} = \rho_{QT}\frac{\sigma_T}{\sigma_Q}, \text{ where} \tag{3.8}$$

$Cov(Q,T)$ denotes the covariance of Q and T and $\rho_{Q,T}$ denotes the correlation coefficient between the measurement Q and the true intake T. The latter correlation $\rho_{Q,T}$ is also commonly referred to as 'validity coefficient' (see also Section 4.2.3). Equation (3.8) can be derived from (3.7) using the assumptions of the classical measurement error model:

$$\lambda_{TQ} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_{\epsilon_Q}^2} = \frac{Cov(Q,T)}{\sigma_Q^2} = \frac{Cov(Q,T)}{\sigma_Q \sigma_T}\frac{\sigma_T}{\sigma_Q} = \rho_{QT}\frac{\sigma_T}{\sigma_Q}.$$

The sign of $\tilde{\alpha}_T$ is always the same as the sign of α_T and in nutritional studies usually $\rho_{Q,T} \geq 0$ and $\sigma_T^2 < \sigma_Q^2$ such that the factor λ_{TQ} lies between 0 and 1. Accordingly, the slope estimate is always biased towards the null when assuming the classical measurement error model, commonly referred to as 'attenuation'.
The attenuation factor can be estimated based on reference data (see also Section 3.3.1). In the presence of independent biomarker information, an unbiased estimate of the attenuation factor can be obtained which can then be used to correct the observed slope $\tilde{\alpha}_T$ for the measurement error. In case a second dietary assessment instrument serves as reference measure, it is likely that the estimate of λ_{TQ} is biased due to the potentially correlated error structures between the reference measure and the exposure Q (cf. Section 4.2.2). However, risk estimates or regression coefficients corrected based on an imperfect estimate of λ_{TQ} will presumably still better reflect the true association compared to effect estimates without any correction.
Analogously, in case of logistic regression, the relationship between the true relative risk RR_T and the observed relative risk $\widetilde{RR}_T$ when using Q instead of T can be derived as

$$RR_T^{\lambda_{TQ}} = \widetilde{RR}_T. \tag{3.9}$$

In logistic regression models, the attenuation of odds ratio estimates is always towards the one assuming the classical measurement error model.

However, the assumption of strictly additive and random within-person errors may be questioned when dealing with dietary data (cf. Section 3.1). In general, the commonly drawn conclusion that measurement errors always bias the slope estimates in linear regression models towards the null and respectively odds ratio estimates in logistic models towards the one ('attenuation') is questionable. The effect of measurement errors depends upon the selected measurement error model as well as on the joint distribution of the measurement error and the other variables. For instance, in a general measurement error model with possibly differential error the slope estimate obtained when naively substituting T by Q is given as

$$\frac{\alpha_T Cov(Q,T) + Cov(Q,\epsilon_T)}{\sigma_Q^2}$$

and thus differs from α_T (Buzas et al., 2004). In this case, the slope of the naïve model could be larger or smaller than α_T which depends on the covariances between Q and T and between Q and ϵ_T as well as on the variance of Q. No general statement on the bias can be given in this situation. The same applies to the residual variance which may also be larger or smaller depending on the mentioned variances and covariances (Buzas et al., 2004). When allowing differential error and thus assuming this least restrictive type of measurement error model, recovery of α_T from the regression of D on Q is only feasible if one knew or was able to estimate $Cov(Q,T)$ as well as $Cov(Q,\epsilon_T)$.

3.3.1 Correction of regression coefficients

Random measurement error in the dependent variable D only reduces the precision of the estimated regression coefficient but does not result in attenuation, so that no correction is needed here. However, random errors in the measured exposure not only tend to attenuate the slope in the regression as described in the previous section but also lead to an increase in the residual variance meaning that the data are more noisy. In the classical measurement error model the observed residual variance, i.e. $Var(D|Q)$, is $\sigma^2_{\epsilon_T} + \lambda_{TQ}\alpha_T^2\sigma^2_{\epsilon_Q}$ instead of $\sigma^2_{\epsilon_T}$ which would be the residual variance in case of no error. The increased variance results in a loss of power. This scenario is commonly referred to as 'double whammy' (Carroll, 1998). Therefore, a correction for the effects of random errors would be desirable. A correction based on the attenuation factor λ_{TQ} is in general not feasible as it is unknown but it can be estimated based on reference information R.

Assuming non-differential errors of the reference measure R and an uncorrelated error structure between R and the measured exposure Q, the relationship between the measured exposure Q and the outcome variable D can be corrected using the information obtained when regressing the reference measure R on the measured exposure Q:

$$R = \lambda_0 + \lambda_{RQ}Q + \epsilon. \tag{3.10}$$

The regression coefficient $\widehat{\widetilde{\alpha}}_T$ obtained when simply substituting T by Q in the disease model (3.5) can be corrected based on the estimated regression coefficient $\widehat{\lambda}_{RQ}$ obtained from Model (3.10) to get an estimate $\widehat{\alpha}_T$ of the true regression coefficient α_T:

$$\widehat{\alpha}_T = \widehat{\lambda}_{RQ}^{-1}\widehat{\widetilde{\alpha}}_T,$$

where $\widehat{\lambda}_{RQ} = \frac{\widehat{\sigma}_Q^2 - \widehat{\sigma}_{\epsilon_Q}^2}{\widehat{\sigma}_Q^2}$.
Most commonly chosen reference measures R are objective biomarkers of intake or replicates of short-term instruments like repeated 24-HDRs. In the latter case, the estimator of the error variance $\sigma_{\epsilon_Q}^2$ can be obtained based on variance decomposition (ANOVA).

Violation of model assumptions

When estimating the attenuation factor λ_{TQ} or the validity coefficient ρ_{QT} based on a reference measure R, it is assumed that the errors ϵ_Q and ϵ_R of Q and R are uncorrelated. When using replicate measurements R_i with $i = 1, \ldots, d$ denoting the number of the repetition, it is further assumed that the covariance $Cov(R_i, R_j)$, for $i \neq j$ and $i, j = 1, \ldots, d$, is zero. Violation of one or both of these assumptions leads to an over- or underestimation of the attenuation factor and of the validity coefficient. For example, if $Cov(\epsilon_Q, \epsilon_R) > 0$, the correlation between Q and R is overestimated and hence the estimated attenuation factor will be closer to 1 compared to the true attenuation factor. This means that the actual attenuation will be underestimated. At the same time the validity coefficient ρ_{QT} is overestimated (cf. Section 4.2.3). In contrast, a positive covariance between replicate measurements results in an underestimation of ρ_{QT} and hence in an overestimation of the actual attenuation (Kaaks, 1997). If both assumptions are violated, it is not possible to determine which of the two biases will predominate in the absence of additional information.

Pragmatic approach: correction of regression coefficients

In the presence of replicate measurements for the measured exposure Q, another simple correction for random error in the independent variable Q, assuming approximate normality and only random within-person variation, was suggested by Beaton et al. (1979) which requires no additional reference measure:

$$\widehat{\alpha}_T = \widehat{\widetilde{\alpha}}_T * (1 + \frac{\widehat{\lambda}_Q}{n_Q}), \text{where} \tag{3.11}$$

$\widehat{\lambda}_Q = \frac{\widehat{\sigma}_w^2}{\widehat{\sigma}_b^2}$ denotes the ratio of the within- to between-person variance of Q which can be estimated e.g. by an ANOVA and n_Q denotes the number of replicates per subject for the measurement Q.

A crucial assumption of this correction is the presence of only random error in the replicate measurements Q. Methods for the correction of regression coefficients based only on the ratio of within- to between-person variation are not valid in case of systematic errors. Here a correction is only possible if a second reference measure R of the exposure Q is available at least in a subgroup. As stated in Section 2.1, validation studies based on recovery biomarkers found evidence for systematic errors in short-term instruments like 24-HDRs such that the assumption of only random within-person errors seems questionable.

Correction of relative risks

Different methods exist to correct univariate models for attenuation of relative risks caused by random and systematic within-person errors assuming non-differential errors. A simple correction of odds ratio estimates derived from 2×2 tables is based on the sensitivity (proportion of truly exposed subjects that are classified as exposed in the study) and specificity (proportion of truly unexposed subjects that are classified as unexposed) of the measurement method. Details are given in Barron (1977) and Diamond & Lilienfeld (1962). Such methods are rarely applied for several reasons: (1) an external validation study is needed to assess the sensitivity and specificity of the measurement method; (2) it does not provide confidence limits; (3) it is only applicable for 2×2 tables without any adjustments.

In case of continuous exposures, odds ratio estimates can be corrected based on the ratio of within- to between-person variation in the univariate case but again such methods can neither control for confounding nor do they provide confidence limits. More sophisticated methods to correct relative risks for measurement error in continuous exposure variables are mainly based on regression calibration (Carroll et al., 1998; Carroll & Stefanski, 1994; Spiegelman et al., 1997) which requires a validation or calibration study at least in a subset of the study participants and is further outlined in the following paragraph.

Regression calibration

Next to the a posteriori correction based on the estimated attenuation factor, another commonly applied correction approach is based on 'regression calibration' that was made popular by Rosner et al. (1989). The idea of calibration in general is the replacement of exposures measured with error by 'adjusted' values using additional reference information (cf. Chapter 4) such as biomarker measurements or data from a second dietary assessment instrument. The unobserved exposure Q is replaced by its predicted values obtained from the regression model (3.10) given on p. 45. Afterwards, a standard statistical analysis is run to obtain the parameter estimates of interest. The resulting standard errors need to be adjusted using either bootstrap or asymptotic methods to account for the substitution of Q by its estimated values (Carroll et al., 1995). The regression calibration model can further be extended by inclusion of covariates in Equation (3.10).

An adapted regression calibration model can be applied to correct relative risk estimates obtained from multiple logistic regression models. These procedures rely on the assumption of non-differential linear measurement error with constant variance or linear random within-person error in case of replicate measurements like e.g. repeated 24-HDRs (Freedman et al., 2011; Kaaks & Riboli, 1997; Kaaks et al., 1995; Spiegelman et al., 1997) and are often applied in epidemiological studies.

However, regression calibration based on imperfect reference data is problematic unless the error correlations between the reference R and the exposure Q are known (Fraser & Shavlik, 2004). There is evidence that dietary reference instruments based on self-reports do not meet the requirement of having error structures independent of those of the measured exposure Q. For this reason, adaptations of the standard regression calibration approach to model group-specific biases related to true intake and correlated error structures between R and Q were proposed by Fraser & Shavlik (2004) and Kipnis et al. (2001).

Apart from that, additional information from reference instruments is rarely available or often available only in subgroups. Nevertheless, even application of the above correction methods based on a representative subsample or based on estimates of the regression coefficient λ_{RQ} obtained from comparable previous studies may still yield better results compared to those obtained when analyzing the data without any correction.

Recent developments

Another sophisticated method to consider measurement error in the dietary exposure when estimating diet-disease associations was recently proposed by Kipnis et al. (2009). It is an extension of the so-called U.S. National Cancer Institute (NCI) Method (Tooze et al., 2006, 2010) which aims to estimate population distributions of usual dietary intakes based on repeated 24-HDRs (see also Section 3.3.4). The NCI-Method relies on the assumption that the intake reported in 24-HDRs is an unbiased estimate of the true intake T. The extension proposed by Kipnis et al. (2009) predicts individual usual intakes which can then be used as exposure measures in diet-disease models. The individual usual intakes are predicted as conditional mean intakes given the reported 24-HDR intakes and further covariates. The approach is based on regression calibration where the calibration model additionally includes person-specific effects to account for subject-specific deviations in intake (cf. 'Kipnis model' in Section 3.2). The model accounts not only for covariates related to the health outcome but additionally for covariates related to usual intake. Such covariate adjustment may consider information obtained from an FFQ which was shown to yield improved estimates of individual usual intakes especially when the interest lies in episodically consumed foods. The advantage of combining two dietary assessment instruments when investigating diet-disease associations over the use of single instruments was recently also confirmed in a study by Carroll et al. (2012).

Another recently proposed method to calculate individual usual intakes which can then be used to investigate associations between diet and health is the so-called Multiple Source Method (MSM) (Haubrock et al., 2011). It is also based on repeated 24-HDR data incorporating additional FFQ information and is applicable to nutrient and food intakes including episodically consumed foods. In this method, a "two-part-shrinkage-technique" is applied to the residuals obtained from two regression models - one for the consumption amount and one for the frequency of consumption - to obtain estimates on individual usual intakes. Details are given in Haubrock et al. (2011).

3.3.2 Correction of correlation coefficients

If the correlation between two variables shall be estimated and one or both of them are measured with random within-person error, the estimated correlation coefficient may be attenuated towards zero. In nutritional epidemiology one is e.g. often interested in the correlation between an assessment instrument Q and

a reference instrument R for validation purposes. A commonly proposed correction method is based on variance decomposition (Willett, 1998) and assumes both variables Q and R to be measured with random within-person error only. Furthermore, the random within-person errors of Q and R are assumed to be independent. The true correlation r_T can then be estimated as follows based on the observed correlation $\widehat{\widehat{r}}_T$ and estimates of the variance components

$$\widehat{r}_T = \widehat{\widehat{r}}_T * \sqrt{(1 + \frac{\widehat{\lambda}_Q}{n_Q}) * (1 + \frac{\widehat{\lambda}_R}{n_R})}, \tag{3.12}$$

where $\widehat{\lambda}_Q = \frac{\widehat{\sigma}^2_{w_Q}}{\widehat{\sigma}^2_{b_Q}}$ and $\widehat{\lambda}_R = \frac{\widehat{\sigma}^2_{w_R}}{\widehat{\sigma}^2_{b_R}}$ denote the estimated ratios of the within- to between-person variances for Q and R and n_Q and n_R denote the numbers of replicates per subject for variables Q and R, respectively.

If only one variable, say R, is measured with random within-person error, the Equation (3.12) reduces to

$$\widehat{r}_T = \widehat{\widehat{r}}_T * \sqrt{(1 + \frac{\widehat{\lambda}_R}{n_R})}. \tag{3.13}$$

The corrected correlation coefficients are commonly referred to as 'de-attenuated correlation coefficients'.

Confidence intervals for the corrected correlation coefficients should reflect errors in the estimated observed correlations and errors in the estimation of within- and between-person variances making such calculations quite complex. A comprehensive discussion on this topic can be found in Rosner & Willett (1988).

In epidemiological studies aiming to obtain the most precise estimate of true correlations based on short-term dietary information, there is always a trade-off between the number of subjects and the number of replicates per subject due to limitations in terms of time and costs. Minimizing the standard error of the corrected correlation can serve as a criterion for optimization (Rosner & Willett, 1988). It was shown that in most situations the variance is minimized with only two replicates per subject. Furthermore, having two replicates for all study subjects was shown to be more efficient compared to having at least two replicates in a subsample only (Willett, 1998).

Systematic within-person error that does not affect all study subjects equally may also attenuate correlation coefficients but up to now there is no formal way to overcome this problem. Systematic errors that affect all subjects to the same

degree do not affect correlations.

3.3.3 Correction of errors in confounding variables

Measurement errors in confounding variables can distort risk estimates in any direction (attenuation, exaggeration, change of sign, observed positive effect though true null effect) (Greenland, 1980; Kupper, 1984). This applies especially in case that the variables measured with error are correlated or in case their errors are correlated. A thorough measurement error analysis is required in such situations. Bias induced by measurement error in one variable, say Z, may not only affect the regression coefficient of Z but also the estimated effect of another variable Y unless Y and Z are independent (Carroll et al., 1985). Different statistical procedures were suggested to correct relative risk estimates for measurement errors in confounding variables or in the exposure variable (Rosner et al., 1990, 1992) assuming either only random within-person error or both, random as well as additive systematic within-person error. The latter is an extension of the regression calibration approach briefly summarized in Section 3.3.1 and requires again validation data at least in a subgroup. It accounts for the errors in mismeasured variables as well as for correlations of errors between variables.

3.3.4 Correction of dietary intake distributions

Next to the investigation of diet-disease associations, another common research question in nutritional epidemiology is the investigation of population groups at risk for dietary inadequacy or excess[4]. Evaluation of the percentage of a population at risk requires knowledge on the population distribution of usual intakes and is based on defined cut-offs for nutritional adequacy. The left side of Figure 3.1 shows a true population distribution of usual intakes. Furthermore, the critical intake level is indicated through which the population at risk of malnutrition is defined in this example. The "true" prevalence is marked in red. When estimating population distributions of dietary intake based on short-term measurements, the presence of random within-person variation increases the standard deviation of the measurements which broadens the estimated intake distribution (see the right hand side of Figure 3.1). The flattening of the distribution will in general

[4] The term 'population group at risk of dietary inadequacy or excess' is used in a rather general manner here and shall include population groups complying/not-complying to intake recommendations as well as population groups at risk for mal- or overnutrition.

result in misleading estimates of the prevalence of low and high intakes (Souverein et al., 2011). Additionally, systematic errors in dietary data may shift the observed intake distribution to the left or right which further biases prevalence estimates. In the example displayed in Figure 3.1, the observed distribution is markedly shifted to the right such that the observed prevalence of malnutrition based on the given critical intake level is zero though the true intake distribution indicates a population group at risk. This example reveals that an adequate estimation of the usual intake distribution is important for the calculation of population attributable risks and for etiological research.

The problem of random day-to-day variation applies in particular when assessing dietary data by means of short-term instruments where the 24-HDR is the preferred instrument when estimating population distributions of intake. Replicate short-term measurements are needed at least in a subsample to correct the observed distribution for the effects of within-person variation. Single 24-HDRs are a poor estimate of usual intake and averaging over few repeated assessments removes only part of the intra-individual variation. This means that percentile estimates will still be biased. Assuming 24-HDRs to be an unbiased estimator of usual intake, the mean over a large number of repeated assessments would approximate the true intake. But as repeated assessments are cost-intense and put a high burden on study participants as well as on study personnel, this approach is not feasible in epidemiological studies. Therefore, several statistical modeling procedures have been proposed to obtain a valid estimate of the distribution of usual intakes based on short-term dietary information (Dodd et al., 2006; Haubrock et al., 2011; Hoffmann et al., 2002; Souverein et al., 2011; Tooze et al., 2006). Such methods primarily aim to correct for the inflated intra-individual variability due to day-to-day variation. Methods recently proposed, which were developed based on predecessor methods, include the NCI-Method (Tooze et al., 2006), the EFCOVAL Consortium Multiple Source Method (MSM) (Haubrock et al., 2011) and the Statistical Program for Age-adjusted Dietary Assessment (SPADE) (Dekkers, 2009). The underlying statistical models differ, but the correction methods share a common structure that can be described in three steps:

- **Description of the assumed relationship between individual 24-HDR data and individual usual intake:**
  The 24-HDR data are transformed to obtain approximately normally distributed data, where log-, Box-Cox- or power-transformations are usually chosen. Depending on the selected method, the 24-HDR is either assumed to be an unbiased estimator of usual intake on the transformed or on the

untransformed scale .

- **Decomposition of variance in intra- and inter-individual variance components:**
  The mean usual intake is estimated on the transformed scale (sometimes as function of age (SPADE) or optionally for different age groups (NCI-Method)) where the correction for the increased intra-individual variation is done based on variance decomposition. To be able to obtain estimates for the intra- and inter-individual variation, at least two repeated 24-HDRs are required for the total study sample or in some methods for a subsample only.

- **Estimation of the usual intake distribution accounting for the intra-individual variation:**
  After elimination of the intra-individual variation the data are back-transformed to the original scale.

Differences between the methods lie in

- the statistical approach (repeated measures models, fractional polynomials to estimate the intake as a function of age,...);
- the possibility to adjust for covariates like sex, socio-economic status or assessment day;
- the possibility to estimate intakes for episodically consumed foods;
- the inclusion of information from a second dietary assessment instrument (e.g. FFQ) as covariate or for the identification of consumers and non-consumers in case of episodically consumed foods;
- the possibility to account for correlations, e.g. between intake quantity and probability of consumption.

In case that no repeated assessments are available, Jahns et al. (2005) proposed a method to correct for the inflated intra-individual variation based on external variance estimates obtained from previous studies. This approach was shown to give satisfactory results though estimates based on internal variance estimates were shown to be superior. For this reason, the correction approach based on external data should only be chosen if no repeated measures are available for the study population under investigation. Nevertheless, Jahns et al. (2005) concluded

that 'any' correction may give better estimates compared to estimates obtained without accounting for the variance inflation in any way.

Most of these statistical procedures assume that 24-HDRs are unbiased for usual intake on the individual level though recent validation studies based on recovery markers made this assumption questionable (Westerterp & Goris, 2002; Kipnis et al., 2003). To date, none of the above mentioned methods based on self-reported data is able to account for systematic measurement errors e.g. resulting from subject-specific misreporting of single food items or underestimation of portion sizes. Therefore, it is likely that the resulting distributions are still biased where it can be hypothesized that the intake distributions are shifted to the left, i.e. the intakes are underestimated on average, as underreporting seems more likely to occur than overreporting (Livingstone & Black, 2003; Maurer et al., 2006).

Due to the violation of the central assumption of 24-HDRs being unbiased for usual intake, Yanetz et al. (2008) proposed a method based on external nutritional biomarker information from previous studies to correct the usual intake distribution not only for within-person variation but also for systematic bias. This method, however, relies on the crucial assumption, that the ratio of the mean true intake to the mean reported intake is equal in the analysis sample and in the external biomarker study. The same is assumed for the ratio of the variances. When applying this method, the estimated usual intake distributions were narrower and less skewed compared to estimates obtained without incorporating the external biomarker information. Estimated median intakes increased by 8% up to 16% after biomarker adjustment which led the authors to the conclusion that this method provides a basis to account for the well-documented problem of underreporting.

Apart from this method, only little is known about how to correct for misreporting - especially in the absence of reference biomarker information. In addition, differential misreporting of single food items poses a challenge and to date there is no formal way to account for it neither when investigating diet-disease associations nor when modeling population distributions of usual intake.

Chapter 4

Reliability and validity of dietary data

This chapter describes basic concepts to check the structure and presence of measurement errors in self- or proxy-reported dietary data based on reference measures of intake which may also include repeated assessments. The optimal reference instrument fulfills the following criteria (Kipnis et al., 1997):

- It is unbiased with regard to the true intake;
- Its errors are not correlated with the true intake;
- Its errors are not correlated with the errors in the reported dietary data.

Unfortunately, 'optimal' reference measures are rare such that researchers often have to fall back to reference measures fulfilling the requirements only to some degree.

4.1 Reliability of dietary information

A dietary assessment instrument is considered reliable if its measurements are reproducible and stable under the different conditions in which it is likely to be used (Rutishauser, 2005). Repeated assessments are needed to evaluate the reliability of an instrument. Statistical methods to investigate the reliability of a measurement are summarized in Atkinson & Nevill (1998). As reliability is decreased by measurement error, test-retest methods may be problematic in the assessment of diet due to the known systematic and random errors, especially due to the day-to-day variation in diet. This problem in particular affects short-term instruments like 24-HDRs. For example total energy intakes assessed on two different days in the same subject via 24-HDRs will presumably not coincide even if the information given by the participant is correct. For this reason, reliability studies in nutritional epidemiology are mainly conducted based on long-term instruments like the FFQ.

4.2 Validation of dietary information

A dietary assessment method is called valid if it actually measures what it is intended to measure ('the truth') (Rutishauser, 2005). In nutritional studies the 'true' exposure means the actual intake over the period under observation. Here, validity is difficult to assess as the true intake is never known with certainty. For dietary assessment instruments a validation requires an additional measure of intake, e.g. a second assessment instrument or biomarker information. The measurement error structure of the self-report instrument is examined then by comparison with the reference measure.

Validation studies may either assess the 'relative' or the 'absolute' validity of an instrument. When measuring relative validity, also often referred to as calibration study, an assessment instrument is compared to another instrument of the same kind, e.g. food records vs. 24-HDRs. The main limitation of this approach is the potentially correlated error structure between the instruments. For example participants that tend to underreport may do this in both assessments (Kipnis et al., 2003).

Absolute validity is assessed by comparison of reported dietary intake with a gold standard measurement. The classical example is the comparison of energy intakes assessed by 24-HDRs with total energy expenditure (TEE) measured by the doubly labeled water (DLW) technique. Such validations are rarely applied or only applied in small samples due to the high burden for both, respondents and study personnel, as well as due to the high costs (Livingstone & Black, 2003; Schoeller & van Santen, 1982). A comprehensive description of the DLW technique is given in Section 5.1.1.

4.2.1 Biomarker-based validation

Nutritional biomarkers are indicators of dietary intake or nutritional status that can be measured in biological media like human tissues, cells or body fluids. Biomarkers can be used to compare measured nutrient intakes to intakes estimated through dietary assessment. Due to the independent error structures between biomarker measurements and reported dietary intakes, biomarkers can provide an objective validation of several assessment instruments. The underlying assumption is that biomarkers of intake respond to intake in a dose-dependent way. The sensitivity to intake, e.g. the bioavailability and the half time of the nutrient, as well as the time integration, which relates to the time needed to respond to intake, differ depending on the chosen biomarker (Willett, 1998). In

particular, two types of biomarkers are commonly distinguished (Livingstone & Black, 2003):

I) Recovery biomarkers
Recovery biomarkers are considered the gold standard for validating a dietary assessment instrument. They are based on a known quantitative relationship between intake and output over a given period of time and can therefore be translated into absolute estimates of intake (Kaaks et al., 2002). Due to the independent error structures, they give an objective measure of intake. Yet, there are only three dietary recovery biomarkers known, namely DLW[5], urinary nitrogen and urinary potassium, which can be assessed under free-living conditions.

II) Concentration biomarkers
These biomarkers measure concentrations of specific components in urine, blood, adipose tissue, or other tissues. Examples are plasma vitamin C, plasma carotenoids or fatty acids. They do not reflect absolute dietary intake as they may be influenced by lifestyle behaviors like e.g. smoking or supplement use. This means that the between-subject variation is not only explained by differences in dietary intakes, but also by "*variations in digestion and absorption, distribution over body compartments, endogenous synthesis and metabolism, and excretion*" (Kaaks et al., 2002). Furthermore, the time period of exposure reflected by the biomarker may differ from that estimated using a dietary assessment instrument. For example, vitamin C intake measured with one 24-HDR will differ from vitamin C measured at the same day in plasma due to the time integration as well as due to lifestyle factors that affect biomarkers of intake. Concentration biomarkers can therefore only be used to interpret lower limits of true validity (McKeown et al., 2001). This means that good agreement between a nutritional biomarker and reported intake suggests good validity, while rather small agreement does not necessarily imply poor validity as it is likely that the observed agreement is attenuated, e.g. due to the different exposure times reflected by the two measurements.

Biological markers that are most commonly used include DLW, urinary nitrogen, urinary sodium and potassium, plasma levels of vitamins and tissue levels of

[5] Validations based on DLW require the assumption that the study subjects are in energy balance.

minerals and fatty acids (Rutishauser, 2005). In validation studies, the intake assessed using a dietary assessment instrument is compared to the corresponding biomarker of intake. Typically, the concordance between both intake measures is evaluated either based on correlation coefficients or based on the difference between group means. The correlation is considered as an index of the accuracy with which the dietary assessment instrument is able to rank individuals according to their intakes whereas the comparison of group means serves as an indicator for the degree of under- or overestimation of intake (see also Chapter 5).

4.2.2 Relative validation

The term 'relative validation' is used if one assessment instrument is validated against another instrument of the same kind. Such studies are problematic due to the potentially correlated errors between the two assessment methods. Both methods are subject to the same random and systematic errors as both rely on the respondents' memory and capability to correctly estimate food intakes (Kaaks, 1997). In relative validation studies, a high correlation coefficient of the measurements may be either a consequence of the agreement between the two methods or may reflect agreement in the errors. Therefore, relative validation studies should be better called 'calibration' studies.
Short-term dietary assessment instruments can further be validated against direct observation of intake (cf. Section 2.1) where ideally the study participant should not know about the observation. Under laboratory conditions it is not possible to determine what people are usually eating as the intake may be strongly affected by the study subject's knowledge on being under study. This may result in observed intakes that differ from those that would have been observed under free-living conditions.

4.2.3 Method of triads

In order to overcome the drawbacks of relative validation studies, another suggested method to validate dietary assessment instruments is the so-called 'method of triads' (Kaaks, 1997; Ocke & Kaaks, 1997; Yokota et al., 2010). It is commonly applied to validate a dietary assessment instrument Q against a superior reference instrument R and a biomarker B. Figure 4.1, which was adapted from Kaaks (1997), illustrates the method of triads graphically. In most applications, the comparison includes an FFQ (Q), repeated short-term measurements (R) like food records and an appropriate biomarker (B), but other combinations like e.g.

comparisons of an FFQ against two different biomarkers are also possible.
The triads method is used to evaluate the correlation between three measurements

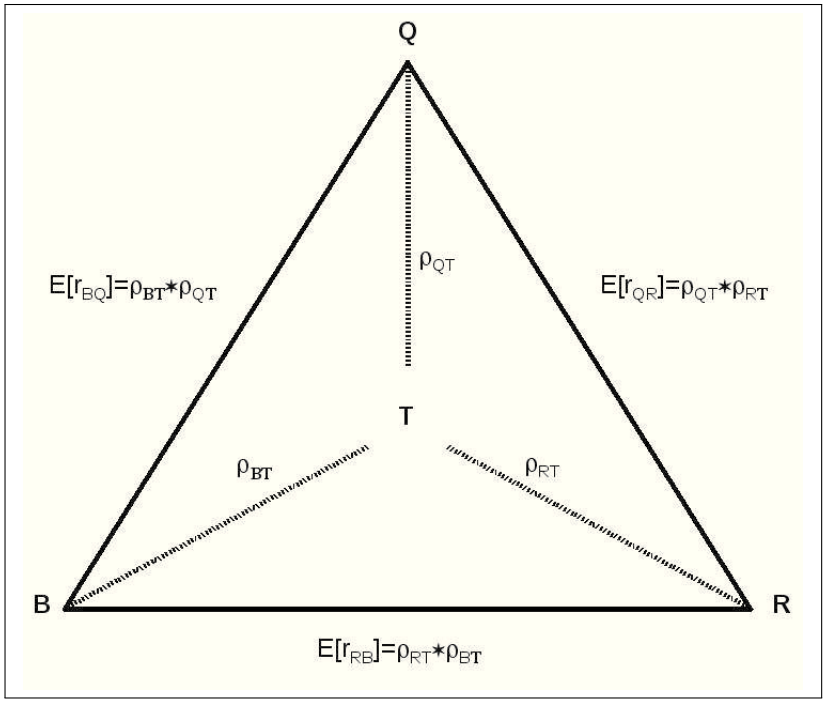


Figure 4.1: Method of triads: Triangular comparison between a dietary assessment instrument Q, a reference method R and a biological marker B. The three correlations between the different measurements are denoted as r_{BQ}, r_{QR} and r_{BR} and ρ_{BT}, ρ_{QT} and ρ_{RT} denote the respective validity coefficients. The correlations between the different measures can each be expressed as the products of the measurements' correlations with the latent true intake T, i.e. as products of the validity coefficients (Kaaks, 1997).

and the latent true intake T and is an application of factor analysis though the equations of this technique can also be derived from structural equation models (Yokota et al., 2010). The basic idea is that the latent intake T can be estimated by means of manifest variables, i.e. measures of dietary intake (Ocke & Kaaks, 1997). The method assumes a linear relationship between the three measurements and the true intake T, such that

$$Q = \beta_{Q0} + \beta_{Q1}T + \epsilon_Q,$$

$$R = \beta_{R0} + \beta_{R1}T + \epsilon_R,$$

$$B = \beta_{B0} + \beta_{B1}T + \epsilon_B,$$

where β_{Q0}, β_{R0}, β_{B0}, β_{Q1}, β_{R1} and β_{B1} denote unknown parameters and ϵ_Q, ϵ_R and ϵ_B error terms. Furthermore, the method assumes independence of the error structures of the three measurements, i.e. independence of the three error terms

ϵ_Q, ϵ_R and ϵ_B.

A triangular comparison between the three measurements is then used to obtain a quantitative estimate of the measurements' correlations with the latent true intake, i.e. the so-called validity coefficients (Kaaks, 1997). These are usually calculated using Pearson correlation coefficients or alternatively Spearman correlations depending on the measurements' scales and distributions. The correlations between the different measures can each be expressed as the products of the validity coefficients (cf. Figure 4.1)

$$E[r_{QR}] = \rho_{QT} * \rho_{RT},$$

$$E[r_{QB}] = \rho_{QT} * \rho_{BT},$$

$$E[r_{BR}] = \rho_{BT} * \rho_{RT},$$

where r_{QR}, r_{QB} and r_{BR} denote the observed correlation coefficients between the three measurements Q, R and B (based on the validation data), and ρ_{QT}, ρ_{RT} and ρ_{BT} denote the validity coefficients between Q, R as well as B and the latent true intake T, respectively.

This allows the validity coefficients to be estimated based on the observed correlations as (Kaaks, 1997)

$$\hat{\rho}_{QT} = \sqrt{\frac{r_{BQ} * r_{QR}}{r_{RB}}}$$

$$\hat{\rho}_{RT} = \sqrt{\frac{r_{QR} * r_{RB}}{r_{BQ}}}$$

$$\hat{\rho}_{BT} = \sqrt{\frac{r_{RM} * r_{BQ}}{r_{QR}}}.$$

Different from correlation coefficients, validity coefficients cannot take negative values due to the square root and range from 0 to 1. High values of ρ_{QT}, ρ_{RT} and ρ_{BT} indicate relatively high correlations between the measurements, i.e. for r_{QR}, r_{QB} and r_{BR}. A small observed correlation between any of the three measurements Q, R and B suggests that at least any of the three measurements is a rather poor estimate of the true intake which results in a small validity coefficient, respectively (Yokota et al., 2010).

A major limitation of this technique is the occurrence of so-called "Heywood cases", i.e. validity coefficients above the value of 1. Such cases appear if the product of two of the three correlation coefficients is larger than the third correlation coefficient, e.g. $r_{BQ} * r_{QR} > r_{RB}$. Main causes for Heywood cases are random sampling variations as well as the violation of any of the model assumptions. Unfortunately, the assumption of independent error structures of the three measurements Q, R and B is often violated. This may e.g. happen if two different dietary assessment instruments are included in the comparison (cf. also Section 4.2.2) which is done in most practical applications due to the limited availability of biomarker data. The most likely model violation is a positive covariance between the errors in different dietary self-report instruments whereas the errors in the biomarker measurements can be assumed to be almost uncorrelated with the errors in the self-report measures (Kaaks, 1997; see also the paragraph on violation of model assumptions in Section 3.3.1). Hence, when including e.g. an FFQ (Q) and 24-HDR (R) in the comparison, it was suggested to interpret the validity coefficient ρ_{QT} for the FFQ as upper limit and the correlation coefficient r_{QB} between the FFQ and the biomarker as lower limit of the FFQ's validity (McNaughton et al., 2005; Dixon et al., 2006). Especially an underestimation of r_{QB} is quite likely which may result from random errors in Q, in B or in both.
Furthermore, the occurrence of empirical negative correlations between any of the three measurements does not allow the calculation of the validity coefficients. Negative correlations may be observed in situations in which the true correlations are close to zero. This problem can be mitigated by adequate choice of the three measurements e.g. based on previous literature before applying the method of triads.
The main advantages of the triads method are the independent error structure of the biomarker measurement compared to the dietary assessment methods as well as the inclusion of information from three different sources.
An adapted version of the method of triads was suggested by Fraser & Shavlik (2004) which does not rely on the assumption of independence of the measurement errors in the different measures of intake. The authors observed correlations between Q and R far from zero such that they concluded that methods based on imperfect reference data are problematic unless the correlations between the errors are known.

Chapter 5

Misreporting

This chapter briefly summarizes the extent, nature and determinants of misreporting and how misreporting may affect estimates when investigating diet-disease associations or usual intake distributions. Furthermore, methods for the identification of misreports are discussed. Different approaches that were suggested to correct for misreporting in the absence of biomarker data are outlined and exemplarily applied in Section 8.2.

Misreporting, i.e. under- and overreporting, is one of the main sources of error in dietary assessments and introduces severe error not only in the estimation of energy intakes but also in the estimation of nutrient intakes. In this context, the term 'underreporting' comprises both, *undereating* and *underrecording* (Poslusna et al., 2009). Undereating means that the study participant eats less than required to maintain body weight. Days of dieting are a typical example for undereating. Underrecording refers to a situation in which the reported energy intake (EI) is smaller than the measured total energy expenditure (TEE) while presuming that the individual is in energy-balance. In energy balance, EI and TEE should coincide. Therefore, reported EI smaller than measured TEE indicates that the study participant did not report all foods consumed or reported smaller amounts than actually consumed, i.e. he/she underrecorded. Underreporting results in erroneously low results for usual intakes (Poslusna et al., 2009), even in case of undereating, as days of dieting are considered as exceptional days and do hence not reflect usual intakes. The term 'overreporting' includes *overeating* and *overrecording*, both defined analogously to undereating and -recording but in the reverse direction.

Furthermore, it is commonly distinguished between *intentional* and *unintentional* misreporting. If a respondent is aware of the misreporting, i.e. he/she intentionally skips or adds food items or reports higher or smaller portions than actually consumed, this situation is called intentional misreporting. If, in contrast, the

respondent is not aware that he/she is misreporting, e.g. due to memory lapses, difficulties to correctly estimate portion sizes or due to unobserved meals in case of proxy-reported data, this is called unintentional misreporting.
It is unlikely that misreporting affects all foods and nutrients in all study participants to the same extent. Several studies hypothesized that misreporting occurs differentially, also referred to as 'selective misreporting' (Poppitt et al., 1998; Lioret et al., 2011), meaning that some food items are more likely to be misreported than others. E.g. food items considered as unhealthy are presumably more likely to be omitted due to social desirability. Furthermore, misreporting was shown to be influenced by the study subjects' characteristics (Murakami et al., 2011; Poppitt et al., 1998). Differential misreporting produces biased effect estimates such that observed diet-disease associations may be attenuated, obscured or even reversed (cf. also Section 3.1.2). The problem is especially challenging if the direction of misreporting is reversed in groups of diseased and non-diseased study subjects. Factors reported to be associated with misreporting include the body mass index (BMI), age, sex, socio-economic status, education, psychological factors as well as eating habits (Lioret et al., 2011; Livingstone & Black, 2003; Poslusna et al., 2009).

5.1 Methods to identify misreports

5.1.1 Doubly labeled water

Doubly labeled water (DLW) is considered the gold standard method to measure TEE under free-living conditions (Schoeller & van Santen, 1982; Westerterp et al., 1995). Water containing enriched quantities of the stable isotopes deuterium (^2H) and oxygen-18 (^{18}O) is orally administered to the study participants. The name of the method ("doubly labeled") results from the fact that both, the hydrogen and oxygen, are labeled. The oxygen-18 is excreted from the body as carbon dioxide ($C^{18}O_2$) and water ($H_2^{18}O$), and the deuterium is excreted as water (^2H_2O). The difference in the excretion rate between these two isotopes serves as a measure of the carbon dioxide production. Using standard equations for indirect calorimetry, energy expenditure can then be calculated from the measured carbon dioxide production (Trabulsi & Schoeller, 2001). The DLW method was shown to be accurate to 1% where the within-subject precision was estimated to lie between 5 and 8% (Schoeller, 2002). Except from the urine collection, the method does not interfere with the study participants' normal activities and therefore with

usual energy expenditure. Nevertheless, the burden for respondents and study personnel is relatively high. Furthermore, the DLW method is quite cost-intense as e.g. the laboratory analysis requires sophisticated equipment. For this reason, energy expenditure is not routinely measured in epidemiological studies. DLW measurements are often obtained in subsamples only for the purpose of dietary assessment validations where sample sizes range from 20 to 500 (Schoeller, 2002). When validating dietary information based on the DLW method, measured TEE is compared to reported EI assuming individuals being in weight balance, i.e. TEE = EI. The difference between both values can serve as an estimate for the magnitude of misreporting (Poslusna et al., 2009).

5.1.2 Goldberg cut-offs

Due to the high costs of the DLW method, the ratio of EI over basal metabolic rate (BMR) is a commonly used alternative to compare reported EI with an independent estimate of calculated energy requirements. Goldberg et al. (1991) defined minimum and maximum plausible levels of EI as multiples of BMR and used the ratio EI/BMR as index for a validity check for bias in EI: If the reported EI is much smaller or higher compared to the BMR value, i.e. the ratio lies below or above the defined plausible limit, the reported EI is likely to be misreported.
BMR can be either measured via indirect calorimetry or calculated from predictive equations that were derived based on data from previous studies. The most commonly applied age- and sex-specific equations for the calculation of BMR were suggested by Schofield (1985) and recommended by FAO/WHO/UNU (1985)[6].
The Goldberg cut-offs consider the duration of dietary assessment, i.e. number of recall days, the sample size as well as intra-individual variations in BMR, physical activity level (PAL) and EI. To account for very low or high reported intakes resulting from random day-to-day variation, the derived cut-off values are less conservative if only one or few recall day/-s is/are available. This means that single intake days that may by chance lie below the normal range of an individual's intake will not be rejected, i.e. classified as misreport, though the intake may be too low to reflect the usual intake.

[6] FAO=Food and Agriculture Organization of the United Nations; WHO=World Health Organization; UNU=United Nations Universtiy

Derivation of the Goldberg cut-offs

In weight-stable individuals, EI must equal TEE on a long-term basis, thus

$$EI = TEE \Leftrightarrow \frac{EI}{TEE} = 1$$

where TEE is usually expressed as product of BMR and PAL

$$TEE = BMR * PAL.$$

Combining both equations, the energy balance equation can be written as

$$\frac{EI}{BMR * PAL} = 1.$$

Accounting for the skewed distributions of PAL and EI by log-transformation to achieve an approximate normal distribution, Goldberg derived the cut-off limits for plausible intake levels as approximate 95% confidence limits (95%-CL) of the agreement between EI and TEE which resulted in the following equation:

$$\text{Lower/upper cut-off} = \text{PAL} * \exp\left[\pm 1.96 * \frac{\frac{S}{100}}{\sqrt{n}}\right], \text{ with} \tag{5.1}$$

$S = \sqrt{\frac{CV^2_{wEI}}{d} + CV^2_{wBMR} + CV^2_{PA}}$, where

CV^2_{wEI} denotes the within-subject coefficient of variation for EI,
CV^2_{wBMR} the within-subject coefficient of variation for BMR,
CV^2_{PA} the coefficient of variation for physical activity,
d the number of assessment days and
n denotes the number of study subjects.

It should be noted that if the dietary assessment method is assumed to measure long-term intake like e.g. the FFQ, the number of days d converges to infinity ($d \to \infty$) and hence $\frac{CV^2_{wEI}}{d} \to 0$. More details on the derivation are given in the appendix of Goldberg et al. (1991).
In practice, the cut-off limits are calculated by insertion of reference values for PAL and for the coefficients of variation of EI, BMR and PA into Equation (5.1)

where the reference values are usually obtained from previous studies. The cut-offs are no longer confidence limits in the statistical sense then.
The originally suggested reference values (cf. Goldberg et al. (1991)) for the different factors in Equation (5.1) were revised by Black (2000) and sometimes further adapted depending on the study population under investigation, e.g. for application to data in children (McCrory et al., 2002; Sichert-Hellert et al., 1998; Börnhorst et al., 2012b).
The Goldberg cut-offs are typically used to classify study subjects in three reporting groups, namely

- Low-energy-report (also often referred to as underreport), if EI/BMR $<$ lower cut-off
- Adequate-energy-report (also referred to as plausible report), if lower cut-off $\leq$ EI/BMR $\leq$ upper cut-off
- High-energy-report (also referred to as overreport), if EI/BMR $>$ upper cut-off.

Throughout this thesis, the term 'report' will be used instead of 'reporter' in order to directly refer to the classification of the recalls but not to the corresponding persons. E.g. a proxy-reported 24-HDR of a child may be classified as 'underreport' but the child is not an 'underreporter' in this example. Here, the proxy provided the dietary information and therefore the term 'underreporter' would be misleading.

5.1.3 Physical activity

The original Goldberg cut-offs were further improved by assignment of individual PA levels to the study subjects and by respective use of different cut-off values depending on the assigned PA level (Black, 2000). Unfortunately, individual information on objectively measured PA was rarely available in previous studies such that this approach is rather uncommon. However, due to the increasing number of studies using e.g. accelerometry to assess PA, it is to be expected that in future the reference value for PAL in Equation (5.1) will be more often replaced by individually calculated PA levels when applying the Goldberg cut-off approach.

5.1.4 Others methods

Besides the methods described above, reported EI may be validated by comparison with actual intakes obtained via direct observation (cf. Section 2.1). As this procedure is quite cost- and time-intense and as the environmental change e.g. due to observation under laboratory conditions may further bias the results, this method is rarely used in epidemiological studies.

Unfortunately, none of the above methods is able to determine selective misreporting of specific foods or macronutrients. Only the total energy intake over a specified period is checked for plausibility based on measurements or estimates of energy requirements. To date, there is no established procedure how to test reported dietary information for selective misreporting. However, typically nutrient or food intakes are compared between groups of low-, adequate- and high-energy-reports to get at least an indication whether selective misreporting may present a problem in the data (Poslusna et al., 2009).

Chapter 6

Two European population-based studies on health in children and adolescents: IDEFICS and HELENA

6.1 The IDEFICS study

This PhD-thesis was written within the framework of the IDEFICS study (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants), which was funded by the 6th EU Framework Programme. The study period was August 2006 until March 2012.
IDEFICS is a longitudinal, multi-center setting-based study that on the one hand aimed to investigate the causes of diet- and lifestyle-induced health effects in children and infants and on the other hand developed, implemented and evaluated primary prevention programs to tackle childhood obesity. The baseline survey was conducted from September 2007 until June 2008 in eight European countries (Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain, and Sweden). In total, more than $31,500$ children were invited out of whom finally $16,225$ participated and fulfilled the inclusion criteria of the IDEFICS study. The inclusion criteria were met, if a child was between 2 and 9 years of age at baseline and if at least information on age, sex, weight and height was recorded. Children were recruited through schools and kindergartens. The survey included interviews with parents on lifestyle habits and dietary intakes as well as anthropometric measurements and physical examinations of the children. Biomarker information was collected via blood, urine and saliva samples. All measurements were taken using standardized procedures by all eight centers participating in the study.
In each country, the participating centers obtained ethical approval from the local ethics committees. Parents provided written informed consent for all examinations. Each child was informed orally about the modules by field workers and

asked for his/her consent immediately before the examination.
More details on the design and objectives of the study can be obtained from Ahrens et al. (2006, 2011) and Bammann et al. (2011).
In IDEFICS repeated short-term as well as long-term dietary information was assessed using 24-HDRs as well as the so-called Children's Eating Habits Questionnaire (CEHQ) including a food frequency part focusing on obesogenic foods (CEHQ-FFQ) (see also Section 6.1.1). These data in combination with the various questions on lifestyle behavior, biomarker data as well as heart rate monitoring and accelerometer measurements offered a great opportunity to gain knowledge on the validity and sources of measurement errors in dietary data in young children.

6.1.1 Dietary information in IDEFICS

24-HDR

The computerized 24-HDR used in the IDEFICS-study called 'SACINA' (Self-Administered Children and Infants Nutrition Assessment) was based on the previously designed and validated HELENA-DIAT (see Section 6.2.1) that was originally developed for Flemish adolescents (Vereecken et al., 2005). SACINA is structured according to six meal occasions (breakfast, morning snack, lunch, afternoon snack, dinner, evening snack) related to a range of chronological daily activities. When selecting a food item, pictures with increasing portion sizes are displayed on the screen to facilitate estimation of portion sizes as illustrated in Figure 2.2 on p. 31. The participant had to specify the consumed quantity in terms of pre-defined standard amounts, e.g. how many spoons of oil (quantity) each with 20ml (pre-defined standard amount) or how many slices of bread (quantity) with 40g each (standard amount). Proxies, mainly the parents, completed the 24-HDR under supervision of field personnel which lasted 20 to 30 minutes. If the child had lunch at school, school meals were additionally assessed by means of direct observation. Teachers and school kitchen teams were interviewed by trained survey personnel and data were documented using special documentation sheets including portion sizes. The assessment procedure was slightly different in the Hungarian study center, where the dietary recalls were not performed via the standardized SACINA software but via paper-pencil 24-HDR registrations that were entered in the SACINA software afterwards. The 24-HDRs were assessed on non-consecutive days over the whole week and over the complete IDEFICS assessment period (see Tables A.1 and A.2 in the Appendix on p. 113f.).

The uniquely coded food items were linked to country-specific food composition tables. Missing quantities for single food items as well as obviously implausible data entries were imputed by country, food group and age-specific median intakes to avoid excessive record exclusions. Approximately 0.01% of all entries were imputed. Up to six repeated 24-HDRs were assessed in 9,774 children (see Table 6.1). Incomplete interviews were excluded, e.g. if the proxy did not know about at least one main meal or in case of missing school meal information[7] (see Tables A.5 and A.6 on p. 117f.). As expected, reported mean energy intakes decreased with increasing number of meals that the parents did not observe. In some cases, the parents indicated that they did not observe a meal but either made a guess on the respective intakes or this meal was assessed in course of the school meal assessment (N=626; see Table A.6 on p. 117). Such interviews were not excluded as mean reported intakes did not significantly differ from intakes obtained based on complete interviews. Intakes from days with missing school meal information, i.e. the parents reported that the child had a meal at school but no school meal was assessed, were significantly lower compared to interviews with complete information (see Table A.5 on p. 116).
Tables A.1 to A.3 on p. 112f. present mean energy intakes (kcal/day) by week-

Table 6.1: Number of repeated SACINA interviews assessed during the IDEFICS baseline survey and mean energy intake (kcal/day) by sex.

	All		Male		Female	
Assessment day	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
1st 24-HDR	9774	1519 (548)	4945	1579 (565)	4829	1457 (522)
2nd 24-HDR	2649	1482 (530)	1321	1532 (549)	1328	1432 (506)
3rd 24-HDR	1276	1390 (510)	597	1443 (555)	679	1343 (463)
4th 24-HDR	52	1276 (530)	23	1223 (467)	29	1319 (580)
5th 24-HDR	14	1438 (581)	5	1659 (529)	9	1315 (602)
6th 24-HDR	8	2108 (928)	3	2408 (1036)	5	1928 (929)

day and separately for days with and without additional school meal assessment including only complete interviews[8]. Here days without school meals are either related to weekend days or to days without lunch at school. Regarding the total

[7] A day with missing school meal means that the child had a meal at school but the information is actually missing. Days without school meal may also relate either to weekend days or to days with lunch at home. Therefore, days without school meal do not necessarily imply missing information. The latter described cases are not considered as interviews with "missing school meal" and not excluded, respectively.

[8] Complete interview means that the parents were able to report the intakes of all meals and that the school meal information was not missing in case the child had a meal at school.

study group, energy intakes were lowest on Fridays and highest on Thursdays and Sundays, though results differed by study center. No significant differences in energy intakes comparing complete interviews with and without school meal assessment were found in the total study group (see Table A.3 on p. 114) but again results differed depending on the study center.
Consistently with previous studies, mean energy intakes for the total study group decreased with increasing number of assessment days, where the highest intakes were reported on the first assessment day (see Table 6.1). This may be explained by decreasing motivation due to the increasing burden for the respondents when repeatedly assessing dietary information. However, after stratification by study center (see Table A.4 on p. 115) this effect was no longer observed where only the intakes of the first 24-HDR were compared to the intakes of the second 24-HDR due to the low numbers of children providing more than two repeated 24-HDRs in most of the study centers. For Hungary and Spain, the mean energy intake was even significantly higher on the second assessment day compared to the first day (t-test: $p < 0.01$). Only in Estonia the mean reported intakes where significantly lower on the second day compared to the first day's intakes in the stratified analysis.

Details on the validation and first analyses using these 24-HDR data are presented in Sections 7.2.1, 7.3, 8.1 and 8.2.

The Children's Eating Habits Questionnaire and its food frequency part

The Children's Eating Habits Questionnaire (CEHQ) contains general questions on eating behaviors, parents' attitudes towards diet and meal frequencies as well as a food frequency questionnaire part (CEHQ-FFQ). The numbers of children whose parents completed a CEHQ in course of the IDEFICS baseline survey are presented in Table 6.2 separately by study center.
The CEHQ-FFQ was developed as a screening instrument to investigate the consumption frequencies of foods that were suspected to be either positively or negatively associated with overweight and obesity in children (Huybrechts et al., 2011). The CEHQ was not designed to provide an estimate of total energy intake or total food intake, and foods less likely to be associated with obesity were not included. The CEHQ questions allow for distinguishing, for instance, between white and whole meal bread, thereby making it possible to compare consumption

of these alternatives. Although different food classification systems are currently available (Ireland et al., 2002), a new food grouping system was developed for the CEHQ-FFQ to meet the specific aims mentioned above. At the same time, the cultural differences in dietary habits between the different IDEFICS centers were taken into account.

The CEHQ-FFQ consists of 43 food items for which parents or other caregivers were asked to report at home or at other people's home[9] consumption frequencies of their children in a typical week of the previous month. Frequency categories ranged from 'Never/less than once a week' up to '4 or more times per day'. Considering the increased burden for the respondent and the lack of accuracy of usual portion size estimates when included in an FFQ (Cade et al., 2002), no attempts were made to assess portion sizes.

The 43 food items were clustered into 14 food groups: breakfast cereals, cheese, drinks, eggs and mayonnaise, fish, fruits, meat, meat replacements, milk, snacks, vegetables, spreadable products, cereal products and yogurt. First results on the reliability and validity of the CEHQ-FFQ are presented in Sections 7.1, 7.2.2 and 7.3.

Table 6.2: Numbers of Children Eating Habit Questionnaires (CEHQ) assessed during the IDEFICS baseline survey by study center.

Study center	CEHQ-FFQ data
Belgium	1862
Cyprus	1672
Estonia	1666
Germany	2014
Hungary	2506
Italy	2250
Spain	1468
Sweden	1759
Total	15197

6.2 The HELENA study

HELENA (Heathy Lifestyle in Europe by Nutrition in Adolescents) is an European multi-center longitudinal study that was funded within the 6th EU Framework Programme (Moreno et al., 2008b). The main objective of the HELENA

[9] For example at the home of grandparents or friends.

study was to obtain reliable and comparable data of a representative sample of European adolescents concerning nutritional and lifestyle status, including amongst others measures of food and nutrient intakes, food choices and preferences, obesity prevalence, physical activity and fitness patterns as well as genetic markers. Data were collected from October 2006 until December 2007. Details on the study design and sampling procedures can be obtained from Moreno et al. (2008).
In brief, participants were sampled conducting a random cluster sampling stratified by geographical location, age and socio-economic status based on boys and girls aged 12.5 to 17.49 years from ten European cities of more than 100,000 inhabitants (Athens (Greece), Dortmund (Germany), Ghent (Belgium), Heraklion (Greece), Lille (France), Pecs (Hungary), Rome (Italy), Stockholm (Sweden), Vienna (Austria), Zaragoza (Spain)). Adolescents were recruited through schools where also the questionnaires were handed out and completed during classes. The total HELENA population consisted of 3,528 eligible adolescents (52.3% females). Both, the adolescents and their parents, gave written informed consent and all study protocols were approved by the local Ethical Committee of the participating study centers.

6.2.1 Dietary information in HELENA

Dietary intake was assessed using a computerized 24-HDR on two non-consecutive days of the week, excluding Fridays and Saturdays. The recalls were completed within a time-span of two weeks where in addition a blood sample was obtained on the day of the first 24-HDR. The 24-HDRs were assessed using HELENA-DIAT (HELENA-Dietary Assessment Tool) which is a validated computer-based tool for self-reported 24-HDRs that is based on a previous version developed for Flemish adolescents (Young Adolescents' Nutrition Assessment on Computer (YANA-C)) (Vereecken et al., 2005).
Furthermore, a short FFQ was completed by the adolescents (Vandevijvere et al., 2012) that was obtained from the Healthy Behaviour in School aged Children (HBSC) study (Currie, 1998). This self-administered FFQ queried consumption frequencies of 15 food items: fruits, vegetables, sweets, soft drinks, light soft drinks, cereals, white bread, brown bread, skimmed milk, whole-fat milk, other milk, cheese, fish, crisps and French fries. The seven answer categories ranged from 'Never' up to 'More than once a day/ every day'.

Chapter 7

Reliability and validity studies in IDEFICS and HELENA

The ability to adequately validate dietary data is addicted to the available reference measurements. Comparing different kinds of validation studies, the informative value varies strongly depending on the reference instrument used (cf. Section 4.2). As gold standard reference measures are rare, in most situations validation studies are only able to provide lower or upper limits of validity, i.e. the minimum or maximum level of validity of the instrument under investigation. Within the framework of the IDEFICS project different studies were conducted to assess the validity of the CEHQ-FFQ and of the 24-HDR SACINA. In addition, a validation study applying the method of triads (cf. Section 4.2.3) was performed to evaluate the dietary assessment methodology used within the framework of the HELENA study (see Section 6.2). These studies are summarized in the following sections where the presented results are mainly based on publications to which I markedly contributed. An exception are the results presented in Section 7.1.

7.1 Reliability of the CEHQ-FFQ used in IDEFICS

The reliability of the CEHQ-FFQ (see Section 6.1.1) was tested in a subsample of 258 children, in which the CEHQ was collected twice to investigate the reproducibility of the questionnaire results (Lanfer et al., 2011). The second administration was done 0 up to 354 days after the first one. Spearman's correlation coefficients and weighted Cohen's kappa coefficients were calculated for each food item of the CEHQ-FFQ to evaluate the agreement between both assessments. Data were analyzed stratified by sex, age group, geographical region and length of the period between the first and second administration.
Significantly positive weighted Cohen's kappa coefficients and Spearman's cor-

relations were observed for all food items under study. Correlation coefficients were comparable to those observed in previous studies on FFQ reproducibility in children and adults. Stratification did not reveal systematic differences in reproducibility by sex and age group but Spearman's correlation coefficients differed significantly between northern and southern European countries for ten food items. As expected, longer time spans (> 128 days) between the first and second administration resulted in lower reproducibility which may not only be explained by memory errors but also by seasonal influences. Several studies suggested that dietary information assessed via long-term instruments may be affected by recent intakes (cf. Section 2.1). Nevertheless, even agreement of responses between the two administrations may not necessarily reflect 'true' responses but also repetition of the same errors e.g. due to underreporting of single food items in both assessments (cf. Section 4.2.2). The latter problem may result in an overestimation of the reliability such that the results should be interpreted as upper limits of reliability only.

The reliability of the 24-HDR SACINA was not tested as repeatedly reported intake data relating to the same day were not assessed. Reported intakes assessed on different days within the same individual are likely to actually differ due to the daily variation in diet. Therefore short-term measures of diet assessed on different days have to be considered as a poor measure to assess reliability.

7.2 Biomarker-based validation in IDEFICS

7.2.1 Recovery biomarker (DLW) vs. SACINA

Though self-reported 24-HDRs were shown to be a valid measure of energy intake (EI) in older children and adolescents (Forrestal, 2011), little is known about the validity of proxy-reported 24-HDRs in young populations yet. Therefore, in a first step the validity of the 24-HDR SACINA (cf. Section 6.1.1) was evaluated by comparison of proxy-reported energy intakes with objective measures of total energy expenditure (TEE) obtained by the DLW technique. Details on the DLW technique are given in Section 5.1.1. This validation study was submitted to the journal Clinical Nutrition and is currently under review (Börnhorst et al., 2012c; see also Appendix B.2).
The study was conducted in a convenience sample of 36 children aged 4 to 10 years from Belgium and Spain in course of the IDEFICS validation study (Bam-

mann et al., 2011). The agreement between EI and TEE was investigated using subgroup analyses and Bland-Altman plots. Groups of low-energy-reports, adequate-energy-reports and high-energy-reports were defined based on the ratio of EI over TEE by application of age- and sex-specific cut-off values. The cut-off values were calculated using the equation suggested by Goldberg (cf. Section 5.1.2) but the reference values used in the calculation of the original cut-offs were substituted by child-specific reference values obtained from previous literature (Black, 2000; Goldberg et al., 1991; Nelson et al., 1989; Torun et al., 1996). Only children with complete information on age, sex, height, weight, two 24-HDRs and DLW measurements were included in the validation.

Regarding the total study group, means of EI (1,500 kcal/day) and TEE (1,523 kcal/day) matched almost exactly though partially large differences between EI and TEE were observed at the individual level. When stratifying by weight status, almost perfect agreement between EI and TEE was observed in thin/normal weight children (EI: 1,511 kcal/day; TEE: 1,513 kcal/day), but also in overweight/obese children the mean difference between EI and TEE was only -86 kcal/day (EI: 1,468 kcal/day; TEE: 1,554 kcal/day) which corresponds to an underreporting of total EI by approximately 4%. Among the participants, 28 (78%) were classified as adequate-energy-reports, five (14%) as high-energy-reports and three (8%) as low-energy-reports. Percentages of EI from fat were lowest in low-energy-reports and highest in high-energy-reports (low-energy-reports: 37.8% EI from fat; adequate-energy-reports: 39.4%; high-energy-reports: 41.8%); the opposite was true for carbohydrates (low-energy-reports: 43.3% EI from carbohydrates; adequate-energy-reports: 42.2%; high-energy-reports: 35.9%).

In a subgroup analysis, the agreement between EI and TEE in an exclusively obese study sample from Sweden recruited through an obesity clinic was investigated. Here, a mean difference of -455 kcal/day between EI and TEE was observed and seven out of the ten Swedish children were classified as low-energy-reports. As these results were likely not to reflect reporting errors but dietary restrictions, the Swedish children were excluded from the main analysis.

In summary, two proxy-reported 24-HDRs turned out to be a valid instrument to assess EI on group level but not on the individual level. The results were in line with previous studies in children based on self-reported dietary data that also observed good validity of the reported intake at least at group level and obesity as a major factor for differences between reported EI and measured TEE (Forrestal, 2011).

If the interest lies in individual usual energy intakes, a higher number of repeated

24-HDRs or additional reference information is required to overcome the problem of day-to-day variation.
Nevertheless, generalization of the results is strongly limited due to the fact that a convenience study sample was investigated and due to the small sample size. For these reasons, the relative validity of the 24-HDR SACINA was additionally assessed by comparison with reported consumption frequencies assessed via the CEHQ-FFQ (see Section 7.3). In this study, all IDEFICS children with CEHQ-FFQ and at least two SACINA measurements were included ($N = 2,033$). Though it would have given the most objective measure of validity, it was not feasible to conduct a validation study based on the DLW method in the total IDEFICS sample due to the high costs and high burden for the participating children, their parents as well as for the study personnel.

7.2.2 Concentration biomarkers vs. CEHQ-FFQ

Due to the fact that the CEHQ-FFQ assessed consumption frequencies, but no portion sizes and respectively no energy, macro- and micronutrient intakes, only a relative validation of single food items either by comparison with appropriate biomarkers or by comparison with information obtained from a second dietary assessment instrument (see the following Section 7.3) was feasible here.
In the biomarker-based validation study, reported milk consumption frequencies were related to urinary calcium and potassium excretions from spot urine samples (Huybrechts et al., 2011). As milk products are a major source of calcium and potassium in children's daily food intakes, urinary calcium and potassium excretions can be used as biological markers to validate reported milk intakes. This validation study was published in the International Journal of Obesity, a peer-reviewed epidemiological journal, and is printed in Appendix B.1.
Data from $10,309$ children aged 2 to 9 years from the IDEFICS baseline survey were included in the analysis. Urinary calcium and potassium excretions were measured in morning spot urine samples and standardized for urinary creatinine excretion. Ratios of urinary calcium over creatinine and urinary potassium over creatinine, respectively, were used in multivariate regression models after logarithmic transformation to obtain approximately normally distributed data. Milk consumption frequencies were obtained from the CEHQ-FFQ. Spearman's correlation coefficients and multivariate regression analyses adjusting for age, sex, study center, soft drink consumption and frequency of main meals consumed at

home were used to investigate the associations between milk consumption frequencies and urinary calcium as well as between milk consumption frequencies and urinary potassium. Crude and partial Spearman's correlations revealed a significant positive correlation between milk consumption frequencies and ratios of potassium over creatinine and a weaker but still significant positive correlation with ratios of calcium over creatinine. Multivariate regression models showed associations between milk consumption frequencies and both urinary biomarkers in the adjusted analyses. Mean ratios of potassium over creatinine increased comparing the first, second and third tertile of the milk consumption frequencies. In addition, children adhering to the recommendation to consume at least two milk servings per day had significantly higher mean ratios of potassium and calcium over creatinine compared to children that consumed less. Large differences in mean ratios of potassium as well as calcium over creatinine were observed between the different study centers.

Summing up, the considered concentration biomarkers reflected the reported milk consumption frequencies to a satisfactory degree where the results should be regarded as lower limits of validity; absolute agreement was not expected as neither reported consumption frequencies nor measures from single spot urine samples reflect absolute intake amounts. However, at least the limitation of only one spot urine sample was partially mitigated by the standardization with urinary creatinine levels which corrects for variations in concentrations of urinary excretions during the day.

Due to the lack of adequate nutritional biomarkers, all other food groups assessed with the CEHQ-FFQ apart from milk were only tested for relative validity by comparison with reported intakes obtained from two repeated 24-HDRs as described in the following Section 7.3 yet.

7.3 Relative validation in IDEFICS: SACINA vs. CEHQ-FFQ

The relative validity of the CEHQ-FFQ was further tested comparing the reported food consumption frequencies with food frequencies obtained from two non-consecutive SACINA interviews (Bel-Serrat et al., 2012). In case of the 24-HDRs, frequencies of intake were equalized to the number of reported portions per recall period. The study included $2,033$ children with complete CEHQ-FFQ, two repeated 24-HDRs and covariate information. In total, 37 food groups were investigated. Agreement between the two instruments was assessed using crude

and de-attenuated Pearson's correlation coefficients (cf. Section 3.3.2), cross-classification analyses and weighted kappa coefficients. Significant differences were observed for the majority of food group intakes estimated from the two methods, except for meat, meat replacement and soy products, and pizza. For most food groups, the CEHQ-FFQ provided higher mean consumption frequency estimates compared to the 24-HDRs except for vegetables, soft drinks, cold cuts, meat, white bread, pasta and rice, pizza, and sweets. De-attenuated Pearson correlations coefficients ranged from 0.02 (milled cereal) to 0.51 (water). The proportion of subjects classified within the same quartile for the CEHQ-FFQ and the 24-HDRs ranged from 27.4% (meat) to 36.3% (fruit), and was within the extreme opposite, i.e. either FFQ frequency classified in the first quartile and 24-HDR frequency in the fourth quartile or vice versa, in less than 12% for all food groups.

In summary, the observed level of agreement was rather low. As the CEHQ-FFQ queries usual consumption frequencies which are not reflected in only two 24-HDR assessments, this was more or less expected. Especially for episodically consumed foods like fish, olives, etc. two 24-HDRs may only provide a poor estimate of usual consumption frequencies. Furthermore, 24-HDRs do not primarily intend to assess consumption frequencies but portions of intake on single days. Equalizing portions reported in 24-HDRs with usual consumption frequencies is a strong limitation of this study which may further explain the low agreement. Moreover, both instruments are subject to measurement errors (cf. Section 2.1) that may have further affected the strength of the association.

7.4 Method of triads in HELENA

The method of triads, which has been described in Section 4.2.3 in detail, was recently applied in European adolescents in course of the HELENA study (Vandevijvere et al., 2012). In the triangular comparison, self-reported consumption frequencies of foods assessed by FFQs and mean food intakes reported in two repeated 24-HDRs were evaluated using different concentration biomarkers. For the evaluation of the food intake assessment, 390 adolescents were included.

First, unadjusted Spearman correlation coefficients were calculated between concentration biomarkers and mean food intakes (24-HDRs), between concentration biomarkers and consumption frequencies (FFQ) as well as between mean food intakes (24-HDRs) and consumption frequencies (FFQ) where the following foods and biomarkers were compared:

- Fruit intake vs. fruit consumption frequency vs. vitamin C status
- Vegetable intake vs. vegetable consumption frequency vs. vitamin C status
- Fruit intake vs. fruit consumption frequency vs. β-carotene status
- Vegetable intake vs. vegetable consumption frequency vs. β-carotene status
- Fish intake vs. fish consumption frequency vs. sum of eicosapentaenoic acid (EPA) and docosahexaenoic acid (DHA)

Highest correlations were observed between food consumption frequencies and mean food intakes (FFQ vs. 24-HDRs), except for the sum of EPA and DHA versus mean fish intake (biomarker vs. 24-HDRs). Correlations were higher between food consumption frequencies and concentration biomarkers (FFQ vs. biomarker) than between mean food intakes and concentration biomarkers (24-HDRs vs. biomarker), especially for the sum of EPA and DHA. In most cases, correlations were higher in girls than in boys. Exclusion of underreporters, which were defined based on the Goldberg cut-offs (cf. Section 5.1.2), resulted in a slight increase in the correlations. Furthermore, mean usual intakes were calculated based on the 24-HDR data using the Multiple Source Method (Haubrock et al., 2011) which corrects for the variance inflation caused by day-to-day variation incorporating FFQ information as covariate (cf. Sections 3.3.1 and 3.3.4). Correlations between mean usual food intakes and concentration biomarkers were higher compared to correlations between mean food intakes without variance correction and concentration biomarkers.

In a next step, the triads method was applied to evaluate the correlation between the three measurements (FFQ, biomarker and 24-HDRs) and the true intake by calculation of validity coefficients (Kaaks, 1997; Ocke & Kaaks, 1997; Yokota et al., 2010). For boys, the highest validity coefficients were found for the frequency of fruit consumption (0.88) and for the sum of DHA and EPA in relation to true intake (0.71). In girls, the highest validity coefficients were found for the frequency of fruit consumption (0.76), frequency of vegetable consumption (0.74), mean fruit intakes (0.90) as well as for the sum of DHA and EPA (0.69) in relation to true intake.

In summary, two 24-HDRs in combination with an FFQ seemed to be useful to rank subjects according to their usual food intakes. Correction for the inflated variance due to day-to-day variation as well as exclusion of underreporters turned out to be beneficial.

Chapter 8

Misreporting in IDEFICS

This chapter summarizes first results of the IDEFICS study concerning misreporting in proxy-reported dietary data and is mainly based on two of my articles that were recently published in the British Journal of Nutrition (Börnhorst et al., 2012b) and in Publish Health Nutrition (Börnhorst et al., 2012a). These articles are printed in Appendices B.3 and B.4.

8.1 Prevalence and determinants of misreporting

A recent review of validation studies based on the doubly labeled water technique in children by Burrows et al. (2010) revealed inconsistent results concerning misreporting. Underreporting ranged from 19% to 41% of reported EI and overreporting from 7% to 11% where data mainly relied on self-reports - partially with parental assistance. As young children lack the cognitive skills to complete dietary assessments (McPherson et al., 2000), 24-HDR data in young children mainly rely on proxy-reporters (cf. Section 2.2). Here, additional problems emerge from meals that are not under parental control, e.g. school meals, leading to unintentional misreporting (Baranowski et al., 1991; Basch et al., 1990; Eck et al., 1989). Whether the accuracy of proxy-reports is comparable to that of self-reports and whether determinants of misreporting coincide for self- vs. proxy-reports is yet unknown. Knowing the degree and direction of misreporting is essential for the assessment of diet-disease relationships as well as for the evaluation of dietary guidelines and nutrition policies. Therefore, the study shortly summarized here was conducted to investigate the prevalence and determinants of misreporting including under- and overreporting in proxy-reported 24-HDRs (Börnhorst et al., 2012b).

The analysis was based on $6,101$ children aged 2 to 9 years with a 24-HDR and complete covariate information. Basal metabolic rate (BMR) was estimated using

the equations published by Schofield (1985) and recommended by FAO/WHO/UNU (1985) taking into account age, sex, body height and weight. Since the original Goldberg cut-offs (see Section 5.1.2) were developed for adults without considering differences in EI due to age and sex, adaptations are required for application in children. Upper and lower cut-off values to identify plausible and implausible reports of EI were calculated substituting Goldberg's single cut-off for actual intake (Goldberg et al., 1991) by age and sex-specific cut-offs for children as suggested previously (McCrory et al., 2002; Sichert-Hellert et al., 1998) using Equation (5.1) on p. 66. The within-subject coefficient of variation for EI (CV_{wEI}), the within-subject coefficient of variation for BMR (CV_{wBMR}) and the coefficient of variation for physical activity (CV_{PA}) were replaced by age and sex-specific reference values as given in Nelson et al. (1989) and Black (2000). The number of days (d) was set to one as the analysis was based on one 24-HDR per child. Goldberg's overall level of 1.55 for physical activity was substituted by age and sex depending levels of light PA (2 to 5 years: 1.45, 6 to 9 years: males 1.55, females 1.50) according to Torun et al. (1996) and Torun (2005).

Backward elimination in course of multi-level logistic regression analyses was conducted to identify factors significantly related to under- and overreporting. Besides characteristics of the children and parents, social factors as well as data on parental concerns and perceptions of their child's weight status were considered. Furthermore, selective misreporting was addressed investigating food group intakes commonly perceived as more or less socially desirable. Proportions of under-, plausible and overreports were 8.0%, 88.6% and 3.4%, respectively. The risk of underreporting increased with age (odds ratio (OR) = 1.19, 95% confidence interval (CI) = $[1.05; 1.83]$), BMI z-score of the child (OR= 1.23, 95% CI=$[1.10; 1.379]$) and household size (OR= 1.12, 95% CI= $[1.01; 1.25]$) and was higher in low/medium income groups (OR= 1.45, 95% CI= $[1.13; 1.86]$). Overreporting was negatively associated with the BMI z-score of the child (OR= 0.78, 95% CI= $[0.69; 0.88]$) and higher in girls (OR= 1.70, 95% CI= $[1.27; 2.28]$). Also parental concerns and perceptions regarding their child's weight status were significantly associated with the reporting status.

Percentages of total EI from chocolate products, soft drinks and sugary products were negatively associated with underreporting, whereas percentages of EI from fruits/vegetables were positively associated with underreporting. This suggests that social desirability may have influenced the reporting behavior. The results indicated the presence of differential misreporting as e.g. underreporting was associated with higher BMI z-scores of the children and at the same time

with higher proxy-reported intakes of healthy foods like fruits and vegetables. This may result in misleading conclusions about size and direction of associations between diet and specific disorders like overweight and obesity (cf. Chapter 3). Unfortunately, it is not possible to exactly determine the magnitude of this problem with available methodologies. In addition, reverse causation cannot be precluded due to the cross-sectional study design. Nevertheless, future studies should try to involve these observed determinants of misreporting to account for differential reporting bias.

8.2 Methods to account for misreporting

Misreporting of dietary variables poses a challenge for epidemiologists when investigating associations between dietary intakes and health outcomes. Several studies including the one presented in the previous section revealed that misreporting is characteristic to specific subjects (Lioret et al., 2011; Livingstone & Black, 2003; Black & Cole, 2001) resulting in differential errors which can typically not be corrected (see Section 3.1.2). Standardized and validated dietary intake assessment methods may help to avoid misreporting at least to a certain degree. However, any remaining and unavoidable misreporting should be considered in the statistical analyses and respective results should be cautiously interpreted accounting for the potential effects of misreporting.
Various procedures have been proposed to screen out implausible dietary recalls (Goldberg et al., 1991; McCrory et al., 2002) but the question how to handle recalls identified as implausible is still open. Misreporting is rarely considered in the statistical analyses though there are different approaches that could be applied and proved for their usefulness (Nielsen & Adair, 2007; Huang et al., 2005):

(a) *Adjustment for covariables related to misreporting*
Inclusion of covariables in a statistical model is a common approach to adjust for confounders that may affect the association between an outcome and an exposure variable but little is known yet on the usefulness of covariate information to correct for misreporting. For example, known determinants of misreporting (cf. Section 8.1) could be included in statistical models to try to account for the misreporting.

(b) *Exclusion of implausible recalls*
The available dietary recalls are categorized in plausible (adequate-energy-reports) and implausible recalls (low-/high-energy-reports) according to any of the methods described in Section 5.1. The analysis is then run based only on the recalls classified as plausible. Though several studies found that exclusion of misreports strengthened diet-obesity relations and although this approach is often practiced (Livingstone & Black, 2003; Mendez et al., 2004; Howarth et al., 2005), data exclusions may introduce a source of unknown bias. It is likely that the characteristics of subjects with low-energy-reports differ from those with adequate-energy-reports or high-energy-reports such that data exclusions have not been recommended (Galli et al., 2005). Differences between the reporting groups with view to the study participants' characteristics were also confirmed by my results presented in Section 8.1.

(c) *Adjustment for the reporting group*
Adjusting for the reporting group, i.e. low-energy-report, adequate-energy-report, high-energy-report, by inclusion of respective dummy variables in the statistical model seems an appropriate alternative to data exclusions. The maximal power is maintained and this approach was shown to yield results that are consistent with those obtained from adequate-energy-reports after exclusion of low- and high-energy-reports (Mendez et al., 2004). However, it has the disadvantage that misclassifications of single recalls are quite likely which may again bias the results. Greenland & Robins (1985) demonstrated that the degree of misclassification needs to be considered when deciding whether to control for a covariable or not. Depending on the degree of misclassifications the bias after adjustment for the variable of interest may even be higher compared to the bias in an unadjusted model.

(d) *Stratified analysis by reporting group*
Another suggested option is to analyze the data stratified by the reporting group (Mendez et al., 2004). Agreement in the effect estimates between groups of low-, adequate- and high-energy-reports would confirm the results observed for the total study group. Furthermore, investigation of differences in the effect estimates between groups of low-, adequate- and high-energy-reports could be used as a part of the uncertainty evaluation (Poslusna et al., 2009). Observed differences may help to determine the degree of bias introduced through misreporting and may serve as an indication for the severeness of the misreporting problem. Furthermore, covariables that may

play different roles depending on the considered reporting group could be identified through this approach. A disadvantage of the stratified analysis is the loss of power.

(e) *Propensity score adjustment*
The propensity score is a common tool to reduce bias by equating groups based on selected covariables. A propensity score reflects the conditional probability of assignment to a particular group given a vector of observed covariables (Rosenbaum & Rubin, 1983). Construction of a propensity score based on variables previously found to be related to misreporting is another option to account for implausible recalls and was applied in this context for the first time in my study presented in the following paragraph. The choice of the relevant covariables is an important step when applying the propensity score approach and may require an in depth literature search as well as additional analyses, e.g. variable selection procedures like backward elimination.

(f) *Energy adjustment*
Different energy adjustment models have been proposed where the so-called 'density-model' is the one that is most often applied (Kipnis et al., 2003). Instead of using single food or nutrient intakes, the intakes are put in relation to total energy intake, e.g. by building the ratio of nutrient intake over total energy intake, and then included in the statistical model. Such approaches would be useful to correct for misreporting if all food items were affected to the same degree which is not likely to be the case (Lafay et al., 2000). Energy adjustment models cannot eliminate bias resulting from differential misreporting (Livingstone & Black, 2003), for example in case food items are misreported in relation to their energy content as commonly observed, i.e. overreporting of food items with low and underreporting of food items with high energy contents. Furthermore, energy intake itself is one of the most difficult parameters to measure accurately in nutrition research such that the energy adjustment may be based on imprecise measures of energy in many situations (Kipnis et al., 2001).

(g) *Adjustment using reference measures*
If reference information is available, e.g. biomarker measurements or data from a second, superior dietary assessment instrument, these can be either used as additional adjustment terms in the statistical model or for regression calibration (see also Section 3.3.1). Common calibration approaches assume

non-differential linear measurement error with constant variance or linear, random within-person error in case of replicate measurements like repeated 24-HDRs (Spiegelman et al., 1997; Freedman et al., 2011; Kaaks et al., 1995) - assumptions that are often violated due to differential misreporting (Black & Cole, 2001; Livingstone & Black, 2003). Moreover, error structures were found to be correlated when assessing dietary information using different instruments based on self-reports. Therefore, the usefulness of regression calibration to correct for misreporting is questionable unless a gold standard reference measure is available which is rarely the case.

In order to compare and evaluate these different approaches to account for misreporting in the statistical analysis, an exploratory study on the association between dietary intakes (total energy intake, percentage of EI from soft drinks, percentage of EI from fruits/vegetables) and overweight/obesity was conducted within the framework of the IDEFICS study (Börnhorst et al., 2012a; see also Appendix B.4).

In total, $5,357$ children who provided one 24-hour dietary recall and complete covariate information on age, sex of the child, net household income, number of persons below 18 in household, day of the interview as well as information on parental concerns and perceptions regarding their child's weight status were included. The covariables were selected based on the results of the study presented in Section 8.1. The 24-hour recalls were classified in three reporting groups based on child- and sex-specific cut-off values calculated according to Goldberg's equation (cf. Sections 5.1.2 and 8.1): low-energy-report, adequate-energy-report or high-energy-report. A basic logistic multi-level model was defined adjusting for age and sex and including the study center as random effect. In the following I will refer to this model as 'basic model'. This basic model was compared to models reflecting the above mentioned approaches (a) to (e) to account for misreporting. Instead of absolute intakes, percentages of total EI from fruits/vegetables as well as percentages of EI from soft drinks were included as exposure variables in the different models (cf. approach (f): energy adjustment). Approach (g) was not applicable due to the lack of appropriate reference information. A dummy variable indicating overweight/obesity according to Cole et al. (2000, 2007) was chosen as outcome variable in all models.

In the basic model the dietary exposures showed no significant association with overweight/obesity (energy intake: $OR = 0.996$, 95% CI= $[0.983; 1.010]$, soft drinks: OR= 0.999, 95% CI= $[0.986; 1.013]$) and revealed even a positive association for fruits/vegetables (OR= 1.009, 95%CI= $[1.001; 1.018]$). Adjustment for

covariables as described in approach (a) revealed similar results, but the association between fruits/vegetables and overweight/obesity was no longer significant (OR= 1.009, 95% CI= [0.998; 1.020]). When excluding low- and high-energy-reports (b), a significantly positive association between EI and overweight/obesity was observed (OR= 1.057, 95%CI=[1.038; 1.076]). Adjustment for the reporting group (c) also revealed a significantly positive association between EI and overweight/obesity that was even slightly more pronounced compared to the model excluding misreports. When adjusting for the propensity score (e), all associations were strengthened with the association between overweight/obesity and fruit/vegetable intake being reversed compared to the basic model. Significant positive associations were found between overweight/obesity and total EI as well as between overweight/obesity and soft drinks. When stratifying the basic model by the reporting group (d), EI was significantly associated with overweight/obesity in all three strata where the associations were even more pronounced in low- and in high-energy-reports compared to adequate-energy-reports. In conclusion, associations between dietary exposures and health outcomes were strongly affected or even masked by measurement errors. Consideration of the reporting group and inclusion of a propensity score for misreporting revealed results that were more consistent with expectations though the true effects remained unknown in the presented analyses due to the lack of reference measures. In the first instance, this analysis demonstrates that it should be acted with caution when modeling an association between diet and an health outcome. Depending on the selected model, results strongly differed where even reversed signs were observed. Therefore, researchers must attempt to identify the 'most correct' analytical strategy with regard to content as well as with regard to the statistical concept to avoid presenting biased effect estimates. Kushi (1992) also observed very differing results in a study on the association between fat intake and breast cancer when applying four different energy-adjustment methods. Such results should raise other researchers' awareness of the importance to carefully select the statistical model and to interpret the results with regard to the selected model incorporating potential limitations and effects of measurement errors.

The study presented in the previous paragraph compared approaches to account for misreporting in the absence of reference data. However, even validation data from a second dietary assessment instrument of the same kind were shown to be rather useless for the purpose to correct for misreporting as it is likely that study subjects that misreport in the first assesment may misreport in the second one as well (Fraser, 2003) which results in correlated error structures (Kipnis et al.,

2003). Although the use of two complementary dietary assessment methods is e.g. recommended when investigating usual intakes (Carroll et al., 2012; de Boer et al., 2011), the benefit of a second assessment instrument to correct for misreporting is questionable (Westerterp & Goris, 2002). Further research is required to investigate how to account for measurement errors resulting from differential misreporting in the absence of independent validation data such as biomarkers. To date, incorporation of validation data with independent error structures and additional incorporation of knowledge on determinants of misreporting based on the study participants' characteristics seem to be the most promising starting points to account for misreporting.

Chapter 9

Discussion and future perspectives

The studies presented in Chapters 7 and 8 confirmed the presence of severe measurement errors in proxy-reported dietary data among young populations. Although in a validation study based on a convenience sample good agreement was observed between measured energy expenditure and proxy-reported energy intake, large deviations were found at the individual level suggesting a large amount of random errors (cf. Section 7.2.1). Random errors may also explain the low agreement that was observed when comparing proxy-reported consumption frequencies obtained from the CEHQ-FFQ with the frequencies obtained from the 24-HDR SACINA. However, in any case, the comparison of different self- or proxy-reported assessment methods for the evaluation of a single instruments' validity is problematic as each has its own error structure (Beaton, 1994).
Furthermore, in the DLW validation study (cf. Section 7.2.1), results strongly differed when investigating a group of exclusively obese children suggesting subject-specific errors. The weight status should especially be taken into account when evaluating the validity of dietary data. On the one hand, obesity was found to be associated with intentional misreporting (Poslusna et al., 2009; Mendez et al., 2004; Poppitt et al., 1998), but on the other hand obesity status may actually lead to a change in the dietary behavior. For the latter reason, even small caloric intakes on single days may be valid e.g. due to dieting. However, such data do not reflect the participants' usual dietary intakes and hence not their actual exposures. This ambiguity complicates the evaluation of self- and proxy-reported dietary data in obese study subjects.
The presence of subject-specific errors was further confirmed when comparing groups of low-, adequate- and high-energy-reports defined based on adapted Goldberg cut-offs (cf. Section 8.1). Study participants' characteristics as well as the characteristics of the proxy-reporters strongly differed between the groups and therefore seemed to influence the reporting behavior. This led to the conclusion

that not only the participants' but also the proxy-reporters' characteristics should be considered as determinants for the validity of dietary data and for the degree of misreporting when data rely on proxy-reporters.
In addition, differences in the contribution of the different macronutrients to the total daily energy intake were observed between the reporting groups which may, amongst others, be explained by selective misreporting of single food items (Lafay et al., 2000). This is especially problematic if such differential reporting bias acts in opposite directions in diseased and non-diseased subjects. There is e.g. some evidence for opposite reporting biases in lean and obese study subjects (Beaton, 1994). This means, when analyzing associations between diet and weight status, the bias is likely to co-vary with the variable of analytical interest which may result in erroneous conclusions. Such differential reporting bias may also explain the contradictory effect estimates observed when exemplarily analyzing the association between overweight/obesity and fruit/vegetable consumption as described in Section 8.2. Without accounting for misreporting in the statistical model, a positive association was observed which was rendered non-significant or pointed even to the reversed direction (negative) when applying different approaches to account for misreporting. In that study, the various measurement errors intrinsic in the proxy-reported data resulted in a very confusing picture. In the absence of objective validation data, it seemed impossible to determine the true association with certainty. This is likely to apply also to other epidemiological studies investigating associations between diet and a health outcome based on self- or proxy-reported data.
Several studies may have serious methodological deficiencies that weaken the validity of their results. For instance, most statistical models applied to analyze an association between dietary intakes and a health outcome rely on the assumption of non-differential measurement errors (Freedman et al., 2011; Kaaks & Riboli, 1997; Kaaks et al., 1995; Spiegelman et al., 1997) which is likely to be violated in case of self- or proxy-reported dietary data. In addition, null effects are often not reported which may distort the overall presented picture of potential diet-disease associations. Even if null effects were published, it is unclear whether they reflect a true null effect or whether they are a consequence of random errors not adequately accounted for in the statistical analysis. The attenuation of measures of association caused by measurement errors in dietary data was shown to be high (cf. Section 3.3), especially for the FFQ, which is often the method of choice in large epidemiological studies due to its simple as well as cost- and time-efficient application (Kipnis et al., 2003). The evaluation of measurement

errors in FFQs and 24-HDRs within the framework of the OPEN biomarker study led Kipnis et al. (2003) even to the conclusion that "*the interpretation of findings from FFQ-based epidemiologic studies of diet-disease associations needs to be reevaluated*".

Negligence or inadequate consideration of measurement errors, but of course also non-consideration of important confounding factors may further explain the high number of inconsistencies and great uncertainty reported in studies on dietary exposures and health outcomes. For example, the association between different dietary factors and colorectal cancer was recently reviewed by Vargas & Thompson (2012). The authors ranked the level of evidence for an association between different dietary factors and colorectal cancer based on previous literature into the categories "Convincing", "Probable" and "Limited, suggestive" where the latter category was chosen most often, i.e. for vitamin D, dietary folate, animal fats, fish consumption, fruits and vegetables and selenium.

Nutritional epidemiology faces a variety of problems. Not only confounding and measurement errors but also the complexity of diet and its relationships with other factors makes it difficult to investigate associations between diet and health outcomes. In many situations, it is difficult to determine whether a specific food item, a specific nutrient or a combination of nutrients available in a specific food is the relevant factor in a detected association. Moreover, nutrient absorption is often influenced by other dietary factors, e.g. iron absorption is known to be reduced by coffee or tea consumption (Zijp et al., 2000). Therefore, investigating a single dietary component is almost infeasible as many interrelations need to be taken into account. Furthermore, lifestyle factors including alcohol consumption, smoking and physical activity as well as sex and age represent further factors that may interfere with the metabolism of dietary intakes. Apart from that, it is likely that the healthiness of a diet as a whole - not the consumption of single foods - may be the essential factor at least when considering diseases like obesity. It was hypothesized that 60 up to 130 questions on food are needed to characterize a person's diet (Fraser, 2003) where each single question is prone to measurement errors. Moreover, mutually intercorrelations exist between the consumption of different foods like bread is often consumed with butter or sausage with ketchup. All these factors add to the complexity of investigating diet as a risk factor such that epidemiological studies on nutritional factors and health based on self- or proxy-reported data should be interpreted with caution. To date, little is known with certainty about the complex associations between diet and health.

In nutritional studies, it is recommended to combine different self-report instru-

ments (Carroll et al., 2012) and in the optimal case to take objective biomarker information additionally into account (Freedman et al., 2011). However, studies including recovery biomarkers often rely on small, convenience study samples limiting the reliability of the results. One attempt to overcome this problem might be to pool different biomarker studies to gain knowledge on measurement error properties of common dietary assessment instruments. In addition, assessment of biomarker information preserves increasing interest in epidemiological studies and biomarker assessment methods become easier and cheaper in application. For this reason, it is likely that objective biomarker information may be available in more studies and in larger samples in future which may help to correct especially for systematic measurement errors in self- or proxy-reported dietary data. Although biomarker measurements may still include random or additive systematic errors e.g. due to random variations in the biological fluid or due to wrong adjustments of the laboratory equipment, respectively, these kinds of errors are better to handle compared to errors resulting from differential misreporting. In addition to this argument, the independence of the error structures of biomarker data compared to data relying on self- or proxy-reports is another advantage (Kaaks, 1997).

Future research

Diet is a special exposure as it is not optional. Hence, diet deserves special attention as a modifiable risk factor. To gain knowledge on the complex associations between diet and health, future research should focus on the understanding of error terms and consequently on the improvement of statistical estimation procedures to cope with these errors. Guidelines for an adequate data analysis and reporting of results with respect to the study population at hand would be desirable. This could help to harmonize different studies, to ease comparisons and to enhance the overall quality of nutritional studies.

Measurement error considerations should already be included when planning a study and not only when starting the data analysis. Future studies on nutrition should put more weight on the embedding of validation studies at least in subgroups. This enables the evaluation of the measurement error structure of self- or proxy-report instruments as well as the calculation of attenuation factors (cf. Section 3.3) which can then be used to correct the observed effect estimates. Depending on the purpose of the study and on the type of data assessed, there are different designs possible to obtain measurement error parameters, e.g. replicate measurements, instrumental variables as well as external or internal validation studies that measure the true exposure and a surrogate (Buzas et al., 2004). An

internal study based on recovery biomarkers will be the best choice in most cases though, of course, also cost-benefit aspects need to be taken into account. Next to the planning of validation studies, sample size calculations should consider the reduced power caused by random errors (Buzas et al., 2004).
In addition, it should be aimed to gain knowledge on the measurement error properties of the different dietary assessment instruments as well as to improve current instruments to prevent measurement errors already in an early stage. Recent developments focus on innovative technologies like PDA-, smartphone-, interactive computer-, web-, camera-, tape-recorder, scan- or sensor-based applications where the underlying methodologies remain mainly the same (Illner et al., 2012). The latter means that e.g in large epidemiological studies 24-HDRs or FFQs may still be used, but via smartphone- or web-applications instead of paper-pencil based. The main advantages of instruments based on innovative technologies lie in the cost- and time-efficiency, higher compliance in certain target groups and lower work-loads of data collection and entry (Illner et al., 2012). Also the assessment over a longer period of time, like e.g. higher numbers of repeated 24-HDRs, may be easier and cheaper to realize with these technologies.
However, improvement of self- or proxy-report dietary assessment instruments is limited and may only contribute to a limited degree when being faced with the problem of measurement errors in dietary data. These innovative technologies will presumably not resolve the problem of individual bias and will thus not provide accurate quantitative estimates of individual intakes (Illner et al., 2012). Although prospective assessments based on PDA or smartphone applications may improve the data quality through real-time recording at the time of the eating occasions, also reactivity will still pose a challenge in future. Nevertheless, newer technologies may be beneficial if the main interest lies in interventions and in the dissemination of dietary recommendations. These instruments allow for prompt feedback on the healthiness of a study participant's diet via email or SMS, which may make the study participants think about their diet and may help them changing their dietary behaviors. Therefore, newer prospective assessment instruments are especially advantageous in studies that seek for reactivity, i.e. for a long-term change in dietary behavior according to an implemented intervention program.
In case of retrospective assessments, measurement errors are likely to be similar and thus unchanged by the technology comparing conventional and recent assessment methods. This was at least suggested by a study comparing web-based and paper-pencil based FFQs (Illner et al., 2012). However, further research is needed to finally evaluate the strengths and weaknesses of innovative technologies

as knowledge is yet limited.
In any case, for the reasons mentioned above, it is questionable whether these innovative technologies will help to assess true associations between dietary intakes and health outcomes, such that future studies will still require additional, objective information on dietary intakes. Nutritional studies including objective biomarker information deliver the most trust-worthy results such that it would be desirable to strive for the exploration of further, yet unknown objective nutritional biomarkers, e.g. from hairs, nails or stool samples, where it would be ideal to have inexpensive reference measures as gold standard.
Another relevant factor for the validity of the results in studies on nutrition is the population under investigation. Sources of error differ depending on the given study population. Memory errors may pose a great challenge when the interest lies in the nutritional behavior of the elderly, whereas unintentional misreporting due to unobserved meals is a major source of errors in proxy-reported data in young populations as explained in Section 2.2. Enhancement of the statistical modeling procedures considering the specificities of the study population at hand is another topic for future research. This may e.g. include the improvement of statistical procedures to estimate usual dietary intakes in young children.
Established methods to estimate the distribution of usual dietary intakes were developed for adults (see Section 3.3.4). To date, no study investigated whether it is appropriate to apply these methods to data in young children that mainly rely on proxy-reports. This thesis led to the conclusion that the error structures differ between self- and proxy-reported data as also the sources of error differ to some degree. In addition, it is likely that the unintentional misreporting is a more serious problem in proxy-reports due to unobserved meals. This may result in underestimations of total intakes which may strongly bias estimates of usual intake distributions in young populations. The distributions will presumably be shifted to the left. It can be hypothesized that this bias is more severe compared to population distribution estimates based on self-reports. Although self-reports in adults were also shown to be strongly affected by misreporting, here the unintentional misreporting, e.g. caused by underestimations of portion sizes, may be less pronounced. Intentional misreporting on the other hand, that was shown to be associated with the study subjects' characteristics (Poslusna et al., 2009; Maurer et al., 2006; Mendez et al., 2004), is one of the main sources of measurement errors in data relying on self-reports (Westerterp & Goris, 2002). Nevertheless, the above results (see also Section 8.1) suggest that intentional misreporting is also a common problem in proxy-reports. Parental reports may be influenced

by social desirability or shame for their child's unhealthy diet or weight status. Little is known on how to handle this problem yet, but it is important to search for approaches to consider misreporting when estimating distributions of usual intakes or investigating associations between diet and health in young populations.

Furthermore, dietary intakes vary widely during childhood. Not only the mean dietary intake varies by age but also the variation in intake. Some studies based on DLW measurements showed higher intra-individual variability in energy intakes in children compared to adults and age-specific changes in the variance (Livingstone et al., 2004; Nelson et al., 1989; Nielsen et al., 2008). Current methods to estimate usual dietary intakes usually assume a constant intra-individual variance. In children, preferably age group-specific variance estimates should be used in the estimation of usual intake distributions. Of course, also studies on diet-disease associations need to consider the strong age-dependencies when analyzing data in young children.

This example demonstrates, that many factors including the specifics of the study population at hand need to be taken into account when analyzing dietary data and many open questions remain. Detection of a "true" association with absolute certainty is almost impossible based on current methodologies. Due to the complexity of diet and difficulties in the assessment, the adequate planning, analysis and interpretation of studies on diet and health outcomes remains a great challenge in nutritional epidemiology.

Bibliography

Ahrens, W., Bammann, K., de Henauw, S., Halford, J., Palou, A., Pigeot, I., Siani, A., & Sjöström, M. (2006). Understanding and preventing childhood obesity and related disorders – IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis*, *16*(4), 302–308.

Ahrens, W., Bammann, K., Siani, A., Buchecker, K., De Henauw, S., Iacoviello, L., Hebestreit, A., Krogh, V., Lissner, L., Marild, S., Molnar, D., Moreno, L., Pitsiladis, Y., Reisch, L., Tornaritis, M., Veidebaum, T., & Pigeot, I. (2011). The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond)*, *35 Suppl 1*, S3–15.

Alinia, S., Hels, O., & Tetens, I. (2009). The potential association between fruit intake and body weight – a review. *Obes Rev*, *10*(6), 639–647.

Andersen, L., Lioret, S., Brants, H., Kaik-Rak, A., de Boer, E., Amiano, P., & Trolle, E. (2011). Recommendations for a trans-European dietary assessment method in children between 4 and 14 years. *Eur J Clin Nutr*, *65*, S58–S64.

Atkinson, G. & Nevill, A. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Med*, *26*(4), 217–238.

Bammann, K., Sioen, I., Huybrechts, I., Casajus, J., Vicente-Rodriguez, R., Cuthill, R., Konstabel, K., Tubic, B., Wawro, N., Rayson, M., Westerterp, K., Marild, S., Pitsiladis, J., Reilly, J., Moreno, L., & de Henauw, S. (2011). The IDEFICS validation study on field methods for assessing physical activity and body composition in children: design and data collection. *Int J Obes (Lond)*, *35 Suppl 1*, S79–87.

Baranowski, T., Sprague, D., Baranowski, J., & Harrison, J. (1991). Accuracy of maternal dietary recall for preschool children. *J Am Diet Assoc*, *91*(6), 669–674.

Barron, B. (1977). The effects of misclassification on the estimation of relative risk. *Biometrics*, *33*(2), 414–418.

Basch, C., Shea, S., Arliss, R., Contento, I., Rips, J., Gutin, B., Irigoyen, M., & Zybert, P. (1990). Validation of mothers' reports of dietary intake by four to seven year-old children. *Am J Public Health*, *80*(11), 1314–1317.

Beaton, G. (1994). Approaches to analysis of dietary data: relationship between planned analyses and choice of methodology. *Am J Clin Nutr*, *59*, 253–261.

Beaton, G., Milner, J., Corey, P., McGuire, V., Cousins, M., Stewart, E., de Ramos, M., Hewitt, D., Grambsch, P., Kassim, N., & Little, J. (1979). Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am J Clin Nutr*, *32*(12), 2546–2559.

Bel-Serrat, S., Mouratidou, T., Pala, V., Huybrechts, I., Börnhorst, C., Fernández-Alvira, J., Hadjigeorgiou, C., Eiben, G., Hebestreit, A., Lissner, L., Molnár, D., Siani, A., Veidebaum, T., Krogh, V., & Moreno, L. (2012). Relative validity of the Children's Eating Habits Questionnaire among young European children: the IDEFICS study. *Public Health Nutr (accepted).*

Black, A. (2000). Critical evaluation of energy intake using the Goldberg cut-off for energy intake: basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes Relat Metab Disord*, *24*(9), 1119–1130.

Black, A. & Cole, T. (2001). Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. *J Am Diet Assoc*, *101*(1), 70–80.

Börnhorst, C., Huybrechts, I., Hebestreit, A., Vanaelst, B., Molnar, D., Bel-Serrat, S., Mouratidou, T., Moreno, L., Pala, V., Eha, M., Kourides, Y., Siani, A., Eiben, G., & Pigeot, I. (2012a). Diet-obesity associations in children - approaches to counteract attenuation caused by misreporting. *Public Health Nutr*, *Epub ahead of print.*

Börnhorst, C., Huybrechts, I., Ahrens, W., Eiben, G., Michels, N., Pala, V., Molnar, D., Russo, P., Barba, G., Bel-Serrat, S., Moreno, L., Papoutsou, S., Veidebaum, T., Loit, H., Lissner, L., & Pigeot, I. (2012b). Prevalence and determinants of misreporting among european children in proxy-reported 24-hour dietary recalls. *Br J Nutr*, *Epub ahead of print.*

Börnhorst, C., Bel-Serrat, S., Pigeot, I., Huybrechts, I., Ottevaere, C., Sioen, I., de Henauw, S., Mouratidou, T., Mensana, M., Westerterp, K., Bammann, K., Lissner, L., Eiben, G., Pala, V., Rayson, M., Krogh, V., & Moreno, L. (2012c). Validity of 24-hour recalls in (pre-)school aged children: Comparison of proxy-reported energy intakes with measured energy expenditure. *(under review)*.

Burrows, T., Martin, R., & Collins, C. (2010). A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc*, *110*(10), 1501–1510.

Buzas, J., Stefanski, L., & Tosteson, T. (2004). Measurement Error. In W. Ahrens & I. Pigeot (Eds.), *Handbook of Epidemiology* (pp. 730–765). Heidelberg: Springer.

Cade, J., Thompson, R., Burley, V., & Warm, D. (2002). Development, validation and utilisation of food-frequency questionnaires - a review. *Public Health Nutr*, *5*(4), 567–587.

Carriquiry, A. (1999). Assessing the prevalence of nutrient inadequacy. *Public Health Nutr*, *2*(1), 23–33.

Carroll, R. (1998). Measurement error in epidemiological studies. In P. Armitage & T. Colton (Eds.), *Encyclopedia of biostatistics*, volume 3 (pp. 2491–2519). Chichester England: John Wiley & Suns.

Carroll, R., Freedman, L., & Kipnis, V. (1998). Measurement error and dietary intake. *Adv Exp Med Biol*, *445*, 139–145.

Carroll, R., Gallo, P., & Gleser, L. (1985). Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. *J Am Stat As*, *80*(392), 929–932.

Carroll, R., Midthune, D., Subar, A., Shumakovich, M., Freedman, L., Thompson, F., & Kipnis, V. (2012). Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol*, *Epub ahead of print*.

Carroll, R., Ruppert, D., & Stefanski, L. (1995). *Measurement error in Nonlinear Models*. London: Chapman & Hall.

Carroll, R. & Stefanski, L. (1994). Measurement error, instrumental variables and corrections for attenuation with applications to meta-analyses. *Stat Med*, *13*(12), 1265–1282.

Cole, T., Bellizzi, M., Flegal, K., & Dietz, W. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*, *320*(7244), 1240–1243.

Cole, T., Flegal, K., Nicholls, D., & Jackson, A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ*, *335*(7612), 194.

Currie, C. (1998). Health behaviour in school-aged children: a WHO cross-national survey (HBSC). Research protocol for the 1997/1998 survey. *Edinburgh, Research Unit in Health and Behavioural Change, University of Edinburgh.*

de Boer, E., Slimani, N., van't Veer, P., Boeing, H., Feinberg, M., Leclercq, C., Trolle, E., Amiano, P., Andersen, L., Freisling, H., Geelen, A., Harttig, U., Huybrechts, I., Kaic-Rak, A., Lafay, L., Lillegaard, I., Ruprich, J., de Vries, J., & Ocke, M. (2011). The European Food Consumption Validation Project: conclusions and recommendations. *Eur J Clin Nutr*, *65 Suppl 1*, S102–S107.

Dekkers, A. (2009). Statistical program for analysis of dietary exposure (SPADE). *Abstract, EFCOVAL Closing Conference, September 2009. Utrecht, The Netherlands.*

Diamond, E. & Lilienfeld, A. (1962). Effects of errors in classification and diagnosis in various types of epidemiological-studies. *Am J Public Health Nations Health*, *52*(7), 1137–1144.

Dixon, L., Subar, A., Wideroff, L., Thompson, F., Kahle, L., & Potischmann, N. (2006). Carotenoid and tocopherol estimates from the NCI diet history questionnaire are valid compared with multiple recalls serum biomarkers. *J Nutr*, *136*(12), 3054–61.

Dodd, K., Guenther, P., Freedman, L., Subar, A., Kipnis, V., Midthune, D., Tooze, J., & Krebs-Smith, S. (2006). Statistical methods for estimating usual intake of nutrients and foods: a review of the theory. *J Am Diet Assoc*, *106*(10), 1640–1650.

Eck, L., Klesges, R., & Hanson, C. (1989). Recall of a child's intake from one meal: are parents accurate? *J Am Diet Assoc*, *89*(6), 784–789.

Ezzati, M. & Riboli, E. (2012). Can noncommunicable diseases be prevented? Lessons from studies of populations and individuals. *J Am Diet Assoc*, *337*(6101), 1482–7.

FAO/WHO/UNU (1985). Energy and Protein Requirements. Report of a Joint FAO/WHO/UNU Expert Consultation. World Health Organization. *Technical Report Series 724*.

Forrestal, S. (2011). Energy intake misreporting among children and adolescents: a literature review. *Matern Child Nutr*, *7*(2), 112–127.

Fraser, G. (2003). A search for truth in dietary epidemiology. *Am J Clin Nutr*, *78 (Suppl)*, 521S–5.

Fraser, G. & Shavlik, D. (2004). Correlations between estimated and true dietary intakes. *Ann Epidemiol*, *14*(4), 287–295.

Freedman, L., Guenther, P., Dodd, K., Krebs-Smith, S., & Midthune, D. (2010). The population distribution of ratios of usual intakes of dietary components that are consumed every day can be estimated from repeated 24-hour recalls. *J Nutr*, *140*(1), 111–116.

Freedman, L., Midthune, D., Carroll, R., Tasevska, N., Schatzkin, A., Mares, J., Tinker, L., Potischman, N., & Kipnis, V. (2011). Using regression calibration equations that combine self-reported intake and biomarker measures to obtain unbiased estimates and more powerful tests of dietary associations. *Am J Epidemiol*, *174*(11), 1238–1245.

Galli, C., Rise, P., Ghezzi, S., & Marangoni, F. (2005). *Principles of Nutritional Assessment* (2 ed.). New York: Oxford University Press.

Goldberg, G., Black, A., Jebb, S., Cole, T., Murgatroyd, P., Coward, W., & Prentice, A. (1991). Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr*, *45*(12), 569–581.

Greenland, S. (1980). The effect of misclassification in the presence of covariates. *Am J Epidemiol*, *112*(4), 564–569.

Greenland, S. & Robins, J. (1985). Confounding and misclassification. *Am J Epidemiol*, *122*(3), 495–504.

Haubrock, J., Nöthlings, U., Volatier, J., Dekkers, A., Ocke, M., Harttig, U., Illner, A., Knuppel, S., Andersen, L., & Boeing, H. (2011). Estimating usual food intake distributions by using the multiple source method in the EPIC-Potsdam Calibration Study. *J Nutr*, *141*(5), 914–920.

Hoffmann, K., Boeing, H., Dufour, A., Volatier, J., Telman, J., Virtanen, M., Becker, W., & De Henauw, S. (2002). Estimating the distribution of usual dietary intake by short-term measurements. *Eur J Clin Nutr*, *56 Suppl 2*, S53–S62.

Howarth, N., Huang, T., Roberts, S., & McCrory, M. (2005). Dietary fiber and fat are associated with excess weight in young and middle-aged US adults. *J Am Diet Assoc*, *105*(9), 1365–1372.

Huang, T., Roberts, S., Howarth, N., & McCrory, M. (2005). Effect of screening out implausible energy intake reports on relationships between diet and BMI. *Obes Res*, *13*(7), 1205–1217.

Huybrechts, I., Börnhorst, C., Pala, V., Moreno, L., Barba, G., Lissner, L., Fraterman, A., Veidebaum, T., Hebestreit, A., Sieri, S., Ottevaere, C., Tornaritis, M., Molnar, D., Ahrens, W., & De Henauw, S. (2011). Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children. *Int J Obes (Lond)*, *35 Suppl 1*, S69–S78.

Illner, A.-K., Freisling, H., Boeing, H., Huybrechts, I., Crispim, S., & Slimani, N. (2012). Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *Int J Epi*, *41*(4), 1187–1203.

Ireland, J., van Erp-Baart, A., Charrondiere, U., Moller, A., Smithers, G., & Trichopoulou, A. (2002). Selection of a food classification system and a food composition database for future food consumption surveys. *Eur J Clin Nutr*, *56 Suppl 2*, S33–S45.

Jahns, L., Arab, L., Carriquiry, A., & Popkin, B. (2005). The use of external within-person variance estimates to adjust nutrient intake distributions over time and across populations. *Public Health Nutr*, *8*(1), 69–76.

Kaaks, R. (1997). Biochemical markers as additional measurements in studies of the accuracy of dietary questionnaire measurements: conceptual issues. *Am J Clin Nutr*, *65*(4 Suppl), 1232S–1239S.

Kaaks, R., Ferrari, P., Ciampi, A., Plummer, M., & Riboli, E. (2002). Uses and limitations of statistical accounting for random error correlations in the validation of dietary questionnaire assessments. *Public Health Nutr*, *5*(6A), 969–976.

Kaaks, R. & Riboli, E. (1997). Validation and calibration of dietary intake measurements in the EPIC project: methodological considerations. European Prospective Investigation into Cancer and Nutrition. *Int J Epidemiol*, *26 Suppl 1*, S15–S25.

Kaaks, R., Riboli, E., & van Staveren, W. (1995). Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol*, *142*(5), 548–556.

Kipnis, V., Freedman, L., Brown, C., Hartman, A., Schatzkin, A., & Wacholder, S. (2001). Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol*, *153*(4), 394–403.

Kipnis, V., Midthune, D., Buckman, D., Dodd, K., Guenther, P., Krebs-Smith, S., Subar, A., Tooze, J., Carroll, R., & Freedman, L. (2009). Modeling data with excess zeros and measurement error: application to evaluating relationships between episodically consumed foods and health outcomes. *Biometrics*, *65*(4), 1003–1010.

Kipnis, V., Midthune, D., Freedman, L., Bingham, S., Day, E. D., Riboli, E., Ferrari, P., & Carroll, R. (2002). Bias in dietary-report instruments and its implications for nutritional epidemiology. *Public Health Nutr*, *5*(6A), 915–923.

Kipnis, V., Midthune, D., Freedman, L., Bingham, S., Schatzkin, A., Subar, A., & Carroll, R. (1997). Empirical evidence of correlated biases in dietary assessment instruments and its implications. *Am J Epidemiol*, *146*(10), 842–855.

Kipnis, V., Subar, A., Midthune, D., Freedman, L., Ballard-Barbash, R., Troiano, R., Bingham, S., Schoeller, D., Schatzkin, A., & Carroll, R. (2003). Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*, *158*(1), 14–21.

Kupper, L. L. (1984). Effects of the use of unreliable surrogate variables on the validity of epidemiologic research studies. *Am J Epidemiol*, *120*(4), 643–648.

Kushi, L. H. (1992). Dietary fat and postmenopausal breast cancer. *J Natl Cancer Inst*, *84*(14), 1091–1099.

Lafay, L., Mennen, L., Basdevant, A., Charles, M., Borys, J., Eschwege, E., & Romon, M. (2000). Does energy intake underreporting involve all kinds of food or only specific food items - Results from the Fleurbaix Laventie Ville Sante (FLVS) study. *Int J Obes Relat Metab Disord*, *24*(11), 1500–1506.

Lanfer, A., Hebestreit, A., Ahrens, W., Krogh, V., Sieri, S., Lissner, L., Eiben, G., Siani, A., Huybrechts, I., Loit, H., Papoutsou, S., Kovacs, E., & Pala, V. (2011). Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study. *Int J Obes (Lond)*, *35 Suppl 1*, S61–S68.

Lioret, S., Touvier, M., Balin, M., Huybrechts, I., Dubuisson, C., Dufour, A., Bertin, M., Maire, B., & Lafay, L. (2011). Characteristics of energy underreporting in children and adolescents. *Br J Nutr*, *105*(11), 1671–80.

Livingstone, M. & Black, A. (2003). Markers of the validity of reported energy intake. *J Nutr*, *133 Suppl 3*, 895S–920S.

Livingstone, M., Prentice, A., Coward, W., Strain, J., Black, A., Davies, P., Stewart, C., McKenna, P., & Whitehead, R. (1992). Validation of estimates of energy intake by weighed dietary record and diet history in children and adolescents. *Am J Clin Nutr*, *56*(1), 29–35.

Livingstone, M. & Robson, P. (2000). Measurement of dietary intake in children. *Proc Nutr Soc*, *59*(2), 279–293.

Livingstone, M., Robson, P., & Wallace, J. (2004). Issues in dietary intake assessment of children and adolescents. *Br J Nutr*, *92 Suppl 2*, S213–S222.

Maurer, J., Taren, D., Teixeira, P., Thomson, C., Lohman, T., Going, S., & Houtkooper, L. (2006). The psychosocial and behavioral characteristics related to energy misreporting. *Nutr Rev*, *64*(2 Pt 1), 53–66.

McCrory, M., Hajduk, C., & Roberts, S. (2002). Procedures for screening out inaccurate reports of dietary energy intake. *Public Health Nutr*, *5*(6A), 873–882.

McKeown, N., Day, N., Welch, A., Runswick, S., Luben, R., Mulligan, A., McTaggart, A., & Bingham, S. (2001). Use of biological markers to validate

self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr*, *74*(2), 188–196.

McNaughton, S., Marks, G., Gaffney, P., Williams, G., & Green, A. (2005). Validation of a food-frequency questionnaire assessment of carotenoid and vitamin E intake using weighed food records and plasma biomarkers: the method of triads. *Eur J Clin Nutr*, *59*(2), 211–208.

McPherson, R., Hoelscher, D., Alexander, M., Scanlon, R., & Serdula, M. (2000). Dietary assessment methods among school-aged children: Validity and reliability. *Preventive Medicine*, *31*(2), 11–33.

Mendez, M., Wynter, S., Wilks, R., & Forrester, T. (2004). Under- and overreporting of energy is related to obesity, lifestyle factors and food group intakes in Jamaican adults. *Public Health Nutr*, *7*(1), 9–19.

Moreno, L., De Henauw, S., González-Gross, M., Kersting, M., Molnár, D., Gottrand, F., Barrios, L., Sjöström, M., Manios, Y., Gilbert, C., & et, a. (2008). Design and implementation of the Healthy Lifestyle in Europe by Nutrition in Adolescence Cross-Sectional Study. *Int J Obes (Lond)*, *32 Suppl*(5), 4–11.

Moreno, L., González-Gross, M., Kersting, M., Molnár, D., de Henauw, S., Beghin, L., Sjöström, M., Hagströmer, M., Manios, Y., Gilbert, C., & et, a. (2008b). Assessing, understanding and modifying nutritional status, eating habits and physical activity in European adolescents: The HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) Study. *Pub Health Nutr*, *11*(3), 288–299.

Murakami, K., Miyake, Y., Sasaki, S., Tanaka, K., & Arakawa, M. (2011). Characteristics of under- and over-reporters of energy intake among Japanese children and adolescents: The Ryukyus Child Health Study. *Nutrition*, *28*(5), 532–38.

Nelson, M., Black, A., Morris, J., & Cole, T. (1989). Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr*, *50*(1), 155–167.

Nielsen, S. & Adair, L. (2007). An alternative to dietary data exclusions. *J Am Diet Assoc*, *107*(5), 792–799.

Nielsen, S., Montgomery, C., Kelly, L., Jackson, D., & Reilly, J. (2008). Energy intake variability in free-living young children. *Arch Dis Child*, *93*(11), 971–973.

Ocke, M. & Kaaks, R. (1997). Biochemical markers as additional measurements in dietary validity studies: application of the method of triads with examples from the European Prospective Investigation into Cancer and Nutrition. *Am J Clin Nutr*, *65*(4 Suppl), 1240S–1245S.

Paeratakul, S., Popkin, B., Kohlmeier, L., Hertz-Picciotto, I., Guo, X., & Edwards, L. (1998). Measurement error in dietary data: implications for the epidemiologic study of the diet-disease relationship. *Eur J Clin Nutr*, *52*(10), 722–727.

Poppitt, S., Swann, D., Black, A., & Prentice, A. (1998). Assessment of selective under-reporting of food intake by both obese and non-obese women in a metabolic facility. *Int J Obes Relat Metab Disord*, *22*(4), 303–311.

Poslusna, K., Ruprich, J., de Vries, J., Jakubikova, M., & van't Veer, P. (2009). Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr*, *101 Suppl 2*, S73–S85.

Rosenbaum, P. & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55.

Rosner, B., Spiegelman, D., & Willett, W. (1990). Correction of logistic regression relative risk estimates and confidence intervals for measurement error: the case of multiple covariates measured with error. *Am J Epidemiol*, *132*(4), 734–745.

Rosner, B., Spiegelman, D., & Willett, W. (1992). Correction of logistic regression relative risk estimates and confidence intervals for random within-person measurement error. *Am J Epidemiol*, *136*(11), 1400–1413.

Rosner, B. & Willett, W. (1988). Interval estimates for correlation coefficients corrected for within-person variation: implications for study design and hypothesis testing. *Am J Epidemiol*, *127*(2), 377–386.

Rosner, B., Willett, W., & Spiegelman, D. (1989). Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Stat Med*, *8*(9), 1051–1069.

Rutishauser, I. (2005). Dietary intake measurements. *Public Health Nutr*, *8*(7A), 1100–1107.

Schoeller, D. (2002). Validation of habitual energy intake. *Public Health Nutr*, *5*(Suppl 6a), 883–888.

Schoeller, D. & van Santen, E. (1982). Measurement of energy expenditure in humans by doubly labeled water method. *J Appl Physiol*, *53*(4), 955–959.

Schofield, W. (1985). Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr*, *39 Suppl 1*, 5–41.

Sempos, C., Briefel, R., Johnson, C., & Woteki, C. (1992). Process and rationale for selecting dietary methods for NHANES III. *Vital Health Stat 4*, (27), 85–90.

Sichert-Hellert, W., Kersting, M., & Schoch, G. (1998). Underreporting of energy intake in 1 to 18 year old German children and adolescents. *Z Ernährungswiss*, *37*(3), 242–251.

Souverein, O., Dekkers, A., Geelen, A., Haubrock, J., de Vries, J., Ocke, M., Harttig, U., Boeing, H., & van't Veer, P. (2011). Comparing four methods to estimate usual intake distributions. *Eur J Clin Nutr*, *65 Suppl 1*, S92–101.

Spiegelman, D., McDermott, A., & Rosner, B. (1997). Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr*, *65*(4 Suppl), 1179S–1186S.

Thompson, F. & Subar, A. (2012). Dietary assessment methodology. In A. Coulson, C. Boushey, & M. Ferruzi (Eds.), *Nutrition in the prevention and treatment of disease*, volume 3 (pp. 3–38). Academic Press Inc.

Tooze, J., Kipnis, V., Buckman, D., Carroll, R., Freedman, L., Guenther, P., Krebs-Smith, S., Subar, A., & Dodd, K. (2010). A mixed-effects model approach for estimating the distribution of usual intake of nutrients: the NCI method. *Stat Med*, *29*(27), 2857–2868.

Tooze, J., Midthune, D., Dodd, K., Freedman, L., Krebs-Smith, S., Subar, A., Guenther, P., Carroll, R., & Kipnis, V. (2006). A new statistical method for estimating the usual intake of episodically consumed foods with application to their distribution. *J Am Diet Assoc*, *106*(10), 1575–1587.

Torun, B. (2005). Energy requirements of children and adolescents. *Public Health Nutr*, *8*(7A), 968–993.

Torun, B., Davies, P., Livingstone, M., Paolisso, M., Sackett, R., & Spurr, G. (1996). Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. *Eur J Clin Nutr*, *50 Suppl 1*, S37–S80.

Trabulsi, J. & Schoeller, D. (2001). Evaluation of dietary assessment instruments against doubly labeled water, a biomarker of habitual energy intake. *Am J Physiol Endocrinol Metab*, *281*(5), 891–899.

Vandevijvere, S., Geelen, A., Gonzalez-Gross, M., Van't Veer, P., Dallongeville, J., Mouratidou, T., Dekkers, A., Börnhorst, C., & et, a. (2012). Evaluation of food and nutrient intake assessment using concentration biomarkers in European adolescents from the Healthy Lifestyle in Europe by Nutrition in Adolescence study. *Br J Nutr*, *Epub ahead of print*.

Vargas, A. & Thompson, P. (2012). Diet and nutrient factors in colorectal cancer risk. *Nutrition in Clinical Practice*, *27*(5), 613–623.

Vereecken, C., Covents, M., Matthys, C., & Maes, L. (2005). Young adolescents' nutrition assessment on computer (YANA-C). *Eur J Clin Nutr*, *59*(5), 658–667.

Westerterp, K. & Goris, A. (2002). Validity of the assessment of dietary intake: problems of misreporting. *Curr Opin Clin Nutr Metab Care*, *5*(5), 489–493.

Westerterp, K., Wouters, L., & van Marken Lichtenbelt, W. (1995). The Maastricht protocol for the measurement of body composition and energy expenditure with labeled water. *Obes Res*, *3 Suppl 1*, 49–57.

Willett, W. (1995). Dietary assessment methods. *Br J Nutr*, *74*(1), 141–143.

Willett, W. (1998). *Nutritional Epidemiology* (2 ed.). New York: Oxford University Press.

Yanetz, R., Kipnis, V., Carroll, R., Dodd, K., Subar, A., Schatzkin, A., & Freedman, L. (2008). Using biomarker data to adjust estimates of the distribution of usual intakes for misreporting: application to energy intake in the US population. *J Am Diet Assoc*, *108*(3), 455–464.

Yokota, R., Miyazaki, E., & Ito, M. (2010). Applying the triads method in the validation of dietary intake using biomarkers. *Cad Saude Publica*, *26*(11), 2027–2037.

Zijp, I., Korver, O., & Tijburg, L. (2000). Effect of tea and other dietary factors on iron absorption. *Crit Rev Food Sci Nutr*, *40*(5), 371–98.

Appendix A

Tables

Table A.1: Mean energy intake (kcal/day) by weekday and study center including only complete interview days.

	Energy in kcal									
Assessment day	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Monday	1987	1546 (542)	409	1766 (532)	377	1654 (482)	133	1290 (459)	85	1367 (487)
Tuesday	2175	1520 (542)	398	1793 (524)	305	1735 (503)	317	1320 (439)	74	1357 (474)
Wednesday	1829	1507 (514)	402	1717 (495)	293	1645 (516)	325	1275 (444)	92	1310 (497)
Thursday	1688	1579 (542)	356	1804 (520)	337	1754 (555)	255	1329 (448)	78	1245 (405)
Friday	718	1258 (558)	0		11	1469 (585)	57	1389 (433)	3	1920 (162)
Saturday	54	1306 (605)	0		5	1196 (718)	18	1156 (550)	0	
Sunday	1323	1565 (568)	410	1767 (617)	7	1263 (621)	110	1228 (391)	53	1446 (464)
	Sweden		Germany		Hungary		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
Monday	220	1558 (476)	389	1498 (583)	225	1237 (499)	149	1580 (503)		
Tuesday	261	1484 (398)	370	1487 (580)	325	1226 (501)	125	1658 (568)		
Wednesday	235	1538 (426)	317	1493 (538)	72	1016 (331)	93	1524 (414)		
Thursday	241	1526 (462)	309	1485 (573)	47	1477 (525)	65	1543 (417)		
Friday	11	1598 (634)	97	1632 (665)	539	1162 (510)	0			
Saturday	1	1000	1	1761	29	1413 (630)	0			
Sunday	257	1431 (477)	317	1466 (529)	9	1884 (304)	160	1728 (557)		

Table A.2: Mean energy intake (kcal/day) comparing working days and weekend days by study center including only complete interview days.

	Energy in kcal									
Assessment day	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Working day	7679	1537 (536)	1565	1769 (518)	1312	1696 (515)	1030	1304 (445)	329	1320 (469)
Weekend (Fr/Sa/Su)	2095	1453 (584)	410	1767 (617)	23	1347 (608)	185	1271 (427)	56	1471 (465)
	Sweden		Germany		Hungary		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
Working day	957	1525 (440)	1385	1491 (569)	669	1225 (495)	432	1585 (495)		
Weekend (Fr/Sa/Su)	269	1436 (484)	415	1505 (567)	577	1186 (523)	160	1728 (557)		

Table A.3: Mean energy intake (kcal/day) by study center comparing days with and without school meals including only complete interview days.

	Energy in kcal									
Assessment day	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
No school meal*	5989	1519 (555)	1603	1757 (559)	173	1481 (590)	1215	1299 (443)	350	1349 (471)
With school meal	3785	1519 (536)	372	1817 (449)	1162	1721 (500)	0		35	1276 (467)
	Sweden		Germany		Hungary		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
No school meal*	446	1399 (453)	1665	1511 (575)	171	1290 (510)	366	1673 (518)		
With school meal	780	1566 (440)	135	1292 (444)	1075	1194 (507)	226	1545 (503)		

*Days without school meal relate either to weekend days or to days without a meal at school but do not include interviews with missing school meals.

Table A.4: Mean energy intake (kcal/day) by study center and interview day including only complete interview days.

	Energy in kcal									
Counter - day	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
1st 24-HDR	9774	1519 (548)	1975	1769 (540)	1335	1690 (519)	1215	1299 (443)	385	1342 (471)
2nd 24-HDR	2649	1482 (530)	395	1768 (531)	17	749 (549)	41	1345 (682)	18	1426 (327)
3rd 24-HDR	1276	1390 (510)	2	1104 (1320)	0		0		0	
4th 24-HDR	52	1276 (530)	0		0		0		0	
5th 24-HDR	14	1438 (581)	0		0		0		0	
6th 24-HDR	8	2108 (928)	0		0		0		0	
	Sweden		Germany		Hungary*		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
1st 24-HDR	1226	1505 (452)	1800	1494 (569)	1246	1207 (508)	592	1624 (516)		
2nd 24-HDR	90	1457 (396)	395	1503 (514)	1483	1376 (493)	210	1753 (519)		
3rd 24-HDR	91	1482 (422)	93	1565 (544)	1088	1367 (509)	2	1867 (201)		
4th 24-HDR	3	1304 (334)	6	1423 (423)	43	1254 (557)	0			
5th 24-HDR	0		3	1498 (182)	11	1422 (657)	0			
6th 24-HDR	0		0		8	2108 (928)	0			

*A first 24-HDR was assessed in 2003 children from Hungary and a second one in 1836 children. This table presents only data from children with complete dietary intake information, i.e. in case of missing information on single meals like school meals the interview was excluded. Due to higher percentages of Hungarian 24-HDRs with missing information in the first assessment, the number of children with complete second 24-HDR is higher here compared to the number of children with complete first 24-HDR.

Table A.5: Comparison of energy intakes between complete 24-HDR interviews and interviews with missing information.

	Energy in kcal									
	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
Complete 24-HDRs*	9984	1508 (551)	1994	1761 (544)	1360	1678 (526)	1215	1299 (443)	392	1334 (476)
Only school meal	1004	415 (281)	2	884 (248)	126	567 (378)	0		200	288 (157)
Missing school meal	673	868 (459)	0		14	1461 (820)	3	978 (303)	76	964 (398)
	Sweden		Germany		Hungary		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
Complete 24-HDRs*	1280	1474 (473)	1862	1483 (569)	1287	1204 (509)	594	1622 (516)		
Only school meal	7	836 (358)	13	395 (266)	524	407 (261)	102	355 (122)		
Missing school meal	257	708 (372)	100	1167 (498)	192	784 (408)	31	1234 (270)		

*A classification as complete interview means that school meal information is available if the child ate at school. These interviews may still be incomplete due to missing information caused by memory errors, unobserved meals, etc.

Table A.6: Number of mainmeals (breakfast, lunch, dinner) that parents did not observe or remember and mean energy intakes.

	Energy in kcal									
Parents did not observe...	Total		Italy		Estonia		Cyprus		Belgium	
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)
0 meals*	9794	1467 (566)	1975	1769 (540)	744	1679 (563)	1218	1298 (442)	461	1280 (480)
1 meal, but data**	626	1681 (475)	0		605	1699 (469)	0		0	
1 meal (no data)	208	1001 (488)	19	971 (333)	23	1043 (486)	0		7	889 (585)
2 meals (no data)	24	497 (216)	0		2	565 (274)	0		0	
3 meals (no data)	5	220 (69)	0		0		0		0	
	Sweden		Germany		Hungary		Spain			
	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)	N	Mean (SD)		
0 meals*	1446	1384 (523)	1896	1478 (570)	1431	1153 (516)	623	1604 (513)		
1 meal, but data**	15	1227 (279)	3	953 (312)	3	1054 (510)	0			
1 meal (no data)	56	730 (410)	61	1176 (459)	40	1121 (535)	2	1071 (404)		
2 meals (no data)	15	497 (233)	2	548 (341)	5	451 (156)	0			
3 meals (no data)	5	220 (69)	0		0		0			

*These numbers may still include interviews with missing school meal information as only the completeness of parental reports was considered here.
**Parents indicated that they did not observe a meal but either made a guess on the respective intakes or this meal was assessed in course of the school meal assessment

Appendix B

Publications

B.1 Evaluation of the CEHQ used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children

This article describes the biomarker-based validation of the CEHQ-FFQ which was published in the peer-reviewed *International Journal of Obesity* (Huybrechts et al., 2011).

International Journal of Obesity (2011) 35, S69–S78

0307-0565/11
www.nature.com/ijo

ORIGINAL ARTICLE

Evaluation of the Children's Eating Habits Questionnaire used in the IDEFICS study by relating urinary calcium and potassium to milk consumption frequencies among European children

I Huybrechts[1], C Börnhorst[2], V Pala[3], LA Moreno[4], G Barba[5], L Lissner[6], A Fraterman[7], T Veidebaum[8], A Hebestreit[2], S Sieri[3], C Ottevaere[1], M Tornaritis[9], D Molnár[10], W Ahrens[2] and S De Henauw[1,11], on behalf of the IDEFICS Consortium

[1]*Department of Public Health, Ghent University, Ghent, Belgium;* [2]*Bremen Institute for Prevention Research and Social Medicine (BIPS), University of Bremen, Bremen, Germany;* [3]*Department of Preventive and Predictive Medicine, Nutritional Epidemiology Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy;* [4]*Growth, Exercise, Nutrition and Development Research Group, EU Ciencias de la Salud, Universidad de Zaragoza, Zaragoza, Spain;* [5]*Unit of Epidemiology and Population Genetics, Institute of Food Science, National Research Council, Avellino, Italy;* [6]*Department of Public Health and Community Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden;* [7]*Medizinisches Versorgungszentrum Dr Eberhard &Partner Dortmund, Laboratoriumsmedizin Dortmund, Dortmund, Germany;* [8]*Department of Chronic Diseases, National Institute for Health Development, Tallinn, Estonia;* [9]*Research and Education Institute of Child Health, Strovolos, Cyprus;* [10]*Department of Paediatrics, University of Pécs, Pécs, Hungary and* [11]*Department of Nutrition and dietetics, Faculty of Health Care Vesalius, University College Ghent, Ghent, Belgium*

Background: Measuring dietary intake in children is notoriously difficult. Therefore, it is crucial to evaluate the performance of dietary intake assessment methods in children. Given the important contribution of milk consumption to calcium (Ca) and potassium (K) intakes, urinary calcium (UCa) and potassium (UK) excretions in spot urine samples could be used for estimating correlations with milk consumption frequencies.
Objective: The aim of this study was to evaluate the assessment of milk consumption frequencies derived from the Food Frequency Questionnaire section of the Children's Eating Habits Questionnaire (CEHQ-FFQ) used in the IDEFICS (Identification and prevention of dietary- and lifestyle induced health effects in children and infants) study by comparing with UCa and UK excretions in spot urine samples.
Design: This study was conducted as a setting-based community-oriented intervention study and results from the first cross-sectional survey have been included in the analysis.
Subjects: A total of 10 309 children aged 2–10 years from eight European countries are included in this analysis.
Methods: UCa and UK excretions were measured in morning spot urine samples. Calcium and potassium urine concentrations were standardised for urinary creatinine (Cr) excretion. Ratios of UCa/Cr and UK/Cr were used for multivariate regression analyses after logarithmic transformation to obtain normal distributions of data. Milk consumption frequencies were obtained from the CEHQ-FFQ. Multivariate regression analyses were used to investigate the effect of milk consumption frequencies on UCa and UK concentrations, adjusting for age, gender, study centre, soft drink consumption and frequency of main meals consumed at home.
Results: A significant positive correlation was found between milk consumption frequencies and ratios of UK/Cr and a weaker but still significant positive correlation with ratios of UCa/Cr, when using crude or partial Spearman's correlations. Multivariate regression analyses showed that milk consumption frequencies were predictive of UCa/Cr and UK/Cr ratios, when adjusted for age, gender, study centre, soft drink consumption and frequency of main meals consumed at home. Mean ratios of UK/Cr for increasing milk consumption frequency tertiles showed a progressive increase in UK/Cr. Children consuming at least two milk servings per day had significantly higher mean UCa/Cr and UK/Cr ratios than children who did not. Large differences in

Correspondence: Dr I Huybrechts, Department of Public Health, Faculty of Medicine and Health Sciences, Ghent University, UZ – 2 Blok A, De Pintelaan 185, B-9000 Ghent, Belgium.
E-mail: inge.huybrechts@ugent.be

correlations between milk consumption frequencies and ratios of UCa/Cr and UK/Cr were found between the different study centres.
Conclusion: Higher milk consumption frequencies resulted in a progressive increase in UK/Cr and UCa/Cr ratios, reflecting the higher Ca and K intakes that coincide with increasing milk consumption, which constitutes a major K and Ca source in children's diet.
International Journal of Obesity (2011) **35,** S69–S78; doi:10.1038/ijo.2011.37

Keywords: children; FFQ; CEHQ-FFQ; milk consumption; urinary potassium; urinary calcium

Introduction

Recent studies showed an inverse association between dietary calcium (Ca) intake, particularly from dairy sources, and body weight among adults[1–3] and children.[4] In 1984, McCaron *et al*.[5] for the first time described an inverse association between Ca intake and body weight and, a few years later, Trevisan *et al*.[6] observed that body mass index was inversely associated with the frequency of milk consumption, but not of cheese, in Italian male adults. More recently, Barba *et al*.[4] showed a significant inverse association between the frequency of milk consumption and body mass index among children. On consideration of these results, further research on this issue was encouraged because it might have important implications for the prevention of obesity among children. However, before investigating associations between milk consumption and disease risk or milk consumption and obesity, it is important to evaluate to what extent dietary intake assessment methods that are currently being used for estimating children's milk consumption frequency can be considered as valid assessment tools.

Although different methods are available for estimating dietary intakes among individuals, previous research has shown that all methods available have strengths and limitations.[7] Although estimation of dietary intakes is difficult in all age groups, the collection of dietary intake data in children is notoriously difficult because of the necessity of relying on proxy records, mostly from parents. Preschool years, for instance, are characterised as a time for developing autonomy, expanding language skills, increasing ability to control behaviour and broadening the social environment by attending preschools or by staying with friends or relatives.[8] Many factors typical among preschool children (for example, illiteracy, short memory and so on) make the assessment of dietary intake and consequently diet-related diseases in this young population very difficult. The most important barrier for measuring dietary intake in children is the fact that they are unable to complete questionnaires on their own and have a limited cognitive ability to recall, estimate and otherwise cooperate. They eat small amounts of food at frequent intervals, and their dietary habits and nutrient intakes may change rapidly with maturation and exposure to new foods.[9,10]

For the above-mentioned reasons, a proxy respondent was used in the IDEFICS (Identification and prevention of dietary- and lifestyle induced health effects in children and infants) study, a European Union-funded project of the Sixth Framework. Even though proxy reports could solve a part of the above-mentioned problems, the fact that children often spend time under the care of several individuals makes it difficult to obtain correct dietary intake assessments of young children.

Therefore, it was considered important to investigate to what extent a convenient method such as the parent/proxy-reported Food Frequency Questionnaire (FFQ) section of the Children's Eating Habits Questionnaire (CEHQ-FFQ) used in the IDEFICS study could be a valid assessment tool for estimating food consumption among children by comparison with urinary excretions of minerals that are known as important components of the food group(s) under study. As mentioned above, milk is not only an important component of children's diet but it might also have a role in the prevention of obesity among children.[4] Therefore, it was decided to focus on evaluating milk consumption frequencies.

Milk and other dairy products are major dietary sources of Ca and potassium (K) in Western diets.[11–13] Milk alone accounts for more than 18 and 17%, respectively, of the total Ca and potassium intake of the adult US population.[11] Recent analyses among Belgian children revealed that milk was the main dietary source of both Ca (50%) and potassium (24%) intakes among children.[14,15] Previous studies also showed significant correlations between Ca intake and urinary potassium (UK) excretions.[16] Because of these reasons, one would expect a positive association between milk consumption frequencies and UCa and UK excretions.

Therefore, the present study evaluated milk consumption frequencies measured by proxy FFQ administrations among children enrolled in the IDEFICS study, by comparing with UCa and UK concentrations derived from spot urine samples. Although the same analyses were performed using other food items of the dairy food group, including milk, yoghurt and cheese, the high between-children and between-food item variations in portion sizes could interfere with and jeopardise possible associations. Therefore, it was decided to focus on milk consumption frequencies only.

Subjects and methods

Study population/setting

IDEFICS is a large-scale multicentre European setting-based community-oriented intervention study, aiming to investigate the causes and consequences of overweight and obesity in 2- to 10-year-old European children in order to prevent obesity. The baseline survey was conducted from September 2007 to May 2008. In total, 31 543 children were contacted and the response rate was 53.5%. In each country, the participating centres obtained ethical approval from the local ethics committees. All participating children and their parents provided oral and/or written informed consent for all examinations and/or for the collection of blood, urine and saliva samples, as well as for subsequent analysis and storage of personal data and collected samples.[17]

Among several other features, interviews with parents pertaining to lifestyle habits and dietary intakes, as well as anthropometric measurements, were included in the survey. Preanalytical sample preparations were conducted using standard procedures by all eight centres participating in the study, which were located in Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain and Sweden. Detailed information on the study procedures can be obtained from previous publications.[18,19]

Subjects with missing information on milk consumption frequency, UCa/UK concentrations or on any covariate in statistical models were excluded from the analyses ($N_{\text{study sample}} = 10\,309$; $N_{\text{original sample}} = 16\,224$). No significant differences were observed when comparing main subject characteristics such as age and gender distributions, prevalences of overweight/obesity and hypertension of the total IDEFICS study population and the study sample included in the statistical analysis of the present study.

Data collection and analyses

The Children's Eating Habits Questionnaire. The IDEFICS CEHQ questionnaire contains general questions on eating behaviours, parents' attitude towards diet and meal frequencies, as well as an FFQ (CEHQ-FFQ). The CEHQ-FFQ is regarded as a screening instrument to investigate the consumption of foods that are shown by consistent evidence to be related, either positively or negatively, to overweight and obesity in children. The CEHQ was not designed to provide an estimate of total energy intake or total food intake, and foods less likely to be associated with obesity were not included. The CEHQ questions allowed us to distinguish, for instance, between white and wholemeal bread, thereby making it possible to compare consumption of these alternatives. Although different food classification systems are currently available,[20] a new food grouping system has been developed for the CEHQ questions to meet the specific aims mentioned above, considering, at the same time, the cultural differences in dietary habits between the different IDEFICS centres.

The CEHQ-FFQ consists of 43 food items on which parents or other caregivers were asked to report the number of meals the children usually consumed at home or at other people's homes, such as of grandparents and friends, in a typical week of the previous month. Frequency categories ranged from 'Never/less than once a week' to '4 or more times per day' and were converted into times per week ranging from 0 to 30. Considering the increased respondent burden and the lack of accuracy found in other studies for usual portion size estimates included in an FFQ,[21] no attempts were made to describe portion sizes. However, Portion size was examined in detail within a separate 24-hour recall module.

The 43 food items were clustered into 14 food groups: vegetables; fruits; drinks; breakfast cereals; milk; yoghurt; fish; meat; eggs and mayonnaise; meat replacements; cheese; spreadable products; cereal products; and snacks. The present analyses have been based on questions included in the milk group in which a differentiation was made between sweetened and unsweetened milk. However, in the analyses, the sum of sweetened and unsweetened milk consumption was used. During the explorative phase of data analyses, the food groups fruits, vegetables, yoghurt, cheese and soft drinks have also been used in the analyses (data not shown).

Urinary concentrations. First morning urine samples were analysed centrally in an International Organization for Standardization 15189 accredited laboratory using a photometric assay for Ca (Orthocresolphtalein, ROCHE, Mannheim, Germany) and creatinine (Cr) (Jaffe-reaction, ROCHE) excretions and an ion-sensitive electrode method for potassium (Integra, ROCHE) excretion. Urinary concentrations were expressed in $\text{mmol}\,\text{l}^{-1}$. Calcium and potassium urinary concentrations were standardised for urinary Cr excretion by using ratios of UCa to Cr (UCa/Cr) and ratios of UK to Cr.

Statistical methods

Descriptive analyses of the study population were performed, as well as prevalence analysis of mean milk consumption frequencies and UCa and UK concentrations.

Differences in UCa/Cr and UK/Cr ratios were compared between children consuming at least two servings per day and those who did not. Country-specific tertiles of total milk consumption frequencies were calculated in order to compare mean UCa/Cr and UK/Cr ratios between the tertiles to allow for estimation of differences in the urinary ratios across the tertiles of milk consumption frequencies. This was carried out both for the study population as a whole as well as separately for children from each study centre. Age, sex, soft drink consumption, the number of meals per week consumed at home or at other people's homes, and the study centre were considered as potential confounders for the associations between milk consumption frequencies and UCa and UK concentrations. Soft drink consumption was used as a confounding factor, as this could influence K/Ca excretion.[22] The number of meals per week consumed

at home or at other people's homes was included as a confounding factor, as this frequency could influence the reported milk consumption frequencies. For instance, if milk was mainly consumed at school, the true consumption frequency would be underreported in the questionnaire if parents were unaware of the milk consumption at school.

Unadjusted Spearman's correlation coefficients and partial correlation coefficients, adjusted for the confounders given above, were calculated for milk consumption frequencies and urinary ratios of Ca/Cr and K/Cr.

By application of different multivariate linear regression models, the effects of milk consumption frequencies on UCa and UK excretions were investigated using ratios of UCa/Cr and UK/Cr as dependent variables. To achieve an approximately normal distribution for these dependent variables, UCa/Cr and UK/Cr were log transformed using the natural logarithm (LN). Total milk consumption frequencies (as the sum of sweetened and unsweetened milk consumption frequencies) were entered in the LN UCa/Cr and LN UK/Cr model, respectively (Model 1, Model 2), to evaluate the effects of milk intake on UCa and UK concentrations. All models were adjusted for the above-mentioned confounding factors and when appropriate for study centre. All analyses were carried out for the whole study group and stratified by study centre.

When considering the large sample size, it was decided to consider a *P*-value less than 0.01 as statistically significant. Data were analysed using the Statistical Analysis System software package (version 9.2; SAS Institute, Cary, NC, USA).

Results

Description of the study population

An in-depth description of the IDEFICS study population is given by Ahrens *et al*.[17] in the same issue of the *Journal*. The mean age of the IDEFICS subsample of 10 309 children included in the present study was 6.1 ± 1.8 years. Boys (50.8%) and girls were almost equally represented.

Milk consumption frequencies

Table 1 shows milk consumption frequencies by countries participating in the IDEFICS study, by considering the children included in the analyses carried out in this study. Considering all countries together, the mean ± s.d. milk frequency was 11 ± 8.2 times per week. The highest frequency was found in Cyprus (15.9 ± 10.5 times per week), whereas Hungary, Belgium and Italy had the lowest with less than 10 times per week. Unsweetened milk was the largest contributor to total milk consumption in all countries except Italy and Spain (Table 1).

Urinary calcium and potassium concentrations

Urinary calcium and potassium concentrations were (mean ± s.d.) 2.8 ± 2.2 and 44.3 ± 24.4 mmol l^{-1}, respectively (Table 2). Ratios of UCa/Cr and UK/Cr were 0.4 ± 0.3 and 6.7 ± 5.3, respectively. These urinary concentrations differed between countries (Table 2). The highest ratios of UCa/Cr

Table 1 Consumption frequencies of types of dairy products (times per week) in the subsample of the IDEFICS population used in this study

Country	N	*Total milk; mean (s.d.)*	*Total yoghurt; mean (s.d.)*	*Total cheese; mean (s.d.)*	*Total dairy; mean (s.d.)*[a]
All	10 042	11.0 (8.2)	4.9 (4.9)	5.2 (5.1)	21.1 (12.1)
Italy	1311	8.9 (6.5)	3.0 (4.5)	5.5 (5.4)	17.4 (10.7)
Estonia	1091	11.9 (8.3)	5.2 (4.8)	5.7 (5.3)	22.8 (12.0)
Cyprus	780	15.8 (10.4)	4.9 (6.6)	4.3 (5.2)	25.1 (14.5)
Belgium	1087	9.2 (7.4)	4.2 (4.0)	3.5 (3.8)	17.0 (10.9)
Sweden	1233	11.9 (7.9)	4.6 (4.0)	4.4 (4.6)	20.8 (10.7)
Germany	1327	11.4 (9.0)	5.7 (4.9)	6.2 (5.5)	23.2 (12.8)
Hungary	2081	9.5 (7.6)	4.6 (4.3)	6.6 (5.6)	20.6 (11.9)
Spain	1132	12.3 (7.0)	7.5 (5.7)	4.1 (3.7)	23.8 (10.6)

Abbreviation: IDEFICS, Identification and prevention of dietary- and lifestyle induced health effects in children and infants. [a]Sum of milk, yoghurt and cheese.

Table 2 Mean urinary calcium and potassium concentrations with and without standardisation for urinary creatinine concentrations

Country	N	*Calcium (mmol l^{-1}); mean (s.d.)*	*Potassium (mmol l^{-1}); mean (s.d.)*	*Calcium to creatinine ratio; mean (s.d.)*	*Potassium to creatinine ratio; mean (s.d.)*
All	10 309	2.8 (2.2)	44.3 (24.4)	0.40 (0.30)	6.70 (5.27)
Italy	1382	3.1 (2.4)	42.1 (22.2)	0.40 (0.31)	5.72 (4.16)
Estonia	1117	3.1 (2.3)	49.0 (26.1)	0.41 (0.30)	7.01 (4.64)
Cyprus	810	2.9 (2.2)	49.0 (27.0)	0.42 (0.32)	7.43 (5.57)
Belgium	1109	2.2 (1.8)	52.9 (26.4)	0.30 (0.25)	8.12 (6.37)
Sweden	1268	3.0 (2.2)	43.6 (22.6)	0.45 (0.32)	7.16 (5.52)
Germany	1373	2.3 (1.9)	46.2 (26.4)	0.30 (0.25)	6.46 (5.29)
Hungary	2100	3.1 (2.2)	36.4 (21.2)	0.41 (0.30)	5.37 (4.24)
Spain	1150	3.0 (1.9)	43.5 (21.6)	0.48 (0.31)	7.89 (6.15)

were found for Spain (0.5 ± 0.3) and the lowest for both Germany and Belgium (0.3 ± 0.3). The highest ratios of UK/Cr were found for Belgium (8.12 ± 6.4) and Spain (7.9 ± 6.1) and the lowest for Italy (5.7 ± 4.2) and Hungary (5.4 ± 4.3).

The group of children who consumed at least two milk servings per day had significantly higher ratios of UCa/Cr and UK/Cr than children who consumed less ($P<0.001$) (Table 3). The differences were significant for all countries for ratios of UK/Cr, whereas ratios of UCa/Cr were only significant for Italy, Cyprus, Hungary, Spain and Belgium (one-sided *t*-test using LN-transformed ratios).

Correlation analyses

The Spearman correlations found between milk consumption frequencies and UCa and UK excretions are presented in Table 4. There was a significant positive correlation between milk consumption frequencies and the ratios of UK/Cr (0.16 ($P<0.001$)) and a weaker but still significant positive correlation with the ratios of UCa/Cr (0.07 ($P<0.001$)) when using crude or partial Spearman correlations. Large differences in correlations between milk consumption frequencies and ratios of UCa/Cr and UK/Cr were found between the different study centres (Table 4). For the ratios of UCa/Cr the highest correlations with milk consumption frequencies were found for Belgium ($r=0.08$), whereas the highest correlations for the ratios of UK/Cr were found for Hungary ($r=0.18$). The lowest correlations between total milk consumption frequencies and ratios of UCa/Cr were found for Germany ($r=0.01$), followed by Hungary and Estonia ($r=0.03$ and $r=0.04$, respectively), and the lowest correlation between total milk consumption frequencies and ratios of UK/Cr was found for Italy (0.09). It is noteworthy that no significant correlations were found between fruit and vegetable consumption and the ratios of UCa/Cr and UK/Cr. For yoghurt and cheese consumption as well, no significant correlations were found with the ratios of UCa/Cr and UK/Cr (data not shown). Therefore, it was decided to exclude the yoghurt, cheese, fruit and vegetable food groups from any further analyses, as they were not confounding factors.

Table 3 Differences in urinary calcium and potassium concentrations comparing groups of children consuming milk at least (⩾2) versus less than (<2) 2 times per day

Country	*Number of children*		*Urinary calcium to creatinine ratio*		P-*value*[a]	*Urinary potassium to creatinine ratio*		P-*value*[a]
	Milk consumption frequencies		*Milk consumption frequencies*[b]			*Milk consumption frequencies*[b]		
	<2 times per day	*⩾2 times per day*	*<2 times per day*	*⩾2 times per day*		*<2 times per day*	*⩾2 times per day*	
All	6016	4293	0.37 (0.29)	0.43 (0.32)	<0.0001	6.09 (4.83)	7.54 (5.71)	<0.0001
Italy	959	423	0.38 (0.29)	0.46 (0.34)	0.001	5.47 (4.18)	6.27 (4.06)	<0.0001
Estonia	617	500	0.40 (0.30)	0.42 (0.30)	0.065	6.65 (4.36)	7.45 (4.93)	0.000
Cyprus	284	526	0.38 (0.29)	0.43 (0.33)	0.018	6.53 (4.75)	7.92 (5.91)	<0.0001
Belgium	734	375	0.28 (0.24)	0.34 (0.28)	0.003	7.34 (5.58)	9.65 (7.46)	<0.0001
Sweden	621	647	0.43 (0.32)	0.47 (0.33)	0.081	6.99 (5.92)	7.33 (5.10)	0.003
Germany	855	518	0.30 (0.25)	0.31 (0.26)	0.470	6.00 (5.05)	7.21 (5.60)	<0.0001
Hungary	1448	652	0.40 (0.29)	0.44 (0.32)	0.014	4.94 (3.94)	6.32 (4.70)	<0.0001
Spain	498	652	0.46 (0.31)	0.50 (0.32)	0.007	6.92 (4.99)	8.63 (6.82)	<0.0001

In order to adjust for multiple testing a *P*-value less than 0.01 was considered statistically significant. [a]One-sided *t*-test performed on log-transformed ratios. [b]Mean (s.d.).

Table 4 Crude and partial Spearman's correlation coefficients of ratios of urinary Ca/Cr and K/Cr and milk consumption frequencies[a]

Country	*Spearman's correlations; urinary Ca/Cr vs total milk consumption*[b]		*Spearman's correlations; urinary K/Cr vs total milk consumption*[b]	
	Crude (P-*value*)	*Adjusted* (P-*value*)[a]	*Crude* (P-*value*)	*Adjusted* (P-*value*)[a]
All	0.07 (<0.0001)	0.07 (<0.0001)	0.17 (<0.0001)	0.16 (<0.0001)
Italy	0.07 (0.008)	0.07 (0.008)	0.14 (<0.0001)	0.09 (0.001)
Estonia	0.04 (0.197)	0.04 (0.167)	0.09 (0.002)	0.11 (0.0001)
Cyprus	0.07 (0.039)	0.08 (0.032)	0.15 (<0.0001)	0.11 (0.002)
Belgium	0.06 (0.042)	0.07 (0.025)	0.19 (<0.0001)	0.14 (<0.0001)
Sweden	0.06 (0.034)	0.07 (0.015)	0.10 (0.001)	0.13 (<0.0001)
Germany	0.00 (0.876)	0.01 (0.635)	0.15 (<0.0001)	0.13 (<0.0001)
Hungary	0.04 (0.079)	0.03 (0.112)	0.19 (<0.0001)	0.18 (<0.0001)
Spain	0.07 (0.020)	0.06 (0.040)	0.15 (<0.0001)	0.11 (0.0001)

Abbreviations: Ca, calcium; Cr, creatinine; K, potassium. [a]Adjusted for age, sex, soft drinks and number of home meals. [b]Sum of sweetened and unsweetened milk. In order to adjust for multiple testing a *P*-value less than 0.01 was considered statistically significant.

Multivariate regression analyses

Results from multivariate linear regression analyses are presented in Table 5. Ratios of UCa/Cr and UK/Cr were entered in the model as dependent variables that were predicted by milk consumption frequencies. Age, gender, soft drink consumption, frequency of meals under parental control and study centre (if appropriate) were the confounders considered in the model. Ratios of UCa/Cr were significantly associated with milk consumption frequencies in the whole group ($P<0.0001$) and in each country, except Estonia, Germany and Hungary (Table 5). Ratios of UK/Cr were significantly associated with milk consumption frequencies in the whole group ($P<0.0001$) and in each country. Other important determinants of both LN UCa/Cr and UK/Cr were gender and soft drink consumption (Table 5).

Comparison of urinary calcium and potassium concentrations by tertiles of milk consumption frequencies

Ratios of UCa/Cr by tertiles of milk consumption frequencies showed a progressive increase with increasing milk consumption frequencies for the whole study group (Figure 1; Table 6). However, the increase was non-monotonic in Belgium, Hungary and Spain. Ratios of UK/Cr by tertiles of milk consumption frequencies showed a progressive increase by increasing milk consumption frequencies for the whole study group and for each study centre except for Sweden (Figure 2; Table 6).

Discussion

The evaluation of the results obtained for milk consumption frequencies from the CEHQ-FFQ in the IDEFICS project revealed that UCa and UK excretions increased with increasing milk consumption frequencies. There was a significant positive correlation between milk consumption frequencies and ratios of UK/Cr and a weaker but still significant positive correlation with ratios of UCa/Cr, which were used to standardise UCa and UK concentrations for Cr excretion. Ratios of UK/Cr showed a progressive increase with increasing milk consumption frequency tertiles. Large differences in correlations between milk consumption frequencies and UCa and UK excretions were found between the different study centres.

Even though no other studies were identified that had investigated the relationship between ratios of UK/Cr and milk consumption, a relationship between ratios of UK/Cr and potassium consumption has been previously documented.[23] McKeown *et al*.[23] reported higher correlations between dietary and UK for 7-day food diaries ($r=0.51$–0.55) than for the FFQ ($r=0.32$–0.34) among adults. In addition, some reports used UK excretions as a surrogate marker for estimating intakes of fruit and vegetables[24] and for poor diet quality[25] among Japanese adults. Therefore, the correlations

Table 5 Multivariate linear regression models: effects of milk consumption frequencies on ratios of urinary Ca/Cr and K/Cr

	All[a]		*Italy*		*Estonia*		*Cyprus*		*Belgium*		*Sweden*		*Germany*		*Hungary*		*Spain*	
	β	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*	*β*	P-*value*
Dependent variable: LN ratios of urinary Ca/Cr (LN UCa/Cr)																		
Model 1																		
Milk	0.005	<0.0001	0.009	0.008	0.003	0.218	0.006	0.032	0.008	0.025	0.004	0.142	0.003	0.180	0.004	0.082	0.005	0.141
Age	−0.023	<0.0001	0.000	0.990	−0.036	0.004	−0.035	0.165	0.045	0.011	−0.053	<0.0001	0.010	0.440	−0.041	<0.0001	−0.031	0.011
Sex	0.083	<0.0001	0.073	0.104	0.059	0.205	0.164	0.009	0.146	0.008	0.062	0.164	0.145	0.001	0.014	0.690	0.064	0.134
Home meals	−0.001	0.569	−0.002	0.753	−0.003	0.695	0.002	0.764	0.008	0.206	−0.002	0.780	−0.013	0.042	−0.004	0.435	0.011	0.129
Soft drinks	−0.004	0.014	−0.008	0.095	−0.015	0.011	0.003	0.796	−0.007	0.232	−0.011	0.319	−0.004	0.148	0.000	0.996	−0.012	0.422
Dependent variable: LN ratios of urinary K/Cr (LN UK/Cr)																		
Model 2																		
Milk	0.008	<0.0001	0.006	0.009	0.009	<0.0001	0.005	0.014	0.012	<0.0001	0.007	0.000	0.008	<0.0001	0.012	<0.0001	0.006	0.005
Age	−0.129	<0.0001	−0.105	<0.0001	−0.133	<0.0001	−0.151	<0.0001	−0.150	<0.0001	−0.134	<0.0001	−0.096	<0.0001	−0.130	<0.0001	−0.160	<0.0001
Sex	−0.080	<0.0001	−0.034	0.227	−0.056	0.073	−0.105	0.009	−0.081	0.016	−0.127	<0.0001	−0.119	0.000	−0.074	0.002	−0.067	0.036
Home meals	−0.002	0.268	0.004	0.296	0.001	0.874	0.007	0.188	−0.012	0.003	0.000	0.975	−0.001	0.858	0.000	0.939	−0.015	0.006
Soft drinks	−0.004	<0.0001	0.000	0.950	−0.009	0.024	0.005	0.424	−0.009	0.008	−0.017	0.026	−0.003	0.099	−0.004	0.037	−0.027	0.015

Abbreviations: Ca, calcium; Cr, creatinine; LN, natural logarithm; K, potassium; UK, urinary potassium. [a]Adjusted for study centre.

Table 6 Results of statistical analyses of ratios of urinary potassium to creatinine (UK/Cr) and ratios of urinary calcium to creatinine (UCa/Cr) by tertiles of milk consumption frequencies for each country as presented in Figures 1 and 2

	Country	*All*	*Italy*	*Estonia*	*Cyprus*	*Belgium*	*Sweden*	*Germany*	*Hungary*	*Spain*
A. Ratios of UK/Cr										
First tertile of milk consumption frequencies	Mean (s.d.)	6.075 (4.814)	5.441 (4.165)	6.652 (4.288)	6.581 (4.806)	7.162 (5.277)	6.815 (5.697)	6.046 (5.404)	4.703 (3.698)	6.898 (5.112)
Second tertile of milk consumption frequencies	Mean (s.d.)	6.927 (5.483)	5.691 (4.423)	7.229 (4.870)	7.307 (5.209)	7.515 (5.878)	7.923 (6.099)	6.411 (4.685)	5.183 (4.151)	8.351 (6.688)
Third tertile of milk consumption frequencies	Mean (s.d.)	7.398 (5.571)	6.276 (4.074)	7.366 (4.932)	8.458 (6.466)	9.645 (7.455)	6.902 (4.494)	7.414 (5.797)	6.319 (4.700)	8.988 (6.476)
	P-value	<0.0001	<0.0001	0.0451	0.0002	<0.0001	0.0022[a]	<0.0001	<0.0001	<0.0001
B. Ratios of UCa/Cr										
First tertile of milk consumption frequencies	Mean (s.d.)	0.380 (0.290)	0.375 (0.286)	0.401 (0.303)	0.383 (0.297)	0.292 (0.233)	0.426 (0.310)	0.298 (0.252)	0.405 (0.287)	0.457 (0.310)
Second tertile of milk consumption frequencies	Mean (s.d.)	0.395 (0.297)	0.424 (0.289)	0.406 (0.286)	0.409 (0.285)	0.273 (0.236)	0.457 (0.336)	0.297 (0.236)	0.400 (0.292)	0.494 (0.317)
Third tertile of milk consumption frequencies	Mean (s.d.)	0.420 (0.322)	0.441 (0.343)	0.422 (0.305)	0.460 (0.360)	0.337 (0.278)	0.466 (0.330)	0.312 (0.274)	0.441 (0.324)	0.491 (0.312)
	P-value	<0.0001	0.0067	0.5210	0.0418	0.0116[a]	0.5307	0.8823	0.0835[a]	0.0620[a]

Abbreviations: Ca, calcium; Cr, creatinine; K, potassium; UCa, urinary calcium; UK, urinary potassium. [a]Non-monotonic. *P*-values for differences between tertiles (ANOVA).

Figure 1 Ratios of urinary potassium to creatinine (mean ± s.d.) by tertiles of milk consumption frequencies.

Figure 2 Ratios of urinary calcium to creatinine (mean ± s.d.) by tertiles of milk consumption frequencies.

between UK/Cr and consumptions of fruit and vegetables were also investigated in the explorative analyses of the present study. However, in contrast to the study on Japanese adults,[24] no relationship was found between UK excretion and intakes of fruit and vegetables in the European children included in the IDEFICS study (data not shown). These findings are in line with recent data from Denmark in which no associations were found between UK excretion and intakes of fruit and vegetables.[26] Taken together, these findings suggest that milk consumption frequency contributes more efficiently to total K intake than does the consumption of fruit and vegetables in children.[14,15]

Previous studies have already used UK excretions as a surrogate marker for K intake for studying associations between K intakes and health outcomes.[27,28] The relationship between ratios of UCa/Cr and Ca intakes and the potential of ratios of UCa/Cr as a surrogate marker for Ca intakes have also been investigated.[29–31] Esbjörner and Jones[29] found no significant correlation between ratios of UCa/Cr and reported milk or Ca intake in Swedish children. More recently, Toren and Norman[30] confirmed this finding in adults and concluded that 24-h UCa

concentrations should not be used as a surrogate marker for dietary Ca intake. Therefore, the very low correlations found between total milk consumption frequencies and ratios of UCa/Cr ($r=0.07$; $P<0.0001$ for the total IDEFICS subsample) might be due to the fact that ratios of UCa/Cr are not a reliable marker for Ca intakes, as it reflects total Ca intake only to a minor extent.

Comparison of the ratios of UCa/Cr of the Swedish children participating in the study by Esbjörner and Jones[29] with the Swedish children participating in the IDEFICS study revealed similar results. For 2- to 6-year-old Swedish children, mean ± s.d. ratios of UCa/Cr were 0.59 ± 0.49 in the study by Esbjörner and Jones and 0.51 ± 0.37 in the IDEFICS study. For 7- to 10-year-old Swedish children, ratios of UCa/Cr were 0.43 ± 0.35 and 0.41 ± 0.27, respectively.

McGill *et al.*[12] recently reported that dairy products (mainly milk) made a significant contribution to dietary K intakes among different US age and gender groups. They also revealed that mean K intakes were significantly higher in those subjects who met dairy intake recommendations compared with those who did not, and this held true for all age groups. When comparing children of the IDEFICS study who consumed milk products at least twice per day with children who did not, significantly higher UK concentrations were found among children consuming milk products at least twice per day (Table 3).

Finally, it should be noted that the significant negative correlations found between soft drink consumption and ratios of UK/Cr as well as UCa/Cr could possibly be explained by the replacement of milk products by soft drinks among children who are frequently consuming soft drinks.[22] However, an investigation of possible correlations between soft drink and milk consumption frequencies reported in the IDEFICS study showed no significant inverse correlations between these two food groups. In contrast, a significant positive correlation was found between soft drink consumption and sweetened milk consumption, whereas for most countries no significant correlations were found with unsweetened milk consumption frequencies (data not shown). These latter findings suggest no replacement effect of milk consumption by soft drink consumption.

An important strength of the IDEFICS study is its large sample size and multicentric nature, which increased the between-children variation in dietary habits. Assessment of fairly homogeneous populations makes it difficult to detect associations between dietary habits (or consumption of particular foods and nutrients) and biomarkers of health and disease outcomes because of the lack of sufficient between-subject variation of the variables under investigation.[32] Therefore, this large multicentre and heterogeneous population might be one of the reasons why significant correlations were found between milk consumption frequencies and ratios of UCa/Cr, which are in contrast with findings from previous studies using smaller and more homogenous study populations.[29,30] Other important strengths of the present study are the strict standardisation procedures that were followed during the data collection of the IDEFICS fieldwork[18,19] and the high quality control procedures applied during the different stages of the project, including, for instance, checks for plausibility that were already implemented in the database and performed during data entry. The fact that all urine samples collected in this survey by the different countries were all analysed in the same laboratory using the same standard protocol, methods and materials reduced analytical variations that might have affected the results obtained, which is considered to be an additional strength of the IDEFICS study. Furthermore, this is the first study investigating the relationship between UK concentrations and milk consumption, which is considered to be the main contributor to K intake among children.[12,13]

Some limitations to be considered are the fact that the study centres that were selected within the IDEFICS study were not necessarily representative for each particular country. Therefore, the descriptive results presented by the participating study centres of eight countries cannot be generalised for the whole participating countries. However, the representativeness of the study sample should not be a major issue for the association analyses reported in this study.

Another limitation is the fact that parental reporting was used to assess the consumption frequencies of the different food groups, as this proxy reporting might strongly rely on the number of meals under parental control.[8] As the number of meals consumed at home and out of home (for example, school lunches) might differ substantially between countries, the accuracy of the consumption frequencies reported by parents could differ between countries. However, to minimise this reporting bias, the authors corrected in the correlation and regression analyses for the number of main meals consumed at home and reported in the FFQ as weekly consumption frequencies. In addition, the lack of portion size information in the FFQ could have further reduced the accuracy of the estimated amount of food consumption. However, as mentioned before, the accuracy of usual portion sizes estimated in (semi-)quantitative FFQs has been shown to be a problem as well, and even more so in young children in whom proxy reporting is required.[21,33–35] Given this lack of accuracy and considering the increased respondent burden for estimating usual portion sizes in addition to consumption frequencies, it was decided not to include portion size information in the FFQ used in the IDEFICS study. Unfortunately, the lack of portion size information did not allow us to calculate any nutrient intakes for the children included in the analyses. Therefore, it was not possible to correct for any dietary factors that possibly interacted with Ca reabsorption, such as phosphorus or protein intake.

Even though the use of one (or even more) 24-h urine collection would have been more accurate than the use of first morning urine samples,[36,37] the authors standardised UCa and UK excretions for Cr excretions by calculating the ratios of UCa and UK to UCr concentrations. A recent study

showed that spot urine samples may be used to monitor trends in dietary sodium and K intakes and to compare subgroups of the population, even if they do not exactly replicate the electrolyte concentrations of 24-h samples.[38] The fact that the means for the ratios of UK/Cr of the spot urine samples followed the same patterns as the 24-h samples suggests that the spot samples could differentiate between subgroups of the population in a similar way as the 24-h sample.[38] More recently, Garde *et al*.[39] investigated the error introduced when using the two types of standardisation procedures on 24-h samples of healthy individuals—that is, spot urine samples versus 24-h urine samples. From this study it could be concluded that the uncertainty of Cr standardisation is increased when studying single voids rather than 24-h urine samples. However, this was partially counteracted by the increased statistical power due to the high number of samples collected when using spot urine samples because of considerably increased convenience for the study participants.[39]

Although UCr excretion was used to standardise the UCa and UK concentrations derived from spot urine samples, it should be noted that, within an age group, the daily volume of urinary Cr was suggested to increase with body size.[40] Therefore, height/Cr could be a better marker of renal function than 1/Cr in childhood populations. Because of this reason, the authors also conducted multivariate regression analyses including the height of the child as a covariate. However, the results derived from these latter analyses were very similar to the results obtained without height correction. To facilitate comparability with the other studies mentioned in this manuscript, the authors decided to report the results from the multivariate regression analyses without height correction.

In addition, it is noteworthy to mention that according to the protocol instructions the 'first morning urine samples' were not necessarily fasted samples. However, if the above-mentioned limitations of the FFQ and the spot urine samples had affected the results of the present study, they would have attenuated the correlations observed.

An in-depth evaluation of the reproducibility of the CEHQ-FFQ is given by Lanfer *et al*.[41] in the same issue of the *Journal*. Furthermore, it would be important to investigate to what extent the data on urinary excretions collected in the IDEFICS study and expressed as ratios of UK/Cr and UCa/Cr could be used as objective biomarkers to calibrate the measurement error in the dietary reports. The introduction of biomarkers to calibrate measurement errors in dietary reports is a significant development in the effort to improve estimates of diet-disease risks within populations.[42]

Conclusion

Increasing milk consumption frequencies were associated with a progressive increase in the ratios of UK/Cr, suggesting that higher K intakes coincide with increasing milk consumption as a main potassium source among children. Further, the weaker but still significant positive correlation found between milk consumption frequencies and ratios of UCa/Cr suggests that higher UCa excretion coincides with increasing milk consumption as the major dietary Ca source in children.

Although a more in-depth validation study using multiple food records or 24-h recalls would be necessary to evaluate the validity and reliability of the IDEFICS CEHQ-FFQ, the current analyses already showed an increase in urinary Ca and K concentrations with increasing milk consumption frequencies.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

This study was conducted as part of the IDEFICS study (http://www.idefics.eu). We gratefully acknowledge the financial support of the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD). We also thank the children and parents who participated in the IDEFICS study.

References

1 Zemel MB. Regulation of adiposity and obesity risk by dietary calcium: mechanisms and implications. *J Am Coll Nutr* 2002; **21**: 146S–151S.
2 Heaney RP. Normalizing calcium intake: projected population effects for body weight. *J Nutr* 2003; **133**: 268S–270S.
3 Teegarden D. Calcium intake and reduction in weight or fat mass. *J Nutr* 2003; **133**: 249S–251S.
4 Barba G, Troiano E, Russo P, Venezia A, Siani A. Inverse association between body mass and frequency of milk consumption in children. *Br J Nutr* 2005; **93**: 15–19.
5 McCarron DA, Morris CD, Henry HJ, Stanton JL. Blood pressure and nutrient intake in the United States. *Science* 1984; **224**: 1392–1398.
6 Trevisan M, Krogh V, Farinaro E, Panico S, Mancini M. Calcium-rich foods and blood pressure: findings from the Italian National Research Council Study (the Nine Communities Study). *Am J Epidemiol* 1988; **127**: 1155–1163.
7 Willett WC. *Nutritional Epidemiology*, 2nd edn. Oxford University Press: New York, NY, 1998.
8 Wooldridge NH. Toddler and preschool nutrition. In: Brown JE (ed). *Nutrition Through the Lifecycle*. Wadsworth/Thompson Learning: Belmont, 2002, pp 241–268.
9 Stein AD, Shea S, Basch CE, Contento IR, Zybert P. Variability and tracking of nutrient intakes of preschool children based on multiple administrations of the 24-h dietary recall. *Am J Epidemiol* 1991; **134**: 1427–1437.
10 Stein AD, Shea S, Basch CE, Contento IR, Zybert P. Consistency of the Willett semi-quantitative food frequency questionnaire and

24-h dietary recalls in estimating nutrient intakes of preschool children. *Am J Epidemiol* 1992; **135**: 667–677.
11 Subar AF, Krebs-Smith SM, Cook A, Kahle LL. Dietary sources of nutrients among US children, 1989–1991. *Pediatrics* 1998; **102** (4 Part 1): 913–923.
12 McGill CR, Fulgoni III VL, DiRienzo D, Huth PJ, Kurilich AC, Miller GD. Contribution of dairy products to dietary potassium intake in the United States population. *J Am Coll Nutr* 2008; **27**: 44–50.
13 Rafferty K, Heaney RP. Nutrient effects on the calcium economy: emphasizing the potassium controversy. *J Nutr* 2008; **138**: 166S–1171.
14 Huybrechts I, De Keyzer W, Lin Y, Vandevijvere S, Vereecken C, Van Oyen H *et al*. Dietary sources and correlates of sodium and potassium intakes in Flemish preschoolers. In review in *PHN*.
15 Huybrechts I, Lin Y, De Keyzer W, Vandevijvere S, Vereecken C, Van Oyen H *et al*. Dietary sources and correlates of calcium and vitamin D intakes in Flemish preschoolers. In review in *Eur J Clin Nutr* 2010.
16 Morikawa Y, Nakagawa H, Okayama A, Mikawa K, Sakata K, Miura K *et al*. A cross-sectional study on association of calcium intake with blood pressure in Japanese population. *J Hum Hypertens* 2002; **16**: 105–110.
17 Ahrens W, Bammann K, Siani A, Buchecker K, De Henauw S, Iacoviello L *et al*. The IDEFICS cohort: design, characteristics and participation in baseline survey. *Int J Obes* 2011; **35** (Suppl 1): S3–S15.
18 Ahrens W, Bammann K, De Henauw S, Halford J, Palou A, Pigeot I *et al*. Understanding and preventing childhood obesity and related disorders–IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis* 2006; **16**: 302–308.
19 Bammann K, Peplies J, Pigeot I, Ahrens W. IDEFICS: a multicenter European project on diet- and lifestyle-related disorders in children. *Med Klin (Munich)* 2007; **102**: 230–235.
20 Ireland J, Erp-Baart AM, Charrondiere UR, Moller A, Smithers G, Trichopoulou A. Selection of a food classification system and a food composition database for future food consumption surveys. *Eur J Clin Nutr* 2002; **56** (Suppl 2): S33–S45.
21 Cade J, Thompson R, Burley V, Warm D. Development, validation and utilisation of food-frequency questionnaires—a review. *Public Health Nutr* 2002; **5**: 567–587.
22 Heaney RP, Rafferty K. Carbonated beverages and urinary calcium excretion. *Am J Clin Nutr* 2001; **74**: 343–347.
23 McKeown NM, Day NE, Welch AA, Runswick SA, Luben RN, Mulligan AA *et al*. Use of biological markers to validate self-reported dietary intake in a random sample of the European Prospective Investigation into Cancer United Kingdom Norfolk cohort. *Am J Clin Nutr* 2001; **74**: 188–196.
24 Tamaki J, Yoshita K, Kikuchi Y, Takebayashi T, Chiba N, Okamura T *et al*. Applicability of the stages of change model for analyzing fruit and vegetable intake in relation to urinary potassium excretion: baseline results from the High-Risk and Population Strategy for Occupational Health Promotion (HIPOP-OHP) Study. *Hypertens Res* 2004; **27**: 843–850.
25 Mente A, Irvine EJ, Honey RJ, Logan AG. Urinary potassium is a clinically useful test to detect a poor quality diet. *J Nutr* 2009; **139**: 743–749.
26 Krogholm KS, Haraldsdottir J, Knuthsen P, Rasmussen SE. Urinary total flavonoid excretion but not 4-pyridoxic acid or potassium can be used as a biomarker for the intake of fruits and vegetables. *J Nutr* 2004; **134**: 445–451.
27 Iseki K, Iseki C, Itoh K, Uezono K, Sanefuji M, Ikemiya Y *et al*. Urinary excretion of sodium and potassium in a screened cohort in Okinawa, Japan. *Hypertens Res* 2002; **25**: 731–736.
28 Dyer A, Elliott P, Chee D, Stamler J. Urinary biochemical markers of dietary intake in the INTERSALT study. *Am J Clin Nutr* 1997; **65** (4 Suppl): 1246S–1253S.
29 Esbjorner E, Jones IL. Urinary calcium excretion in Swedish children. *Acta Paediatr* 1995; **84**: 156–159.
30 Toren PJ, Norman RW. Is 24-h urinary calcium a surrogate marker for dietary calcium intake? *Urology* 2005; **65**: 459–462.
31 O'Brien KO, Abrams SA, Stuff JE, Liang LK, Welch TR. Variables related to urinary calcium excretion in young girls. *J Pediatr Gastroenterol Nutr* 1996; **23**: 8–12.
32 Michels KB. Nutritional epidemiology–past, present, future. *Int J Epidemiol* 2003; **32**: 486–488.
33 Schlundt DG, Buchowski MS, Hargreaves MK, Hankin JH, Signorello LB, Blot WJ. Separate estimates of portion size were not essential for energy and nutrient estimation: results from the Southern Community Cohort food-frequency questionnaire pilot study. *Public Health Nutr* 2007; **10**: 245–251.
34 Noethlings U, Hoffmann K, Bergmann MM, Boeing H. Portion size adds limited information on variance in food intake of participants in the EPIC-Potsdam study. *J Nutr* 2003; **133**: 510–515.
35 Tjonneland A, Haraldsdottir J, Overvad K, Stripp C, Ewertz M, Jensen OM. Influence of individually estimated portion size data on the validity of a semi-quantitative food frequency questionnaire. *Int J Epidemiol* 1992; **21**: 770–777.
36 Tanaka T, Okamura T, Miura K, Kadowaki T, Ueshima H, Nakagawa H *et al*. A simple method to estimate populational 24-h urinary sodium and potassium excretion using a casual urine specimen. *J Hum Hypertens* 2002; **16**: 97–103.
37 Siani A, Iacoviello L, Giorgione N, Iacone R, Strazzullo P. Comparison of variability of urinary sodium, potassium, and calcium in free-living men. *Hypertension* 1989; **13**: 38–42.
38 Joint Health Surveys Unit. *A Survey of 24 h and Spot Urinary Sodium and Potassium Excretion in a Representative Sample of the Scottish Population*. National Center for Social Research: London, UK, 2007.
39 Garde AH, Hansen AM, Kristiansen J, Knudsen LE. Comparison of uncertainties related to standardization of urine samples with volume and creatinine concentration. *Ann Occup Hyg* 2004; **48**: 171–179.
40 Mage DT, Allen RH, Kodali A. Creatinine corrections for estimating children's and adult's pesticide intake doses in equilibrium with urinary pesticide and creatinine concentrations. *J Expo Sci Environ Epidemiol* 2008; **18**: 360–368.
41 Lanfer A, Hebestreit A, Ahrens W, Krogh V, Sieri S, Lissner L *et al*. Reproducibility of food consumption frequencies derived from the Children's Eating Habits Questionnaire used in the IDEFICS study. *Int J Obes* 2011; **35** (Suppl 1): S61–S68.
42 Bingham SA. Biomarkers in nutritional epidemiology. *Public Health Nutr* 2002; **5**: 821–827.

B.2 Validity of 24-hour recalls in (pre-)school aged children: Comparison of proxy-reported energy intakes with measured energy expenditure

This article describes the validation study of the 24-HDR SACINA based on objective measurements of energy expenditure obtained by the DLW method (cf. Section 5.1.1). As the manuscript was still under review when publishing this thesis, the paper draft was removed from the published version of the thesis. Only the abstract is given here. The complete manuscript can be obtained from the author on request.

Validity of 24-hour recalls in (pre-)school aged children: Comparison of proxy-reported energy intakes with measured energy expenditure

C. Börnhorst[1*], S. Bel-Serrat[2*], I. Pigeot[1], I. Huybrechts[3,4], C. Ottavaere[3], I. Sioen[3], S De Henauw[3,5], T. Mouratidou[2], M. I. Mesana Graffe[2], K. Westerterp[6], K. Bammann[1,7], L. Lissner[8], G. Eiben[8], V. Pala[9], M. Rayson[10], V. Krogh[9], L.A. Moreno[2], on behalf of the IDEFICS consortium.

1 BIPS - Institute for Epidemiology and Prevention Research, Achterstr. 30, 28359 Bremen, Germany
2 GENUD (Growth, Exercise, Nutrition and Development) research group. E.U. Ciencias de la Salud, Universidad de Zaragoza, Corona de Aragón 42, 2nd floor, 50009 Zaragoza, Spain
3 Department of Public Health, Ghent University, 2blokA De Pintelaan 185, 9000 Ghent, Belgium
4 Dietary Exposure Assessment Groups, International Agency for Research on Cancer, Lyon, France
5 University College Ghent, Department of Nutrition and Dietetics, Faculty of Health Care "Vesalius", De Pintelaan 185, 9000 Ghent, Belgium
6 Department of Human Biology, Maastricht University, The Netherlands
7 Institute for Public Health and Nursing Research, University of Bremen, Bremen, Germany
8 Department of Public Health and Community Medicine, University of Gothenburg, Box 454, 40530 Gothenburg, Sweden
9 Department of Preventive and Predictive Medicine, Nutritional Epidemiology Unit, Fondazione IRCSS Istituto Nazionale dei Tumori, Via Venezian 1, 20133 Milan, Italy
10 BioTel Ltd. Clifton, 26 B Oakfield Road, Clifton, Bristol, BS8 2AT, United Kingdom

Correspondence:
Luis A. Moreno
E.U. de Ciencias de la Salud
University of Zaragoza
C/ Domingo Miral, s/n
50009 Zaragoza (Spain)
Phone: +34 976 76 10 00 Ext. 4457
E-mail: lmoreno@unizar.es

Abstract

Background & Aims: Little is known about the validity of proxy-reported 24-HDR in young children yet. This study aimed to evaluate the validity of proxy-reported EI compared with total energy expenditure (TEE) measured by the double labeled water (DLW) technique.
Methods: The agreement between EI and TEE was investigated in 36 children aged 4-10 years from Belgium and Spain using subgroup analyses and Bland-Altman plots. Groups of low-energy-reporters (LER), adequate-energy-reporters (AER) and high-energy-reporters (HER) were defined from the ratio of EI over TEE by application of age- and sex-specific cut-off values.
Results: Means of EI (1500 kcal/day) and TEE (1523 kcal/day) matched almost exactly at group level though partially large differences were observed at the individual level. Almost perfect agreement between EI and TEE was observed in thin/normal weight children (EI: 1511 kcal/day; TEE: 1513 kcal/day). In overweight/obese children the mean difference between EI and TEE was only -86 kcal/day which corresponds to an underreporting of total EI by approximately 4%. Among the participants, 28 (78%) were classified as AER, five (14%) as HER and three (8%) as LER.
Conclusion: Two proxy-reported 24-HDRs were a valid instrument to assess EI on group level but not on the individual level.

B.3 Prevalence and determinants of misreporting among European children in proxy reported 24 h dietary recalls

This study investigated the prevalence and determinants of misreporting in proxy-reported 24-HDRs and was published in the peer-reviewed *British Journal of Nutrition* (Börnhorst et al., 2012b).

Prevalence and determinants of misreporting among European children in proxy-reported 24 h dietary recalls

C. Börnhorst, I. Huybrechts, W. Ahrens, G. Eiben, N. Michels, V. Pala, D. Molnár, P. Russo, G. Barba, S. Bel-Serrat, L. A. Moreno, S. Papoutsou, T. Veidebaum, H.-M. Loit, L. Lissner and I. Pigeot

British Journal of Nutrition / *FirstView* Article / September 2012, pp 1 - 9
DOI: 10.1017/S0007114512003194, Published online: 06 August 2012

British Journal of Nutrition, page 1 of 9
doi:10.1017/S0007114512003194

Prevalence and determinants of misreporting among European children in proxy-reported 24 h dietary recalls

C. Börnhorst[1], I. Huybrechts[2,3], W. Ahrens[1], G. Eiben[4], N. Michels[2], V. Pala[5], D. Molnár[6], P. Russo[7], G. Barba[7], S. Bel-Serrat[8], L. A. Moreno[8], S. Papoutsou[9], T. Veidebaum[10], H.-M. Loit[10], L. Lissner[3] and I. Pigeot[1]* on behalf of the IDEFICS consortium

[1]*BIPS – Institute for Epidemiology and Prevention Research, Achterstraße 30, 28359 Bremen, Germany*
[2]*Department of Public Health, Ghent University, 2 BlokA De Pintelaan 185, 9000 Ghent, Belgium*
[3]*Dietary Exposure Assessment Groups, International Agency for Research on Cancer, Lyon, France*
[4]*Department of Public Health and Community Medicine, University of Gothenburg, Box 454, 40530 Gothenburg, Sweden*
[5]*Nutritional Epidemiology Unit, Department of Preventive and Predictive Medicine, Fondazione IRCSS Istituto Nazionale dei Tumori, Via Venezian 1, 20133 Milan, Italy*
[6]*Department of Pediatrics, Medical Faculty, University of Pécs, József A.u.7 H-7623, Pécs, Hungary*
[7]*Institute of Food Sciences, CNR, Via Roma 64, 83100 Avellino, Italy*
[8]*GENUD (Growth, Exercise, Nutrition and Development) Research Group, Escuela Universitaria de Ciencias de la Salud, Universidad de Zaragoza, Corona de Aragón 42, 2nd floor, 50009 Zaragoza, Spain*
[9]*Research and Education Institute of Child Health, 8 Attikis Street, 2027 Strovolos, Cyprus*
[10]*Center of Health and Behavioral Science, National Institute for Health Development, Hiiu 42, 11619 Tallinn, Estonia*

(Submitted 19 January 2012 – Final revision received 25 May 2012 – Accepted 19 June 2012)

Abstract

Dietary assessment is strongly affected by misreporting (both under- and over-reporting), which results in measurement error. Knowledge about misreporting is essential to correctly interpret potentially biased associations between diet and health outcomes. In young children, dietary data mainly rely on proxy respondents but little is known about determinants of misreporting here. The present analysis was conducted within the framework of the multi-centre IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) study and is based on 6101 children aged 2–9 years with 24 h dietary recall (24-HDR) and complete covariate information. Adapted Goldberg cut-offs were applied to classify the 24-HDR as 'over-report', 'plausible report' or 'under-report'. Backward elimination in the course of multi-level logistic regression analyses was conducted to identify factors significantly related to under- and over-reporting. Next to characteristics of the children and parents, social factors and parental concerns/perceptions concerning their child's weight status were considered. Further selective misreporting was addressed, investigating food group intakes commonly perceived as more or less socially desirable. Proportions of under-, plausible and over-reports were 8·0, 88·6 and 3·4 %, respectively. The risk of under-reporting increased with age (OR 1·19, 95 % CI 1·05, 1·83), BMI *z*-score of the child (OR 1·23, 95 % CI 1·10, 1·37) and household size (OR 1·12, 95 % CI 1·01, 1·25), and was higher in low/medium income groups (OR 1·45, 95 % CI 1·13, 1·86). Over-reporting was negatively associated with BMI *z*-scores of the child (OR 0·78, 95 % CI 0·69, 0·88) and higher in girls (OR 1·70, 95 % CI 1·27, 2·28). Further social desirability and parental concerns/perceptions seemed to influence the reporting behaviour. Future studies should involve these determinants of misreporting when investigating diet–disease relationships in children to correct for the differential reporting bias.

Key words: **Energy intake: Goldberg cut-off: Parental perceptions: Social desirability**

Due to its low respondent burden and easy application, the 24 h dietary recall (24-HDR) is often the method of choice for short-term assessment of dietary intakes in large epidemiological studies. However, numerous sources of measurement error have been encountered when operating with 24-HDR data. Memory of consumption, estimation of portion sizes, decompositions of mixed dishes (unknown recipes), supplement use as well as instrument-based biases are common problems that researchers are confronted with[(1)]. As young children lack the cognitive skills to complete dietary assessments[(2)], 24-HDR data in children younger than 7 years old usually rely on proxy reporters, mainly the

Abbreviations: 24-HDR, 24 h dietary recall; IDEFICS, Identification and prevention of dietary- and lifestyle-induced health effects in children and infants; EI, energy intake; OVR, over-reports; PA, physical activity; PAL, physical activity level; PLR, plausible reports; SACINA, Self-Administered Children and Infants Nutrition Assessment; UNR, under-reports.

* **Corresponding author:** I. Pigeot, fax +49 421 218 56 941, email pigeot@bips.uni-bremen.de

parents[3]. Here additional problems emerge from meals that are not under parental control (e.g. school meals), leading to unintentional misreporting[4–6].

Among these difficulties, biased assessment of energy intake (EI) is often a consequence of intentional under- or over-reporting attributed to specific groups. Anthropometry, for example the actual weight status of the study subject, is a well-known determinant of misreporting[7,8]. Age, sex, socio-economic status, psychosocial and behavioural characteristics are further factors that were found to be related to misreporting[9–11]. The validity of proxy-reported EI might additionally be affected by parental characteristics as well as by psychological factors such as parental perception of their child's weight status[3,12–14]. Further social desirability may result in over-reporting of food items perceived to be healthy while unhealthy/energy-dense food items might be under-reported at the same time[15,16]. Intentional misreporting introduces differential error that may attenuate or even hide associations between dietary factors and health outcomes, whereas non-differential error may distort such associations in any direction.

Recent validation studies based on the doubly labelled water technique in children have revealed inconsistent results concerning misreporting (under-reporting from 19 to 41 %; over-reporting from 7 to 11 % of reported EI)[17] where data relied mainly on self-reports – partially with parental assistance.

Whether the accuracy of proxy reports is comparable to that of self-reports and whether determinants of misreporting coincide for self- *v.* proxy-reports is yet unknown. Several studies have only addressed under-reporting; the nature and extent of over-reporting have rarely been addressed in young populations[10,18,19]. Knowing the degree and direction of misreporting is essential for the assessment of diet–disease relationships as well as for the evaluation of dietary guidelines and nutrition policies. Therefore, the present study aimed to investigate the prevalence and determinants of misreporting (including under- and over-reporting) in a large sample of European children.

Methods

Study population

The IDEFICS (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) study is a multi-centre setting-based study aiming to prevent and investigate the causes of diet- and lifestyle-related diseases such as overweight and obesity in 2–9-year-old European children. The baseline survey was conducted from September 2007 until June 2008; more than 31 500 children were invited, out of whom, finally, 16 220 participated and fulfilled the inclusion criteria of the IDEFICS study. Details on the design and objectives of the study have been given elsewhere[20–22]. Briefly, children were recruited through schools/kindergartens. Interviews with parents concerning lifestyle habits and dietary intakes as well as anthropometric measurements and examinations of the children were included in the survey. Biomarker information was collected via blood, urine and saliva samples. All measurements were conducted using standardised procedures by all eight centres participating in the study (Italy, Estonia, Cyprus, Belgium, Sweden, Germany, Hungary and Spain).

The present study was conducted according to the guidelines laid down in the Declaration of Helsinki and all procedures involving human subjects/patients were approved by the local ethics committees in each participating country. Parents provided written informed consent for all examinations and for the collection of blood, urine, saliva samples as well as subsequent analysis and storage. Each child was informed orally about the modules by fieldworkers and asked for its consent immediately before examination[22]. Verbal consent was witnessed and formally recorded.

Dietary data

Dietary data were assessed using the computerised 24-HDR 'SACINA' (Self-Administered Children and Infants Nutrition Assessment) based on the previously designed and validated YANA-C ('Young Adolescents' Nutrition Assessment on Computer') developed for Flemish adolescents and further adapted to European adolescents in the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) study[23,24]. The SACINA is structured according to six meal occasions (breakfast, morning snack, lunch, afternoon snack, dinner and evening snack) embedded in questions related to a range of chronological daily activities. Proxies, mainly the parents, completed the 24-HDR under the supervision of fieldwork personnel in about 20–30 min. Except for Cyprus, where school ends at 13.00 hours, school meals were additionally assessed by means of direct observation. Teachers and school kitchen staff were interviewed by trained survey personnel and data were documented using special documentation sheets including portion sizes. School meal data were merged with the parentally reported 24-HDR data to enhance completeness of dietary intakes. The 24-HDR were assessed on non-consecutive days over the whole week and over the complete IDEFICS assessment period. The assessment procedure in Hungary slightly differed from the other study centres. Here dietary information was recorded on documentation sheets and entered into the SACINA program afterwards.

The uniquely coded food items were linked to country-specific food composition tables. Missing quantities for single food items as well as obviously implausible data entries were imputed by country, food group and age-specific median intakes (0·01 % of the entries) to avoid excessive record exclusions. Although up to six repeated 24-HDR were carried out in a smaller sample, only the first recall day was included in the present analysis (including weekdays and weekend days) to obtain an equal number of 24-HDR for each child and to achieve an adequate statistical power for a cross-country analysis. Incomplete interviews were excluded, for example if the proxy did not know about at least one main meal or in case of missing school meal information. Further, intakes of more than 16 736 kJ/d (4000 kcal/d) were excluded (n 10).

Anthropometry

Height (cm) of the children was measured to the nearest 0·1 cm with a calibrated stadiometer (model: telescopic height measuring instrument SECA 225 Stadiometer; SECA).

British Journal of Nutrition

Body weight (kg) was measured in the fasting state in light underwear on a calibrated scale accurate to 0·1 kg (model: electronic scale TANITA BC 420 SMA with an adapter; Tarita Europe GmbH).

Covariables

A set of covariables previously found to be related to misreporting in adults or expected to be relevant in children[9,11,23,24] was defined to explore the determinants of misreporting in this young study population: age, sex, BMI *z*-scores according to Cole *et al.*[25,26] and average audio-visual media time (h/week) of the child, age, sex and self-reported BMI of the proxy, educational level (maximum of both parents, dummy: high *v.* medium/low) according to the International Standard Classification of Education (ISCED 1997, UNESCO Institute for Statistics: Montreal, 2006), net household income (dummy: high *v.* medium/low) and number of persons below 18 years of age in the household as indicators for socio-economic status, the interview day (dummy: weekday *v.* Saturday/Sunday), assessment of a school meal (dummy: yes *v.* no) and the use of a day-care service or babysitter (dummy: yes *v.* no) were considered.

The following information on parental concerns/perceptions of their child's weight status was included where the questions were obtained from previously validated questionnaires[27,28] and slightly modified for use in IDEFICS: 'How concerned are you about your child' – (1) 'eating too much when you are not around him/her?'; (2) 'having to diet to maintain desirable weight?'; (3) 'becoming overweight?'; (4) 'becoming underweight?' (answer categories: 'Unconcerned', 'A little concerned', 'Concerned' or 'Very concerned'); 'Do you think your child is' – (1) 'Much too underweight?'; (2) 'Slightly too underweight?'; (3) 'Proper weight?'; (4) 'Slightly too overweight'; (5) 'Much too overweight?' (answer categories: 'Yes' or 'No'). The rationale behind this was the assumption of parental concerns/perceptions being associated with misreporting. Furthermore, the question 'Do you sit down with your child when he/she eats meals?' (answer categories: 'Never', 'Rarely', 'Sometimes', 'Often' or 'Always') was included as an indicator for family meal behaviours.

To investigate the degree to which given answers were influenced by social desirability, intakes of the following food items commonly perceived to be healthy/unhealthy were included as predictors for misreporting in a second step: chocolate products; other sugary products (e.g. cakes, biscuits, ice cream); carbonated soft drinks; fruits/vegetables; milk (all as the percentage from total EI per d); water (g/d).

Statistical methods

BMR was estimated using the equations published by Schofield[29] and recommended by the FAO/WHO/UNU (1985) taking into account age, sex, body height and weight. Goldberg *et al.*[30] defined cut-off values to classify the 24-HDR in under-reports (UNR), plausible reports (PLR) and over-reports (OVR), respectively. The cut-offs make allowance for the errors associated with the duration of dietary assessment (number of recall days), the sample size as well as the variation in BMR, physical activity (PA) level (PAL) and EI. Minimum/maximum plausible levels of EI are defined as multiples of BMR. Since these cut-offs were developed for adults without considering differences in EI due to age and sex, adaptations are required for application in children.

Upper and lower cut-off values to identify plausible/implausible reports of EI were calculated substituting Goldberg's single cut-off 2[30] by age- and sex-specific cut-offs for children, as suggested previously[11,31], using the following formula:

$$\text{Cut-off} = \text{PAL} \times \exp\left[\pm 1{\cdot}96 \times \frac{(S/100)}{\sqrt{n}}\right].$$

where

$$S = \sqrt{\frac{\text{CV}_{\text{wEI}}^2}{d} + \text{CV}_{\text{wBMR}}^2 + \text{CV}_{\text{PA}}^2}.$$

The within-subject CV for EI (CV_{wEI}), the within-subject CV for BMR (CV_{wBMR}) and the CV for PA (CV_{PA}) were replaced by age- and sex-specific reference values as given in Nelson *et al.*[32] and Black[33]. The number of days (d) was set to one as the analysis is based on one 24-HDR per child. Goldberg's overall level of 1·55 for PA was substituted by age- and sex-dependent levels of light PA (2–5 years: 1·45; 6–10 years: males 1·55, females 1·50) according to Torun *et al.*[34]. All reference values used are summarised in Table 1. The resulting age- and sex-specific cut-off values to define UNR, PLR and OVR are given in Table 2. Records were classified as UNR, PLR and OVR according to the recalculated cut-off values.

Multi-level logistic regression analysis was conducted to identify factors statistically significantly associated with misreporting. Determinants for UNR and OVR were investigated in separate models (model 1a: outcome UNR, reference PLR; model 2a: outcome OVR, reference PLR). In the model addressing UNR, records classified as OVR were excluded and the other way around. All covariables mentioned earlier were entered into the two models except for the dietary variables and the backward selection procedure was applied to screen out the relevant factors. Under this approach, one starts fitting a model that contains all covariables. The least significant one is dropped except if it is significant at the critical level of 0·05. The reduced models are successively refitted applying the same rule until all the remaining variables are statistically significant.

Table 1. Reference values to recalculate the Goldberg cut-offs for application in children

Age (years)	Sex	CV_{wEI}* (%)	CV_{wBMR}† (%)	PAL‡	CV_{PAL}§ (%)
2 to <6	Boys	24·0	6·8	1·45	23·8
2 to <6	Girls	24·0	7·6	1·45	19·1
6 to <10	Boys	22·5	6·8	1·55	12·6
6 to <10	Girls	21·3	7·6	1·50	9·5

EI, energy intake; PAL, physical activity level.
* Within-subject CV of energy intake; values obtained from Nelson *et al.*[32].
† Within-subject CV of BMR; values obtained from Black[33].
‡ PAL; values obtained from Torun *et al.*[34].
§ CV of PAL; values obtained from Black[33].

Table 2. Lower and upper cut-off limits to classify 1 d 24 h dietary recalls (24-HDR) as under-, plausible and over-reports based on the ratio of energy intake (EI*):BMR†

Age (years)	Sex	Under-report	Plausible report	Over-report
2 to <6	Boys	EI:BMR ≤ 0·74	0·74 < EI:BMR < 2·85	2·85 ≤ EI:BMR
2 to <6	Girls	EI:BMR ≤ 0·78	0·78 < EI:BMR < 2·69	2·69 ≤ EI:BMR
6 to <10	Boys	EI:BMR ≤ 0·92	0·92 < EI:BMR < 2·61	2·61 ≤ EI:BMR
6 to <10	Girls	EI:BMR ≤ 0·93	0·93 < EI:BMR < 2·43	2·43 ≤ EI:BMR

* EI estimated from 24-HDR.
† BMR estimated from Schofield equations[29].

In a next step, the dietary variables were added to the resulting models (including only the relevant covariables now) to investigate their predictive power for misreporting (model 1b: outcome UNR; model 2b: outcome OVR). Random effects for the study centre and setting (schools/kindergartens) were entered in all models to account for the clustered study design.

The present analysis only includes children with 24-HDR and complete covariable information (*n* 6101).

All analyses were performed using the statistical software package SAS (version 9.1; SAS Institute).

Results

Both the prevalence of UNR (1·2–16·4 %) and OVR (1·5–5·4 %) strongly differed between the study centres (Table 3). UNR was highest in the Hungarian study centre, OVR in the Italian one. UNR and OVR were higher in girls and UNR was higher in older children. Regarding the total study group, 8·0 % of the reports were classified as UNR and 3·4 % as OVR.

Descriptive statistics of all covariables can be obtained from Tables 4 and 5 stratified by reporting group (UNR, PLR and OVR). The mean BMI of children and their proxies were highest in UNR, whereas the percentage of proxies with a high income or educational level was highest in PLR. In UNR, a higher percentage of proxies were male and the use of day-care services was less frequent. The percentage of recalls assessed on weekends was highest in OVR. Furthermore, proxies of UNR were more likely to perceive their child as overweight/obese and stated more often to be concerned about their child becoming overweight, whereas proxies of OVR were more concerned about their child becoming underweight. Percentages of daily EI from chocolate products and sugary products increased with reporting group (lowest in UNR and highest in OVR), whereas percentages of EI from fruits/vegetables decreased with reporting group (Table 5).

Application of the backward selection procedure including all covariables except the dietary ones revealed that different factors were significantly associated with UNR compared with the model addressing OVR (models 1a and 2a; Table 6). The risk of UNR increased with age (OR 1·19, 95 % CI 1·11, 1·27), BMI *z*-score of the child (OR 1·23, 95 % CI 1·10, 1·37), the number of persons below 18 years of age in the household (OR 1·12, 95 % CI 1·01, 1·25) and was higher in the low/medium income group (OR 1·45, 95 % CI 1·13, 1·86; reference: high income group) as well as on interview days without additional school meal assessment (OR 1·58, 95 % CI 1·17, 2·13). Sitting always (OR 0·61, 95 % CI 0·43, 0·85) or often down while eating (OR 0·62, 95 % CI 0·44, 0·87; reference: sitting sometimes down while eating) turned out to be negatively associated with UNR. Proxies perceiving their child as slightly (OR 1·63, 95 % CI 1·03, 2·56) or much too overweight (OR 3·30, 95 % CI 1·51, 7·18; reference: slightly too underweight) were more likely to under-report. On the other hand, OVR was higher in female children (OR 1·70, 95 % CI 1·27, 2·28; reference: male children). BMI *z*-scores of children (OR 0·78, 95 % CI 0·69, 0·88) were negatively associated with

Table 3. Prevalence of misreporting by study centre, sex and age group (Total numbers and percentages)

	Under-report		Plausible report		Over-report		Total study group
	n	%	*n*	%	*n*	%	*n*
Study centre							
Belgium	26	9·7	239	88·8	4	1·5	269
Cyprus	50	16·1	256	82·3	5	1·6	311
Estonia	24	4·9	446	90·8	21	4·3	491
Germany	137	10·3	1149	86·3	46	3·5	1332
Hungary	144	16·4	708	80·8	24	2·7	876
Italy	69	5·0	1239	89·7	74	5·4	1382
Spain	6	1·2	459	94·6	20	4·1	485
Sweden	30	3·1	911	95·4	14	1·5	955
Boys	218	7·1	2779	90·3	79	2·6	3076
Girls	268	8·9	2628	86·9	129	4·3	3025
2 to <6 years	130	4·9	2442	91·5	98	3·7	2670
6 to <10 years	356	10·4	2965	86·4	110	3·2	3431
Total study group	486	8·0	5407	88·6	208	3·4	6101

Table 4. Descriptive analysis of categorical covariates by reporting group
(Percentages and total numbers)

	Under-report (*n* 486)		Plausible report (*n* 5407)		Over-report (*n* 208)		Total study group (*n* 6101)	
	%	*n*	%	*n*	%	*n*	%	*n*
Sex of the child								
Male	44·9	218	51·4	2779	38·0	79	50·4	3076
Female	55·1	268	48·6	2628	62·0	129	49·6	3025
Sex of the proxy								
Male	18·5	90	17·2	931	16·3	34	17·3	1055
Female	81·5	396	82·8	4476	83·7	174	82·7	5046
Income level								
Low/medium	75·3	366	68·4	3698	77·9	162	69,3	4226
High	24·7	120	31·6	1709	22·1	46	30·7	1875
ISCED level*								
Low/medium	53·3	259	46·7	2523	59·6	124	47·6	2906
High	46·7	227	53·3	2884	40·4	84	52·4	3195
Use of day-care service or babysitter								
Yes	28·6	139	42·7	2311	41·3	86	41·6	2536
No	71·4	347	57·3	3096	58·7	122	58·4	3565
Day of the interview								
Weekday	81·7	397	82·9	4484	78·8	164	82·7	5045
Saturday/Sunday	18·3	89	17·1	923	21·2	44	17·3	1056
School meal assessment†								
No school meal	69·8	339	63·1	3414	67·3	140	63·8	3893
With school meal	30·2	147	36·9	1993	32·7	68	36·2	2208
Parents sit down with the child when eating								
Never	2·9	14	2·2	121	1·9	4	2·3	139
Rarely	2·5	12	2·5	134	1·0	2	2·4	148
Sometimes	12·8	62	7·6	411	11·5	24	8·1	497
Often	35·0	170	33·9	1833	35·6	74	34·0	2077
Always	46·9	228	53·8	2908	50·0	104	53·1	3240
Perception of the child's weight								
Much too underweight	1·4	7	1·3	72	2·4	5	1·4	84
Slightly too underweight	14·8	72	16·8	910	27·4	57	17·0	1039
Proper weight	55·6	270	70·8	3826	63·5	132	69·3	4228
Slightly too overweight	24·5	119	10·3	558	5·3	11	11·3	688
Much too overweight	3·7	18	0·8	41	1·4	3	1·0	62
Concerned – child eating too much when parents not around								
Unconcerned	57·8	281	62·5	3377	57·2	119	61·9	3777
A little concerned	17·3	84	19·9	1074	26·9	56	19·9	1214
Concerned	15·4	75	12·4	668	12·5	26	12·6	769
Very concerned	9·5	46	5·3	288	3·4	7	5·6	341
Concerned – child having a diet to maintain desirable weight								
Unconcerned	58·2	283	66·6	3602	64·9	135	65·9	4020
A little concerned	16·7	81	13·6	738	17·8	37	14·0	856
Concerned	16·3	79	12·6	681	11·5	24	12·9	784
Very concerned	8·8	43	7·1	386	5·8	12	7·2	441
Concerned – child becoming overweight								
Unconcerned	43·0	209	55·6	3004	53·8	112	54·5	3325
A little concerned	18·7	91	16·7	903	23·1	48	17·1	1042
Concerned	20·2	98	15·1	815	9·6	20	15·3	933
Very concerned	18·1	88	12·7	685	13·5	28	13·1	801
Concerned – child becoming underweight								
Unconcerned	55·1	268	52·0	2809	37·0	77	51·7	3154
A little concerned	16·7	81	16·6	899	18·3	38	16·7	1018
Concerned	13·4	65	14·6	787	25·0	52	14·8	904
Very concerned	14·8	72	16·9	912	19·7	41	16·8	1025

ISCED, International Standard Classification of Education.
* Low/medium education is defined as ISCED levels 1–3; high education is defined as ISCED levels 4 and 5 (ISCED 1997, UNESCO Institute for Statistics: Montreal, 2006).
† Days without school meals relate either to weekend days or to working days where the child had no lunch or lunch at home.

OVR. Being very concerned about the child becoming overweight (OR 0·44, 95 % CI 0·23, 0·84) decreased the risk for OVR, whereas being very concerned about the child becoming underweight increased the risk (OR 1·77, 95 % CI 1·10, 2·85).

Adding the dietary variables to the models showed that percentages of total EI from chocolate products, soft drinks and sugary products were negatively associated with the risk of UNR, whereas percentages of EI from fruits/vegetables were

Table 5. Descriptive analysis of continuous covariates and dietary intakes by reporting group
(Mean values and standard deviations)

	Under-report (*n* 560)		Plausible report (*n* 5308)		Over-report (*n* 228)		Total study group (*n* 6096)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age of the child (years)	6·8	1·5	6·0	1·8	6·0	1·8	6·1	1·8
BMI *z*-score of the child*	0·8	1·6	0·2	1·3	-0·2	1·2	0·2	1·3
Media consumption time (h/week)	12·6	7·6	11·5	7·1	11·0	6·5	11·6	7·1
Age of the proxy	35·3	5·1	35·9	5·2	35·6	5·1	35·8	5·2
BMI of the proxy	25·1	4·8	24·2	4·4	24·0	4·5	24·3	4·4
Number of persons <18 years of age in household	2·1	1·1	2·0	0·8	1·9	0·9	2·0	0·9
Energy (kcal/d)	774·1	220·5	1563·0	425·5	2757·0	429·9	1541·0	517·9
Energy (kJ/d)	3238·8	922·6	6539·6	1780·3	11535·3	1798·7	6447·5	2166·9
Water (g/d)	310·5	330·9	337·9	347·4	406·5	450·6	338·1	350·4
Chocolate products (% of total EI)	2·5	6·6	3·1	5·9	3·7	7·1	3·1	6·0
Milk (% of total EI)	9·6	13·3	10·6	9·8	8·2	8·3	10·4	10·1
Soft drinks (% of total EI)	2·8	6·6	2·8	5·8	2·3	4·1	2·8	5·8
Sugary products (% of total EI)	6·7	11·2	9·8	11·6	12·6	12·4	9·6	11·7
Fruits/vegetables (% of total EI)	10·4	12·4	8·4	7·8	7·3	6·3	8·6	8·2

EI, energy intake.
* According to Cole *et al.*[25,26].

positively associated (model 1b). The OR for the other covariables changed only slightly in model 1b compared with model 1a. Inclusion of the dietary variables in model 2a revealed no significant associations between the percentages of EI and OVR for any of the considered food items except for milk (OR 0·97, 95 % CI 0·95, 0·98) and sugary products (OR 1·01, 95 % CI 1·00, 1·02).

Discussion

In general, proportions reported for UNR and OVR vary widely between publications (UNR 2–85 % and OVR 3–46 %, obtained from a current review including children and adolescents[18]) where the proportion of UNR is usually higher than that of OVR. The proportions of UNR (8·0 %) and OVR (3·4 %) found in the present study sample are

Table 6. Results of the multi-level logistic regression applying backward selection: factors significantly associated with under-reports/over-reports (models 1a and 2a) and predictive value of selected food items for misreporting (models 1b and 2b)*
(Odds ratios and 95 % confidence intervals)

	OR for under-reports (*n* 5893)				OR for over-reports (*n* 5615)			
	Model 1a: backward selection		Model 1b: adding food items to model 1a		Model 2a: backward selection		Model 2b: adding food items to model 2a	
Covariates	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI
Age of the child (years)†	1·19	1·11, 1·27	1·20	1·12, 1·29				
Sex of the child: female *v.* male					1·70	1·27, 2·28	1·69	1·26, 2·26
Sex of the relative: female *v.* male	1·38	1·05, 1·83	1·44	1·09, 1·91				
BMI *z*-score of the child (Cole)†	1·23	1·10, 1·37	1·23	1·10, 1·37	0·78	0·69, 0·88	0·77	0·68, 0·87
Income level: low/medium *v.* high	1·45	1·13, 1·86	1·48	1·15, 1·91				
Number of persons <18 years of age in household†	1·12	1·01, 1·25	1·11	1·00, 1·24				
School meal assessed‡: no *v.* yes	1·58	1·17, 2·13	1·63	1·21, 2·21				
Sitting down while eating: always *v.* sometimes	0·61	0·43, 0·85	0·59	0·42, 0·83				
Sitting down while eating: often *v.* sometimes	0·62	0·44, 0·87	0·62	0·44, 0·87				
Perception: much too overweight *v.* slightly too underweight	3·30	1·51, 7·18	3·26	1·49, 7·16				
Perception: slightly too overweight *v.* slightly too underweight	1·63	1·03, 2·56	1·56	0·99, 2·46				
Child becoming overweight: concerned *v.* very concerned					0·44	0·23, 0·84	0·43	0·22, 0·83
Child becoming underweight: concerned *v.* very concerned					1·77	1·10, 2·85	1·71	1·06, 2·76
Water (g/d)†			1·00	1·00, 1·00			1·00	1·00, 1·00
Chocolate products (% of total EI)†			0·97	0·95, 0·99			1·01	0·99, 1·03
Milk (% of total EI)†			0·99	0·98, 1·00			0·98	0·96, 0·99
Soft drinks (% of total EI)†			0·98	0·96, 0·99			0·97	0·94, 1·01
Sugary products (% of total EI)†			0·98	0·97, 0·99			1·01	1·00, 1·02
Fruits/vegetables (% of total EI)†			1·02	1·00, 1·03			0·98	0·96, 1·00

EI, energy intake.
* All models include random effects for study centre and setting.
† Effects of continuous variables are assessed as one unit offsets from the mean.
‡ Days without school meals relate either to weekend days or to working days where the child had no lunch or lunch at home.

British Journal of Nutrition

difficult to compare with other studies due to differences in age groups, number of assessment days, cut-off values applied and the respondent status (self *v.* proxy). The relatively low proportions of UNR and OVR in the present data may be a consequence of cooperation with parents/caregivers, which has been shown to be associated with a lower risk of UNR/OVR previously[35]. The present data revealed a decreased risk of UNR on days with additional school meal assessment (days without school meal assessment relate either to weekend days or to working days where the child had no lunch or lunch at home). Lioret *et al.*[11] reported proportions for UNR and OVR of 4·9 and 1·4 %, respectively, in 3–10-year-old French children. This study was similar to the present one in terms of sample size, cut-off values and instruments applied. In the study by Murakami *et al.*[36], UNR ranged from 2·9 to 28·0 % and OVR from 3·0 to 28·1 % depending on the considered age group (children aged 6–15 years, stratified by 1-year age groups). UNR increased and OVR decreased with age, which agrees with the present results (Table 3). The increase in UNR with age may be explained by reduced parental control and a higher frequency of out-of-home meals in older children.

The notably high proportion of UNR in the Hungarian study centre (16·4 %) may be a consequence of the slightly different study protocol. As opposed to the paper-based assessment in Hungary, the computerised SACINA program used in the other study centres included reminders for certain foods and already some checks for plausibility. Further pictures with increasing portion sizes were displayed to facilitate the estimation of portion sizes. These differences in the assessment procedure may explain the discrepancy between the proportion of UNR in Hungary and the other study centres. In Cyprus, schools do not offer meals and therefore no additional information on school meals was assessed, which may explain the high percentage of UNR (16·1 %) in this study centre.

Over-reporting was found to be higher in children and adolescents compared with adults, which has been suggested to be rather a consequence of intrusion of foods that were not actually consumed than errors in portion size estimation[18,37]. However, these studies relied on self-reports. In proxy reports, over-reporting could be suspected to be either a result of intrusion due to the lack of parental control or a result of over-eating due to increased energy requirements during growth. The latter would result in misclassifications of records, for example OVR in spite of correct parental reports. Difficulties such as decreasing metabolic costs of movement during maturation and heavier children spending more energy at the same intensity of PA than peers may further affect classifications of UNR/OVR[38]. Moreover, it cannot be precluded that the 24-HDR was assessed on an exceptional day resulting in very high or low reported intakes (for example, child was ill).

Though the mean BMI of the child and of its proxy were both highest in the group of UNR in the descriptive analysis, the multivariate analysis revealed only the BMI of the child being significantly associated with misreporting. It is likely that similar dietary patterns within a family as well as shared genetic/environmental factors lead to correlations between parental and children's BMI[39], which may explain the bivariate association between UNR and parental BMI. Previous findings on the association between parental obesity and misreporting are inconsistent in children and adolescents[3,14,40]. Nevertheless, the strong association between parental concerns/perceptions of their child's weight status and misreporting rather suggests that, actually, the BMI of the child is the determining factor. To date, no other study has examined parental concerns/perceptions in relation to misreporting of EI by proxy respondents.

UNR was higher in low/medium income groups, whereas educational level was neither found to be associated with UNR nor with OVR. Previous studies in children reported no association between misreporting and income level[11,14,40]. In adults also inconsistent results concerning socio-economic variables have been reported in a review investigating markers on the validity of reported EI[9]. The authors assumed that poor literacy skills in the less well-educated group and better health and diet consciousness in the better-educated group might both result in misreporting leading to contradictory results. To the authors' knowledge, the effect of household size on misreporting, which may either serve as an indicator for socio-economic status or for parental control, has rarely been addressed in children. The present data suggest a positive relationship with UNR. Opposed to the present results, Vagstrand *et al.*[41] found UNR based on self-reports to be more likely in adolescents from one-child families. Nevertheless, in proxy reports, the impact of household size may be different as a high number of children may reduce parental control over each single child's food intakes. Parental control may also explain the effect of the 'sitting while eating' variable. Children sitting often or always down while eating had a reduced risk to be classified as UNR.

The present analysis revealed that OVR was higher in girls, whereas UNR was not associated with the sex of the child. It is likely that the determinants of misreporting may differ by sex and also by age group. Unfortunately, stratified analyses were not possible since corresponding models did not converge due to the high number of covariables and the comparably low number of UNR/OVR.

In a literature review mainly relying on self-reports, sex and social desirability have been reported to be consistent predictors for misreporting in adults but not in children and adolescents[18]. When adding the dietary variables to our models, results pointed to intentional, selective misreporting in the UNR group reflecting socially desirable answer behaviour. Food items commonly perceived to be unhealthy were negatively associated with UNR (chocolate products, sugary products and soft drinks), whereas fruit/vegetable intake showed a positive association. Although the SACINA instrument (retrospective) does not influence the child's eating behaviour, it seems to encourage socially desirable answers of the parents.

Some studies have already applied adapted validation procedures in children substituting Goldberg's single cut-off 2 by individual limits for children[11,31,42,43]. Sichert-Hellert *et al.*[31,43], for example, applied recalculated Goldberg cut-offs based on three assessment days in a German sample of 695 children aged 1–18 years but only addressing under-reporting. UNR ranged from 1·2 to 19·2 % depending on the

British Journal of Nutrition

respective age group and was lower compared with the proportions obtained when applying the original Goldberg cut-offs[30] (1 d cut-off; UNR, EI:BMR ≤0·9; OVR, EI:BMR ≥2·68). Also in the present study, the recalculated cut-offs revealed a slightly lower proportion of UNR (8·0 *v.* 8·3 %) and a higher proportion of OVR (3·4 *v.* 3·1 %) compared with the original Goldberg cut-offs, as expected (data not shown in the tables).

Limitations and strengths

Only one record day per child was used in the present analysis, which does not reflect usual intakes. Black & Cole[7] found that under-reporting is subject-specific, concluding that subjects who under-report on the first 24-HDR accordingly tend to under-report on additionally assessed 24-HDR as well. Therefore, a single 24-HDR can be considered as a reliable instrument for the identification of determinants of misreporting. Nevertheless, an additional analysis was run including only children with at least two 24-HDR. The study sample was markedly reduced (*n* 6101 *v.* 1644) and the number of study subjects strongly differed between the study centres (for example, Estonia *n* 3; Hungary *n* 828), which resulted in unstable model estimates. This corroborated the decision to include all children with at least one 24-HDR where only the first recall day was used in the analysis.

Sensitivity of the cut-off technique is limited, as it aims only to identify UNR resulting in physiologically implausible low EI[44]. By the application of the cut-off technique, varying degrees of misreporting cannot be distinguished, for example under-reporting from a high intake level such that the ratio of EI:BMR does not fall below the cut-off will not be detected. Further cut-off values were calculated assuming light PAL for all children which may result in misclassifications. The likelihood to classify a record as UNR increases with increasing energy expenditure of the child[9]. As PA is a relevant determinant of energy expenditure, classification into UNR, PLR and OVR should consider individual PAL by applying different cut-off values depending on a child's PAL. Unfortunately, due to the lack of valid PAL information, this approach was not feasible. Moreover, differentiation between undereaters (EI actually lower than energy expenditure) and under-reporters is not possible so that some part of UNR may be attributed to undereaters[9]. The same applies analogously to overeaters. Future research should include special questions for the identification of low/high eaters.

The large study sample, the additional assessment of school meals and measured anthropometry can be considered as strengths of the present study. Further, the huge number of covariables should be highlighted, as it facilitated a comprehensive analysis of the determinants of misreporting covering various aspects.

Conclusion

Misreports differ from plausible proxy reports with respect to children's characteristics (age, sex and weight status) as well as social factors (number of persons below 18 years of age in household and net household income). Determinants for UNR and OVR only partly agree where UNR seems to be strongly affected by social desirability. Furthermore, parental concerns/perceptions of their child's weight status had a strong impact on misreporting. Researchers should bear this differential reporting bias in mind when investigating diet–disease relationships in children. Identification of influencing factors may help to improve study designs and to interpret potentially biased results.

Acknowledgements

This study was done as part of the IDEFICS study (www.idefics.eu) and is published on behalf of its European Consortium. We gratefully acknowledge the financial support of the European Community within the Sixth Research and Technological Development Framework Programme Contract no. 016181 (FOOD). The information in this paper reflects the authors' view and is provided as is. S. B.-S. is funded by a grant from the Aragón's Regional Government (Diputación General de Aragón). Each author saw and approved the contents of the submitted manuscript. All authors contributed to the conception and design, the acquisition of the data, and the analysis or interpretation of the data, and approved the final version of the manuscript. C. B. drafted the manuscript; W. A., G. E., D. M., L. A. M. and T. V. conceived the study and participated in its design and coordination; I. H. and I. P. helped with the drafting of the manuscript and the interpretation of the data; all the authors revised the article critically for important intellectual content. None of the authors has any conflicts of interest.

References

1. Arab L & Akbar J (2002) Biomarkers and the measurement of fatty acids. *Public Health Nutr* **5**, 865–871.
2. McPherson RS, Hoelscher DM, Alexander M, *et al.* (2000) Dietary assessment methods among school-aged children: validity and reliability. *Preventive Med* **31**, 11–33.
3. Livingstone MB & Robson PJ (2000) Measurement of dietary intake in children. *Proc Nutr Soc* **59**, 279–293.
4. Eck LH, Klesges RC & Hanson CL (1989) Recall of a child's intake from one meal: are parents accurate? *J Am Diet Assoc* **89**, 784–789.
5. Baranowski T, Sprague D, Baranowski JH, *et al.* (1991) Accuracy of maternal dietary recall for preschool children. *J Am Diet Assoc* **91**, 669–674.
6. Basch CE, Shea S, Arliss R, *et al.* (1990) Validation of mothers' reports of dietary intake by four to seven year-old children. *Am J Public Health* **80**, 1314–1317.
7. Black AE & Cole TJ (2001) Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. *J Am Diet Assoc* **101**, 70–80.
8. Poppitt SD, Swann D, Black AE, *et al.* (1998) Assessment of selective under-reporting of food intake by both obese and non-obese women in a metabolic facility. *Int J Obes Relat Metab Disord* **22**, 303–311.
9. Livingstone MB & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133**, Suppl. 3, 895S–920S.

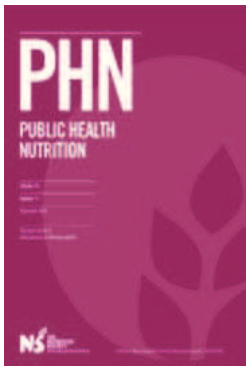
British Journal of Nutrition

10. Maurer J, Taren DL, Teixeira PJ, *et al.* (2006) The psychosocial and behavioral characteristics related to energy misreporting. *Nutr Rev* **64**, 53–66.
11. Lioret S, Touvier M, Balin M, *et al.* (2011) Characteristics of energy under-reporting in children and adolescents. *Br J Nutr* **105**, 1671–1680.
12. McGloin AF, Livingstone MB, Greene LC, *et al.* (2002) Energy and fat intake in obese and lean children at varying risk of obesity. *Int J Obes Relat Metab Disord* **26**, 200–207.
13. Johnson RK, Driscoll P & Goran MI (1996) Comparison of multiple-pass 24-hour recall estimates of energy intake with total energy expenditure determined by the doubly labeled water method in young children. *J Am Diet Assoc* **96**, 1140–1144.
14. Bandini LG, Cyr H, Must A, *et al.* (1997) Validity of reported energy intake in preadolescent girls. *Am J Clin Nutr* **65**, 1138S–1141S.
15. Tooze JA, Subar AF, Thompson FE, *et al.* (2004) Psychosocial predictors of energy underreporting in a large doubly labeled water study. *Am J Clin Nutr* **79**, 795–804.
16. Lissner L (2002) Measuring food intake in studies of obesity. *Public Health Nutr* **5**, 889–892.
17. Burrows TL, Martin RJ & Collins CE (2010) A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc* **110**, 1501–1510.
18. Forrestal SG (2011) Energy intake misreporting among children and adolescents: a literature review. *Matern Child Nutr* **7**, 112–127.
19. Poslusna K, Ruprich J, de Vries JH, *et al.* (2009) Misreporting of energy and micronutrient intake estimated by food records and 24 hour recalls, control and adjustment methods in practice. *Br J Nutr* **101**, Suppl. 2, S73–S85.
20. Ahrens W, Bammann K, de Henauw S, *et al.* (2006) Understanding and preventing childhood obesity and related disorders-IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis* **16**, 302–308.
21. Bammann K, Peplies J, Pigeot I, *et al.* (2007) [IDEFICS: a multicenter European project on diet- and lifestyle-related disorders in children]. *Med Klin (Munich)* **102**, 230–235.
22. Ahrens W, Bammann K, Siani A, *et al.* (2011) The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond)* **35**, Suppl. 1, S3–S15.
23. Vereecken CA, Covents M, Sichert-Hellert W, *et al.* (2008) Development and evaluation of a self-administered computerized 24-h dietary recall method for adolescents in Europe. *Int J Obes (Lond)* **32**, Suppl. 5, S26–S34.
24. Vereecken CA, Covents M, Matthys C, *et al.* (2005) Young Adolescents' Nutrition Assessment on Computer (YANA-C). *Eur J Clin Nutr* **59**, 658–667.
25. Cole TJ, Freeman JV & Preece MA (1998) British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* **17**, 407–429.
26. Cole TJ, Freeman JV & Preece MA (1995) Body mass index reference curves for the UK, 1990. *Arch Dis Child* **73**, 25–29.
27. Birch LL, Fisher JO, Grimm-Thomas K, *et al.* (2001) Confirmatory factor analysis of the Child Feeding Questionnaire: a measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. *Appetite* **36**, 201–210.
28. Mensink GB, Kleiser C & Richter A (2007) [Food consumption of children and adolescents in Germany. Results of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS)]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **50**, 609–623.
29. Schofield WN (1985) Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr* **39**, Suppl. 1, 5–41.
30. Goldberg GR, Black AE, Jebb SA, *et al.* (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 569–581.
31. Sichert-Hellert W, Kersting M & Schoch G (1998) Underreporting of energy intake in 1 to 18 year old German children and adolescents. *Z Ernahrungswiss* **37**, 242–251.
32. Nelson M, Black AE, Morris JA, *et al.* (1989) Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr* **50**, 155–167.
33. Black AE (2000) Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes Relat Metab Disord* **24**, 1119–1130.
34. Torun B, Davies PS, Livingstone MB, *et al.* (1996) Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. *Eur J Clin Nutr* **50**, Suppl. 1, S37–S80.
35. Kimm SY, Glynn NW, Obarzanek E, *et al.* (2006) Racial differences in correlates of misreporting of energy intake in adolescent females. *Obesity (Silver Spring)* **14**, 156–164.
36. Murakami K, Miyake Y, Sasaki S, *et al.* (2011) Characteristics of under- and over-reporters of energy intake among Japanese children and adolescents: The Ryukyus Child Health Study. *Nutrition* **28**, 532–538.
37. Baxter SD, Smith AF, Litaker MS, *et al.* (2004) Recency affects reporting accuracy of children's dietary recalls. *Ann Epidemiol* **14**, 385–390.
38. Freedson P, Pober D & Janz KF (2005) Calibration of accelerometer output for children. *Med Sci Sports Exerc* **37**, S523–S530.
39. Svensson V, Jacobsson JA, Fredriksson R, *et al.* (2011) Associations between severity of obesity in childhood and adolescence, obesity onset and parental BMI: a longitudinal cohort study. *Int J Obes (Lond)* **35**, 46–52.
40. Lanctot JQ, Klesges RC, Stockton MB, *et al.* (2008) Prevalence and characteristics of energy underreporting in African-American girls. *Obesity (Silver Spring)* **16**, 1407–1412.
41. Vagstrand K, Lindroos AK & Linne Y (2009) Characteristics of high and low energy reporting teenagers and their relationship to low energy reporting mothers. *Public Health Nutr* **12**, 188–196.
42. Savage JS, Mitchell DC, Smiciklas-Wright H, *et al.* (2008) Plausible reports of energy intake may predict body mass index in pre-adolescent girls. *J Am Diet Assoc* **108**, 131–135.
43. Sichert-Hellert W, Kersting M, Manz F, *et al.* (1994) Energy intake of children and adolescents aged 1–18 years: nutrition survey versus recommendation. *Bibl Nutr Dieta* 45–48.
44. Haraldsdottir J & Sandstrom B (1994) Detection of underestimated energy intake in young adults. *Int J Epidemiol* **23**, 577–582.

British Journal of Nutrition

B.4 Diet-obesity associations in children: approaches to counteract attenuation caused by misreporting

This study compared different approaches to account for misreporting in dietary data when investigating associations between dietary intakes and overweight/obesity. It was published in the peer-reviewed journal *Public Health Nutrition* (Börnhorst et al., 2012a).



Diet–obesity associations in children: approaches to counteract attenuation caused by misreporting

Claudia Börnhorst, Inge Huybrechts, Antje Hebestreit, Barbara Vanaelst, Dénes Molnár, Silvia Bel-Serrat, Theodora Mouratidou, Luis A Moreno, Valeria Pala, Marge Eha, Yiannis A Kourides, Alfonso Siani, Gabriele Eiben and Iris Pigeot

Public Health Nutrition / *FirstView* Article / January 2006, pp 1 - 11
DOI: 10.1017/S1368980012004491, Published online:

Public Health Nutrition: page 1 of 11 doi:10.1017/S1368980012004491

Diet–obesity associations in children: approaches to counteract attenuation caused by misreporting

Claudia Börnhorst[1], Inge Huybrechts[2,3], Antje Hebestreit[1], Barbara Vanaelst[2,4], Dénes Molnár[5], Silvia Bel-Serrat[6], Theodora Mouratidou[6], Luis A Moreno[6], Valeria Pala[7], Marge Eha[8], Yiannis A Kourides[9], Alfonso Siani[10], Gabriele Eiben[11] and Iris Pigeot[1,*], on behalf of the IDEFICS consortium

[1]BIPS – Institute for Epidemiology and Prevention Research, Achterstrasse 30, 28359 Bremen, Germany: [2]Department of Public Health, Ghent University, Ghent, Belgium: [3]International Agency for Research on Cancer (IARC), Dietary Exposure Assessment Group (DEX), Lyon, France: [4]Research Foundation–Flanders (FWO), Brussels, Belgium: [5]Medical Faculty, Department of Pediatrics, University of Pécs, Pécs, Hungary: [6]GENUD (Growth, Exercise, Nutrition and Development) Research Group, University of Zaragoza, Zaragoza, Spain: [7]Fondazione IRCSS, Istituto Nazionale dei Tumori, Department of Preventive and Predictive Medicine, Nutritional Epidemiology Unit, Milan, Italy: [8]Department of Surveillance and Evaluation, National Institute for Health Development, Tallinn, Estonia: [9]Research and Education Foundation of Child Health, Paphos, Cyprus: [10]Institute of Food Sciences, CNR, Avellino, Italy: [11]Department of Public Health and Community Medicine, University of Gothenburg, Gothenburg, Sweden

Submitted 23 February 2012: Final revision received 8 August 2012: Accepted 24 August 2012

Abstract

Objective: Measurement errors in dietary data lead to attenuated estimates of associations between dietary exposures and health outcomes. The present study aimed to compare and evaluate different approaches of handling implausible reports by exemplary analysis of the association between dietary intakes (total energy, soft drinks, fruits/vegetables) and overweight/obesity in children.
Design: Cross-sectional multicentre study.
Setting: Kindergartens/schools from eight European countries participating in the IDEFICS Study.
Subjects: Children (n 5357) aged 2–9 years who provided one 24 h dietary recall and complete covariate information.
Results: The 24 h recalls were classified into three reporting groups according to adapted Goldberg cut-offs: under-report, plausible report or over-report. In the basic logistic multilevel model (adjusted for age and sex, including study centre as random effect), the dietary exposures showed no significant association with overweight/obesity (energy intake: OR=0·996 (95 % CI 0·983, 1·010); soft drinks: OR = 0·999 (95 % CI 0·986, 1·013)) and revealed even a positive association for fruits/vegetables (OR = 1·009 (95 % CI 1·001, 1·018)). When adding the reporting group (dummy variables) and a propensity score for misreporting as adjustment terms, associations became significant for energy intake as well as soft drinks (energy: OR = 1·074 (95 % CI 1·053, 1·096); soft drinks: OR = 1·015 (95 % CI 1·000, 1·031)) and the association between fruits/vegetables and overweight/obesity pointed to the reverse direction compared with the basic model (OR = 0·993 (95 % CI 0·984, 1·002)).
Conclusions: Associations between dietary exposures and health outcomes are strongly affected or even masked by measurement errors. In the present analysis consideration of the reporting group and inclusion of a propensity score for misreporting turned out to be useful tools to counteract attenuation of effect estimates.

Keywords
Adjustment
Cross-sectional
Propensity score
Reporting group
24 h Dietary recall

Measurement errors in dietary variables pose a challenge for epidemiologists when investigating associations between dietary intakes and health outcomes[1]. Problems in particular emerge from misreporting, which comprises under-reporting and over-reporting. Several studies have revealed that misreporting is characteristic to specific individuals and results in differential errors[2–4]. Differential errors are related to the outcome of interest and induce bias such that associations between dietary factors and health outcomes may be attenuated, exaggerated or hidden[5],

*Corresponding author: Email pigeot@bips.uni-bremen.de

whereas non-differential (random) errors tend to attenuate associations. Various procedures have been proposed to screen out implausible dietary recalls[6,7] but the question how to handle recalls identified as implausible is still open.

Researchers commonly refer to validation studies that confirm the accuracy/reliability of their assessment instruments but do not consider misreporting in the later analyses, although there are different procedures that could be applied[8,9]: (i) exclusion of inaccurate recalls; (ii) adjustment for the reporting group (under-report, plausible report, over-report); (iii) stratified analysis by reporting group; and (iv) propensity score adjustment.

Despite several studies having found that exclusion of under-reports strengthened diet–obesity relationships[3,10,11], data exclusions may introduce a source of unknown bias and has not been recommended[12]. Adjusting for the reporting group seems an appropriate alternative to data exclusions and was shown to yield consistent results compared with those obtained from plausible reports in stratified analyses[10]. Although not applied in this context yet, the propensity score is a common tool to reduce bias by equating groups based on selected covariables. A propensity score reflects the conditional probability of assignment to a particular group given a vector of observed covariables[13]. Construction of a propensity score based on variables previously found to be related to misreporting could be another option to account for implausible recalls.

Studies in adults investigating the handling of implausible recalls are rare[8,9,14]. To the authors' knowledge, no study to date has addressed this issue in children. As dietary recalls in young children often rely on proxy reports[15], it is likely that misreporting is triggered by different factors compared with adults (e.g. unintentional under-reporting due to lack of parental control). The present study aimed to evaluate the four different approaches to account for misreporting in the statistical analysis mentioned above and finally to give recommendations on how to handle the problem of inaccurate reports in future studies on dietary behaviour in children.

Materials and methods

Study population

IDEFICS (Identification and prevention of Dietary- and lifestyle-induced health EFfects In Children and infantS) is a multicentre, setting-based study aiming to prevent and investigate the causes of diet- and lifestyle-related diseases like overweight and obesity in European children aged 2–9 years. The baseline survey was conducted from September 2007 to June 2008; more than 31 500 children were contacted, out of whom finally 16 220 participated and fulfilled the inclusion criteria of the IDEFICS Study. Children were recruited through kindergartens/schools. In addition to self-completion questionnaires, interviews with parents concerning lifestyle habits and dietary intakes as well as anthropometric measurements and examinations of the children were conducted in examination centres, which were the settings in most countries. All measurements were taken by trained study personnel using standardised procedures in all eight study centres (Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain and Sweden). Details on the design and objectives of the study are given elsewhere[16,17].

Ethics approval

Applicable institutional and governmental regulations regarding the ethical use of human volunteers were followed during this research. Approval of the appropriate ethics committees was obtained by each of the eight participating centres carrying out the fieldwork (Belgium: Ethics Committee, University Hospital, Ghent; Cyprus: Cyprus National Bioethics Committee; Estonia: Tallinn Medical Research Ethics Committee; Germany: Ethics Committee, Universtiy of Bremen; Hungary: Egészségügyi Tudományos Tanács, Pécs; Italy: Comitato Etico, Avellino; Spain: Comité Ético de Investigación, Clínica de Aragón (CEICA); Sweden: Regional Ethics Review Board, University of Gothenburg).

Parents provided written informed consent for all examinations. Each child was informed orally about the modules by field workers and asked for his/her consent immediately before examination[17]. Study children did not undergo any procedure before both they and their parents gave consent for examinations, collection of samples, subsequent analysis and storage of personal data and collected samples. Participants and their parents could consent to single components of the study while abstaining from others.

Anthropometry

Height (centimetres) of the children was measured to the nearest 0·1 cm with a calibrated statiometer (Seca 225; Seca, Birmingham, UK); body weight (kilograms) was measured in light underwear on a calibrated scale accurate to 0·1 kg (Tanita BC 420 SMA; Tanita Europe GmbH, Sindelfingen, Germany). BMI was calculated as weight divided by height squared and the children were categorised according to the International Obesity Taskforce criteria[18,19]. According to these criteria, centile curves corresponding to a BMI of 25 kg/m^2 and 30 kg/m^2 at age 18 years are chosen as extrapolation into childhood of the well-accepted adult cut-offs to define overweight/obesity, respectively. Thin and normal-weight children, as well as overweight and obese children, were combined into one category each to construct a binary outcome measure to be included in the logistic model.

Dietary data

Dietary data were assessed using the computerised 24 h dietary recall (24-HDR) SACINA (Self-Administered Children and Infants Nutrition Assessment), which is

based on the previously designed and validated HELENA-DIAT[20] instrument that was originally developed for Flemish adolescents[21]. SACINA is structured according to six meal occasions (breakfast, morning snack, lunch, afternoon snack, dinner, evening snack) related to a range of chronological daily activities. For each food item the participant selects the consumed quantity by means of pictures with increasing portion sizes (based on predefined standard amounts) that are displayed on the screen to facilitate estimation of portion sizes. The intake of the food item is calculated then as the product of the reported quantity and the standard amount (e.g. 4 spoons of sauce at 15 g = 60 g). Proxies, mainly the parents, completed the 24-HDR under supervision of field personnel which lasted 20–30 min. In case the child had lunch at school on weekdays, school meals were additionally assessed by means of direct observation. Trained observers, teachers or caregivers entered portion sizes of all consumed foods and drinks on predefined assessment sheets. The uniquely coded food items were linked to country-specific food composition tables. Missing quantities for single food items as well as obviously implausible data entries were imputed by country-, food group- and age-specific median intakes (0·01 % of the entries) to avoid excessive recall exclusions. Incomplete interviews were excluded, e.g. if the proxy did not know about at least one main meal or in the case of missing school meal information (*n* 2518). Furthermore, intakes of energy >16 736 kJ/d (>4000 kcal/d) which seemed to be a result of computer or data-entry errors rather than of misreporting (e.g. several repeated entries for the same food item) were excluded (*n* 10). Although up to six repeated 24-HDR were carried out in a smaller sample, only the first recall day was included in the current analysis (including weekdays and weekend days) to obtain an equal number of 24-HDR for each child. The assessment procedure was slightly different in the Hungarian study centre, where dietary recalls were not performed via the standardised SACINA software but via paper-and-pencil 24-HDR registrations that were entered in the SACINA software afterwards. As this increased data heterogeneity and further seemed to affect the misreporting behaviour, data from Hungary were not considered in the present analyses. A study sample based on equal procedures and standardised assessment instruments was needed for this exploratory methodological study.

Energy intake (EI; kJ/d), fruit/vegetable intake and soft drink intake (as a percentage of total daily EI; %EI) were used as exposure measures in the different models as these were repeatedly proposed to be associated with overweight/obesity[22–24].

Statistical methods

Classification of 24 h dietary recalls

The BMR was estimated from the equations published by Schofield[25] and recommended by the FAO/WHO/United Nations University (1985) taking into account age, sex, body height and weight. To determine whether reported EI was consistent with energy requirements, the ratio of proxy-reported EI to predicted BMR was used to classify the 24-HDR into under-reports (UdR), plausible reports (PR) and over-reports (OvR) according to Goldberg *et al.*[6]. Since the original Goldberg cut-offs were developed for adults and do not consider differences in EI due to age and sex, cut-off values were re-calculated for application in children as suggested previously[2,26] using the formula:

$$\text{Cut-off} = \text{PAL} \times \exp\left[\pm 1.96 \times \frac{(S/100)}{\sqrt{n}}\right],$$

where

$$S = \sqrt{\frac{\text{CV}_{\text{wEI}}^2}{d} + \text{CV}_{\text{wBMR}}^2 + \text{CV}_{\text{PA}}^2}.$$

The within-subject CV for EI (CV_{wEI}), the within-subject CV for BMR (CV_{wBMR}) and the CV for physical activity (CV_{PA}) were replaced by age- and sex-specific values as given in Nelson *et al.*[27] and Black *et al.*[28]. Goldberg's overall physical activity level (PAL) of 1·55 was substituted by age- and sex-dependent levels of light physical activity (2–5 years: 1·45; 6–10 years: males 1·55, females 1·50) according to Torun *et al.*[29]. The number of days (*d*) was set to 1 (one 24-HDR per child) to account for the large day-to-day variation in diet. Cut-off limits need to be wider if only one or few recall days are available as these may not reflect usual intakes but exceptional days. The resulting age- and sex-specific cut-off values to define UdR, PR and OvR are given in Table 1, which were then used to classify the recalls accordingly.

Calculation of the propensity score

In a previous study based on the IDEFICS data[30], backward elimination in the course of multilevel logistic regression analysis was applied to identify factors significantly related to misreporting in proxy reports for young children. The covariables that turned out to be significantly associated with misreporting were used in the construction of the propensity score: age and sex of the child[31,32], net household income (dummy: high *v.* medium/low), number of persons below 18 years of age in the household and day

Table 1 Lower and upper cut-off limits to classify 1 d 24-HDR as UdR or OvR based on EI:BMR

Age (years)	Sex	Lower cut-off (UdR)	Upper cut-off (OvR)
2–<6	Boys	0·74	2·85
2–<6	Girls	0·78	2·69
6–<10	Boys	0·92	2·61
6–<10	Girls	0·93	2·43

24-HDR, 24 h dietary recall; UdR, under-report; OvR, over-report; EI, energy intake.
PR (plausible report) has EI:BMR within the cut-offs.

of the interview (dummy: weekday *v.* Saturday/Sunday). The following information on parental concerns and perception of their child's weight status obtained from a self-administered proxy questionnaire was included: 'How concerned are you about your child… (i) becoming overweight?'; (ii) becoming underweight?' (response categories were 'unconcerned', 'a little concerned', 'concerned' and 'very concerned'); 'Do you think your child is… (i) 'much too underweight?'; (ii) 'slightly too underweight?'; (iii) 'proper weight?'; (iv) 'slightly too overweight'; (v) 'much too overweight?' (response categories were 'yes' and 'no'). Further intakes from the following food items commonly perceived to be healthy/unhealthy were considered as predictors for misreporting: chocolate products, other sugary products (e.g. cakes, biscuits, ice cream), soft drinks, fruits/vegetables, milk (all as %EI) and water (g/d). Although BMI is a repeatedly shown predictor of misreporting, it was not included in the construction of the propensity score as the weight status is the outcome variable in the present analysis.

The conditional probability (propensity score) of being classified as UdR given the mentioned covariables was calculated applying a logistic multilevel regression model including all covariates mentioned above as fixed effects and the study centre as random effect:

$$\text{Propensity score} = \text{estimated } P(\text{UdR} \mid \text{covariates}).$$

Fruit/vegetable intake was not included as a covariable in the propensity score calculation when investigating diet–obesity models using fruit/vegetable intake as exposure variable. Analogously, soft drink intake was not considered in the construction of the propensity score when investigating models using soft drink intake as exposure.

Model building

Associations between overweight/obesity and dietary intakes were exemplarily analysed to investigate different procedures of handling implausible dietary recalls. Logistic multilevel regression analyses were conducted using a dummy indicating overweight/obesity as outcome and the three dietary variables as exposure measures: EI in kJ/d (models labelled with 'a'), %EI from fruits/vegetables (labelled with 'b') and %EI from soft drinks (labelled with 'c').

The first model (basic model) included only adjustment terms for age and sex and a random effect for the study centre to account for the clustered study design (Model 1a–c). The basic model was also run adding all variables used in the calculation of the propensity score as potential confounders (Model 2a–c). Model 3 was identical to the basic model but here recalls classified as UdR and OvR were excluded. Further, the basic model was run adjusting additionally for the reporting group (Model 4a–c), for the propensity score (Model 5a–c) or for both (Model 6a–c). In addition, the basic model was analysed stratified by reporting group (Model 7a–c) as well as stratified by reporting group and at the same time adjusted for the propensity score (Model 8a–c).

The current analysis includes only children with 24-HDR and complete covariate information (n 5962). All analyses were performed using the statistical software package SAS version 9·1.

Results

Descriptive analyses of the study population and all covariables used for the construction of the propensity score are presented in Table 2 (categorical variables) and Table 3 (continuous variables). Regarding the total study group, 6·7 % (n 402) of the proxy reports were classified as UdR and 4·0 % (n 241) as OvR. Both UdR and OvR were slightly higher in girls compared with boys and higher in the low/medium compared with the high income group. Percentages of UdR were higher in overweight/obese children, in the older age group (6 to <10 years), on weekend days and if proxies were concerned about their child becoming overweight or perceived their child to be slightly/much too overweight. OvR, on the other hand, was higher in thin/normal-weight children, on weekend days or if proxies were concerned about their child becoming underweight. %EI from fruits/vegetables was highest in UdR whereas %EI from chocolate and other sugary products were highest in OvR. Soft drink consumption was slightly lower in the OvR group compared with the UdR and PR groups.

Tables 4 and 5 show the odds ratios and 95 % confidence intervals obtained from the different models for the association between overweight/obesity and the three dietary exposures. Effects of continuous variables are assessed as 1-unit offsets from the mean; e.g. the OR for the association between overweight/obesity and %EI from fruits/vegetables indicates the increase in risk when increasing %EI from fruits/vegetables by 1 % compared with the mean of the total study population.

In the basic model (Table 4, Models 1a–c), odds ratios were not significant for EI and soft drink intake and indicated even a significant positive association between overweight/obesity and fruit/vegetable intake (OR = 1·009, 95 % CI 1·001, 1·018). Adjustment for covariables (Models 2a–c) revealed similar results, but the association between fruits/vegetables and overweight/obesity was rendered insignificant here (OR = 1·009 (95 % CI 0·998, 1·020)). When excluding UdR and OvR (Models 3a–c), a significantly positive association between EI and overweight/obesity was observed (OR = 1·057, 95 % CI 1·038, 1·076). Adjustment for the reporting group (Models 4a–c) also revealed a significantly positive association between EI and overweight/obesity that was even slightly more pronounced compared with the model excluding misreports. When adjusting for the propensity score, all associations were strengthened (Models 5a–c) with the association

Public Health Nutrition

Table 2 Descriptive analyses of categorical covariables stratified by reporting group (total numbers and row percentages): children aged 2–9 years, IDEFICS Study

	Total	UdR		PR		OvR	
	n	*n*	%	*n*	%	*n*	%
All	5962	402	6·7	5319	89·2	241	4·0
Sex of the child							
Male	3029	187	6·2	2747	90·7	95	3·1
Female	2933	215	7·3	2572	87·7	146	5·0
Age groups							
2–<6 years	2625	120	4·6	2388	91·0	117	4·5
6–<10 years	3337	282	8·5	2931	87·8	124	3·7
Weight status*							
Thin/normal weight	4721	249	5·3	4263	90·3	209	4·4
Overweight/obese	1241	153	12·3	1056	85·1	32	2·6
Study centre							
Belgium	310	29	9·4	274	88·4	7	2·3
Cyprus	403	63	15·6	335	83·1	5	1·2
Estonia	602	35	5·8	537	89·2	30	5·0
Germany	1504	159	10·6	1290	85·8	55	3·7
Italy	1492	68	4·6	1320	88·5	104	7·0
Spain	525	7	1·3	492	93·7	26	5·0
Sweden	1126	41	3·6	1071	95·1	14	1·2
Income							
Low/medium	4304	322	7·5	3786	88·0	196	4·6
High	1658	80	4·8	1533	92·5	45	2·7
Day of the interview							
Weekday	4925	319	6·5	4415	89·6	191	3·9
Saturday/Sunday	1037	83	8·0	904	87·2	50	4·8
Concerned: child becoming underweight							
Unconcerned	3109	230	7·4	2796	89·9	83	2·7
A little concerned	923	57	6·2	825	89·4	41	4·4
Concerned	863	52	6·0	751	87·0	60	7·0
Very concerned	1067	63	5·9	947	88·8	57	5·3
Concerned: child becoming overweight							
Unconcerned	3299	182	5·5	2996	90·8	121	3·7
A little concerned	1001	73	7·3	879	87·8	49	4·9
Concerned	878	70	8·0	774	88·2	34	3·9
Very concerned	784	77	9·8	670	85·5	37	4·7
Health: child's weight							
Much too underweight	77	6	7·8	66	85·7	5	6·5
Slightly too underweight	944	48	5·1	836	88·6	60	6·4
Proper weight	4204	234	5·6	3812	90·7	158	3·8
Slightly too overweight	679	100	14·7	564	83·1	15	2·2
Much too overweight	58	14	24·1	41	70·7	3	5·2

UdR, under-report; PR, plausible report; OvR, over-report.
*Weight categories according to International Obesity Taskforce criteria[18,19].

Table 3 Descriptive analyses of continuous covariables stratified by reporting group (means and standard deviations): children aged 2–9 years, IDEFICS Study

	Total group (*n* 5962)		UdR (*n* 402)		PR (*n* 5319)		OvR (*n* 241)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age (years)	6·06	1·82	6·64	1·54	6·02	1·84	5·96	1·76
BMI *Z*-score*	0·31	1·34	0·82	1·60	0·29	1·31	−0·01	1·24
EI (kJ/d)	6602	2218	3197	1021	6632	1807	11 590	1833
EI (kcal/d)	1578	530	764	244	1585	432	2770	438
Water intake (g/d)	319	357	284	346	317	352	419	462
%EI from chocolate	3·2	5·9	2·7	6·5	3·2	5·9	3·5	6·7
%EI from milk	10·1	9·4	8·5	11·3	10·3	9·2	7·9	8·7
%EI from soft drinks	2·7	5·7	2·7	6.7	2·7	5·7	2·2	4·0
%EI from sugary products	9·8	11·7	7·4	11·8	9·8	11·6	12·7	12·4
%EI from fruits/vegetables	8·5	8·2	11·1	13·0	8·4	7·7	6·7	6·0

UdR, under-report; PR, plausible report; OvR, over-report; EI, energy intake; %EI, percentage of energy intake.
*According to Cole *et al.*[31,32].

Table 4 OR and 95 % CI for the associations between overweight/obesity and EI (Model 1a to 6a), %EI from fruits/vegetables (Model 1b to 6b) and %EI from soft drinks (Model 1c to 6c) in different models: children aged 2–9 years, IDEFICS Study

	Basic model		Basic model adjusted for covariables		Exclusion of misreports		Adjustment for reporting group		Adjustment for propensity score		Adjustment for reporting group and propensity score	
	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI
	Model 1a*		Model 2a†		Model 3a‡		Model 4a§		Model 5a‖		Model 6a¶	
EI (1 unit = 418·4 kJ (100 kcal))	0·996	0·983, 1·010	1·013	0·995, 1·031	1·057	1·038, 1·076	1·068	1·049, 1·086	1·019	1·005, 1·034	1·074	1·054, 1·095
PR *v.* UdR							0·205	0·155, 0.271			0·390	0·280, 0·542
OvR *v.* UdR							0·041	0·023, 0·073			0·076	0·041, 0·142
Propensity score									1·222	1·202, 1·243	1·217	1·038, 1·402
	Model 1b*		Model 2b†		Model 3b‡		Model 4b§		Model 5b‖		Model 6b¶	
%EI from fruits/vegetables	1·009	1·001, 1·018	1·009	0·998, 1·020	1·007	0·998, 1·017	1·006	0·997, 1·014	0·994	0·985, 1·003	0·993	0·984, 1·002
PR *v.* UdR							0·365	0·289, 0·461			0·710	0·532, 0·948
OvR *v.* UdR							0·154	0·099, 0·242			0·298	0·181, 0·491
Propensity score									1·250	1·227, 1·274	1·245	1·222, 1.269
	Model 1c*		Model 2c†		Model 3c‡		Model 4c§		Model 5c‖		Model 6c¶	
%EI from soft drinks	0·999	0·986, 1·013	0·996	0·982, 1·011	0·996	0·982, 1·011	1·001	0·988, 1·015	1·016	1·000, 1·031	1·015	1·000, 1·031
PR *v.* UdR							0·359	0·285, 0·453			0·692	0·520, 0·921
OvR *v.* UdR							0·151	0·097, 0·237			0·307	0·188, 0·504
Propensity score									1·231	1·210, 1·253	1·226	1·205, 1·248

EI, energy intake; %EI, percentage of energy intake; PR, plausible report; UdR, under-report; OvR, over-report.
Effects of continuous variables are assessed as 1-unit offsets from the mean. Due to the small scale of the propensity score, 0.01-unit offsets from mean were chosen here.
*Basic model: logistic multilevel regression model; OR for the association between overweight/obesity and food intake adjusted for age and sex and including the study centre as random effect (*n* 5962).
†Basic model additionally adjusted for net household income (dummy: high *v.* medium/low), number of persons below 18 years of age in the household, day of the interview (dummy: weekday *v.* Saturday/Sunday), information on parental concerns and perception regarding their child's weight status and reported intakes from food groups associated with misreporting.
‡Basic model, but excluding UdR and OvR (*n* 5319).
§Basic model adjusted for the reporting group (UdR, PR, OvR).
‖Basic model adjusted for a propensity score for misreporting.
¶Basic model adjusted for the reporting group and for the propensity score for misreporting.

Table 5 OR and 95 % CI for the association between overweight/obesity and EI (Model 7a, 8a), %EI from fruits/vegetables (Model 7b, 8b) and %EI from soft drinks (Model 7c, 8c) in different models stratified by reporting group (UdR, PR, OvR): children aged 2–9 years, IDEFICS Study

	Stratification						Stratification and adjustment for propensity score					
	UdR		PR		OvR		UdR		PR		OvR	
	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI	OR	95 % CI
	Model 7a*						Model 8a†					
EI (1 unit = 4184 kJ (100 kcal))	1·422	1·237, 1·634	1·057	1·038, 1·076	1·268	1·142, 1·407	1·521	1·291, 1·793	1·064	1·043, 1·085	1·310	1·167, 1·470
Propensity score							1·147	1·108, 1·188	1·242	1·217, 1·268	1·208	1·093, 1·336
	Model 7b*						Model 8b†					
%EI from fruits/vegetables	1·006	0·988, 1·024	1·007	0·998, 1·017	0·921	0·845, 1·003	0·995	0·976, 1·014	0·993	0·983, 1·004	0·895	0·814, 0·983
Propensity score							1·180	1·133, 1·229	1·267	1·240, 1·295	1·218	1·101, 1·348
	Model 7c*						Model 8c†					
%EI from soft drinks	1·027	0·993, 1·061	0·996	0·982, 1·011	1·010	0·896, 1·139	1·042	1·004, 1·081	1·010	0·993, 1·027	1·039	0·914, 1·182
Propensity score							1·150	1·110, 1·191	1·250	1·224, 1·276	1·189	1·085, 1·303

EI, energy intake; %EI, percentage of energy intake; UdR, under-report; PR, plausible report; OvR, over-report.
Effects of continuous variables are assessed as 1-unit offsets from the mean. Due to the small scale of the propensity score, 0·01-unit offsets from mean were chosen here.
*Basic model: logistic multi-level regression model stratified by reporting group (UdR, PR, OvR); OR for the association between overweight/obesity and dietary intakes adjusted for age and sex and including the study centre as random effect.
†Basic model adding the propensity score as adjustment term.

between overweight/obesity and fruit/vegetable intake being reversed compared with the basic model. Significant associations were found between overweight/obesity and EI as well as soft drink intake. Finally, adjustment for the reporting group and propensity score at the same time strengthened the association between overweight/obesity and EI whereas the other associations remained nearly unchanged (Models 6a–c) compared with the model adjusting only for the propensity score.

When stratifying the basic model by the reporting group (Table 5, Model 7a–c), only EI was significantly related to overweight/obesity in all three strata. Additional adjustment for the propensity score (Model 8a–c) strengthened associations between all three dietary exposures and overweight/obesity. Here a significant reverse association between fruit/vegetable intake and overweight/obesity was observed in OvR and a positive association was found between soft drinks and overweight/obesity in UdR. The relationship between overweight/obesity and EI was much stronger in the UdR and OvR groups compared with PR.

Discussion

To the authors' knowledge, the present study is the first one in children applying and comparing several statistical approaches to counteract attenuation of risk estimates caused by misreporting of dietary information. Negligence of misreporting in the statistical model revealed insignificant or even (unexpected) reversed diet–obesity associations. Consistent with previous findings on differential misreporting by weight status[33], the UdR group had higher mean BMI *Z*-scores but reported lower (implausible) EI compared with PR. The opposite was true for the OvR group. Such reporting bias may obscure positive relationships between diet and weight status. Researchers should be aware that results may differ strongly depending on the statistical model selected and that the choice of an adequate model needs to be taken thoroughly. Consideration of misreporting in any way yielded results more consistent with hypotheses relating food intake to overweight/obesity[34,35]. However, the true effects remained unknown due to the lack of validation data. A recent study reported that not excluding implausible reports resulted in weak, non-significant or even misleading associations between BMI and diet[9], whereas Nielsen and Adair stated that examining all data but stratifying by level of intake may be more informative for population nutrient intake than exclusion of misreports[8]. Savage *et al.* found a significant association between BMI and reported EI in the PR of pre-adolescent girls, but neither in the total study group nor when analysing only misreports (combining UdR and OvR into one group)[36]. This agrees with our results for the total study group (basic model). Nevertheless, our stratified analysis revealed statistically significant associations between overweight/obesity and EI in all three reporting

groups, being even stronger in UdR and OvR compared with PR. This may be explained by either: (i) differences in the mean intake levels to which the effects are put into relation (mean EI: 3197 kJ/d (764 kcal/d) in UdR, 6632 kJ/d (1585 kcal/d) in PR, 11 590 kJ/d (2770 kcal/d) in OvR); or (ii) differences between the reporting groups in terms of participants' characteristics (e.g. prevalence of overweight/obesity: 38·0 % in UdR, 19·9 % in PR, 13·2 % in OvR). Our results argue against combining UdR and OvR into one group in stratified analyses as determinants of misreporting and participants' characteristics are likely to differ[30]. Moreover, the differences between the groups of UdR, PR and OvR suggest that data exclusions may actually introduce a selection bias, so that exclusion of misreports is not recommended. However, the reduced sample sizes resulting from both data exclusions and stratification go along with limited statistical power especially in the (smaller) groups of UdR and OvR. Adjustment for the reporting group does not affect the statistical power to such a degree and shifted associations between overweight/obesity and all three dietary exposures to the expected directions (Models 4a–c). These results agree with those from a study by Mendez *et al.*[10] where associations between different food groups and overweight/obesity became stronger after inclusion of dummy variables identifying under- and over-reports. In that study, dummy adjustment revealed results similar to those obtained when limiting the analysis sample to plausible reports, as observed in our study. However, this approach has the disadvantage of misclassifications of single recalls being quite likely, which may again bias the results[37].

After adjustment for the propensity score, which combined various indicators for misreporting into one summary measure, associations between overweight/obesity and soft drink as well as fruit/vegetable intakes increased markedly. To correct for selective reporting of single food items, also dietary variables commonly associated with misreporting were included when constructing the propensity score. This approach strived for an effect similar to regression calibration[38] although both procedures differ. The idea of calibration in general is the replacement of exposures measured with error by 'adjusted' values using additional information obtained from biomarker measurements or from a second dietary assessment instrument. Common calibration approaches assume (non-differential) linear measurement error with constant variance or linear random within-person error in the case of replicate measurements (e.g. repeated 24-HDR)[38–40] – assumptions that are often violated due to differential misreporting[4,41]. Moreover, error structures were found to be correlated when assessing dietary information via different assessment methods (e.g. FFQ and 24-HDR)[42]. Although the use of two complementary dietary assessment methods is recommended e.g. when investigating usual intakes[43,44], the benefit of a second assessment instrument to correct for misreporting is questionable[45]. Further studies are needed to explore and compare the calibration and propensity score approach. However, it can be suspected that statistical adjustment of relative risks based on biomarker data with independent error structures (e.g. doubly labelled water for EI) incorporating characteristics of misreporters should be preferred if such data exist[1,39,46]. In the absence of validation data, the propensity score seems to be a useful, cost-effective alternative to account for misreporting.

In our models, intakes from soft drinks and fruits/vegetables were examined in relation to total daily intake of energy (expressed as %EI) instead of including absolute amounts (g/d). Use of absolute amounts would result in lower effects in high energy consumers compared with low energy consumers[3,47]. To overcome this problem, different energy adjustment models have been proposed next to the one applied here[48]. But again energy adjustment cannot eliminate differential biases[3] and is therefore not sufficient to correct for subject-specific and selective misreporting of certain foods/macronutrients[45,49]. The advantage of additional incorporation of the propensity score over simple energy-adjustment methods is that the propensity score is a comprehensive approach to account for several covariables related to misreporting instead of considering only the level of EI. Under-reporting is difficult to distinguish from undereating (defined as eating less than required to maintain body weight, accompanied by weight loss) but both are treated equally in energy-adjustment models, while it can be hypothesised that subject characteristics and therefore propensity scores differ between undereaters and under-reporters. Nevertheless, in the case of non-differential errors energy-adjustment methods were shown to be a good approach to counteract underestimation of relative risks and reduction of statistical power[1].

Several sensitivity analyses were carried out (e.g. including only children with two repeated 24-HDR (n 904), excluding OvR (n 241), excluding UdR (n 402), excluding 24-HDR with at least one imputed value (n 69), excluding thin (n 556) or obese children (n 430)). When including only children with two repeated 24-HDR, model estimates became unstable due to the reduction in sample size. In all other cases, results remained nearly unchanged compared with the results given here. Details can be obtained from the author on request.

The present analysis is based on data in children relying on proxy reports. Here misreporting may result not only from intentional misreporting, e.g. caused by social desirability or parental concerns about their child's weight status, but also from unintentional misreporting due to lack of parental control (out-of-home meals). Our discussion mainly refers to studies in adolescents/adults as related studies are lacking in children. Although determinants for misreporting may differ between children and adolescent/adult populations, previous studies and the present one reveal similar results concerning the statistical approaches

of data exclusions, stratification or adjustment for the reporting group. Nevertheless, results of the newly applied propensity score approach should not simply be transferred. When applying the propensity score approach in future studies, variables for the construction of the score should be selected depending on the study population under investigation, which may require a pre-study to identify the relevant determinants of misreporting. The analysis of the usefulness of the propensity score adjustment in adolescent/adult populations is a task for future research.

Limitations and strengths

Only one recall day per child was used in the present analysis which does not reflect usual intakes due to the day-to-day variation that characterises dietary data in general[50]. Day-to-day variation results in random (non-differential) errors that may have weakened associations between dietary factors and overweight/obesity. In addition, extreme intakes may not necessarily reflect misreporting but rather specific diets (e.g. energy restricted) or exceptional days (e.g. the child was ill or extremely physically active). Reverse causation cannot be precluded as obesity may even cause low intakes due to dieting or change in eating behaviour. Causal inference is limited owing to the cross-sectional study design.

Sensitivity of the cut-off technique to correctly classify UdR and OvR is limited as it aims only to identify misreports resulting in physiologically implausibly low/high EI[6]. By application of the cut-off technique distinction between varying degrees of misreporting is not feasible; e.g. under-reporting from a high intake level may not be detected as the reported intake may still be such high that EI:BMR does not fall below the cut-off. Furthermore, not considering individual physical activity levels of the children when classifying the 24-HDR is a limitation. Physically inactive children may have a very low daily energy expenditure making even low reported intakes plausible, whereas physically active children have an increased likelihood to be misclassified as OvR. Child-specific reference PAL were used in the calculation of the cut-offs to compensate for the lack of sufficient individual information on physical activity.

The study was a first exploratory approach to investigate the usefulness of propensity scores in the context of dietary misreporting in children. The authors are aware that there are several different ways to construct a propensity score by inclusion of additional/different variables, e.g. physical activity, number of daily meals, etc. The rather exploratory character of the paper should be underlined here. However, the application of the new propensity score approach, along with the large sample size, the variety of covariables and the standardised assessment procedures suggest that the present study provides important knowledge on methods to handle misreporting in future research, while also highlighting gaps in knowledge as starting points for further analyses.

Conclusions

Associations between dietary exposures and health outcomes are strongly affected or even masked or reversed by measurement errors. Instead of data exclusions that may result in unknown bias, misreporting should rather be addressed in the model building process including adjustment terms for misreporting. Dummy adjustment for the reporting group revealed associations more consistent with expectations, which was most pronounced considering the association between EI and overweight/obesity. However, more sophisticated adjustments seem to be necessary to counteract the effect of selective misreporting of other food groups. In this respect, the propensity score adjustment turned out to be a useful tool to correct for subject-specific misreporting as it combines various variables associated with misreporting into one scalar and should be further investigated in future studies.

Acknowledgements

Sources of funding: This work was done as part of the IDEFICS Study (www.idefics.eu) and is published on behalf of its European Consortium. Financial support was provided by the European Community within the Sixth RTD Framework Programme Contract No. 016181 (FOOD). S.B.-S. was funded by a grant from the Aragón's Regional Government (Diputación General de Aragón, DGA). B.V. was funded by the Research Foundation–Flanders (FWO). *Conflict of interest:* All the authors declare that there are no conflicts of interest. *Authors' contributions:* All authors contributed to conception and design, acquisition of data, analysis or interpretation of data. Each author has seen and approved the contents of the submitted manuscript. Final approval of the version published was given by all authors.

References

1. Freedman LS, Schatzkin A, Midthune D *et al.* (2011) Dealing with dietary measurement error in nutritional cohort studies. *J Natl Cancer Inst* **103**, 1086–1092.
2. Lioret S, Touvier M, Balin M *et al.* (2011) Characteristics of energy under-reporting in children and adolescents. *Br J Nutr* **105**, 1671–1680.
3. Livingston MB & Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* **133**, Suppl. 3, 895S–920S.
4. Black AE & Cole TJ (2001) Biased over- or under-reporting is characteristic of individuals whether over time or by different assessment methods. *J Am Diet Assoc* **101**, 70–80.
5. Shai I, Rosner BA, Shahar DR *et al.* (2005) Dietary evaluation and attenuation of relative risk: multiple comparisons between blood and urinary biomarkers, food frequency, and 24-hour recall questionnaires: the DEARR study. *J Nutr* **135**, 573–579.
6. Goldberg GR, Black AE, Jebb SA *et al.* (1991) Critical evaluation of energy intake data using fundamental principles of energy physiology: 1. Derivation of cut-off limits to identify under-recording. *Eur J Clin Nutr* **45**, 569–581.

7. McCrory MA, Hajduk CL & Roberts SB (2002) Procedures for screening out inaccurate reports of dietary energy intake. *Public Health Nutr* **5**, 873–882.
8. Nielsen SJ & Adair L (2007) An alternative to dietary data exclusions. *J Am Diet Assoc* **107**, 792–799.
9. Huang TT, Roberts SB, Howarth NC *et al.* (2005) Effect of screening out implausible energy intake reports on relationships between diet and BMI. *Obes Res* **13**, 1205–1217.
10. Mendez MA, Wynter S, Wilks R *et al.* (2004) Under- and overreporting of energy is related to obesity, lifestyle factors and food group intakes in Jamaican adults. *Public Health Nutr* **7**, 9–19.
11. Howarth NC, Huang TT, Roberts SB *et al.* (2005) Dietary fiber and fat are associated with excess weight in young and middle-aged US adults. *J Am Diet Assoc* **105**, 1365–1372.
12. Gibson RS (2005) *Principles of Nutritional Assessment*, 2nd ed. New York: Oxford University Press.
13. Rosenbaum PR & Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
14. Mendez MA, Popkin BM, Buckland G *et al.* (2011) Alternative methods of accounting for underreporting and overreporting when measuring dietary intake–obesity relations. *Am J Epidemiol* **173**, 448–458.
15. Livingstone MB & Robson PJ (2000) Measurement of dietary intake in children. *Proc Nutr Soc* **59**, 279–293.
16. Ahrens W, Bammann K, de Henauw S *et al.* (2006) Understanding and preventing childhood obesity and related disorders – IDEFICS: a European multilevel epidemiological approach. *Nutr Metab Cardiovasc Dis* **16**, 302–308.
17. Ahrens W, Bammann K, Siani A *et al.* (2011) The IDEFICS cohort: design, characteristics and participation in the baseline survey. *Int J Obes (Lond)* **35**, Suppl 1, S3–S15.
18. Cole TJ, Bellizzi MC, Flegal KM *et al.* (2000) Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ* **320**, 1240–1243.
19. Cole TJ, Flegal KM, Nicholls D *et al.* (2007) Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ* **335**, 194.
20. Vereecken CA, Covents M, Sichert-Hellert W *et al.* (2008) Development and evaluation of a self-administered computerized 24-h dietary recall method for adolescents in Europe. *Int J Obes (Lond)* **32**, Suppl. 5, S26–S34.
21. Vereecken CA, Covents M, Matthys C *et al.* (2005) Young adolescents' nutrition assessment on computer (YANA-C). *Eur J Clin Nutr* **59**, 658–667.
22. Zurriaga O, Perez-Panades J, Quiles Izquierdo J *et al.* (2011) Factors associated with childhood obesity in Spain. The OBICE study: a case–control study based on sentinel networks. *Public Health Nutr* **14**, 1105–1113.
23. O'Connor TM, Yang SJ & Nicklas TA (2006) Beverage intake among preschool children and its effect on weight status. *Pediatrics* **118**, e1010–e1018.
24. Alinia S, Hels O & Tetens I (2009) The potential association between fruit intake and body weight – a review. *Obes Rev* **10**, 639–647.
25. Schofield WN (1985) Predicting basal metabolic rate, new standards and review of previous work. *Hum Nutr Clin Nutr* **39**, Suppl. 1, 5–41.
26. Sichert-Hellert W, Kersting M & Schoch G (1998) Underreporting of energy intake in 1 to 18 year old German children and adolescents. *Z Ernahrungswiss* **37**, 242–251.
27. Nelson M, Black AE, Morris JA *et al.* (1989) Between- and within-subject variation in nutrient intake from infancy to old age: estimating the number of days required to rank dietary intakes with desired precision. *Am J Clin Nutr* **50**, 155–167.
28. Black AE (2000) Critical evaluation of energy intake using the Goldberg cut-off for energy intake:basal metabolic rate. A practical guide to its calculation, use and limitations. *Int J Obes Relat Metab Disord* **24**, 1119–1130.
29. Torun B, Davies PS, Livingstone MB *et al.* (1996) Energy requirements and dietary energy recommendations for children and adolescents 1 to 18 years old. *Eur J Clin Nutr* **50**, Suppl.1, S37–S80.
30. Börnhorst C, Huybrechts I, Ahrens W *et al.* (2012) Prevalence and determinants of misreporting among European children in proxy-reported 24-hour dietary recalls. *Br J Nutr* (Epublication ahead of print version).
31. Cole TJ, Freeman JV & Preece MA (1998) British 1990 growth reference centiles for weight, height, body mass index and head circumference fitted by maximum penalized likelihood. *Stat Med* **17**, 407–429.
32. Cole TJ, Freeman JV & Preece MA (1995) Body mass index reference curves for the UK, 1990. *Arch Dis Child* **73**, 25–29.
33. Heitmann BL, Lissner L & Osler M (2000) Do we eat less fat, or just report so? *Int J Obes Relat Metab Disord* **24**, 435–442.
34. Tohill BC, Seymour J, Serdula M *et al.* (2004) What epidemiologic studies tell us about the relationship between fruit and vegetable consumption and body weight. *Nutr Rev* **62**, 365–374.
35. Rolls BJ, Drewnowski A & Ledikwe JH (2005) Changing the energy density of the diet as a strategy for weight management. *J Am Diet Assoc* **105**, 5 Suppl. 1, S98–S103.
36. Savage JS, Mitchell DC, Smiciklas-Wright H *et al.* (2008) Plausible reports of energy intake may predict body mass index in pre-adolescent girls. *J Am Diet Assoc* **108**, 131–135.
37. Greenland S & Robins JM (1985) Confounding and misclassification. *Am J Epidemiol* **122**, 495–506.
38. Spiegelman D, McDermott A & Rosner B (1997) Regression calibration method for correcting measurement-error bias in nutritional epidemiology. *Am J Clin Nutr* **65**, 4 Suppl., 1179S–1186S.
39. Freedman LS, Midthune D, Carroll RJ *et al.* (2011) Using regression calibration equations that combine self-reported intake and biomarker measures to obtain unbiased estimates and more powerful tests of dietary associations. *Am J Epidemiol* **174**, 1238–1245.
40. Kaaks R, Riboli E & van Staveren W (1995) Calibration of dietary intake measurements in prospective cohort studies. *Am J Epidemiol* **142**, 548–556.
41. Livingstone MB, Robson PJ, Black AE *et al.* (2003) An evaluation of the sensitivity and specificity of energy expenditure measured by heart rate and the Goldberg cut-off for energy intake: basal metabolic rate for identifying mis-reporting of energy intake by adults and children: a retrospective analysis. *Eur J Clin Nutr* **57**, 455–463.
42. Kipnis V, Subar AF, Midthune D *et al.* (2003) Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol* **158**, 14–21.
43. Carroll RJ, Midthune D, Subar AF *et al.* (2012) Taking advantage of the strengths of 2 different dietary assessment instruments to improve intake estimates for nutritional epidemiology. *Am J Epidemiol* **175**, 340–347.
44. de Boer EJ, Slimani N, van't Veer P *et al.* (2011) The European Food Consumption Validation Project: conclusions and recommendations. *Eur J Clin Nutr* **65**, Suppl.1, S102–S107.
45. Westerterp KR & Goris AH (2002) Validity of the assessment of dietary intake: problems of misreporting. *Curr Opin Clin Nutr Metab Care* **5**, 489–493.
46. Freedman LS, Tasevska N, Kipnis V *et al.* (2010) Gains in statistical power from using a dietary biomarker in combination with self-reported intake to strengthen the

Public Health Nutrition

analysis of a diet–disease association: an example from CAREDS. *Am J Epidemiol* **172**, 836–842.
47. Willett W & Stampfer MJ (1986) Total energy intake: implications for epidemiologic analyses. *Am J Epidemiol* **124**, 17–27.
48. Kipnis V, Freedman LS, Brown CC *et al.* (1997) Effect of measurement error on energy-adjustment models in nutritional epidemiology. *Am J Epidemiol* **146**, 842–855.
49. Lafay L, Mennen L, Basdevant A *et al.* (2000) Does energy intake underreporting involve all kinds of food or only specific food items? Results from the Fleurbaix Laventie Ville Sante (FLVS) study. *Int J Obes Relat Metab Disord* **24**, 1500–1506.
50. Nielsen SB, Montgomery C, Kelly LA *et al.* (2008) Energy intake variability in free-living young children. *Arch Dis Child* **93**, 971–973.

Appendix C

Curriculum vitae

Claudia Börnhorst boern@bips.uni-bremen.de

Curriculum vitae Claudia Börnhorst

Current position: **Research Associate at the BIPS – Institute for Epidemiology and Prevention Research GmbH**
Department: Biometry and Data Management
Unit: Statistical Methods in Epidemiology

Educational background

June 2003:	General qualification for university entrance (Abitur) at the Gymnasium Bad Iburg
Okt. 2003:	Academic studies of Economical Mathematics, University of Bielefeld
Aug. 2005:	Pre-diploma in Economical Mathematics; grade: 1 (very good)
2005 - 2008:	Student research assistant at the Chair of Macroeconomics and at the Chair of Economic Policy, University of Bielefeld
WS 2006/2007:	Exchange scholarship (Erasmus) at the Johannes Kepler University of Linz, Austria
Sept. 2008:	Diploma in Economical Mathematics; grade: 1+ (with distinction)

Professional career

Okt. 2008 – Nov. 2011: Research assistant at ZWE BIPS, University of Bremen

since Dez. 2011: Research assistant at BIPS – Institute for Epidemiology and Prevention Research, Bremen

Membership in scientific associations

since Okt. 2011: German Society for Epidemiology (DGEpi)

Claudia Börnhorst boern@bips.uni-bremen.de

List of publications:

Publications in scientific journals with peer-review

- **Börnhorst C**, Huybrechts I, Hebestreit A, Vanaelst B, Molnar D, Bel-Serrat S, Mouratidou T, Moreno L, Pala V, Eha M, Kourides Y, Siani A, Eiben G, Pigeot I (2012). *Diet-obesity associations in children: Approaches to counteract attenuation caused by misreporting*. Public Health Nutr [Epub ahead of print]
- **Börnhorst C**, Huybrechts I, Ahrens W, Eiben G, Michels N, Pala V, Molnar D, Russo P, Barba G, Bel-Serrat S, Moreno LA, Papoutsou S, Veidebaum T, Loit HM, Lissner L, Pigeot I (2012). *Prevalence and determinants of misreporting among European children in proxy-reported 24 h dietary recalls*. Br J Nutr, 6: 1-9 [Epub ahead of print]
- Vyncke K, Cruz Fernandez E, Fajó-Pascual M, Cuenca-García M, De Keyzer W, Gonzalez-Gross M, Moreno LA, Beghin L, Breidenassel C, Kersting M, Albers U, Diethelm K, Mouratidou T, Grammatikaki E, De Vriendt T, Marcos A, Bammann K, **Börnhorst C**, Leclercq C, Manios Y, Dallongeville J, Vereecken C, Maes L, Gwozdz W, Van Winckel M, Gottrand F, Sjöström M, Díaz LE, Geelen A, Hallström L, Widhalm K, Kafatos A, Molnár D, De Henauw S, Huybrechts I (2012). *Validation of the Diet Quality Index for Adolescents (DQI-A) by comparison with biomarkers, nutrient and food intakes: the HELENA study.* Br J Nutr (accepted)
- Vandevijvere S, Geelen A, Gonzalez-Gross M, van't Veer P, Dallongeville J, Mouratidou T, Dekkers A**, Börnhorst C**, Breidenassel C, Crispim SP, Moreno LA, Cuenca-Garcia M, Vyncke K, Beghin L, Grammatikaki E, De HS, Catasta G, Hallstrom L, Sjostrom M, Warnberg J, Esperanza L, Slimani N, Manios Y, Molnar D, Gilbert CC, Kafatos A, Stehle P, Huybrechts I (2012). *Evaluation of food and nutrient intake assessment using concentration biomarkers in European adolescents from the Healthy Lifestyle in Europe by Nutrition in Adolescence study*. Br J Nutr, 23: 1-12 [Epub ahead of print]
- **Börnhorst C,** Hense S, Ahrens W, Hebestreit A, Reisch L, Barba G, Von Kries R, Bayer O, on behalf of the IDEFICS consortium (2012). *From sleep duration to childhood obesity - What are the pathways?* Eur J Pediatr, 171 (7):1029 - 1038.
- Huybrechts I, **Börnhorst C,** Pala V, Moreno LA, Barba G, Lissner L, Fraterman A, Veidebaum T, Hebestreit A, Sieri S, Krogh V, Ottevaere C, Tornaritis, M, Molnar D, Ahrens W, De Henauw S, on behalf of the IDEFICS consortium (2011). *Evaluation of the Children"s Eating Habits Questionnaire used in the IDEFICS study by using urinary calcium and potassium excretion in spot urine samples for estimating correlations with milk consumption frequencies among European children*. Int J Obes, 35 (Suppl.1):S69 - S78.
- **Börnhorst C**, Bel-Serrat S, Pigeot I, Huybrechts I, Ottevaere C, Sioen I, de Henauw S, Mouratidou T, Mensana MI, Westerterp K, Bammann K, Lissner L, Eiben G, Pala V, Rayson M, Krogh V, Moreno LA (2012). *Validity of 24-hour recalls in (pre-)school aged children: Comparison of proxy-reported energy intakes with measured energy expenditure*. Clin Nutr (submitted)
- Bel-Serrat, S, Mouratidou T, **Börnhorst C**, Peplies J, De Henauw S, Marild S, Molnár D, Siani A, Tornaritis M, Veidebaum T, Krogh V, Moreno LA (2012). *Food consumption and cardiovascular risk factors in European children: the IDEFICS study.* Pediatr Obes (accepted)
- Bel-Serrat S, Mouratidou T, Pala V, Huybrechts I, **Börnhorst C**, Fernández-Alvira JM, Hadjigeorgiou C, Eiben G, Hebestreit A, Lissner L, Molnár D, Siani A, Veidebaum T,

Krogh V, Moreno LA (2012). *Relative validity of the Children's Eating Habits Questionnaire among young European children: the IDEFICS study.* Public Health Nutr (accepted)

- Hebestreit A, **Börnhorst C**, Barba G, Siani A, Huybrechts I, Lissner L, Eiben G, Moreno LA, Fernández Alvira JM, Loit HM, Kovács E, Tornaritis M, Krogh V, on behalf of the IDEFICS Consortium (2012). *The impact of portion size, energy intake and density on childhood obesity.* (submitted)

Selected presentations

- **Börnhorst C**, Huybrechts I, Eiben G, Molnar D, Moreno LA, Pala V, Eha M, Kourides Y, Siani A, Pigeot I, on behalf of the IDEFICS consortium (2012). Diet-obesity relationships in children: Approaches to handle implausible dietary recalls. International Conference on Diet and Activity Methods (ICDAM), May 2012, Rome, Italy
- **Börnhorst C**, Bel-Serrat S, Bammann K, Moreno LA, Huybrechts I, Eiben G, Pigeot I (2011). Validität von Ernährungsdaten bei Kleinkindern: Vergleich zwischen 24-Stunden Erinnerungsprotokollen und objektiv gemessenem Energieverbrauch. "Biometrie, Epidemiologie und Informatik - Gemeinsam forschen für Gesundheit". 56. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) und 6. Jahrestagung der Deutschen Gesellschaft für Epidemiologie (DGEpi), September 2011, Mainz, Germany
- **Börnhorst C**, Günther F, Jahn I (2011). Wie wird geschlechtersensible Forschung in der Epidemiologie aktuell "gemacht"? Erste Ergebnisse einer Suche nach guten Beispielen im Projekt Epi goes Gender. "Prävention sozial und nachhaltig gestalten". Gemeinsamer Kongress der Deutschen Gesellschaft für Sozialmedizin und Prävention (DGSMP) und der Deutschen Gesellschaft für Medizinische Soziologie (DGMS) in Zusammenarbeit mit dem Medizinischen Dienst der Krankenversicherungen (MDK), September 2011, Bremen, Germany