Investigating participation in population-based cohort studies using paradata

Dissertation

Zur Erlangung des Grades eines Doktors der Naturwissenschaften

-Dr. rer. nat.-

vorgelegt von

Malte Langeheine

Universität Bremen

Fachbereich Mathematik / Informatik

Februar 2020

1. Gutachter: Prof. Dr. W. Ahrens

2. Gutachter: Prof. Dr. V. Didelez

Defense of the dissertation: 20.05.2020

Hiermit erkläre ich, Malte Langeheine, geboren am 29. 07. 1986, dass für das Verfassen der vorliegenden Dissertation **Investigating participation in population-based cohort studies using paradata** folgende drei Aussagen zutreffen:

1. Ich habe die Arbeit ohne unerlaubte fremde Hilfe angefertigt.

2. Ich habe keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt.

3. Ich habe die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht.

Bremen, 12. Februar 2020

Malte Langeheine

Acknowledgements

This thesis would not have been completed without the outstanding support I received. In particular, I would like to thank

Prof. Dr. Wolfgang Ahrens for making me realize the thesis, his helpful suggestions and criticism,

Prof. Dr. Vanessa Didelez and Prof. Dr. Kate Tilling for their guidance and sharing their knowledge on missing data,

Dr. Stefan Rach und Dr. Hermann Pohlabeln for their constant helpful advises, guidance, comments, and critical discussions,

the Unit Field Work at BIPS for all their support over the years, in particular Ramona Siebels und Marina Resnikov for their patience on all the questions related to the fieldwork in the IDEFICS/I.Family study,

the entire Leibniz Institute for Prevention Research and Epidemiology - BIPS, in particular, Dr. Achim Reineke for his support on the MODYS data, Dr. Christoph Buck, Tim Intemann, and Dr. Ronja Foreita for their critical discussions and Dr. Florence Samkange-Zeeb for proofreading the thesis,

Dr. Sabrina Hense for writing an inspiring doctoral thesis,

the library of BIPS for their constant support to get access to the necessary scientific literature,

all co-authors for contributing to the articles included in this thesis and the IDEFICS and I.Family consortia as well as the German National Cohort,

Dr. Rolf Langeheine inspiring me to follow his footprints,

Stuart Braithwaite, Bruce Dickinson for motivating me, and

Sara Langeheine for her emotional support.

Revision

It has to be noted that this manuscript contains a slightly revised version of the chapter “1.3 Potential consequences of nonresponse and attrition” compared to the previously submitted version. We revised the definitions stated by missing data theory, highlighted that we illustrated the consequences of missing data in the special case of a regression analysis, and slightly reworded the caption of Figure 1.2.

Summary

This thesis consists of seven main chapters. In the introduction (Chapter 1) the topics under study are conceptualized within a general framework, the Total Study Error, that classifies errors, linking the error sources to the steps of study design, conduct (data collection) and the estimation of prevalences, incidences or associations. This thesis focuses on non-participation (also called nonresponse or at follow-ups attrition) in population-based cohort studies at baseline or follow-up and investigates mainly topics related to the Total Study Error component *nonresponse error* (Chapters 2 to 5) and to a minor extent to the component related to *measurement error* (Chapter 6) mainly including paradata (i.e., information about the process of data collection) (Chapters 3, 5, and 6) in the analysis. While Chapters 2, 3, 5, and 6 are published peer-reviewed scientific papers, Chapter 4 is currently under review. Non-participation is not only a phenomenon of the past, but is a source of error that future cohort studies will have to deal with. Hence, it is imperative to address nonresponse.

Attrition may lead to bias in epidemiological cohorts, since participants who are healthier and have a higher social position are less likely to drop out. In Chapter 2, we investigated possible selection effects regarding key exposures and outcomes in the IDEFICS/I.Family study, a large European cohort on the etiology of overweight, obesity and related disorders during childhood and adulthood. In particular, we investigated associations of attrition with sociodemographic variables, weight status, and study compliance and assessed attrition across time regarding children's weight status and variations of attrition across participating countries. We investigated selection effects with regard to social position, adherence to key messages concerning a healthy lifestyle, and children's weight status. Attrition was associated with a higher weight status of children, lower children's study compliance, older age, lower parental education, and parent's migration background, consistent across time and participating countries. Although overweight (odds ratio 1.17, 99% confidence interval 1.05-1.29) or obese children (odds ratio 1.18, 99% confidence interval 1.03-1.36) were more prone to drop-out, attrition only seemed to slightly distort the distribution of children's BMI at the upper tail. Restricting the sample to subgroups with different attrition characteristics only marginally affected exposure-outcome associations.

Declining response proportions in population-based studies are often countered by extended recruitment efforts at baseline that may, however, result in higher attrition in a subsequent follow-up. In Chapter 3, we analyzed the effect of extended recruitment efforts on attrition at the first follow-up of the child cohort IDEFICS. We used paradata from the German IDEFICS cohort to quantify recruitment effort and classify respondents as completing the recruitment early vs. late for baseline and follow-up separately. Individuals who were late respondents at baseline and early respondents at the follow-up had a higher chance of attrition (odds ratio 1.65, 95% confidence interval 1.19, 2.28) as compared to

other groups. An investigation of reasons for non-participation revealed that members of this group were more likely to be not reachable by phone.

Although missing data are a major concern in epidemiology, missing data may not necessarily result in biased estimates. Analyzing only the individuals with complete data on the outcome, the exposure, and all explanatory variables, is a very simple and therefore popular strategy of handling missing data known as complete-case analysis. In a regression analysis, complete-case analysis is consistent in general when missingness is unrelated to the outcome given the explanatory variables included in the regression model. However, it appears not widely appreciated that this validity of complete-case analysis critically hinges on correct specifications of the analysis model and a misspecified model might re-introduce bias. In Chapter 4, we illustrated with a simulation how different modeling choices can affect our conclusions even when a complete-case analysis is in principle valid. We based our study on empirical data from the IDEFICS study and simulate only the missingness mechanism assuming different association strengths and different frequencies of missingness. In each scenario, we investigated the performance of three different analysis models using complete-case analysis as well as multiple imputation and inverse probability weighting as methods to correct for missing data. Our results suggest that model misspecification can lead to considerable bias when data contain missing values, even in a scenario where an ideal complete-case analysis is known to be consistent. This bias equally affects multiple imputation, and to a lesser extent, at the cost of precision, inverse probability weighting which requires a correct weighting model. In our example, basic model diagnostics were sufficient to alert us to the misspecification of the simple analysis model with regard to the functional form of the exposure; this was detectable even for the most extreme missingness mechanisms.

Another aspect of this thesis was whether we could enhance the response to study invitations (Chapter 5). We therefore conducted a trial embedded within the German National Cohort comparing the responses to study invitations sent in recycled envelopes of grey color vs. envelopes of white color. We analyzed paradata for reactions to the invitation letters by potential subjects, the duration between mailing date of the invitation and active responses, and study participation. Grey envelopes only slightly increased the chance of active responses (odds ratio 1.16, 95% confidence interval: 0.83, 1.62) to the invitation letter. Potential study subjects with German nationality (odds ratio 3.75, 95% confidence interval: 2.07, 7.66) and age groups above 50 years (50-59: odds ratio 1.78, 95% confidence interval: 1.19, 2.69; 60-69: odds ratio 2.25, 95% confidence interval: 1.48, 3.43) were more likely to actively respond to the invitation letter. The duration between mailing date of the invitation and active response was not associated with envelope color, sex, nationality, or age.

In Chapter 6, we touched upon a kind of participation mostly regarded as undesired. We analyzed factors associated with the presence of an intimate partner during face-to-face interviews using data from the first wave of the German Family Panel (pairfam). Although the intimate partner is most likely

to be the third person present during the interview, an examination of the association of the relationship quality and the presence of the intimate partner is lacking in the literature. Our descriptive analysis revealed that an intimate partner was present in every seventh interview. The opportunity structure, such as the couple's living arrangements or their employment status, had the greatest influence on the presence of both female and male partners while aspects of the relationship quality were to a minor extent associated with the partner's presence.

Chapter 7 summarizes the main findings of this thesis in light of previous literature and discusses methodological considerations related to the scientific articles presented in this thesis. In conclusion, paradata are valuable information for the investigation of nonresponse and attrition in cases where they capture all the information required to answer a given research question. Paradata are often considered only as a 'by-product', but provide considerable scientific benefit and should be already defined in the planning phase of a study.

Contents

Abbreviations

BIC	Bayesian information criterion
BMBF	Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung)
BMI	Body mass index
CAPI	Computer-assisted personal interview
CASI	Computer-assisted self-administered interview
CCA	Complete-case analysis
CI	Confidence interval
CPE	Change in point estimates
DAG	Directed acyclic graph
FDCT-N	Dyadic coping questionnaire (Fragebogen zur Erfassung des Dyadischen Copings)
GIMD	German index of multiple deprivation
GNC	German National Cohort
HOMA-IR	Homeostasis model assessment to quantify insulin resistance
IDEFICS	Identification and prevention of Dietary- and lifestyle-induced health EFfects In Children and infantS
IPW	Inverse probability weighting
ISCED	International Standard Classification of Education
KiGGS	German Health Interview and Examination Survey for Children and Adolescents (Studie zur Gesundheit von Kindern und Jugendlichen in Deutschland)
KS test	Kolmogorov-Smirnov test
LOESS	Locally weighted scatterplot smoothing
MAR	Missing at random
MCAR	Missing completely at random
MI	Multiple imputation
MNAR	Missing not at random
MODYS	Modular control and documentation system

MVPA	Moderate to vigorous physical activity
NRI	Network of relationships inventory
OLS	Ordinary least squares
OR	Odds ratio
PA	Physical activity
pairfam	German family panel
PIN	Personal identity number
Q-Q plot	Quantile-quantile plot
RE model	Random-effects meta-analysis
zBMI	z-scores of body mass index
zHOMA	z-score of HOMA

Figures

Tables

1 Introduction

The accumulation of knowledge about diseases and their causes is a central aim of epidemiology, often making use of population-based studies. In many instances, however, subjects invited to participate in these population-based studies often do not provide the requested information (Bethlehem et al., 2011). Missing information is a major concern, as it may introduce bias (Howe et al., 2013). In the following section, we contextualize the topic under study, that is, missing information in population-based cohort studies, within a broader range of other potential error sources, briefly describe the potential consequences of missing information, and outline approaches to investigate missing information.

1.1 Population-based cohort studies

A particular population-based study design in epidemiology is the cohort study (Miller et al., 2014). In a cohort study, an event such as the occurrence of a disease or death in a study population is prospectively observed for a given period of time from a certain time point, t_0, until a set time point t_x (Kreienbrock et al., 2012). Cohort studies thus entail the repeated observation of study participants over time (Pearce, 2012). The distinctive feature of a cohort study is that it allows the assessment of change at the individual level (Lynn, 2009), and thereby theoretically captures the causal action of an exposure on the subsequent development of diseases resulting from that exposure (Rothman, 2012). Unfortunately, despite this distinctive feature, cohort studies face fundamental methodological challenges as population-based studies are always prone to some kind of error (Bethlehem et al., 2011).

1.2 The Total Study Error framework

In this thesis, we use the Total Study Error framework to contextualize the topics under study. From the quality perspective of conducting an epidemiological study, all errors can be summarized into a single framework, the Total Study Error framework. The framework stems from methodological considerations in the context of survey methodology, in which it is called the Total *Survey* Error framework (Groves and Lyberg, 2010). We aim to translate the Total *Survey* Error framework into Epidemiology, as the aims and concepts of Epidemiology and Survey Methodology (respectively Social Sciences) may differ. For example, Rothman (2012) pointed out that generalizations from an epidemiological study do not necessarily involve making inferences about a target population as they can be based on the understanding of the underlying biology, which is not dependent on statistical sampling.

The Total Study Error (Figure 1.1) is a framework that classifies errors, linking the error sources to the steps of study design, conduct (data collection) and the estimation of prevalences, incidences or associations (Groves and Lyberg, 2010; Lavrakas, 2008). In the Total Study Error framework an error is

divided into bias which is a systematic (non-random) error, and variance, a random error (Lavrakas, 2008). It has to be noted that the Total Study Error framework is a concept to guide researchers. The idea of the concept is that the sum of all errors results in a discrepancy between the "true" prevalence, incidence or associations to be estimated and the estimate obtained by a study, that is, the Total Study Error (Bethlehem et al., 2011). As pointed out by Lynn and Lugtig (2017), the conceptual components of the Total Study Error framework are the same in longitudinal studies as in cross-sectional studies, but the process leading to an error and the nature of an error may differ. Two components of error sources can be distinguished: errors related to measurement and errors related to the selection of study subjects (Groves et al., 2009).

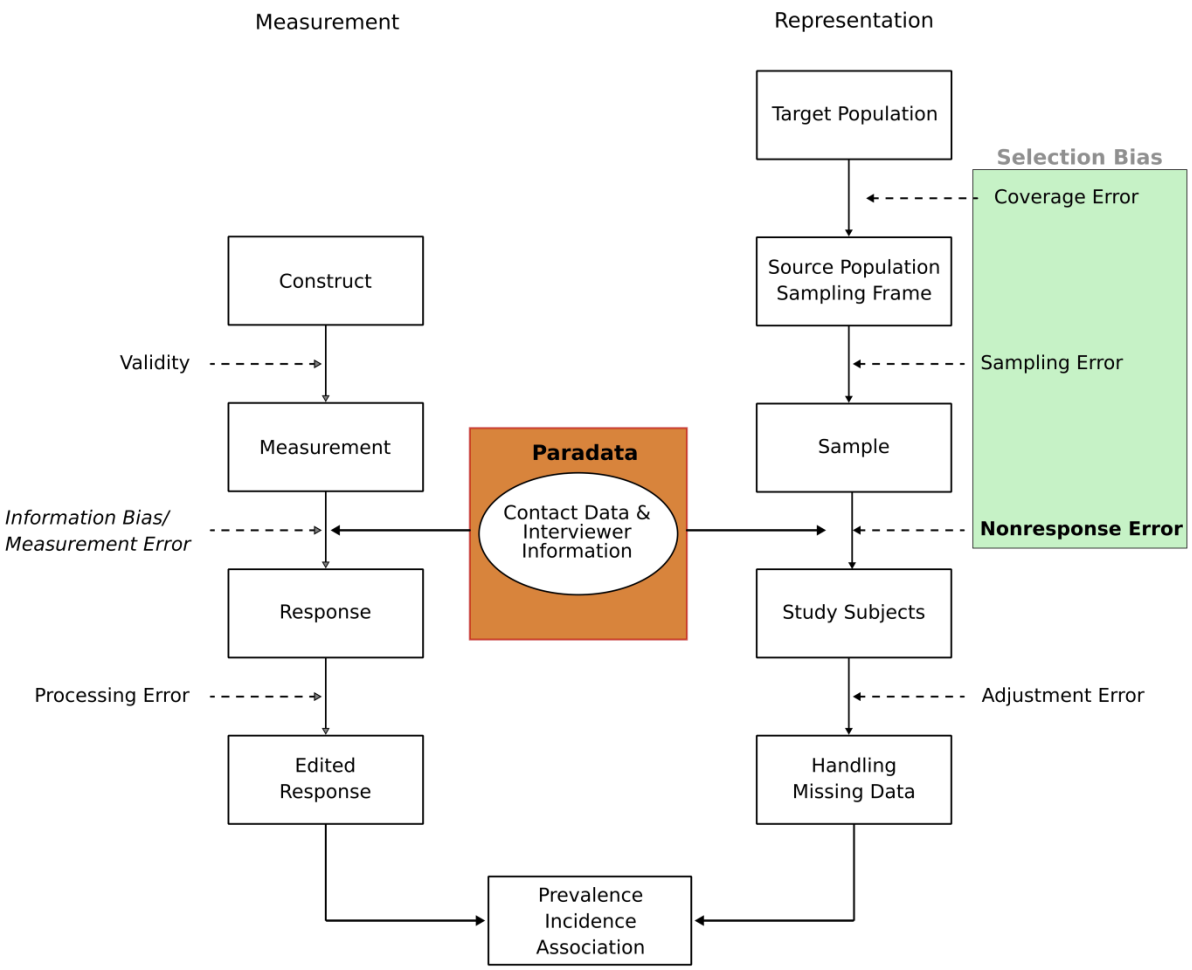

Figure 1.1 The Total Study Error framework adapted from (Groves et al., 2009; Kreuter and Casas-Cordero, 2010)

First, the Total Study Error framework (Figure 1.1) contains a branch of three errors related to the response of a single study subject to a study instrument (e.g. question in questionnaire, collection of biosamples): The difference between a theoretical construct and its measurement (*validity*), the difference between a "true" value to be measured and the value provided by a study subject (*information bias/measurement error*), and the difference between the value provided by a study subject and the value used in the final analysis (*processing error*) (Groves et al., 2009). An example is illustrated in Table 1.1.

Table 1.1 Fictional example of measurement of adiposity (Kuskowska-Wolk and Rössner, 2009)

Construct	Adiposity
Measurement	Body mass index (BMI, kg/m^2)
Response	Self-reported weight and height via a paper and pencil questionnaire
Edited Response	Based on a defined set of rules, information on self-reported weight and height is edited to ensure readability and unambiguousness prior to transferring the information into a database

Second, the Total Study Error framework (Figure 1.1) contains a branch of four errors related to the selection of study subjects: The difference between the target population and the sampling frame (*coverage error*), the difference between the sampling frame (contact list for prospective study population, e.g. addresses from a population register, list of Kindergartens and schools in study region) and the sample (*sampling error*), the difference between the full sample and the measured study subjects (*nonresponse error*), and an error resulting from methods that aim to compensate for missing data, for example by weighting or imputation (*adjustment error*). An example is detailed in Table 1.2.

Table 1.2 Example of the selection of study subjects in the German Health Interview and Examination Survey for Children and Adolescents (KiGGS baseline examination) (Kamtsiuris et al., 2007)

Target population	Children and adolescents aged 0-17 years living in the Federal Republic of Germany, with a main residence registered in the population registers
Sampling frame	1. Total number of political municipalities in Germany 2. Addresses from population registers of 167 municipalities (sample points)
Sample	An equal number of 24 personal addresses were drawn from the population registers in each of the selected municipalities. A total of 26 899 children and teens were selected and invited (unadjusted gross sample)
Study subjects	17 641 children and adolescents (net sample) participated
Handling of missing data	Not explicitly reported

Coverage bias, sampling bias, and nonresponse bias can be summarized by the term *selection bias,* broadly defined by Rothman (2012) as:

> "[…] a systematic error in a study that stems from the procedures used to select subjects and from factors that influence study participation. It comes about when the association between exposure and disease differs for those who participate and those who do not participate in a study." (Rothman, 2012, p. 133)

This thesis focuses on one of the four branches of errors depicted in the Total Study Error framework that are related to the selection of participants, that is, nonresponse. Unfortunately, not all subjects invited to take part in a study are successfully measured as some study subjects do not provide the requested information (Bethlehem et al., 2011). A *nonresponse error* (Groves et al., 2009) occurs if an estimate based on a subgroup of measured study subjects of the sample differs from the "true"

estimate of the full sample. Nonresponse can be divided into item- and unit-nonresponse. Study subjects may provide parts of the requested information (item-nonreponse) or no information at all (unit-nonresponse) (Bethlehem et al., 2011). As this thesis does not cover aspects of item-nonresponse, the term nonresponse will be used to refer to unit-nonresponse throughout the rest of the thesis. In the epidemiological context, nonresponse is more pronounced among the less advantaged and less healthy (Howe et al., 2013; Miller et al., 2014), and is therefore likely to induce bias. In longitudinal studies, nonresponse may occur not only at baseline, but baseline participants may completely drop out of a study at subsequent follow-ups. The latter is termed cohort attrition (also panel attrition or loss to follow-up), and is said to be one of the most important error sources in longitudinal studies (Lugtig, 2014; Lynn and Lugtig, 2017). Lynn and Lugtig (2017) stress two aspects unique to longitudinal studies that affect nonresponse. Firstly, the mobility of subjects can result in non-location, leading to nonresponse at each subsequent wave. Secondly, the experience of having taken part at a previous time point in a longitudinal study is likely to influence participation at further time points, due to the length of the measurement or the content of the study (Lynn, 2014).

1.3 Potential consequences of nonresponse and attrition

The consequence of nonresponse and attrition is missing information, shortly missingness. Missingness is classified by the missing data theory (Rubin, 1987), whereby missing data is classified as (1) Missing Completely At Random (MCAR), if missingness does not depend on any observed or missing variables at all, (2) as Missing At Random (MAR), if missingness is independent of the missing values given the observed values or (3), as Missing Not At Random (MNAR), if missingness is still associated with missing values itself, given the observed variables (Little and Rubin, 2002; Spieß, 2010).

The consequences of missing data can be difficult to understand, in particular regarding associations. Associations have the special feature that missingness does not necessarily induce bias. Hence, the consequences of missing data are better understood when visualized. Figure 1.2 [adapted from Daniel et al. (2011)], depicts artificial data of an explanatory variable *A* and a response variable *Y*, and linear regression lines for both a scenario with full data (all circles and solid red line) and one with incomplete data (only filled black circles and blue dashed line). For simplicity, we illustrate missingness in *A* or *Y*, although missingness may also occur in *A* and *Y*.

In the special case of a regression analysis we describe the conditional distribution of the outcome Y given the explanatory variable *A*. If missingness is independent of *Y* given *A* then a CCA can still give a consistent estimate of the regression coefficient.

In a scenario in which missing values occur randomly in the explanatory variable *A* (Figure 1.2, Top row, Column 1) or missing values occur randomly in the outcome variable *Y* (Figure 1.2, Bottom row, Column 1) the complete-case analyses remain unaffected by missingness as can be seen from the aligning regression lines.

Furthermore, in a scenario in which missing values occur above a certain cutoff point of the explanatory variable *A* (*Y* missing at random given *A*), the mean of *A* decreases, but the complete-case analysis of *A* and *Y* remains unaffected (Figure 1.2, Top row, Column 2), as can be seen from the aligning regression lines. In contrast, if missing values occur above a certain cutoff point of the dependent variable *Y* (*Y* missing dependent on *Y*; cf. missing not at random), the mean of *Y* decreases and the regression lines do not align, indicating the bias of the complete-case analysis (Figure 1.2, Bottom row, Column 2) (Daniel et al., 2011).

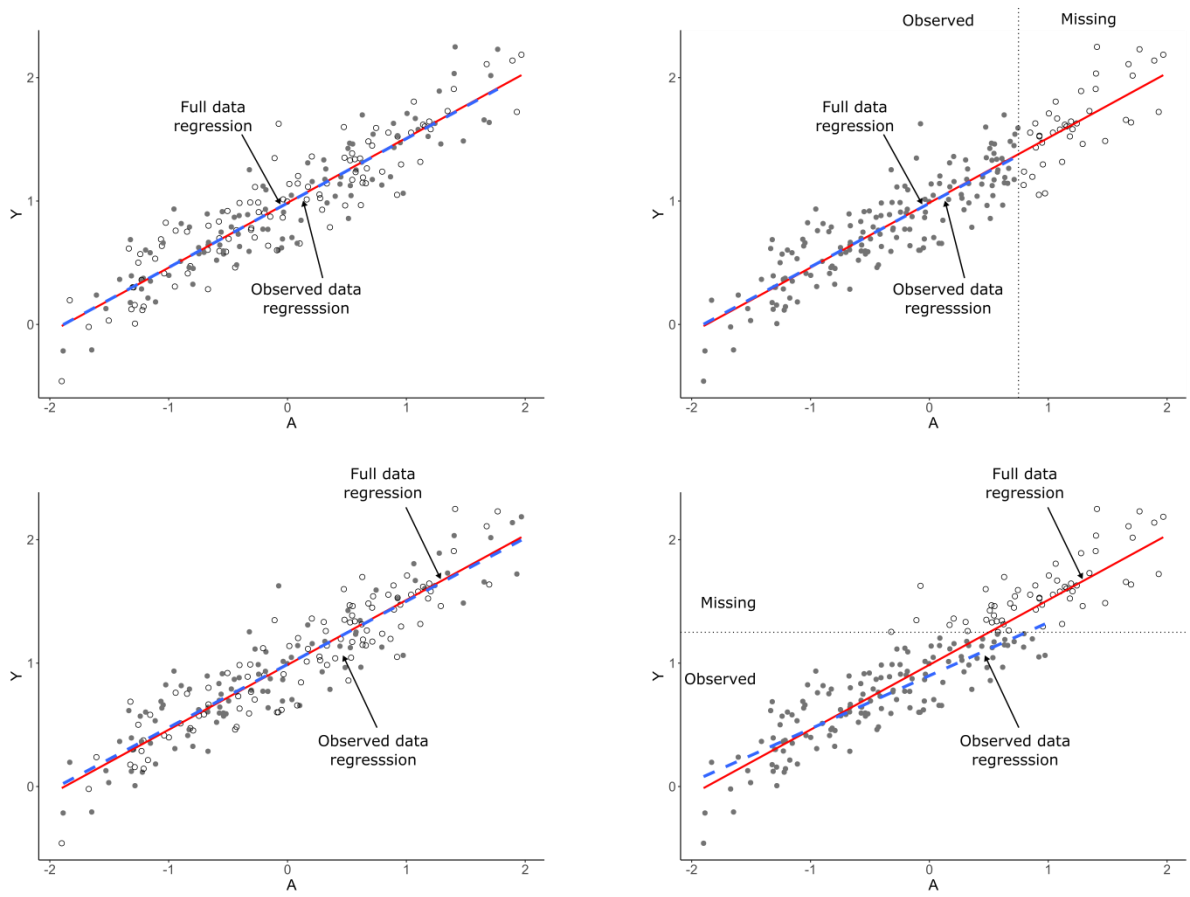

Figure 1.2 Illustration of bias in bivariate regression using complete-case analysis when the missing mechanism is Missing Completely At Random (top-left and bottom-left), Missing At Random (Column 2, Top row, missingness depends on *A* only) or Missing Not At Random (Column 2, Bottom row, missingness depends on *Y* only) for a scenario in which the bivariate association is linear (Daniel et al., 2011). Full data regression (red solid line), observed data regression (blue dashed line), observed data (dots), missing data (circles)

The mechanism introducing selection bias to an association due to nonresponse or attrition is called conditioning on a common effect (Hernán et al., 2004), also known as conditioning on a collider. Conditioning refers to a restriction (Cole et al., 2010), such as nonresponse in cohort studies resulting

in a subset of invited subjects, while attrition results in a subset of baseline participants. Consider the following example: Suppose a certain college gives scholarships either to sports talents or to applicants with extremely high grade point averages. For simplicity, assume that sports talent and scholastic achievement are not associated in the general population. Within the subgroup of students with scholarships, a selected person with a sports talent is likely to have a lower grade point average and vice versa (Elwert, 2013; Pearl et al., 2016). Hence, if two non-associated variables A (sports talent) and B (grades) have a common effect C (scholarship), they will be conditionally associated if the measure of association is computed within levels of the common effect C. Causal diagrams (Pearl, 1995) allow the identification of scenarios in which nonresponse or attrition induce a selection bias (Daniel et al., 2011). The theory of causal diagrams uses directed acyclic graphs (DAGs) to represent the causal relationship of an exposure-outcome association, including covariates and a missingness indicator.

1.4 Investigating nonresponse and attrition

The question regarding the characteristics of those not taking part in a study and the potential consequences has a long tradition (Hansen and Hurwitz, 1946; Suchman and McCandless, 1940). Different methods to assess potential selection bias have since been developed, each with its own advantages and disadvantages (Groves, 2006). Nonresponse at baseline of a cohort study is particularly challenging to investigate, as invited subjects who decline to participate do not provide any information. Therefore, additional sources of individual-level data that can be added to the collected study data not only offer the potential to investigate nonresponse at baseline, but are also helpful for investigating attrition. Sources of such data include information provided by the sampling frame, population registers, and paradata.

Sampling frames containing addresses may include additional sociodemographic variables such as age, sex, or nationality, for example from population registers. The strength of such data is that identical measurements for participants and nonrespondents are available (Groves, 2006). Unfortunately, the variables available in the sampling frame are often not of key interest to the study and may also have missing values (Groves, 2006).

Population registers may contain a rich source of medical individual-level data (Ludvigsson et al., 2009). In several Scandinavian countries, it is possible to combine data from multiple registers at an individual-level, via a personal identity number (PIN). For example, in Sweden, since 1947, all individuals residing in the country are accorded a PIN by the National Tax Board. Amongst others, the number is used by different authorities and the health care system (Ludvigsson et al., 2009). For example, Regber et al. (2013) used medical data from population registers to investigate selection bias. To this end, they compared a sample of children in their data with referent children from the general

population at municipal level by matching a child from the general population to each child in their sample based on age and gender. Further, Greene et al. (2011) used Danish medical data from a population register to obtain end point information for all members of a birth cohort study and assessed the presence and magnitude of bias induced by attrition. The availability of comparable data, the access to the data, and a personal identification throughout multiple sources is however limited outside Scandinavia. In Germany, the allocation of uniform personal identifiers across registers is prohibited (Jacobs et al., 2015).

To help manage various study processes such as data collection, editing or coding, process data, called paradata, are often collected (Groves and Heeringa, 2006). As pointed out by Kreuter and Casas-Cordero (2010), paradata can be used to investigate different components of the Total Study Error, including nonresponse. In general, paradata are auxiliary data about the study process and mostly considered to be a by-product. Paradata are collected in various study modes for example mail, telephone and web. They are useful because they allow the statistical assessment of the study processes, and thus the investigation of quality control. For completed studies, paradata can be used ex-post to carry out analyses to describe the final recruitment effort, estimate recruitment costs, calculate response proportions, or assess the quality of face-to-face interviews. The collection of paradata may be highly automated, using computer-aided data collection such as keystrokes in web surveys, or mainly manual, using designated software and eventually requiring a detailed and elaborate documentation by field staff. During an ongoing study, paradata are used to monitor and manage the fieldwork (Kreuter and Casas-Cordero, 2010). The growing reluctance of the general population to take part in studies challenges study practitioners (de Leeuw and de Heer, 2002), and has led to increased efforts being put into obtaining interviews. Study designs that use paradata and are responsive to increased study costs and take cost-quality tradeoff decisions in real-time into account, have hence been developed (Groves and Heeringa, 2006).

An example of a software to monitor and manage the fieldwork is the **mo**dular control and **d**ocumentation s**ys**tem (MODYS) used, for instance, in the German National Cohort (GNC). MODYS is an electronic recruitment system developed at the Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany, designed to support the field staff during the recruitment of study participants in population-based studies (Reineke et al., 2018). The software enables the documentation and management of the recruitment, and guides the field staff throughout all the steps of the pre-defined recruitment pathways, on a case-by-case basis. MODYS stores information necessary for recruiting purposes (to send letters or to make phone calls) including names, addresses, telephone numbers or email addresses of potential study subjects, and only the field staff involved in the recruitment has access to the contact data.

MODYS records all events (e.g. contact outcomes) and creates an individual recruitment history for each potential study subject. For each potential study subject, the software records paradata that contain a date and time stamp for an event as well as a description of the event, the subject with whom contact was established, the outcome of the action, a code for non-participation, and optional free text notes (Reineke et al., 2018). Analyzing MODYS paradata is not necessarily straightforward because the data collected are a chronological record of all recruitment events for a subject and not a mere summary of recruitment results. For example, for the calculation of response proportions, it is not sufficient to simply tabulate drop-out codes and participation status, because multiple events that are not necessarily conflicting may be documented for a single subject. For instance, some subjects may take part in the examination and afterwards withdraw from the study, or others may, for some reason, have been assigned a drop-out code and afterwards take part. Even an event which should logically be documented only once may be documented several times due to human error, for example be assigned multiple drop-out codes. Thus, to calculate exact response proportions, the chronological history of events for each potential study subject has to be taken into account and rules for assigning a subject to a final processing status have to be defined. In cases where the assignment of codes is not clear, it may be beneficial to consult field staff to understand how documentation resulted in the final data.

A commonly used strategy to investigate nonresponse is to take the recruitment effort of all respondents captured by paradata and divide the sample into early and late respondents. Based on the continuum of resistance (Ellis et al., 1970), assuming an underlying latent variable ranging from highly motivated to unmotivated to participate (Filion, 1976), it is argued that late respondents are most similar to nonrespondents. Although this method is appealing as it is easy to implement, it suffers from the fact that no direct information about nonrespondents is available (Groves, 2006). Further, it has been shown that late respondents are different from nonrespondents (Lin and Schaeffer, 1995).

A further method to investigate nonresponse is to compare distributions of the collected data with data from other more accurate sources (Groves, 2006). In Germany, the Microcensus (Schimpl-Neimanns and Weiss, 2014) is considered to be a potential benchmark as it is a compulsory household survey (Koch and Blohm, 2015). It however mainly captures sociodemographics, thus comparisons with key study variables, for example in a study to determine the aetiology of overweight, obesity and related disorders in children (Ahrens et al., 2011; Ahrens et al., 2017), are likely to be impossible.

Compared to nonresponse at the baseline of a cohort study, depending on the level of item-nonresponse and its underlying missing mechanism, studying attrition at least allows information collected prior to the dropout to be considered. Therefore, a commonly used strategy to investigate attrition is to pool all wave-on-wave attrition patterns (Lugtig, 2014) and compare participants to those who drop out. Furthermore, to explore potential consequences of attrition, baseline distributions or

exposure-outcome estimates of the full cohort could be compared to those of subsets of dropouts at later waves of the cohort (Howe et al., 2013).

Regarding the empirical findings stated in the literature, there appears to be general consensus that nonresponse and attrition are more likely to occur among the less advantaged and less healthy subjects, resulting in study populations mainly comprising relatively healthy and wealthy subpopulations (Howe et al., 2013; Watson and Wooden, 2009).

In epidemiological literature, nonresponse has been reported to be associated with living in disadvantaged neighborhoods (Gaertner et al., 2016), having foreign citizenship (Gaertner et al., 2016), being non-white (Corry et al., 2017), lower educational level (Corry et al., 2017; Gaertner et al., 2016), lower social class (Martikainen et al., 2016), lower health-related quality of life (Gaertner et al., 2016), and age varying by studies (Gaertner et al., 2016; Martikainen et al., 2016; Stuart et al., 2019).

Attrition has also been reported to be associated with various sociodemographics including age, in particular younger and older population members (Dijkema et al., 2005; Grigoletto et al., 1994; Long et al., 2019; May et al., 2012; Mein et al., 2012; Rinsky et al., 2017), lower educational level (Brilleman et al., 2010; Dijkema et al., 2005; Eaton et al., 1992; Gubhaju et al., 2016; Long et al., 2019; May et al., 2012; Rinsky et al., 2017; Slymen et al., 1996), lower income (Gubhaju et al., 2016; Long et al., 2019; Slymen et al., 1996; Spiers et al., 2018), non-white study subjects (Eaton et al., 1992; Long et al., 2019), non-native speakers (Eaton et al., 1992), and not married or not partnered study subjects (Alonso et al., 2006; Gubhaju et al., 2016; Rinsky et al., 2017; Slymen et al., 1996). Several health-related variables such as a lower health status (Brilleman et al., 2010; Long et al., 2019; Slymen et al., 1996), physical inactivity (Brilleman et al., 2010; May et al., 2012), obesity (Alonso et al., 2006; Gubhaju et al., 2016; May et al., 2012), underweight (Gubhaju et al., 2016; May et al., 2012), mental health problems such as depression (Eaton et al., 1992; Siddiqi and Holmes-Rovner, 2006; Slymen et al., 1996) and schizophrenia (Martin et al., 2016), erectile dysfunction (Gades et al., 2006), ischemic attack (Long et al., 2019), diabetes (Long et al., 2019), and smoking (Alonso et al., 2006; Brilleman et al., 2010; Gubhaju et al., 2016; Long et al., 2019) have also been reported to be associated with attrition. Furthermore, weather conditions (higher sunshine duration at the day before the appointment or day of appointment rainy) may influence attrition (Wallborn et al., 2013). Some causes of attrition are very study specific. For example, in a study on “the assessment of risk factors for motor-related injuries” (Alonso et al., 2006, p. 352), the history of previous motor vehicle accidents leading to hospitalization (Alonso et al., 2006) was associated with higher attrition. The topic of a study might be the reason for mixed results for causes of attrition. For example, some studies reported a higher chance of attrition among males (Eaton et al., 1992; Eerola et al., 2005; Grigoletto et al., 1994; May et al., 2012; Mein et al., 2012) and some among females (Dijkema et al., 2005; Long et al., 2019; Slymen et al., 1996).

Different study results of associates with attrition may in some instances be explained by subjective impairment of health and individual risk perceptions (Guey et al., 2008).

1.5 Research questions

As summarized in the Total Study Error framework, population-based cohort studies face fundamental methodological challenges. This thesis investigates selected aspects of the Total Study Error framework that are mainly related to the topic *nonresponse error,* and to a minor extent to *measurement error*. Nonresponse and attrition are not only a phenomenon of the past, but are sources of error that future cohort studies will have to deal with. Hence, it is imperative that both are addressed (Stang, 2003). Although MODYS (Reineke et al., 2018), the electronic recruitment system developed at the Leibniz Institute for Prevention Research and Epidemiology – BIPS, has been in use since the mid-1990s, research using the MODYS paradata to investigate nonresponse and attrition is scarce. Hence there is a need to explore the potentials of MODYS paradata for research on nonresponse and attrition. As this thesis was mainly driven by the use of paradata, we will also investigate paradata related to the discussion of *measurement error*. The following questions will be addressed:

1) What are factors are associated with attrition in the European child cohort IDEFICS/I.Family? (Chapter 2)
2) Are extended recruitment strategies at baseline panelized at follow-up? (Chapter 3)
3) How might modeling in regression analysis bias regression estimates when there is attrition? (Chapter 4)
4) Do study invitations with envelopes made from recycled paper increase the likelihood of active responses or study participation? (Chapter 5)
5) Is information on the relationship quality associated with the presence of the intimate partner during face-to-face interviews? (Chapter 6)

1.6 Outline of the thesis

This thesis comprises seven main chapters as follows: an introductory chapter (Chapter 1), five manuscripts, four of those are peer-reviewed (Chapters 2, 3, 5 and 6) and one is currently under review (Chapter 4), and finally, a general discussion (Chapter 7). All the manuscripts included in this thesis investigate aspects of participation in three particular population-based cohort studies, namely, IDEFICS/I.Family (Ahrens et al., 2017), the German National Cohort (GNC) (German National Cohort, 2014), and the German family panel (pairfam) (Huinink et al., 2011), based on the Total Study Error framework (Table 1.3).

The European prospective cohort study IDEFICS/I.Family (see Chapters 2, 3, and 5) was initiated in 2007/08 and included children from eight countries (Belgium, Cyprus, Estonia, Germany, Hungary,

Italy, Spain, and Sweden). The study investigated the association between dietary, behavioral and socioeconomic factors and non-communicable chronic diseases and disorders, with a focus on overweight and obesity. In each country, data were collected in two or more selected communities. All children aged 2–9.9 years attending kindergarten or primary school in each community were eligible for participation and a total of 31 643 children were invited to participate. Children who took part in the baseline examination were invited to the first follow-up (T1) between September 2009 and May 2010 and to a further follow-up (I.Family, T3), conducted between 2013 and 2014 (Ahrens et al., 2017).
The GNC is to date the biggest epidemiological multi-center cohort study in Germany. The recruitment of the GNC (see Chapter 5) started in 2013 (pilot phase) and the aim of the study is to investigate the causes for the development of major chronic diseases, for example cardiovascular diseases, cancer, and diabetes. Across Germany, a random sample of the general population that includes a total of 100 000 women and 100 000 men aged 20–69 years, was drawn by 18 regional study centers. All participants are invited for re-assessment every 4-5 years. The baseline assessments included an extensive interview and self-completion questionnaires, a wide range of medical examinations and the collection of various biomaterials (German National Cohort, 2014).
The German family panel (pairfam) (see Chapter 6) is an annual study which started in 2008/09. Using face-to-face interviews, pairfam covers various topics of family research, including couple dynamics and partnership stability, childbearing, parenting and child development, and intergenerational relationships. About 12 000 randomly selected respondents of three birth cohorts, 1991-1993, 1981-1983, and 1971-1973, called anchor persons, were invited at baseline. From wave 2 onwards, the so called alteri of the anchor person (the partner, children, and parents) were also invited to take part (Huinink et al., 2011). Huinink et al. (2011) defined the anchor population as “all people living in Germany in private households who have sufficient mastery of the German language to follow the interview“ (Huinink et al., 2011, p. 90).

Table 1.3 Overview of the aspects investigated in this thesis

	Cohort study	Total Study Error component	Paradata involved	General research topic
Chapter 2	IDEFICS/I.Family	Nonresponse error	discussed	Participation at follow-up
Chapter 3	IDEFICS/I.Family	Reduce nonresponse	yes	Participation at follow-up
Chapter 4	IDEFICS/I.Family	Nonresponse error/Analysis	no	Participation at follow-up
Chapter 5	German National Cohort	Reduce nonresponse	yes	Participation atbaseline
Chapter 6	pairfam	Measurement error	yes	Participation at baseline

The manuscripts presented in this work can be divided into two groups according to their focus. Chapters 2 to 4, which investigate participation at follow-up, comprise the first group, while Chapters 5 and 6, which focus on participation at baseline, comprise the second. The main goal of Chapters 2 and 3 is to investigate factors associated with attrition, focusing on the social position of families, child health, and recruitment effort in IDEFICS/I.Family. Chapter 2 uses data from all participating countries, investigates attrition for all IDEFICS/I.Family follow-ups, and discusses the consequences of attrition on selected associations. Chapter 3 on the other hand only uses data from the German cohort, including available paradata, to investigate the consequences of an extended recruitment effort on participation at follow-up. Chapter 4 does not focus on factors associated with attrition, but simulates how modelling choices can bias estimates of regression models for different levels of attrition, using data from IDEFICS/I.Family. With regard to the Total Study Error framework, Chapters 2 and 3 can be linked to the component nonresponse error, whereas Chapter 4 can be linked to the final step of producing a study statistic, the analysis. Regarding the second group focusing on participation at baseline, Chapter 5 investigates whether study invitations with envelopes made from recycled paper increase the likelihood of active responses or study participation at the study center of the German National Cohort in Bremen. To this end paradata were scanned to derive the main outcome. In comparison to the other chapters, Chapter 6 is exceptional as it investigates factors associated with undesired participation of the intimate partner during face-to-face interviews in pairfam, using paradata obtained from interviewer observations. Whereas Chapter 5 can be linked to the Total Study Error component *nonresponse error*, Chapter 6 stems from the discussion of *measurement error* due to the presence of a third person during face-to-face interviews.

Finally, in the general discussion (Chapter 7), the main findings of Chapters 2 to 6 are summarized and discussed in light of previous research. This is followed by a discussion of methodological considerations and recommendations for potential future research.

2 Attrition in the European child cohort IDEFICS/I.Family: Exploring associations between attrition and body mass index

Malte Langeheine, Hermann Pohlabeln, Fabio Lauria, Toomas Veidebaum, Michael Tornaritis, Denes Molnar, Gabriele Eiben, Stefaan de Henauw, Luis A. Moreno, Garrath Williams, Wolfgang Ahrens, and Stefan Rach, on behalf of the IDEFICS and I.Family Consortia[1]

Abstract

Attrition may lead to bias in epidemiological cohorts, since participants who are healthier and have a higher social position are less likely to drop out. We investigated possible selection effects regarding key exposures and outcomes in the IDEFICS/I.Family study, a large European cohort on the etiology of overweight, obesity and related disorders during childhood and adulthood. We applied multilevel logistic regression to investigate associations of attrition with sociodemographic variables, weight status, and study compliance and assessed attrition across time regarding children's weight status and variations of attrition across participating countries. We investigated selection effects with regard to social position, adherence to key messages concerning a healthy lifestyle, and children's weight status. Attrition was associated a higher weight status of children, lower children's study compliance, older age, lower parental education, and parent's migration background, consistent across time and participating countries. Although overweight (odds ratio 1.17, 99% confidence interval 1.05-1.29) or obese children (odds ratio 1.18, 99% confidence interval 1.03-1.36) were more prone to drop-out, attrition only seemed to slightly distort the distribution of children's BMI at the upper tail. Restricting the sample to subgroups with different attrition characteristics only marginally affected exposure-outcome associations. Our results suggest that IDEFICS/I.Family provides valid estimates of relations between socio-economic position, health-related behaviors, and weight status.

Introduction

Epidemiological cohort studies are not only prone to non-response at baseline, but also to drop-out of participants during follow-up (Lynn, 2009), called cohort attrition. Since non-response and drop-out are more likely among less healthy and disadvantaged study participants (Howe et al., 2013; Melton et al., 1993; Vinther-Larsen et al., 2010), it is especially important for cohort studies to assess selection effects. IDEFICS (Identification and prevention of dietary and lifestyle-induced health effects in children and infants) (Ahrens et al., 2011) and I.Family (IDEFICS/I.Family cohort) (Ahrens et al., 2017) is a large

[1] Langeheine M, Pohlabeln H, Lauria F, Veidebaum T, Tornaritis M, Molnar D et al. Attrition in the European Child Cohort IDEFICS/I.Family: Exploring Associations Between Attrition and Body Mass Index. Frontiers in Pediatrics. 2018;6(212). doi:10.3389/fped.2018.00212.

European prospective cohort including children from eight countries (Belgium, Cyprus, Estonia, Germany, Hungary, Italy, Spain, and Sweden) that has been investigating dietary, behavioral and socioeconomic factors in relation to non-communicable chronic diseases and disorders with a focus on overweight and obesity (Ahrens et al., 2011; Ahrens et al., 2017). In IDEFICS/I.Family, a total of 16 228 children and their parents took part in up to three physical examinations between 2007 and 2014 and completed questionnaires on medical history, dietary behavior and other aspects of children's life. The present analysis complements the IDEFICS/I.Family cohort profile (Ahrens et al., 2011; Ahrens et al., 2017). We extend the attrition analysis that included only the first follow-up examination (Hense et al., 2013) and we build on the observed selection effects at baseline (Regber et al., 2013) and the association between recruitment effort and drop-out (Langeheine et al., 2017). Here we investigate the association of cohort attrition with sociodemographic characteristics, weight status, and study compliance in IDEFICS/I.Family ("study compliance" marks how far child and parents undertook all the requested measures and questionnaires). We also consider variations of attrition across the first and second follow-up and between the participating countries, focusing on selection effects by children's weight status.

Methods

Analysis group

In IDEFICS/I.Family, data were collected in each country in two or more selected communities. The sociodemographic profile and infrastructure of the communities were similar and typical for their region. All children aged 2 to 9.9 years attending kindergarten or primary school within each community were eligible. Parents of potential study subjects were either approached directly by mail or by letters delivered through teachers and caretakers in kindergartens and schools. They were asked for consent to examine their children as well as to answer a number of questionnaires. Children and parents were informed about all aspects of the study. Parents gave their written informed consent prior to inclusion into the study; children 12 years or older signed a simplified consent form. Immediately before each examination, a study nurse informed each child orally about the module using a simplified preformulated text. Children were informed that they do not have to participate if they don't want to and examinations were only performed if children assented and parents consented. Consent could be given to single components of the study while refusing others.

All procedures performed in IDEFICS/I.Family were in accordance with the ethical standards of the institutional committee and the 1964 Declaration of Helsinki and its later amendments. Approval was obtained by each of the centers engaged in the fieldwork by its appropriate ethics committees. To ensure that data collection and study parameters were similar between countries, a common manual

of operations containing standard operating procedures for all examinations was developed, and site visits were conducted in all study centers by a central quality control to ensure compliance.

In total, 16 228 children participated in the IDEFICS baseline examination (T0), carried out between September 2007 and May 2008 (Figure 2.1). All children who took part in the baseline examination were invited to the first follow-up (T1) between September 2009 and May 2010 where 11 041 children participated. Baseline and first follow-up included identical examination modules.

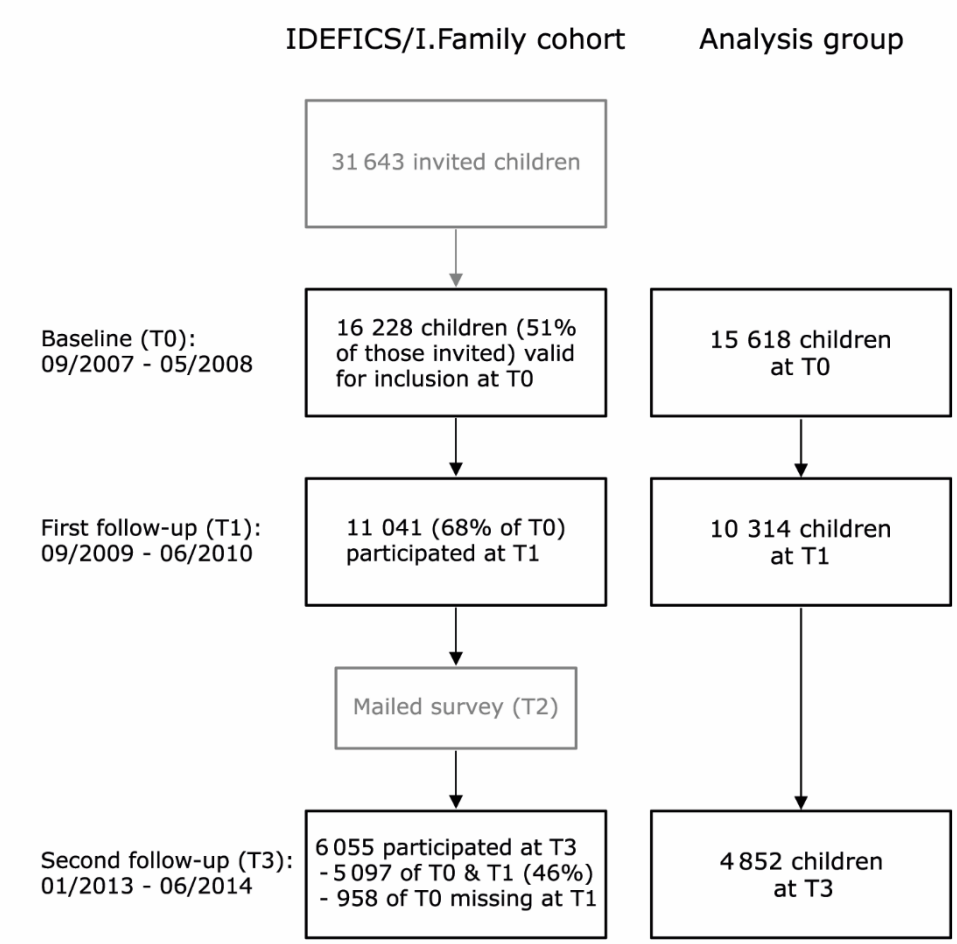

Figure 2.1 Flow chart of the participation in the IDEFICS baseline examination (T0), the first IDEFICS follow-up examination (T1), and the second follow-up examination in the I.Family study (T3). The refreshment samples at the first and the second follow-up (Ahrens et al. 2016) were excluded from this analysis

A second follow-up examination (I.Family, T3) was conducted between 2013 and 2014, again with similar examination modules (Ahrens et al., 2017). Children who participated at baseline, their siblings, and their parents were invited to take part in I.Family, and a total of 6 055 IDEFICS children were examined. Of the 11 041 children examined at the first follow-up, 5 097 children took part in I.Family. In addition, 958 children took part in I.Family who participated at baseline, but not in the first follow-up. Due to model constraints, these children were considered first follow-up drop-outs, that is, only baseline data were included in the analysis. A complete-cases analysis reduced the sample size to 15 618 children at baseline, 10 314 children at the first follow-up, and 4 852 children at the second follow-up. This resulted in a total of 25 932 person-wave observations at baseline and the first follow-up being included in the analysis. Baseline characteristics of the IDEFICS/I.Family baseline sample and the subsamples that participated in the two follow-ups are summarized in Table 2.1.

Table 2.1 Baseline characteristics of the IDEFICS/I.Family baseline sample and the subsamples that participated in the two follow-ups

	Baseline participants		Participated at T1				Participated at T3			
			Yes		No		Yes		No	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age child (years)	6.0	1.8	6.0	1.8	6.0	1.8	5.9	1.8	6.1	1.8
Compliance score of child	6.3	1.1	6.4	1.1	6.2	1.3	6.4	1.0	6.4	1.1
Compliance score parent(s)	3.5	0.8	3.5	0.8	3.5	0.8	3.5	0.7	3.4	0.8
Mother's age (years)	35.2	5.3	35.5	5.3	34.5	5.4	35.7	5.2	35.4	5.3
	n	%	N	%	n	%	n	%	n	%
Sex of child										
Male	7 928	50.8	5 418	50.6	2 510	51.1	2 511	50.8	2 907	50.4
Female	7 690	49.2	5 291	49.4	2 399	48.9	2 433	49.2	2 858	49.6
Weight status child										
Normal weight	12 543	80.3	8 724	81.5	3 819	77.8	4 002	80.9	4 722	81.9
Overweight	1 963	12.6	1 275	11.9	688	14.0	611	12.4	664	11.5
Obese	1 112	7.1	710	6.6	402	8.2	331	6.7	379	6.6
Weight status parents										
No parent overweight	3 943	25.2	2 836	26.5	1 107	22.6	1 355	27.4	1 481	25.7
At least one parent overweight	10 059	64.4	6 909	64.5	3 150	64.2	3 162	64.0	3 747	65.0
Missing	1 616	10.3	964	9.0	652	13.3	427	8.6	537	9.3

	Baseline participants		Participated at T1				Participated at T3			
			Yes		No		Yes		No	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Migration background										
No	12 906	82.6	8 972	83.8	3 934	80.1	4 149	83.9	4 823	83.7
Partly	1 384	8.9	939	8.8	445	9.1	475	9.6	464	8.0
Full	1 079	6.9	656	6.1	423	8.6	265	5.4	391	6.8
Missing	249	1.6	142	1.3	107	2.2	55	1.1	87	1.5
Educational level										
Low education	7 335	47.0	5 387	50.3	1 948	39.7	2 582	52.2	2 805	48.7
Medium education	7 010	44.9	4554	42.5	2 456	50.0	2 062	41.7	2492	43.2
High education	1 073	6.9	651	6.1	422	8.6	251	5.1	400	6.9
Missing	200	1.3	117	1.1	83	1.7	49	1.0	68	1.2
Number of adults in household										
One	1 255	8.0	763	7.1	492	10.0	332	6.7	431	7.5
Two	11 032	70.6	7 847	73.3	3 185	64.9	3 633	73.5	4 214	73.1
Three	1 081	6.9	740	6.9	341	6.9	372	7.5	368	6.4
Four or more	424	2.7	270	2.5	154	3.1	134	2.7	136	2.4
Missing										
Siblings aged < 18 years	1 826	11.7	1 089	10.2	737	15.0	473	9.6	616	10.7
Yes	3 667	23.5	2 364	22.1	1 303	26.5	1 063	21.5	1 301	22.6
No	10 364	66.4	7 412	69.2	2 952	60.1	3 482	70.4	3 930	68.2
Missing	1 587	10.2	933	8.7	654	13.3	399	8.1	534	9.3
Region										
Intervention	8 075	51.7	5 525	51.6	2 550	51.9	2 607	52.7	2 918	50.6
Control	7 543	48.3	5 184	48.4	2 359	48.1	2337	47.3	2 847	49.4
Country										
Spain	1 480	9.5	1 203	11.2	277	5.6	429	8.7	774	13.4
Hungary	2 496	16.0	1 218	11.4	1 278	26.0	468	9.5	750	13.0
Germany	2 008	12.9	1 165	10.9	843	17.2	637	12.9	528	9.2
Cyprus	2 111	13.5	1 589	14.8	522	10.6	862	17.4	727	12.6
Estonia	1 650	10.6	1 284	12	366	7.5	721	14.6	563	9.8
Belgium	1 884	12.1	1 236	11.5	648	13.2	242	4.9	994	17.2
Italy	2 241	14.3	1 543	14.4	698	14.2	873	17.7	670	11.6
Sweden	1 748	11.2	1 471	13.7	277	5.6	712	14.4	759	13.2
N	15 618	100	10 709	100	4 909	100	4 944	100	5 765	100

Abbreviations: SD, standard deviation

Outcome

The outcome *cohort attrition* was defined with respect to participation in the first (T1) and the second (T3) follow-up examination (0: participation vs. 1: dropout).

Exposures

The social position of families was classified according to the International Standard Classification of Education (ISCED) (UNESCO, 2012) using the highest educational attainment of mother or father (low: ISCED levels 0-2; medium: ISCED levels 3-4; high: ISCED levels 5 and higher). The household composition was described as the presence of non-adult siblings besides the participating child (yes vs. no) and the number of adults (age 18 or older) living in the household. The place of birth of parents served to define the migration background (full migrant: both parents foreign-born; partly migrant: one parent foreign-born; not migrant: otherwise). Children's age and mother's age on the day of the examination was recorded in years. For drop-outs at the first or second follow-up, children's and mother's age was estimated by adding the mean duration between two examinations to the age at the previous examination. Because of collinearity and a higher percentage of missing values, the father's age was not considered in the analysis. The weight status was determined using the body mass index (BMI). Children's weight status (thin and normal weight, overweight, obese) was categorized according to Cole and Lobstein (Cole and Lobstein, 2012). Parent's weight status (self-reported) was categorized as "no parent overweight", "at least one parent overweight", and "missing". Overweight was defined as having a BMI ≥25. A score of study compliance was constructed separately for children and parents based on the number of key examination modules they participated in at baseline and at first follow-up (Table 2.2). This was done by counting the number of completed modules (0: module not completed; 1: module completed). For children, key modules were blood pressure, bioelectrical impedance analysis (fasting state), waist-to-hip ratio, skinfold thickness (subscapularis and triceps), blood sample (fasting state), morning urine and saliva. Parents (respectively mother or father) provided key modules included the general questionnaire, food frequency questionnaire, medical history, and the 24-h dietary recall. At the first follow-up, the collection of saliva was restricted to children without a saliva sample at baseline. Therefore saliva was defined as being available at first follow-up if a sample was available at baseline or first follow-up.

Table 2.2 Number and percentage of children and parents participating in examination modules at baseline and at first follow-up

	Examination module	Participation	
		Baseline n (%)	First follow-up n (%)
Children	Blood pressure	14 752 (90.9)	10 563 (95.7)
	Bioelectrical impedance analysis (fasting state)	15 720 (96.9)	10 795 (97.8)
	Waist-to-hip ratio	15 551 (95.8)	10 731 (97.2)
	Skinfold thickness (subscapularis and triceps)	15 160 (93.4)	10 567 (95.7)
	Venous or capillary blood (fasting state)	12 855 (79.2)	8 528 (77.2)
	Morning urine	13 945 (85.9)	8 845 (80.1)
	Saliva[a]	14 273 (88.0)	188 (1.7)
Parents	General questionnaire (children parents)	16 117 (99.3)	10 539 (95.5)
	Food frequency questionnaire	15 199 (93.7)	9 963 (90.2)
	Medical history	12 418 (76.5)	8 978 (81.3)
	24-h dietary recall	11 671 (71.9)	5 520 (50.0)

[a] The collection of saliva at the first follow-up was restricted to children without a saliva sample at baseline. Therefore saliva was defined as being available at first follow-up if a sample was available at baseline or first follow-up.

Statistical analysis

The association between attrition and sociodemographic variables, weight status, and study compliance was assessed by estimating odds ratios (ORs) and 99% confidence intervals (CIs) using a multivariable multilevel logistic regression with respondents as the second-level variable and country as the third-level to account for clustering (Snijders and Bosker, 1999). To avoid that meaningless associations become statistically significant just because of the large sample size and to account for multiple testing of associations a more stringent criterion for statistical significance ($\alpha = 0.01$) was chosen. Data were transformed such that each unit of analysis represented a person-wave observation (Allison, 1985; Watson, 2003). Variables included in the model were either time constant (e.g., sex of the child), or time-variant predictors (e.g., weight status of the child). Time-variant predictors were modeled as lagged covariates, that is, information at baseline was regressed on attrition at first follow-up and information at the first follow-up was regressed on attrition at the second follow-up. Sensitivity analyses were carried out to check for non-independence of siblings in the sample. Random sampling (n = 100) was used to select one child of each family and calculate a random intercept logistic regression model for each sample to obtain a mean odds ratio and a corresponding confidence interval for each predictor. The odds ratios of a logistic regression model with all children and the mean odds

ratios for the 100 samples did not differ substantially. To assess the variation of attrition across time in separate models all possible interaction terms between potential predictors of attrition and time point of follow-up examination were calculated (time x (sex of child, age child, weight status child, compliance score of child, compliance score parent(s), mother's age, weight status parents, migration background, educational level, number of adults in household, siblings aged < 18 years, and region)). The heterogeneity between the countries was investigated by means of meta-analyses: Country-stratified logistic regression models with attrition as the dependent variable and the same predictors as in the random intercept logistic regression model were fitted and a random-effects meta-analysis (RE model) (Riley et al., 2011) was calculated for each predictor of the country-stratified logistic regression models. To evaluate the heterogeneity of attrition between the countries, the percentage of variation that is due to heterogeneity, I^2 (Higgins, 2003), and forest plots were used. Selection effects on children's BMI across time were assessed with quantile-quantile plots (Q-Q plots) and Kolmogorov-Smirnov tests (KS test) (Wang et al., 2003). We explored the impact of selection effects on the cross-sectional association of social position and weight status. Children's weight status was converted into a binary variable (0: normal weight including thin vs. 1: obese including overweight) further referred to as *overweight/obesity*. Social position included educational level (as described above) and income level (low, low/medium, medium, medium/high, vs. high income). We estimated baseline associations and then estimated identical associations with subsamples restricted to first follow-up participants (T1) and second follow-up participants (T3) as well as associations at the first follow-up (T1) and the restricted sample of second follow-up participants (T3). In addition, we explored selection effects on the association between adherence to key messages of a healthy lifestyle promoted by IDEFICS/I.Family and overweight/obesity published by Kovacs et al. (2015). In this analysis we included total screen time, moderate to vigorous physical activity (MVPA), and sleep duration as measures of adherence (see Kovacs et al. (2015) for detailed information on instruments and operationalization). In accordance with Kovacs et al. (2015), we calculated a binary indicator for adherence on respective cut points for screen time, MVPA, and sleep. We estimated the baseline association of adherence and overweight/obesity and then estimated the identical association with subsamples restricted to first follow-up participants (T1) and second follow-up participants (T3). For the exposure-outcome association of adherence to key messages of a healthy lifestyle and overweight/obesity, as well as social position and overweight/obesity we estimated odds ratios and confidence intervals with multivariable multilevel logistic regression models. For the sake of comparability we used 95% confidence intervals in the analysis reproducing the association of adherence to key messages and overweight/obesity published by. Kovacs et al. (2015) (described above)., All other analyses, as pointed out above, utilized 99% confidence intervals.

To quantify a potential bias we calculated the percent change in point estimates (CPE = $OR_{subsample}$ / $OR_{full\ sample}$ x 100 – 100). We considered a CPE of above 10% as indicator of a bias. For Table 2.6 we stratified overweight/obesity by the combination of adherence to key messages regarding media consumption, physical activity and sleep. Children who *did not adhere* to the recommendations of screen time and physical activity and sleep duration were assigned to the group - - - (1 666 children in T0 full sample). In contrast, children who *did adhere to all* recommendations of screen time and physical activity and sleep duration were assigned to the group *+ + +* (263 children in T0 full sample). Children who adhered only to some of the recommendations were assigned accordingly. A full description of the analysis is given in Kovacs et al. (2015). Analyses were performed using R version 3.3.3 (http://www.r-project.org/).

Results

The multilevel logistic regression model with cohort attrition as dependent variable (Table 2.3) revealed that children's age in years was positively associated with attrition (OR 1.05, 99% CI 1.02-1.07). Compared to normal weight children, overweight (OR 1.17, 99% CI 1.05-1.29) or obese (OR 1.18, 99% CI 1.03-1.36) children had a higher chance of attrition. Higher study compliance of children was associated with lower attrition (OR 0.84, 99% CI 0.81-0.87), as was higher mother's age (OR 0.98, 99% CI 0.97-0.99). Children with a partly (OR 1.13, 99% CI 1.00-1.28) or full migrant background (OR 1.41, 99% CI 1.21-1.63) had a higher chance of attrition, as had children of parents with a low (OR 1.49, 99% CI 1.27-1.74) or medium (OR 1.19, 99% CI 1.10-1.29) educational level.

Table 2.3 Odds ratios with 99% confidence intervals for cohort attrition

	Cohort attrition				
	No		Yes		OR[a] (99% CI)
	n	%	n	%	
Time					
First follow-up (T1)	10 709	68.6	4 909	31.4	ref.
Second follow-up (T3)	4 852	47.0	5 462	53.0	2.62 (2.32-2.96)
Sex of child[b]					
Male	7 883	60.1	5 244	39.9	ref.
Female	7 678	60.0	5 127	40.0	0.99 (0.93-1.07)
Age child (years)[c]					1.05 (1.02-1.07)
Weight status child[c]					
Normal weight	12 418	60.9	7 978	39.1	ref.
Overweight	2 059	56.5	1 588	43.5	1.17 (1.05-1.29)
Obese	1 084	57.4	805	42.6	1.18 (1.03-1.36)
Compliance score of child[c]					0.84 (0.81-0.87)
Compliance score parent(s)[c]					0.93 (0.88-0.98)
Mother's age (years)[c]					0.98 (0.97-0.99)
Weight status parents[c]					
No parent overweight	4 010	62.7	2 388	37.3	ref.
At least one parent overweight	10 074	60.2	6 660	39.8	1.05 (0.96-1.14)
Missing	1 477	52.8	1 323	47.2	1.23 (1.08-1.40)
Migration background[b]					
No	12 961	60.6	8 427	39.4	ref.
Partly	1 395	61.6	871	38.4	1.13 (1.00-1.28)
Full	905	53.8	778	46.2	1.41 (1.21-1.63)
Missing	300	50.4	295	49.6	1.35 (0.96-1.91)
Educational level[b]					
Low education	7 965	63.2	4 639	36.8	1.49 (1.27-1.74)
Medium education	6 656	57.9	4 837	42.1	1.19 (1.10-1.29)
High education	912	53.4	796	46.6	ref.
Missing	28	22.0	99	78.0	1.12 (0.77-1.62)
Number of adults in household[c]					
One	1 122	54.2	949	45.8	ref.
Two	11 401	61.5	7 149	38.5	0.89 (0.78-1.02)
Three	1 109	59.5	755	40.5	0.98 (0.82-1.18)
Four or more	414	57.3	308	42.7	1.02 (0.80-1.30)
Missing	1 515	55.6	1 210	44.4	1.02 (0.72-1.44)
Siblings aged < 18 years[c]					
Yes	10 897	61.6	6 796	38.4	0.90 (0.83-0.99)
No	3 334	57.3	2 480	42.7	ref.
Missing	1 330	54.8	1 095	45.2	0.94 (0.67-1.34)
Region[b]					
Intervention	8 072	60.3	5 322	39.7	ref.
Control	7 489	59.7	5 049	40.3	1.05 (0.98-1.13)
Country (third-level)[b]					
Spain	1 622	61.4	1 020	38.6	
Hungary	1 622	46.5	1 868	53.5	
Germany	1 808	57.2	1 352	42.8	
Cyprus	2 455	66.2	1 252	33.8	
Estonia	2 009	68.4	927	31.6	
Belgium	1 480	48.1	1 597	51.9	
Italy	2 414	63.8	1 368	36.2	
Sweden	2 151	68.5	987	31.5	
N	15 561	60.0	10 371	40.0	

Abbreviations: OR, odds ratio; CI, confidence interval; ref., reference category

[a] Adjusted for country.

[b] Time invariant variable using information from baseline (T0).

[c] Time variant variable using information from baseline (T0) and first follow-up (T1).

Variations of attrition across time are depicted in Figure 2.2 as probabilities predicted from separate random intercept logistic regression models containing interaction terms between potential predictors of attrition and time point of follow-up examination. Age of child was not associated with attrition at the first follow-up but was positively associated with attrition at the second follow-up. Higher parent's study compliance was associated with lower attrition at the second follow-up, but was not associated with attrition at the first follow-up. A higher age of the mother was associated with lower attrition at first follow-up but not at the second follow-up.

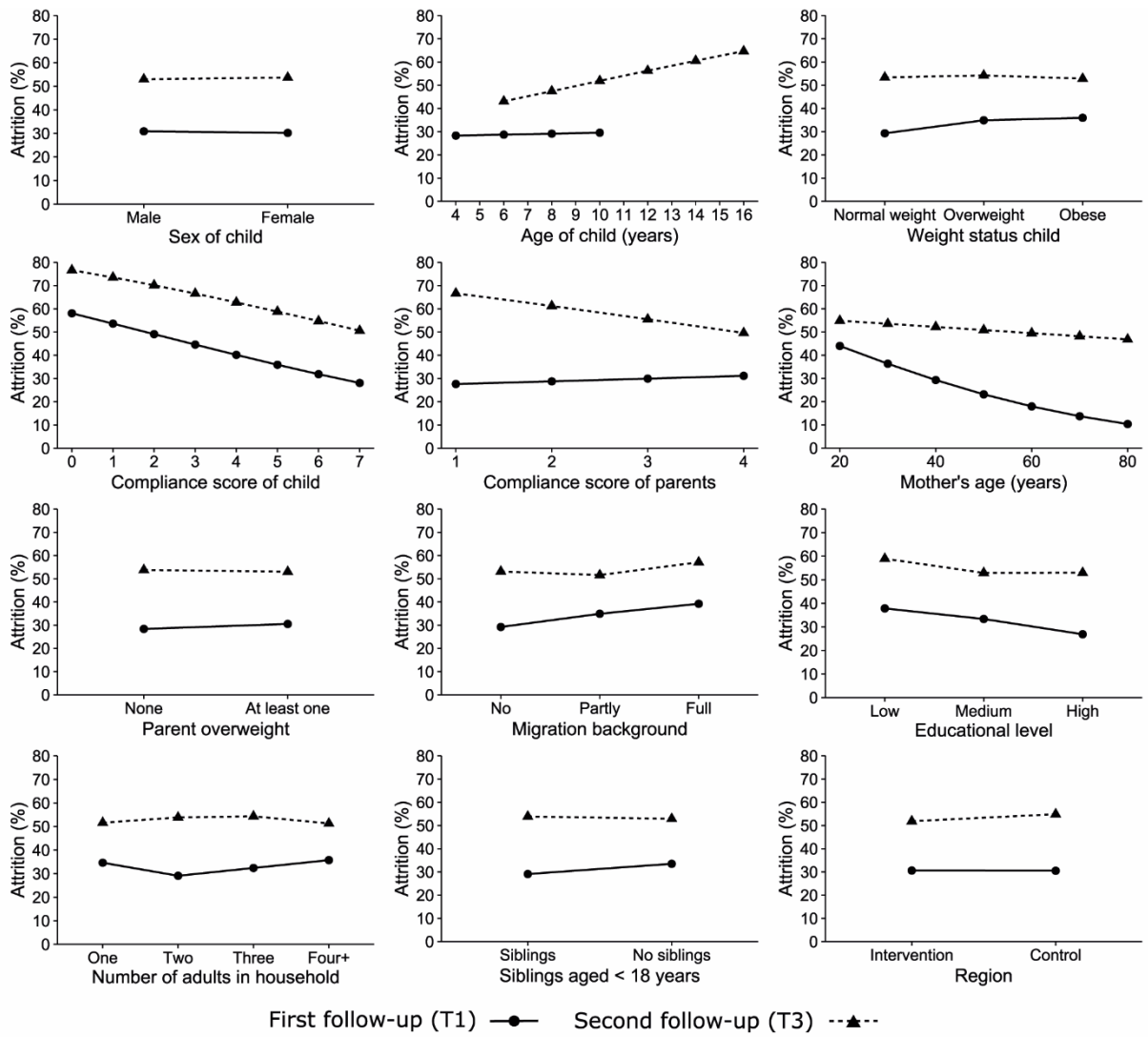

Figure 2.2 Predicted probabilities of separately modeled interactions of the predictors and time in the random intercept logistic regression. Time variant variables are age of child (years), weight status of child, compliance score of child and of parents, mother's age (years), and parent's overweight

To assess how well the model represented data of individual countries, we explored with forest plots whether single countries differed notably from the overall pattern, that is, whether the sign of a countries' odds ratio for a given exposure variable differed from the pooled estimate (Figure 2.3). For

14 out of 17 predictors, estimates for all countries were in line with the pooled estimate. Female children in Belgium had a lower chance of attrition, whereas no association of sex was found for the pooled estimate. A medium educational level was associated with a lower chance of attrition in Italy, while the pooled estimate indicated a higher chance of attrition for a low or high educational level. Further, children from the control region in Belgium had a higher chance of attrition while no association for the region was evident in the pooled estimate. Between countries, substantial heterogeneity was observed for study compliance of children, weight status (overweight / obese; I^2 from 50% to 70%), age of the child, study compliance of parents, full migrant status, low or medium education, and control region (I^2 from 70% to 100%). Sensitivity analyses showed that exclusion of country-stratified odds ratios identified as exceptions attenuated I^2: Excluding Belgium decreased I^2 to zero for the predictor female and decreased I^2 for the control region; excluding Italy decreased the I^2 of low education.

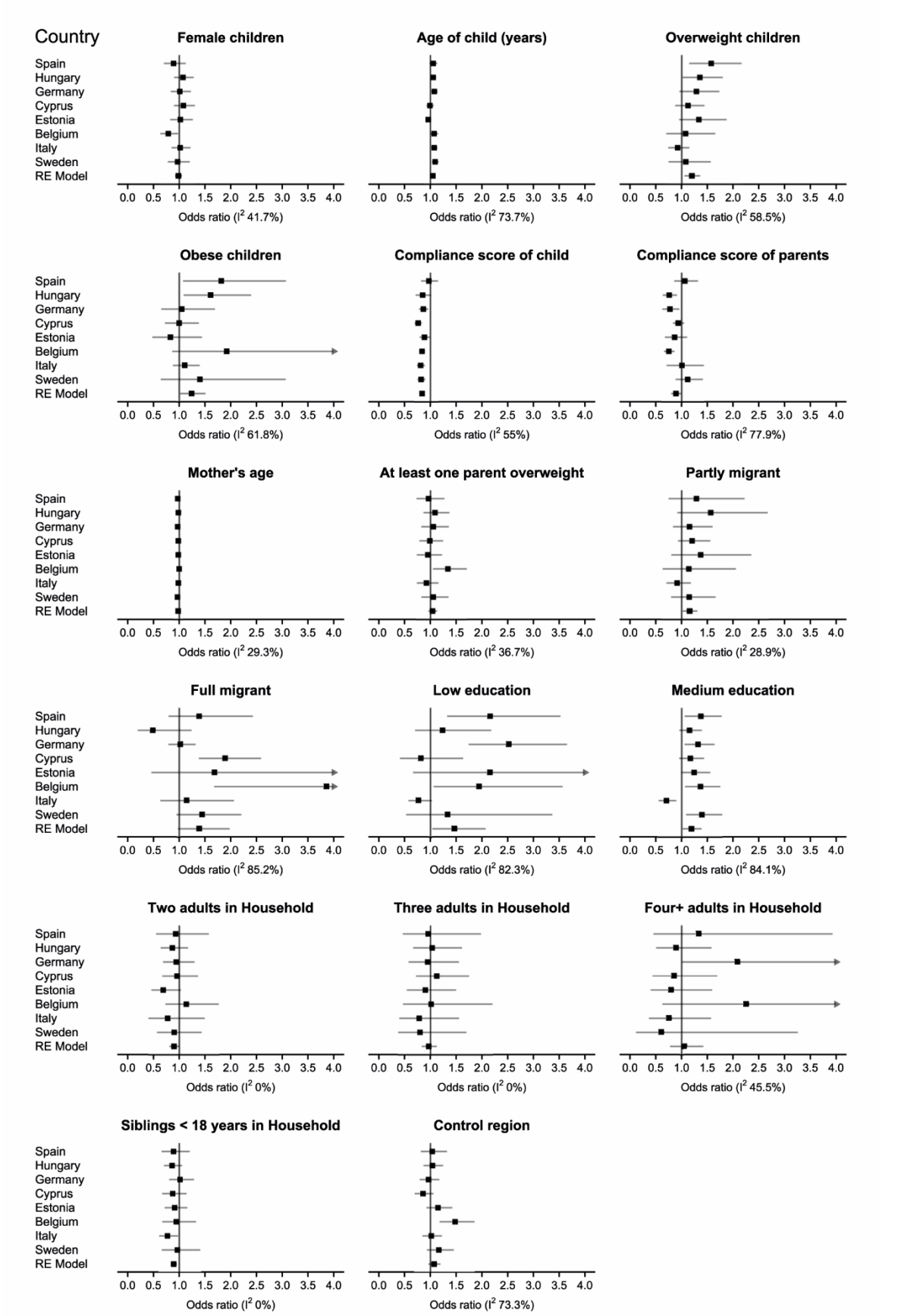

Figure 2.3 Odds ratios for attrition (with 99% confidence intervals) from country-stratified logistic regression models (ordered by baseline response) with attrition as dependent variable and same predictors as in the random intercept logistic regression model (Table 2.3) as well as a pooled estimate

of a random-effects meta-analyses (RE model) and I^2 (%) as a measure of heterogeneity between the countries. I^2 values above 50% indicating substantial heterogeneity were observed for 9 out of 17 variables. Arrows at the upper limit of a confidence interval indicate that the confidence interval extends past four

Since IDEFICS/I.Family was a multi-purpose cohort focusing on overweight and obesity, we further investigated selection effects of children's BMI. BMI distributions for all children and the corresponding BMI distributions for children that did not drop out at a particular follow-up are displayed in Figure 2.4, column 1 to 3. The histograms of BMI at baseline and BMI at baseline without the children that dropped out at the first follow-up differed in the number of observations per bin but the Q-Q plot as well as the KS test (2 sided *P* value of 0.30) (Figure 2.4b, column 1) indicated equal distributions. Similar results were obtained for the distribution of BMI at the first follow-up and the resulting distribution when second follow-up drop-outs were excluded (KS test: 2 sided *P* value of 0.73) (Figure 2.4b, column 2) as well as for the distribution of BMI at baseline and the corresponding distribution without second follow-up dropouts (KS test: 2 sided *P* value of 0.76) (Figure 2.4b, column 3).

Density scatter plots with children's BMI at baseline plotted against BMI at the first follow-up (respectively BMI at first follow-up vs. BMI at second follow-up; BMI at baseline vs. BMI at second follow-up) and ß coefficients of linear regression models were used to evaluate selection effects of BMI across time (Figure 2.4c). The correlation between children's BMI at different time points was consistent across time, both in the shape of the scatter plot and the ß coefficients (baseline vs. first follow-up: ß = 1.15, R^2 = 0.79; first follow-up vs. second follow-up: ß = 1.14, R^2 = 0.76; baseline vs. second follow-up: ß = 1.28, R^2 = 0.57).

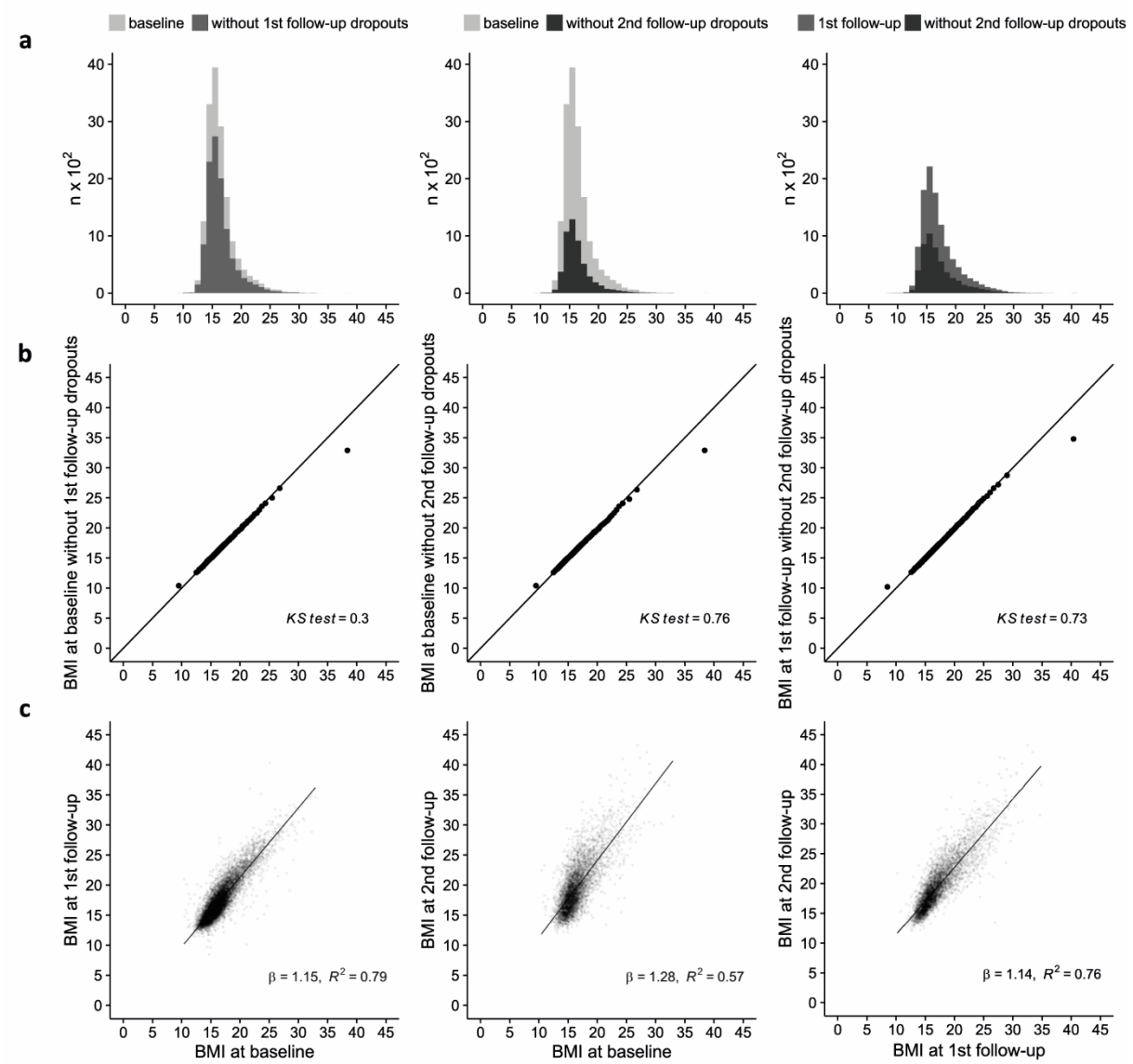

Figure 2.4 Effect of cohort attrition on the distribution of children's body mass index (BMI). (a) Histograms of children's BMI at baseline for different subgroups show no evidence for differences in distributions. Left: full baseline sample vs. baseline sample without attrition at first follow-up; middle: full baseline sample vs. baseline sample without attrition at second follow-up; right: first follow-up sample vs. first follow-up sample without attrition at second follow-up.(b) Quantile-quantile plot (Q-Q plot) and Kolmogorov-Smirnov test (KS test, 2 sided P value) of children's baseline BMI for different subgroups show no evidence for differences in distributions. Results of KS tests (all $p \geq 0.3$) also indicate insufficient evidence to reject the null hypothesis that the respective distributions are the same. Left: full baseline sample vs. baseline sample without attrition at first follow-up; middle: full baseline sample vs. baseline sample without attrition at second follow-up; right: first follow-up sample vs. first follow-up sample without attrition at second follow-up (c) Scatter plot and results of linear regressions (ß coefficients and R^2) between children's BMI at different time points indicate that the correlation was consistent across time. Left: full baseline sample vs. baseline vs. first follow-up, middle: baseline vs. second follow-up; right: first follow-up vs. second follow-up.

We explored the impact of selection effects due to the association between childhood overweight and social position (e.g., (Ahrens et al., 2014); for a review, (Knai et al., 2012)) at baseline and both follow-ups. To this end, we estimated associations between BMI and social variables in the complete baseline sample and compared them to associations between the same variables in two subsamples restricted to participants of the first follow-up and participants of the second follow-up, respectively (Table 2.4). We repeated this procedure with data from the first follow-up for all participants of the first follow-up and a subsample restricted to all participants of the second follow-up (Table 2.4).

At all time points, a lower income level was associated with a higher chance of overweight/obesity. Restricting the baseline association (T0) of income level to T1 participants marginally affected odds ratios indicated by a CPE of less than 10% but led to bigger confidence intervals (e.g. low income at baseline (T0): full sample (OR 1.43, 99% CI 1.12-1.82) vs. T1 participants (OR 1.34, 99% CI 1.00-1.81) vs. T3 participants (OR 1.55, 99% CI 1.02-2.38)). A restriction to T3 participants resulted in a CPE for medium income level of 11%. For the restricted subsample at first follow-up (T1 full sample restricted to T3 participants), odds ratios tended to be higher as compared to other estimates of income level. Apart from medium income, CPE of income level was well above 10%.

A lower educational level was associated with a higher chance of overweight/obesity at baseline and first follow-up. Restricting the association of educational level and overweight/obesity did not affect the trend of this association, with a CPE for the baseline association restricted to T3 participants of 14.8% for medium educational level and the confidence intervals.

Table 2.4 Association between overweight/obesity and social position (odds ratios with 99% confidence intervals)

	Baseline association					First follow-up association		
	T0 full sample OR[a] (99% CI)	T1 subsample OR[a] (99% CI)	% change in OR (T1 vs. T0)	T3 subsample OR[a] (99% CI)	% change in OR (T3 vs. T0)	T1 full sample OR[a] (99% CI)	T3 subsample OR[a] (99% CI)	% change in OR (T3 vs. T1)
Income level[c]								
Low	1.43 (1.12-1.82)	1.34 (1.00-1.81)	-6.3	1.55 (1.02-2.38)	8.4	1.47 (1.13-1.92)	1.79 (1.22-2.63)	21.8
Low/medium	1.38 (1.09-1.74)	1.36 (1.02-1.80)	-1.5	1.33 (0.87-2.03)	-3.6	1.43 (1.12-1.84)	1.77 (1.23-2.57)	23.8
Medium	1.36 (1.10-1.68)	1.30 (1.01-1.68)	-4.4	1.51 (1.04-2.19)	11.0	1.39 (1.12-1.73)	1.53 (1.11-2.11)	10.1
Medium/high	1.25 (1.00-1.57)	1.18 (0.90-1.55)	-5.6	1.29 (0.86-1.92)	3.2	1.16 (0.92-1.48)	1.40 (0.98-1.98)	20.7
High	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Educational level[b]								
Low	1.55 (1.21-1.97)	1.60 (1.18-2.17)	1.6	1.78 (1.12-2.84)	9.0	1.72 (1.27-2.32)	1.69 (1.06-2.69)	-8.2
Medium	1.22 (1.06-1.40)	1.24 (1.05-1.47)	3.2	1.33 (1.04-1.70)	14.8	1.22 (1.04-1.43)	1.12 (0.88-1.41)	-1.7
High	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Age child (years)[c]	1.24 (1.20-1.29)	1.25 (1.20-1.31)	0.8	1.23 (1.16-1.31)	-0.8	1.18 (1.14-1.23)	1.19 (1.12-1.26)	0.9
Sex of child[b]								
Male	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Female	1.21 (1.07-1.36)	1.20 (1.04-1.39)	-0.8	1.14 (0.92-1.40)	-5.8	1.11 (0.97-1.27)	1.14 (0.93-1.38)	2.7
Region[b]								
Intervention	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
Control	0.94 (0.84-1.06)	0.95 (0.82-1.10)	1.1	0.89 (0.72-1.09)	-5.3	0.97 (0.84-1.11)	0.89 (0.73-1.09)	-8.3
N	13 855	9 604		4 461		9 068	4 340	

Abbreviations: OR, odds ratio; CI, confidence interval; ref., reference category

[a] Odds ratios and confidence intervals for overweight/obesityadjusted for country.

[b] Time invariant variable using information from baseline (T0).

[c] Time variant variable using information from baseline (T0), first follow-up (T1) or second follow-up (T3).

IDEFICS/I.Family covered multiple topics including diet, physical activity, sleep, and stress. Results from the baseline examination showed that adherence to key behaviors of a healthy lifestyle was associated with a lower chance of overweight/obesity (Kovacs et al., 2015). We checked whether BMI related selection effects changed the results of Kovacs *et al.* (Kovacs et al., 2015) if the sample was restricted to participants of the first follow-up and the second follow-up, respectively. At baseline, adherence to the key messages total screen time, MVPA and sleep duration (Table 2.5) was associated with a lower chance of overweight/obesity (OR 0.76, 95% CI 0.70-0.83; OR 0.70, 95% CI 0.58-0.84; OR 0.85, 95% CI 0.74-0.96; reported in Kovacs et al. (2015)), their Table 4, rightmost column). Estimating the association of key messages and overweight/obesity restricted to a subsample of T1 participants affected odds ratios and confidence intervals marginally with a CPE of less than 5%. But a restriction to a subsample of T3 participants resulted in confidence intervals for MVPA and sleep duration that included the reference category. Nevertheless, only the CPE for MVPA (22.9%) exceeded 10%.

Table 2.5 Association between overweight/obesity and adherence to key messages concerning a healthy lifestyle (odds ratios with 95% confidence intervals)

	Baseline association				
	T0 full sample	T1 subsample	% change in	T3 subsample	% change in
	OR[a] (95% CI)	OR[a] (95% CI)	OR (T1 vs. T0)	OR[a] (95% CI)	OR (T3 vs. T0)
Total screen time[b]	0.76 (0.70-0.83)	0.79 (0.71-0.88)	4.0	0.80 (0.68-0.93)	5.3
N	15 084	10 374		4 791	
MVPA >60 min per day[c]	0.70 (0.58-0.84)	0.69 (0.55-0.86)	-1.4	0.86 (0.61-1.20)	22.9
N	7 447	5 219		2 421	
Sleep duration[d]	0.85 (0.74-0.96)	0.82 (0.70-0.96)	-3.5	0.90 (0.72-1.14)	5.9
N	10 495	7 370		3 562	

Abbreviations: OR, odds ratio; CI, confidence interval

[a]Odds ratios and confidence intervals for overweight/obesityof single models adjusted for age, sex, and country.

[b] <1 h in pre-school and <2 h in school children (reference category: ≥1 h in pre-school and ≥2 h in school children).

[c] Reference category: MVPA ≤60 min per day.

[d] ≥11 h in pre-school and ≥10 h in school children (reference category: <11 h in pre-school and <10 h in school children).

We found similar results for a detailed examination of the association of adherence and overweight/obesity stratified by the combination of adherence to key messages published in Kovacs *et al.* (Kovacs et al., 2015) (their Table 6, rightmost columns) and the identical analysis restricted to subsamples of T1 or T3 participants (Table 2.6). However, the majority of CPEs for a restricted subsample of T3 participants were well above 10%.

Table 2.6 Odds ratios with 95% confidence intervals for overweight/obesity stratified by the combination of adherence to key messages regarding media consumption, physical activity, and sleep

			Baseline association										
			T0 full sample			T1 subsample			% change in OR (T1 vs. T0)	T3 subsample			% change in OR (T3 vs. T0)
TV	PA	Sleep	Normal[a]	Obese[b]	OR[c] (95% CI)	Normal[a]	Obese[b]	OR[c](95% CI)		Normal[a]	Obese[b]	OR[c] (95% CI)	
-	-	-	1 289	377	ref.	933	259	ref.	ref.	498	128	ref.	ref.
+	-	-	1 168	317	0.73 (0.61-0.88)	834	215	0.72 (0.58-0.91)	-1.4	444	96	0.66 (0.48-0.92)	-9.6
-	+	-	227	44	0.62 (0.44-0.90)	174	28	0.56 (0.36-0.87)	-9.7	89	16	0.72 (0.40-1.30)	16.3
-	-	+	550	132	0.91 (0.70-1.18)	367	77	0.85 (0.61-1.17)	-6.6	150	35	1.01 (0.62-1.64)	11.0
+	+	-	166	30	0.54 (0.35-0.82)	117	24	0.68 (0.42-1.10)	25.9	54	13	0.92 (0.47-1.78)	70.4
+	-	+	938	180	0.63 (0.50-0.80)	671	120	0.65 (0.49-0.87)	3.2	296	56	0.75 (0.49-1.15)	19.1
-	+	+	123	16	0.48 (0.27-0.83)	81	10	0.49 (0.24-0.98)	2.1	36	6	0.74 (0.29-1.90)	54.2
+	+	+	234	29	0.42 (0.28-0.66)	164	16	0.38 (0.21-0.66)	-9.5	68	6	0.39 (0.16-0.95)	-7.1
N			4 695	1 125		3 341	749			1 635	356		

Abbreviations: OR, odds ratio; CI, confidence interval; ref. reference category

[a]Normal weight including thin.

[b] Obese including overweight.

[c] Odds ratios and confidence intervals for overweight/obesityof single models adjusted for age, sex, and country.

+ Adherence (-: non-adherence) to key messages: TV (total screen time), PA (physical activity), and sleep (sleep duration).

Discussion

In line with earlier research our results suggest that higher attrition at the follow-ups was associated with a higher weight status of children, lower children's study compliance, older age, lower parental education, and parent's migration background (Behr et al., 2005; Howe et al., 2013; Lange et al., 2014; Vinther-Larsen et al., 2010). For a multi-purpose cohort focusing on overweight and obesity, the observed association between weight status and attrition was perhaps to be expected. For instance, children with higher BMI might have felt more uncomfortable having their weight measured (in underwear) at baseline, causing them to refuse participation in follow-ups. Or participation in IDEFICS/I.Family might not have met the expectations of children and/or parents concerning a health study, leading them to leave the cohort that "did not work out for them". However, while selection effects on children's BMI did occur, they appeared to only slightly distort the distribution at the upper tail, mainly above the 99% percentile.

We found that older children were less likely to take part in the second follow-up as compared to younger ones. In contrast to studies on adults, the consent of both parents and children was required for inclusion into this study, and it has been shown that this makes recruitment particularly challenging (Fletcher and Hunter, 2003). In particular it remains unclear to which degree the opinion of parents and/or children were decisive for participating. It is reasonable to assume that, as they get older, children act more autonomously and hence have more say regarding whether or not to participate. As children transit into puberty, they might find epidemiological studies less interesting or, might get increasingly uncomfortable with getting examined in underwear. Unfortunately, although puberty status was part of the study protocol at the second follow-up, it was not included at baseline and at first follow-up, rendering it impossible to investigate links between puberty status and attrition.

Furthermore, the association between children's age and attrition might also be influenced by residential mobility, which has been shown to be highly associated with attrition as it can lead to invalid contact data (Behr et al., 2005; Lepkowski and Couper, 2002; Watson, 2003). In most of the participating countries, the transition from primary to secondary school happens when children are between 10 to 12 years old (except for Estonia's and Sweden's single structure school systems) and for many of the children, this transition took place between the first and second follow-up. Hence many families might have used this opportunity to relocate, possibly leading to dropouts if the family moved out of the study region or their contact data became invalid.

As participants were free to decide whether or not to take part in individual study modules, we used the study compliance of both parents and children as proxy-measures of motivation. The fact that both parent's and children's study compliance clustered among high values indicates that once people made the decision to take part they completed the study program as a whole. Nevertheless, children's participation was noticeably lower for the collection of the invasive biosamples. Parents were more

likely to complete all modules that took place in the study center (general questionnaire, food frequency questionnaire, and medical history), and less likely to complete the take-home questionnaires (24-h dietary recall). This could be due to the fact that the latter questionnaires were more time consuming and involved setting aside additional time for the study.

Unfortunately, it cannot be ruled out that parents of children with certain diagnoses covered by the medical questionnaire might have been more reluctant to complete it to avoid stigmatization. Previous results from the cohort published elsewhere showed that, for instance, the prevalence of ADHD in the cohort was somewhat lower as compared to the whole population (Pohlabeln et al., 2017).

Heterogeneity analyses revealed that countries differed considerably in how well the overall model captured the influence of different predictors on attrition. However, although the heterogeneity between the countries was high in terms of I^2, a closer look using country stratified forest plots revealed that for many predictors all countries showed similar trends. For some predictors, high heterogeneity estimates appeared to be caused by single outliers because excluding these outliers improved I^2 considerably. While there are plausible explanations for some of the deviations from the general trend, for others there are none. For instance, Italy's estimates for the influence of educational level probably deviated because of the small proportion of parents with high educational level in their sample. However, it is not clear why female children in Belgium were more likely to take part in the follow-ups, whereas no such association was obvious for other countries. Similarly, we cannot explain why attrition differed for control and intervention regions in Belgium, but not in other participating countries. Often such inconsistencies can be explained by investigating paradata recorded during recruitment (i.e., information about the process of the data collection (Groves and Heeringa, 2006)) with dedicated documentation systems (e.g., (Langeheine et al., 2017)). Unfortunately paradata were only available for the German study cohort (Langeheine et al., 2017), rendering an analysis for the whole cohort impossible. The collection of paradata might thus be especially crucial in multicenter cohort studies, where documentation is often difficult to coordinate between different survey teams operating over long periods of time.

Analysis of selection effects on cross-sectional exposure-outcome associations revealed few effects on point estimates when restricting the full sample at baseline (T0) to participants of the first follow-up (T1). Results on CPEs after restricting the exposure-outcome associations to a subsample of second follow-up participants (T3) were mixed. In particular in the detailed analysis of adherence and overweight/obesity CPEs exceeded 10%, potentially caused by a sharp decline in the number of observations for the subgroups.

Strengths and limitations

Strengths of our study include the large sample size from an international population and the highly standardized procedures for data collection that were enforced by a central quality control. As noted previously, interpretation of our results would have benefitted if information about puberty status would have been gathered at each time point and more centers would have collected paradata.

Conclusion

Potential bias in cohort studies induced by attrition may vary according to exposure and outcome (Greene et al., 2011) and even a high level of attrition may have a limited effect on estimates of associations between exposure and outcome (Greene et al., 2011; Howe et al., 2013; Powers et al., 2015). Our results, however suggest that the IDEFICS/I.Family cohort gives valid estimates of the associations of interest.

3 Consequences of an extended recruitment on participation in the follow-up of a child study: Results from the German IDEFICS cohort

Malte Langeheine, Hermann Pohlabeln, Wolfgang Ahrens, and Stefan Rach, on behalf of the IDEFICS consortium[2]

Abstract

Background: Declining response proportions in population-based studies are often countered by extended recruitment efforts at baseline that may, however, result in higher attrition in a subsequent follow-up. This study analysed the effect of extended recruitment efforts on attrition at the first follow-up of a child cohort.

Methods: We used paradata (i.e., information about the process of data collection) from the German IDEFICS cohort investigating dietary- and lifestyle-induced health effects on children to quantify recruitment effort and classify respondents as completing the recruitment early vs. late for baseline and follow-up separately. Multilevel logistic regression models were used to investigate the association between recruitment effort and attrition at follow-up (loss to follow-up) adjusted for sociodemographic and health related variables.

Results: Individuals who were late respondents at baseline and early respondents at the follow-up had a higher chance of attrition (odds ratio 1.65, 95% confidence interval 1.19, 2.28) as compared to other groups. An investigation of reasons for non-participation revealed that members of this group were more likely to be not reachable by phone.

Conclusions: An extended recruitment effort at baseline of a child cohort study is not per se associated with a higher chance of attrition at follow-up. Much care should be taken to collect valid telephone numbers.

Facing declining response proportions in population-based studies (Battaglia et al., 2007; Curtin et al., 2005; de Leeuw and de Heer, 2002), extended recruitment efforts have become a common practice (Hall et al., 2013; Lacey and Savage, 2016). Since an extended recruitment may influence the composition of the recruited sample, a typical strategy is to check for differences between participants with a low and a high number of contact attempts (so called early and late respondents). It has been shown that early and late respondents indeed can differ in relevant characteristics (Hall et al., 2013).

[2] Langeheine M, Pohlabeln H, Ahrens W, Rach S. Consequences of an Extended Recruitment on Participation in the Follow-Up of a Child Study: Results from the German IDEFICS Cohort. Paediatric and Perinatal Epidemiology. 2017;31(1):76-86. doi:10.1111/ppe.12328.

For instance, a study on drinking and alcohol-related experiences revealed that late respondents are more likely to be male, younger, living in more deprived areas, and to have a higher prevalence of drinking (Maclennan et al., 2012). Studer et al. (Studer et al., 2013) reported that late respondents show a pattern of higher substance use like tobacco and cannabis. Other studies suggest that late respondents tend to be less educated (Cohen et al., 2000; Haring et al., 2009).

For cohort studies, however, the sample composition is not the only concern, since extended recruitment efforts during the baseline study may differentially influence the loss of respondents during follow-up (Lynn, 2009). For instance, previous research showed that late respondents at baseline were more prone to drop out during follow-up as compared to early respondents (Cohen et al., 2000; Haring et al., 2009; Nederhof et al., 2012). It is not clear, however, whether all results from adult cohort studies can be generalized to child/youth cohorts. The recruitment for child/youth cohorts is especially challenging, because the consent of two parties (child and parents) is required for inclusion in the study. Since it has been reported that parent's failures to provide consent are more often caused by failures to respond rather than deliberate refusals (Fletcher and Hunter, 2003), extended recruitment strategies that aim at decreasing parental nonresponse are being developed and tested (e.g., (Schilpzand et al., 2015; Wolfenden et al., 2009)). In light of the literature, however, it is likely that an extended recruitment effort at baseline is penalized at follow-up, that is, the additional participants gained by extending the recruitment are more likely to dropout at a later stage of the study as compared to participants recruited with lower efforts.

We analysed paradata (i.e., information about the process of data collection (Groves and Heeringa, 2006)) from a child cohort investigating dietary- and lifestyle-induced health effects (IDEFICS cohort) (Ahrens et al., 2011). We quantified recruitment effort using paradata from the baseline and the first follow-up examination and assessed their associations with sociodemographic and health-related variables. Furthermore, we investigated the influence of the recruitment effort on attrition at follow-up and on the reasons for dropout.

Methods

The IDEFICS study is an international population-based multicenter child cohort study across eight European countries (Ahrens et al., 2011). For our analysis, only data from the German children of the IDEFICS cohort were included, because paradata were not available for other countries. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional committee and with the 1964 Helsinki declaration and its later amendments. Approval was obtained by the German center engaged in the fieldwork at the Ethics Committee of the University Bremen (Germany).

All children aged 2 to 9.9 years living in the German municipalities of Delmenhorst and Wilhelmshaven attending primary school or kindergarten were eligible. In total, 4,434 German children were invited to take part in the IDEFICS baseline examination (T0) between September 2007 and May 2008. All children who took part in the examination at baseline were invited to the follow-up examination (T1) between September 2009 and May 2010.

For the current analysis, children were only included if (i) they were invited at baseline, (ii) at least one contact attempt was documented, and (iii) their contact data included a telephone number at baseline. Children with unknown telephone numbers at baseline were excluded from the analysis because they could not pass the same recruitment procedure as children with known telephone numbers at baseline. According to these criteria, 3,135 children were included in the analysis, of which 1,691 took part in the baseline examination. Of these 1,691 children, 992 completed the follow-up examinations. For the regression analyses, only the 1,691 baseline children were included.

Children and parents were informed about all aspects of the study and parents gave their written informed consent prior to inclusion to the study. Immediately before each examination a study nurse informed each child orally about the modules using a simplified preformulated text. Children were informed that they do not have to participate if they don't want to and examinations were only performed if children assented and parents consented. Children and their parents could consent to single components of the study while refusing others.

Recruitment procedure

At baseline, an invitation letter was sent to the parents of all eligible children. A first reminder letter was sent 10 days after (median 12 days) and was followed by a second reminder letter another 10 days later (median 11 days), and, if families had not replied yet, telephone calls followed afterwards. For each potential participant, up to 10 telephone calls were made across different weekdays (Monday to Saturday) and daytimes to maximize contact chances. If available, short preformulated messages were left on answering machines. If a first contact was established within the 10 call attempts, but the person was not sure yet whether or not to take part, additional call attempts to re-establish contact exceeding the threshold of 10 could be made. Telephone numbers were collected from public telephone directories. If a telephone number turned out to be invalid, it was removed from the dataset. Upon enrolment, all families were asked to provide their current telephone number when signing the informed consent form.

At follow-up, an invitation letter was sent to the families of children who had participated at baseline. This invitation was followed by a first reminder letter after 10 days (median 11 days). If recipients did not respond, telephone calls were made based on the same procedure as described above. Again, telephone numbers were removed from the dataset if they turned out to be invalid. With enrolling

families, telephone numbers in the dataset were checked and updated when necessary. The recruitment was controlled and documented by the use of MODYS, a software developed by the Leibniz Institute for Prevention Research and Epidemiology - BIPS as a documentation system for the recruitment of study participants in population-based studies (Reineke et al., 2006). For each participant, MODYS recorded all recruitment steps and contact attempts with time stamps. For non-participants, reasons for non-participation were recorded with a dropout code according to pre-defined categories. For this study, dropout codes were categorized into the three groups as not-eligible, (e.g., chronic disease, language problems, moved out of the study region), refusal (e.g., no time, not convinced of study, privacy concerns) and no-contact (i.e., contact to target person was never established).

Outcomes

The recruitment effort was quantified based on the number of contact attempts (mail and telephone; successful and unsuccessful) until the recruitment was completed for a potential participant; that is, either the examination had taken place or a dropout code had been assigned. For baseline and follow-up separately, families were considered as completing the recruitment early, if their number of contact attempts was below the median across all families (early recruitment), and as completing late otherwise (late recruitment). The outcome cohort attrition was defined with respect to participation at follow-up (0: participation vs. 1: dropout).

Sociodemographic and health variables

The social position of families was assessed by the monthly household net income in Euro (low: 0-2,249; medium: 2,250-3,499; high: more than 3,500; missing) and the highest educational level of mother or father according to the International Standard Classification of Education (ISCED) (UNESCO, 2012) (low: ISCED levels 0-2; medium: ISCED levels 3-4; high: ISCED levels 5-6). The household composition was described by the number of adults (age 18 or older) in the household (one; two; three; four or more) and the existence of non-adult siblings besides the participating child (yes versus no). The telephone numbers collected at baseline were categorized with regard to type and quantity (one mobile; one landline; two landlines; landline and mobile). Residential mobility was assessed from changes in the postal zip-code from baseline to follow-up (not moved; moved within study region; moved outside study region). The migration background was defined by the place of birth of the parents (no migrant: both parents born in Germany; partly migrant: one parent not born in Germany; full migrant: otherwise). The age of the mother at baseline was measured in years. Due to collinearity and a higher percentage of missing values, the father's age was not considered in the analysis.

Children's age was measured in years at baseline (2 to 9.9 years). The weight status was assessed by means of body mass index (BMI, kg/m^2). For parents, the weight status was categorized as "no parent overweight", "at least one parent overweight", and "missing" based on the cut point BMI ≥25. Children's weight status (underweight, normal weight, overweight) was categorized according to Cole and Lobstein (2012). Because of a relatively small number of obese children the categories overweight and obese were combined.

Statistical Analysis

The association between recruitment effort and sociodemographic and health variables was assessed by estimating odds ratios (ORs) and 95% confidence intervals (CIs) with multivariable multilevel logistic regression models (Hox et al., 2010). The relationship between cohort attrition and recruitment effort was assessed by fitting multivariable multilevel logistic regression models, which were adjusted for income level, educational level, number of adults in household, number of siblings, migration background, mother's age (years), weight status parents, weight status of the child, and sex of the child. In all multilevel models information about kindergarten or school was used as second level variable to account for children clustering within kindergarten or school. A possible non-independence of siblings in the data was checked via sensitivity analysis of models with and without siblings. Results did not differ with and without siblings.

Bivariate associations between recruitment effort and reasons for dropout were assessed with Fisher's exact test. To quantify deviations between observed and expected frequencies we calculated standardised residuals (Agresti, 2007). To assess differences between follow-up dropouts grouped by baseline/follow-up recruitment effort a multinomial logistic regression model with group membership as outcome and standard errors clustered by kindergarten or school was calculated (Rogers, 1993).

Results

For the 3,135 children included in this study, a total of 14,900 contact attempts were recorded for the baseline examination, and 9,990 contact attempts for the follow-up. The cumulative distribution of the number of contact attempts for all potential participants, as well as for actual participants together with medians is given in Figure 1 separately for baseline and follow-up. At baseline, less than three contact attempts were classified as the early group, whereas, at follow-up, less than four were required.

The maximum number of contact attempts for participants was 24 at baseline and 46 at follow-up. The maximum number of contact attempts for non-participants was 44 at baseline and 39 at follow-up. At baseline, 98.5% of the 1,691 participants were recruited with 12 contact attempts or less, while at follow-up only 85.1% of the 992 participants were recruited within 12 contact attempts.

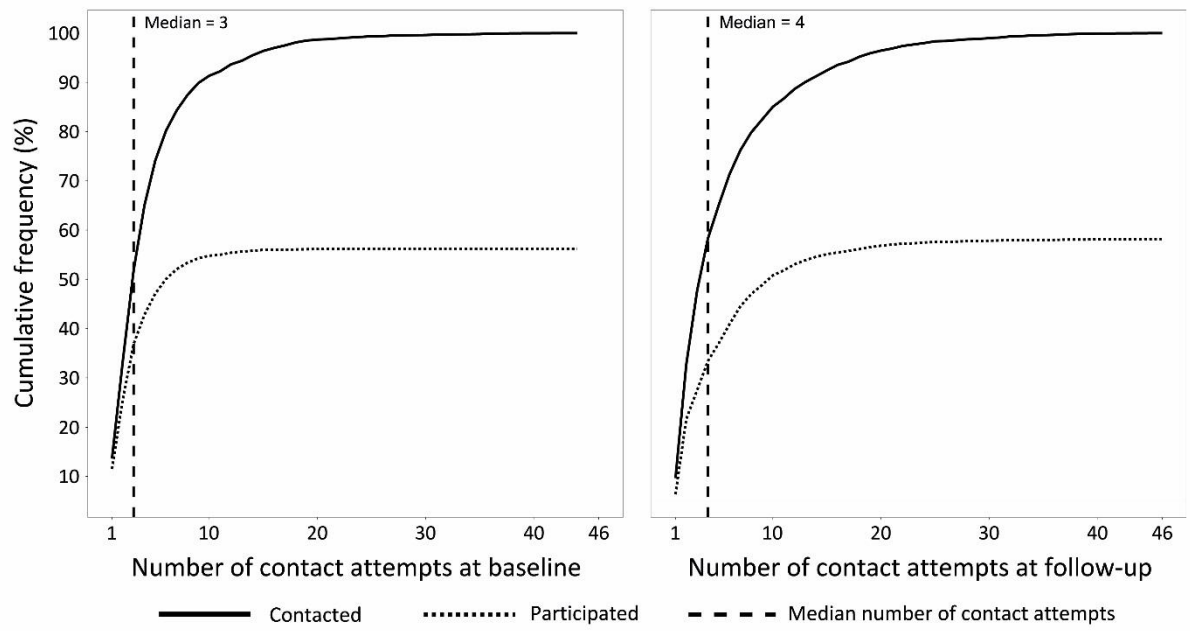

Figure 3.1 Cumulative percentage of contacted (curve), participated (dashed curve) and the median number (vertical line) of contact attempts of the German IDEFICS baseline examination (left panel) and follow-up (right panel)

Recruitment effort

A multilevel logistic regression model with recruitment effort at baseline as the dependent variable (Table 3.1) showed that age of the mother was inversely associated with the chance of being classified as late baseline recruitment (OR 0.96, 95% CI 0.94, 0.98). Age of the child was positively associated with the chance of being classified as late baseline recruitment (OR 1.14, 95% CI 1.05, 1.23).

Table 3.1 Odds ratios (95% CIs) for being categorized as completing the recruitment LATE at the baseline examination and the follow-up

	Recruitment effort at baseline					Recruitment effort at follow-up				
	Early baseline		Late baseline		Odds ratio	Early follow-up		Late follow-up		Odds ratio
	n	%	n	%	(95% confidence interval)	n	%	n	%	(95% confidence interval)
Recruitment effort at baseline										
2 contact attempts at early baseline						413	52.9	368	47.1	1.00 (Reference)
≥3 contact attempts at late baseline						394	43.3	516	56.7	1.28 (1.04, 1.57)
Sex child										
Male	403	47.4	447	52.6	1.00 (Reference)	412	48.5	438	51.5	1.00 (Reference)
Female	378	45.0	463	55.0	1.12 (0.91, 1.38)	395	47.0	446	53.0	1.03 (0.84, 1.26)
Age child at baseline (years)[a]					1.14 (1.05, 1.23)					0.98 (0.92, 1.05)
Weight status child at baseline										
Under weight	88	50.9	85	49.1	0.85 (0.60, 1.20)	84	48.5	89	51.5	0.98 (0.70, 1.37)
Normal weight	580	46.3	673	53.7	1.00 (Reference)	605	48.3	648	51.7	
Overweight	113	42.6	152	57.4	1.02 (0.76, 1.37)	118	44.5	147	55.5	0.99 (0.74, 1.32)
Mother's age at baseline (years)[a]					0.96 (0.94, 0.98)					0.98 (0.96, 1.00)
Weight status parents at baseline										
No parent overweight	172	49.6	175	50.4	1.00 (Reference)	174	50.1	173	49.9	1.00 (Reference)
At least one parent overweight	527	46.9	596	53.1	1.05 (0.81, 1.37)	544	48.4	579	51.6	0.95 (0.74, 1.23)
Missing	82	37.1	139	62.9	1.31 (0.88, 1.96)	89	40.3	132	59.7	1.24 (0.84, 1.82)
Migrationbackground at baseline										
No	533	48.8	560	51.2	1.00 (Reference)	598	54.7	495	45.3	1.00 (Reference)
Partly	96	52.2	88	47.8	0.75 (0.53, 1.07)	75	40.8	109	59.2	1.50 (1.08, 2.09)
Full	152	37.5	253	62.5	1.26 (0.94, 1.68)	133	32.8	272	67.2	1.74 (1.33, 2.30)
Missing	0	0	9	100		1	11.1	8	88.9	5.77 (0.68, 48.69)
Income level at baseline										
Low income	335	40.5	493	59.5	1.00 (Reference)	329	39.7	499	60.3	1.00 (Reference)
Medium income	283	52.4	257	47.6	0.79 (0.61, 1.03)	314	58.1	226	41.9	0.60 (0.46, 0.77)
High income	94	54.0	80	46.0	0.88 (0.58, 1.35)	108	62.1	66	37.9	0.54 (0.36, 0.82)
Missing	69	46.3	80	53.7	0.77 (0.52, 1.15)	56	37.6	93	62.4	1.14 (0.77, 1.67)
Educational level at baseline										
Low education	136	36.3	239	63.7	1.00 (Reference)	132	35.2	243	64.8	
Medium education	468	48.2	503	51.8	0.80 (0.60 1.07)	492	50.7	479	49.3	0.79 (0.60, 1.03)
High education	170	53.3	149	46.7	0.78 (0.53, 1.14)	172	53.9	147	46.1	0.95 (0.65, 1.37)
Missing	7	26.9	19	73.1	1.52 (0.56, 4.10)	11	42.3	15	57.7	0.56 (0.24, 1.30)
Number of adults in household at baseline										
One	86	38.6	137	61.4	1.00 (Reference)	102	45.7	121	54.3	1.00 (Reference)
Two	631	48.6	668	51.4	0.79 (0.55, 1.13)	644	49.6	655	50.4	1.20 (0.85, 1.68)
Three	29	33.7	57	66.3	1.68 (0.93, 3.02)	26	30.2	60	69.8	2.42 (1.36, 4.30)
Four or more	11	39.3	17	60.7	1.05 (0.42, 2.60)	7	25.0	21	75.0	2.73 (1.05, 7.09)
Missing	24	43.6	31	56.4	0.78 (0.20, 2.98)	28	50.9	27	49.1	1.11 (0.32, 3.92)
Siblings aged <18 years at baseline										
Yes	197	46.7	225	53.3	0.98 (0.76, 1.26)	214	50.7	208	49.3	1.13 (0.88, 1.43)
No	565	46.1	660	53.9	1.00 (Reference)	570	46.5	655	53.5	1.00 (Reference)
Missing	19	43.2	25	56.8	0.94 (0.22, 4.08)	23	52.3	21	47.7	0.71 (0.18, 2.83)
All	781	46.2	910	53.8		807	47.7	884	52.3	

[a]Odds ratios correspond to an 1-year increase in age

A multilevel logistic regression model with the dependent variable recruitment effort at follow-up (Table 3.1) revealed that having been in the group late baseline recruitment was associated with a higher chance of being also in the late follow-up recruitment group. Medium or high income (OR was associated with a lower chance of being classified as late follow-up recruitment.

Compared to families with one adult, households with three or four and more adults had a higher chance to belong to the late follow-up recruitment group. Furthermore, if children were classified as partly migrants or full migrants they had a higher chance to be in the group late follow-up recruitment. Similar to baseline, an increasing mother's age was associated with a lower chance of being in the late follow-up recruitment group.

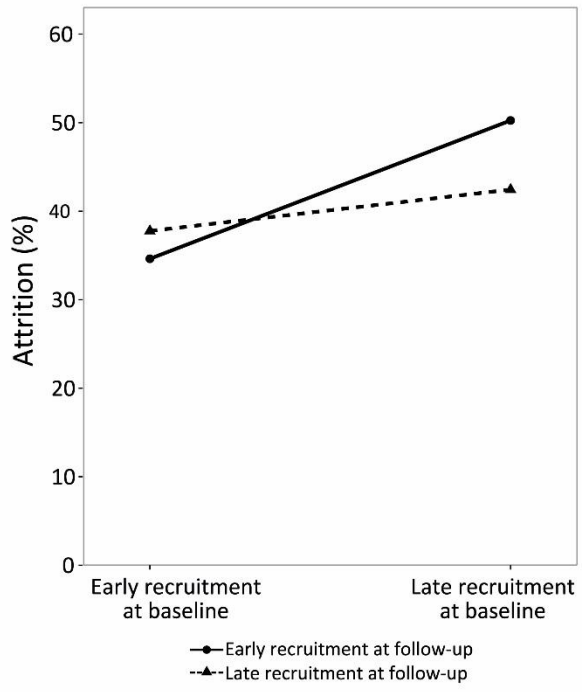

Figure 3.2 Attrition at follow-up as a function of recruitment effort at baseline and follow-up

Cohort attrition

To investigate whether or not the extend of the recruitment influences cohort attrition, the recruitment efforts at baseline and follow-up were included as independent variables in a multivariable multilevel logistic regression model with cohort attrition at follow-up as outcome. A descriptive analysis depicted in Figure 3.2 suggested an interaction effect of the recruitment effort, as being in the group late baseline recruitment indeed seemed to result in a higher attrition than being in the group early baseline recruitment, but this difference was larger in the early follow-up recruitment group as compared to the late follow-up recruitment group.

Table 3.2 Odds ratios (with 95% CIs) for attrition at the follow-up as predicted by a model with interaction

	Attrition at follow-up				Odds ratio (95% CI)
	No		Yes		
	n	%	N	%	
Grouped baseline/follow-up recruitment effort					
Early baseline recruitment & early follow-up recruitment	270	65.4	143	34.6	1.00 (Reference)
Early baseline recruitment & late follow-up recruitment	229	62.2	139	37.8	1.13 (0.81, 1.58)
Late baseline recruitment & early follow-up recruitment	196	49.8	198	50.2	1.65 (1.19, 2.28)
Late baseline recruitment & late follow-up recruitment	297	57.6	219	42.4	1.11 (0.81, 1.52)
Telephone phone numbers at baseline					
One landline number	755	59.7	510	40.3	1.00 (Reference)
Two landline numbers	30	66.7	15	33.3	1.01 (0.49, 2.09)
One mobile number	65	42.8	87	57.2	1.51 (1.01, 2.27)
Landline and mobile	142	62.0	87	38.0	0.76 (0.54, 1.06)
Sex child (baseline)					
Male	509	59.9	341	40.1	1.00 (Reference)
Female	483	57.4	358	42.6	1.14 (0.91, 1.42)
Age child at baseline (years)[a]					1.22 (1.11, 1.33)
Weight status child (baseline)					
Under weight	112	64.7	61	35.3	0.74 (0.51, 1.08)
Normal weight	746	59.5	507	40.5	1.00 (Reference)
Overweight	134	50.6	131	49.4	1.29 (0.95, 1.75)
Mother's age at baseline (years) [a]					0.96 (0.94 ,0.98)
Weight status parents (baseline)					
No parent overweight	216	62.2	131	37.8	1.00 (Reference)
At least one parent overweight	672	59.8	451	40.2	1.01 (0.76 ,1.34)
Missing	104	47.1	117	52.9	1.30 (0.86 ,1.96)
Migration background (baseline)					
No	672	61.5	421	38.5	1.00 (Reference)
Partly	101	54.9	83	45.1	1.04 (0.72, 1.50)
Full	217	53.6	188	46.4	1.02 (0.75, 1.38)
Missing	2	22.2	7	77.8	1.84 (0.33, 10.36)
Income level (baseline)					
Low income	448	54.1	380	45.9	1.00 (Reference)
Medium income	353	65.4	187	34.6	0.83 (0.62, 1.10)
High income	120	69.0	54	31.0	0.95 (0.60, 1.52)
Missing	71	47.6	78	52.4	1.53 (1.02, 2.29)

	Attrition at follow-up				Odds ratio (95% CI)
	No		Yes		
	n	%	N	%	
Educational level (baseline)					
Low education	184	49.1	191	50.9	1.00 (Reference)
Medium education	579	59.6	392	40.4	0.85 (0.63, 1.15)
High education	219	68.7	100	31.4	0.60 (0.39, 0.90)
Missing	10	38.5	16	61.5	1.26 (0.50, 3.15)
Number of adults in household (baseline)					
One	116	52.0	107	48.0	1.00 (Reference)
Two	788	60.7	511	39.3	0.95 (0.66, 1.38)
Three	52	60.5	34	39.5	1.05 (0.58, 1.90)
Four or more	13	46.4	15	53.6	1.69 (0.68, 4.18)
Missing	23	41.8	32	58.2	1.18 (0.31, 4.45)
Siblings aged <18 years (baseline)					
Yes	260	61.6	162	38.4	1.28 (0.98, 1.67)
No	715	58.4	510	41.6	1.00 (Reference)
Missing	17	38.6	27	61.4	1.62 (0.38, 6.96)
Moved					
Stayer	936	60.0	624	40.0	1.00 (Reference)
Mover within study region	49	50.5	48	49.5	1.36 (0.85, 2.18)
Mover outside study region	6	20.7	23	79.3	2.98 (0.95, 9.38)
Missing	1	20.0	4	80.0	4.50 (0.41, 49.26)
All	992	58.7	699	41.3	

Odds ratios are adjusted for sex child (baseline), age child at baseline (years), weight status child (baseline), mother's age at baseline (years), weight status parents (baseline), migration background (baseline), income level (baseline), educational level (baseline), number of adults in household (baseline), siblings aged <18 years (baseline)
[a] Odds ratios correspond to a 1-year increase in age

To test this interaction, a multilevel logistic regression model with three dummy variables early baseline recruitment & late follow-up recruitment, late baseline recruitment & early follow-up recruitment, and late baseline recruitment & late follow-up recruitment was calculated (Table 3.2) to disentangle and compare the effects of early and late recruitment at both surveys with the reference category early baseline recruitment & early follow-up recruitment. The model revealed a higher chance of cohort attrition for the group late baseline recruitment & early follow-up recruitment as compared to the early baseline recruitment & early follow-up recruitment. The model also showed a higher chance of cohort attrition for baseline participants with only one known mobile number compared to baseline participants with one landline number available. Families with high education had a lower chance of cohort attrition. Age of the child was positively associated while mother's age was inversely associated with cohort attrition.

Table 3.3 Observed and expected frequencies of dropout codes at follow-up and standardised residuals as a function of recruitment effort

	Neutral dropout[a]	Refusal[b]	No-contact[c]	All
Early baseline recruitment & early follow-up recruitment				
Observed	5	109	29	143
Expected	*5.1*	*103.9*	*34.0*	
Standardised residuals	-0.06	1.07	-1.09	
Early baseline recruitment & late follow-up recruitment				
Observed	9	106	24	139
Expected	*5.0*	*101.0*	*33.0*	
Standardised residuals	2.06	1.06	-2.01	
Late baseline recruitment & early follow-up recruitment				
Observed	5	131	62	198
Expected	*7.1*	*143.9*	*47.0*	
Standardised residuals	-0.94	-2.43	2.95	
Late baseline recruitment & late follow-up recruitment				
Observed	6	162	51	219
Expected	*7.8*	*159.2*	*52.0*	
Standardised residuals	-0.8	0.52	-0.19	
All	25	508	166	699

P=0.04

[a] Address is unknown; not sufficient language knowledge; contact person in nursing care, or handicapped/invalidity; moved out of survey region

[b] Lack of time; following medical expert suggestion/health reasons; thinks it's senseless; concerns about data protection; other reasons; without giving reasons

[c] At no time personal contact or telephone contact with target person; no information about reasons

To further explore the differences in cohort attrition between the aforementioned groups, we investigated the reasons for dropouts at follow-up. Indeed, the distribution of dropout codes differed between groups, (P = 0.04; Table 3.3). The comparison of standardised residuals (Agresti, 2007) revealed that, among the late baseline recruitment & early follow-up recruitment group, the dropout code “no-contact” was assigned much more frequently than expected (2.95 standard deviations above expectation; Table 3.3). Entering all baseline participants that did not participate at follow-up into a multinomial logistic regression model with the grouped baseline/follow-up recruitment effort as outcome (early baseline recruitment & early follow-up recruitment, early baseline recruitment & late follow-up recruitment, late baseline recruitment & early follow-up recruitment, late baseline recruitment & late follow-up recruitment) did not reveal socioeconomic differences between groups (Table 3.4). An inspection of the quantity and type of the telephone numbers available at baseline revealed that members of the late baseline recruitment group were more likely to have provided only one mobile number as compared to the early baseline recruitment group (11.54% versus 6.02 %; Table 3.5). Mobile numbers, however, had a higher chance of turning out to be invalid at follow-up as compared to landline numbers (13.16 % versus 9.17 %, Table 3.6) when comparing respondents with only one telephone number at baseline.

Table 3.4 Relative risk ratios (with 95% CIs) of follow-up dropouts grouped by baseline/follow-up recruitment effort

	Grouped by baseline/follow-up recruitment effort											
	early baseline recruitment & early follow-up recruitment			early baseline recruitment & late follow-up recruitment			late baseline recruitment & early follow-up recruitment (Reference)			late baseline recruitment & late follow-up recruitment		
	n	%	RRR[a](95%-CI)	n	%	RRR[a](95%-CI)	n	%		n	%	RRR[a](95%-CI)
Sex child (baseline)												
Male	73	21.41	Reference	67	19.65	Reference	96	28.15		105	30.79	Reference
Female	70	19.55	0.90 (0.59, 1.38)	72	20.11	1.01 (0.64, 1.61)	102	28.49		114	31.84	1.03 (0.68, 1.55)
Age child at baseline (years)	Mean = 6.27		0.92 (0.79, 1.07)	Mean = 6.08		0.88 (0.77, 1.01)		Mean = 6.32		Mean = 6.58		1.09 (0.97, 1.23)
Weight status child (baseline)												
Under weight	108	21.30	0.68 (0.26, 1.79)	98	19.33	1.61 (0.73, 3.55)	143	28.21		158	31.16	1.09 (0.50, 2.38)
Normal weight	9	14.75	Reference	18	29.51	Reference	16	26.23		18	29.51	Reference
Overweight	26	19.85	0.97 (0.51, 1.86)	23	17.56	0.95 (0.58, 1.57)	39	29.77		43	32.82	0.83 (0.48, 1.43)
Mother's age at baseline (years)		Mean = 35.71	1.07 (1.02, 1.12)	Mean = 34.38		1.04 (0.99, 1.09)		Mean =33.73		Mean = 33.51		1.00 (0.96, 1.04)
Weight status parents (baseline)												
No parent overweight	30	22.90	Reference	32	24.43	Reference	38	29.01		31	23.66	Reference
At least one parent overweight	96	21.29	0.94 (0.53, 1.68)	82	18.18	0.75 (0.39, 1.42)	130	28.82		143	31.71	1.24 (0.63, 2.45)
Missing	17	14.53	0.85 (0.40, 1.83)	25	21.37	1.32 (0.61, 2.85)	30	25.64		45	38.46	1.72 (0.75, 3.92)
Migration background (baseline)												
No	100	23.75	Reference	79	18.76	Reference	126	29.93		116	27.55	Reference
Partly	20	24.10	1.72 (0.86, 3.45)	20	24.10	1.89 (0.89, 4.00)	19	22.89		24	28.92	1.31 (0.66, 2.62)
Full	23	12.23	0.77 (0.34, 1.74)	40	21.28	1.44 (0.79, 2.60)	52	27.66		73	38.83	1.32 (0.74, 2.37)
Missing	0	0.00	0.00 (0.00, 0.00)	0	0.00	0.00 (0.00, 0.00)	1	14.29		6	85.71	5.00 (0.42, 60.08)

	Grouped by baseline/follow-up recruitment effort										
	early baseline recruitment & early follow-up recruitment			early baseline recruitment & late follow-up recruitment			late baseline recruitment & early follow-up recruitment (Reference)		late baseline recruitment & late follow-up recruitment		
	n	%	RRR[a](95%-CI)	n	%	RRR[a](95%-CI)	n	%	n	%	RRR[a](95%-CI)
Income level (baseline)											
Low income	60	15.79	Reference	73	19.21	Reference	110	28.95	137	36.05	Reference
Medium income	59	31.55	1.64 (0.87, 3.07)	36	19.25	1.01 (0.51, 1.99)	56	29.95	36	19.25	0.63 (0.39, 1.04)
High income	11	20.37	1.22 (0.44, 3.39)	11	20.37	1.18 (0.46, 3.08)	15	27.78	17	31.48	1.35 (0.46, 3.94)
Missing	13	16.67	1.27 (0.66, 2.44)	19	24.36	1.94 (0.85 4.40)	17	21.79	29	37.18	1.31 (0.65, 2.65)
Educational level (baseline)											
Low education	25	13.09	Reference	34	17.80	Reference	51	26.70	81	42.41	Reference
Medium education	93	23.72	1.37 (0.81, 2.33)	86	21.94	1.31 (0.71, 2.42)	107	27.30	106	27.04	0.76 (0.47, 1.23)
High education	21	21.00	0.69 (0.29, 1.61)	19	19.00	0.79 (0.30, 2.04)	34	34.00	26	26.00	0.63 (0.29, 1.38)
Missing	4	25.00	1.14 (0.25, 5.25)	0	0.00	0.00 (0.00, 0.00)	6	37.50	6	37.50	0.57 (0.16, 2.03)
Number of adults in household (baseline)											
One	17	15.89	Reference	18	16.82	Reference	33	30.84	39	36.45	Reference
Two	112	21.92	1.14 (0.55, 2.40)	103	20.16	1.37 (0.55, 3.44)	148	28.96	148	28.96	1.14 (0.64, 2.03)
Three	3	8.82	0.71 (0.19, 2.65)	7	20.59	1.83 (0.47, 7.10)	6	17.65	18	52.94	2.72 (0.89, 8.32)
Four or more	2	13.33	1.61 (0.26, 10.15)	6	40.00	5.73 (0.86, 38.16)	2	13.33	5	33.33	2.10 (0.36, 12.36)
Missing	9	28.13	0.71 (0.09, 5.52)	5	15.63	0.57 (0.07, 4.79)	9	28.13	9	28.13	0.39 (0.03, 5.70)
Siblings aged < 18 years (baseline)											
Yes	35	21.60	Reference	29	17.90	Reference	47	29.01	51	31.48	Reference
No	100	19.61	0.98 (0.49, 1.96)	106	20.78	1.26 (0.68, 2.32)	144	28.24	160	31.37	0.94 (0.58, 1.51)
Missing	8	29.63	3.39 (0.35, 32.77)	4	14.81	2.29 (0.16, 31.88)	7	2.93	8	29.63	2.24 (0.07, 69.30)
	143	20.46		139	19.89		198	28.33	219	31.33	

[a] Relative risk ratio

Table 3.5 Number and column percentages of available telephone numbers at baseline of early baseline recruitment and late baseline recruitment

	early baseline recruitment		late baseline recruitment		All	
	n	%	n	%	n	%
One landline number	610	78.10	655	71.98	1265	74.81
Two landline numbers	27	3.46	18	1.98	45	2.66
One mobile number	47	6.02	105	11.54	152	8.99
Landline and mobile number	97	12.42	132	14.51	229	13.54
All	781	100.00	910	100.00	1691	100.00

Table 3.6 Number and row percentages of the available telephone numbers at baseline and follow-up

	Follow-up											
	No phone number		One landline number		Two landline numbers		One mobile number		Landline and mobile number		All	
Baseline	n	%	n	%	n	%	n	%	n	%	n	%
One landline number	116	9.17	978	77.31	23	1.82	16	1.26	132	10.43	1265	74.81
Two landline numbers	0	0	8	17.78	31	68.89	0	0	6	13.33	45	2.66
One mobile number	20	13.16	3	1.97	1	0.66	91	59.87	37	24.34	152	8.99
Landline and mobile number	16	6.99	11	4.80	0	0	24	10.48	178	77.73	229	13.54
All	152	8.99	1000	59.14	55	3.25	131	7.75	353	20.88	1691	100.00

Comment

In line with earlier research, our results indicate that late respondents are in general more likely to have a higher level of social deprivation (Cohen et al., 2000; Haring et al., 2009; Studer et al., 2013). Furthermore, the observed association between being a late respondent at baseline and at follow-up suggests that a core group of late respondents might exist that is hard to reach with conventional contact modes like land mail and phone. Replicating results from other epidemiological cohort studies we found evidence for a systematic dropout of socioeconomically disadvantaged participants at

follow-up that might bias associations of social position and health outcomes if not properly addressed (Howe et al., 2013).

This study expands the existing literature by not only separately investigating recruitment efforts at baseline and follow-up, but also exploring the interaction between recruitment efforts at baseline and follow-up, revealing a complex relationship between recruitment effort and attrition. According to our results, the observed interaction was mainly driven by individuals who were late respondents at baseline and early respondents at follow-up (group late baseline recruitment & early follow-up recruitment), because their risk to dropout at follow-up was more than 1.5 times higher as compared to the other groups. Hence, being a late respondent at baseline was only associated with a higher chance of attrition, if the follow-up recruitment was completed early.

To understand what might differentiate the early & late groups, we investigated the dropout codes of all persons who did not participate at follow-up and found that their distribution was dependent on the recruitment effort. In particular, among the group late baseline recruitment & early follow-up recruitment, the category no-contact was observed much more frequently than expected by chance. This observation was rather surprising given that, at follow-up, only individuals with up to three contact attempts were categorized as early respondents, whereas the recruitment protocol at follow-up comprised two invitation letters followed by up to ten contact attempts by phone. A plausible explanation is that these individuals did not respond to either of the two letters and their telephone number turned out to be invalid at the first telephone call, rendering additional call attempts futile. The results of the multilevel logistic regression indicated that providing only one mobile number was positively associated with attrition. As indicated by the additional descriptive findings, the higher percentage of mobile numbers among late baseline recruitment and the higher percentage of invalid mobile numbers at follow-up might have caused the problem of no-contact. Interestingly, a higher number of early no-contact dropouts at follow-up was only observed for individuals who were late respondents at baseline (late baseline recruitment & early follow-up recruitment group), but not for early respondents at baseline (early baseline recruitment & early follow-up recruitment group). Unfortunately, the available paradata do not allow to compare the quality of telephone numbers between different groups, or to determine whether a particular phone number was supplied by the participant or collected from public telephone directories. Nevertheless, our analyses did not indicate socioeconomic differences between these groups and we do not see any other obvious explanations why the quality of telephone numbers in the group late baseline recruitment & early follow-up recruitment should be worse than that of other subgroups. We therefore hypothesize that the high number of no-contact dropouts in this group might, at least partially, result from a stronger reluctance to respond to invitation letters, rather than just a higher unwillingness to participate.

These results provide empirical evidence for the everyday wisdom that asking participants to provide more than one telephone number (e.g., landline, mobile, number of the partner) upon enrolment might actually help to attenuate attrition at follow-up. Furthermore, the results of this analysis led to changes in the documentation system MODYS which now records whether a telephone number was provided by participants or taken from public telephone registers, and all changes made to the telephone numbers in a dataset.

Although the standard recruitment procedure for this study consisted of 13 contact attempts at baseline and 12 at follow-up, this upper limit was often exceeded during baseline and follow-up recruitment. At baseline nearly all participants were recruited with 12 contact attempts or less and it might be argued whether or not any further contact attempts exceeding 12 should have been omitted for the sake of reduced recruitment costs. Such a conclusion, however, would be biased by hindsight because during the ongoing recruitment it could not be foreseen in how many additional participants the continued recruitment would result. At the follow-up, for instance, only 85% of all follow-up participants were recruited within 12 contact attempts, that is, the additional contact attempts turned out to be much more effective as compared to baseline, but again, only in hindsight. These results clearly demonstrate that more research on paradata is necessary to develop recruitment schemes that satisfy both, scientific rigor and cost effectiveness.

Conclusion

Extended recruitment efforts at baseline in the German IDEFICS child cohort were not penalized per se at follow-up, as being a late respondent at baseline was only partly associated with a higher attrition at follow-up. Among the dropouts at follow-up, late respondents at baseline were more likely to belong to the “no-contact” dropout category. As the ability to establish contact at follow-up is based on valid telephone numbers, special care has to be taken at baseline while collecting telephone numbers. The collection of paradata during the recruitment is crucial for the development and evaluation of new recruitment strategies.

4 An illustration of the bias due to model misspecification under incomplete data: A simulation based on the IDEFICS study

Malte Langeheine, Kate Tilling, Ronja Foraita, Wolfgang Ahrens, Vanessa Didelez, and Stefan Rach, on behalf of the IDEFICS Consortium[3]

Abstract

Background

Missing data are a major concern in epidemiology. Analyzing only the individuals with complete data on the outcome, the exposure, and all explanatory variables, is a very simple and therefore popular strategy of handling missing data known as complete-case analysis. In a regression analysis, CCA is consistent in general when missingness is unrelated to the outcome given the explanatory variables included in the regression model. However, it appears not widely appreciated that this validity of complete-case analysis critically hinges on correct specifications of the analysis model and a misspecified model might re-introduce bias.

Methods

We illustrate with a simulation how different modeling choices can affect our conclusions even when a complete-case analysis is in principle valid. We base our study on empirical data from the IDEFICS study and simulate only the missingness mechanism assuming different association strengths and different frequencies of missingness. In each scenario, we investigated the performance of three different analysis models using complete-case analysis as well as multiple imputation and inverse probability weighting as methods to correct for missing data.

Results

Model misspecification can lead to considerable bias when data contain missing values, even in a scenario where an ideal complete-case analysis is known to be consistent. This bias equally affects multiple imputation, and to a lesser extent, at the cost of precision, inverse probability weighting which requires a correct weighting model. In our example, basic model diagnostics were sufficient to alert us to the misspecification of the simple analysis model with regard to the functional form of the exposure; this was detectable even for the most extreme missingness mechanisms.

[3] Currently under review

Conclusions

We suggest that researchers carefully consider their choice of analysis model. Misspecification will obviously always pose a risk of bias, but while this may be small with fully observed data and a moderately misspecified model, missing data may seriously amplify the bias.

Background

Missing data are ubiquitous in epidemiological cohort studies and a major concern as inadequate handling of missing data may invalidate conclusions (Hernán et al., 2004). Complete-case analysis (CCA), that is, analyzing only the individuals with complete data on the outcome, the exposure, and all explanatory variables, is a very simple and therefore popular strategy of handling missing data. While CCA is often biased, it is well-known that in a regression analysis CCA is consistent in general when missingness is unrelated to the outcome given the explanatory variables included in the regression model (Giorgi et al., 2008) (this scenario is an instance of Missing At Random, MAR (Little and Rubin, 2002)). However, it appears not widely appreciated that this validity of CCA critically hinges on correct specifications of the analysis model; a misspecified model might re-introduce bias, for example, due to an omitted non-linear effect of an explanatory variable (Fox, 2016). In the present paper we discuss and illustrate with a simulation how different modeling choices can affect our conclusions even when a CCA is in principle valid. Our results highlight the need to pay more careful attention to the choice of the analysis model than is usually the case.

For our simulation study we chose a non-standard approach. Typically, simulations utilize artificial data so that the true data generating model is fully known and controlled. However, we decided to base our study on empirical data and to simulate only the missingness mechanism. With this approach we intend to trade off the advantage of a well-defined and controlled test scenario against an increased degree of realism; in particular we wanted to avoid unrealistically simple models often used in simulations. In brief, our rationale was as follows: Based on data from the first two waves of the IDEFICS study (Ahrens et al., 2017) (Identification and prevention of dietary- and lifestyle-induced health effects in children and infants) we extracted a data set without any missing values (referred to as full data set henceforth) and then deleted observations of the outcome variable based on specified missingness mechanisms determining which values would be observed and which would be missing. The mechanisms were chosen such that missingness of the outcome only depended on the exposure of interest but not on the outcome itself. Moreover, the mechanisms covered a range of different strengths of association between missingness and exposure, as well as different overall frequencies of missingness in the data. As the true data generating model for the empirical data is unknown, we used three linear regression models of increasing complexity as analysis models. These were then combined with three different ways of handling the missing data. With the 'truth' being unknown, the results of

these analyses were compared with the corresponding analyses on the full data set. The three missing data methods were (1) CCA, which, for our choice of missingness mechanisms, should be consistent if the analysis model was correct; (2) imputing missing values (multiple imputation (Schafer, 1999; Sterne et al., 2009)), which has the advantage that information in addition to the variables in the analysis model can be exploited; and (3) weighting of the observed data according to their sampling probability (inverse probability weighting (Li et al., 2013)), which relies on a correctly specified model for the missingness. Finally, we investigated whether popular diagnostics can detect misspecification of the analysis model using the complete data only.

Methodological excursus

It is not always intuitive why missingness can result in biased regression coefficients in some scenarios, while a bias can be theoretically ruled out in others. Missing data is classified as (1) Missing Completely At Random (MCAR) (Little and Rubin, 2002), if missingness does not depend on any observed or unobserved variables, (2) MAR, if missingness depends only on observed variables or (3) Missing Not At Random (MNAR), if missingness is still associated with the missing value given the observed data. In the particular case of a regression analysis, where the outcome variable is sometimes missing but explanatory variables are fully observed, it is obvious that, for complete case analyses, a bias due to missingness can be ruled out under MCAR. It might be less clear why a bias can be avoided under MAR, but will typically occur under MNAR. This difference, however, can be explained by an example (adapted from (Daniel et al., 2011)): scatterplots in Figure 4.1 depict artificial data of an explanatory variable *A* and an outcome variable *Y* with the two panels in the upper row depicting a linear relationship between *A* and *Y*, and a quadratic relationship in the bottom panels. In each of the four panels a scenario with full data (all circles) and one with incomplete data (only light grey circles) are depicted. Regression lines are dashed black for the full data set and solid orange for the incomplete data. In a scenario in which missing values occur above a certain cutoff of the explanatory variable *A* (*Y* missing at random given *A*), the CCA of *A* and *Y* remains unaffected (Figure 4.1, top left panel) as can be seen from the aligned regression lines. In contrast, if missing values in *Y* occur above a certain cutoff of the dependent variable *Y* (*Y* missing dependent on *Y;* cf. *Y* is Missing Not At Random (Little and Rubin, 2002)) the regression lines do not align indicating the bias of the CCA (Figure 4.1, topright panel) (Daniel et al., 2011). This also holds if the association between *A* and *Y* is non-linear, provided that the analysis model reflects the non-linear relationship (Figure 4.1, bottom panels).

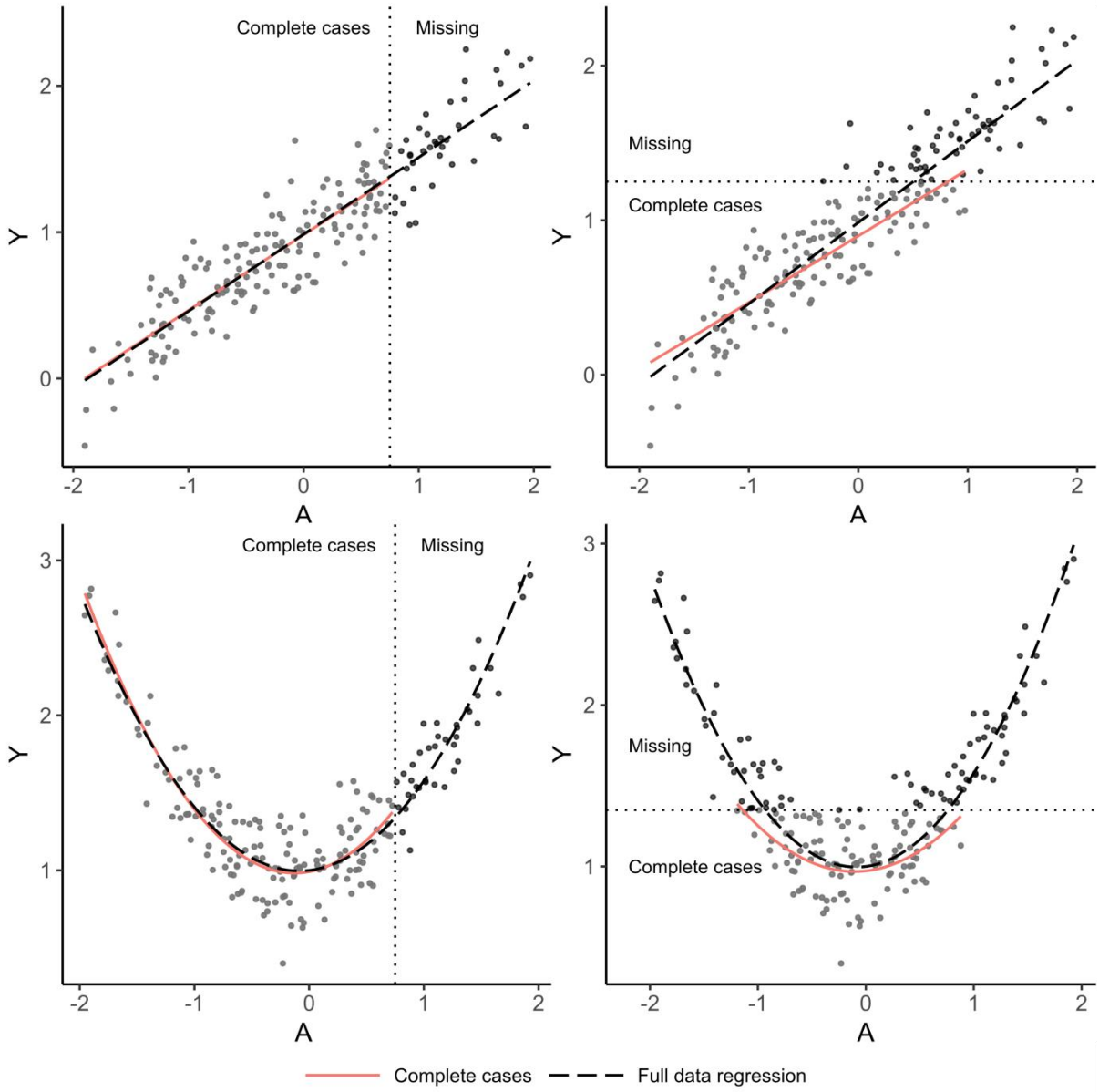

Figure 4.1 Illustration of bias in bivariate regression using complete-case analysis when the missing mechanism is Missing At Random (Left Column) or Missing Not At Random (Right Column) for a scenario in which the bivariate association is linear (Top row) or non-linear (Bottom row).

Methods

The analytic approach of the present study consisted of three steps: Based on the data of (Peplies et al., 2016), we first constructed realistic data containing no missing values at all which we call the *full data set*. With the full data set we simulated missing values in our outcome variable dependent on an explanatory variable (i.e., a MAR scenario) assuming different association strengths and different frequencies of missingness. In each scenario, we then investigated the performance of three different

analysis models using CCA as well as multiple imputation and inverse probability weighting as methods to correct for missing data.

Data

The IDEFICS study is a large European cohort on the etiology of overweight, obesity and related disorders during child- and adulthood (for a cohort profile, see Ahrens et al., 2011; Ahrens et al., 2017). Data used in this study were previously analyzed in Peplies et al. (2016) including a detailed description of data and methods regarding the association of HOMA-IR (homeostasis model assessment to quantify insulin resistance; we will use the abbreviation HOMA throughout the manuscript) with lifestyle-related factors and weight status. Note, however, that the scope of the current study is not the investigation of HOMA, but the methodological issues outlined in the introduction.

For this study, we selected 5023 children with complete data on all exposure variables at the baseline examination, including 2081 children with missing values on HOMA measured at the follow-up examination because they declined to participate. Using a matching algorithm (see Additional file 1 below for details), children with missing HOMA data were replaced by matches with complete data, resulting in a *full data* set of 4138 children, which was used for the simulations.

Outcomes and exposures

For all models the outcome of interest was z-score of HOMA (zHOMA). As predictors of zHOMA, we used the following variables: z-scores of body mass index (zBMI), physical activity (PA), educational level, audio-visual media time (AVM), children's age, and sex of child. zBMI was calculated according to (Cole and Lobstein, 2012). Uniaxial accelerometers (ActiGraph® GT1M or Actitrainer, LLC, Pensacola, FL, USA) were used to measure physical activity (h/day), that is, time spent performing either moderate or vigorous intensity physical activity according to (Evenson et al., 2008). A detailed description of accelerometry in IDEFICS can be found elsewhere (Konstabel et al., 2014). The educational level was classified according to the International Standard Classification of Education (ISCED) (UNESCO, 2012) using the highest educational attainment of mother or father (low: ISCED levels 0-2; medium: ISCED levels 3-4; high: ISCED levels 5 and higher). Furthermore, the time spent with audio-visual media was measured in h/day and children's age at the day of the examination in years. All continuous variables were mean centered.

Simulation of missingness

In our simulation we stipulated that the probability of not participating in the follow-up examination, resulting in missing zHOMA, depended on zBMI. The following sigmoid function was used to generate a subject-specific probability δ_i of missing zHOMA values (Wolke et al., 2009):

$$\delta_i = \frac{1}{1 + exp\big(\tau * (-zBMI_i + a)\big)}$$

where $\tau \in \{0.2, 1, 2, 10\}$ is a scaling parameter quantifying the strength of dependence of missingness on zBMI and a constant a that was determined by iterative search such that the frequency of missingness was kept fixed at 30%, 50% or 70%. Figure 4.2 shows how the missing mechanism translated into S-shaped probabilities of missingness for each τ by frequency of missingness. zHOMA was set to missing based on an independent draw for each subject from a binomial distribution with probability δ_i. For each combination of frequency of missingness and τ, we generated 300 data sets with missing data.

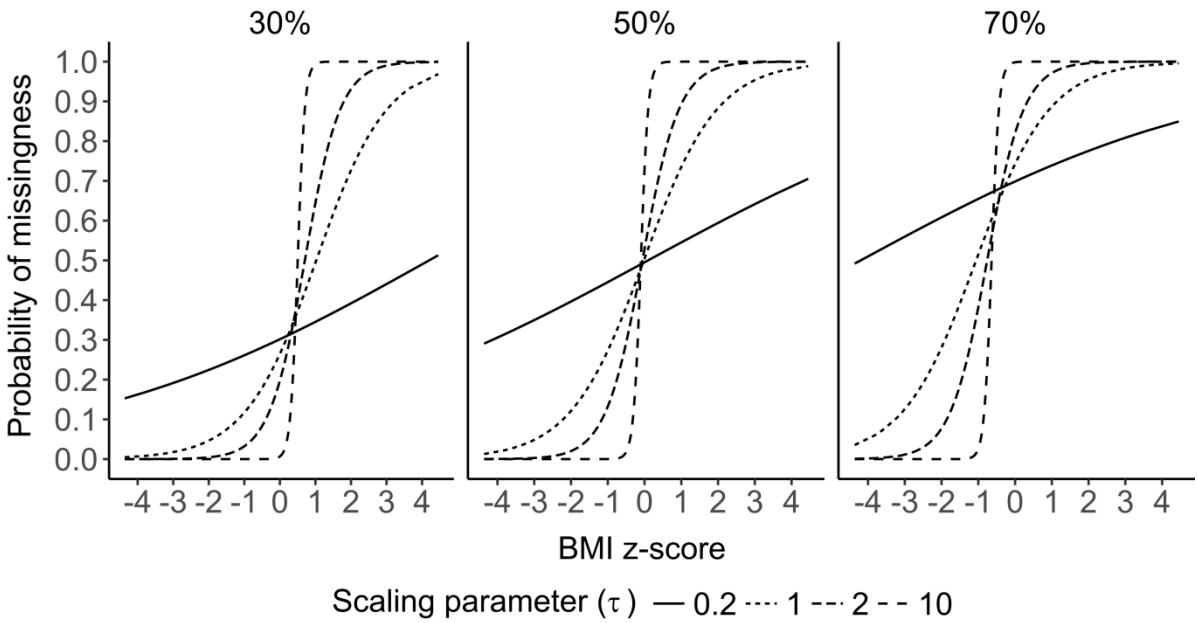

Figure 4.2 Probability of missingness for different frequencies of missingness (30%, 50% or 70%) and scaling parameters $\tau \in \{0.2, 1, 2, 10\}$

Figure 4.3 illustrates the difference between the zBMI distribution in the full data set and the zBMI distribution of the complete cases. Depicted are histograms of zBMI in the full data set ($N_{\text{full data set}}$ = 4138) and mean aggregated histograms of zBMI of the complete cases of the 300 simulated samples in each scenario. Note that a larger scaling parameter τ meant more missing values

at the upper tail of the zBMI distribution; in a deterministic scenario, at τ = 10, hardly any individuals with a zBMI $\geq$ 0 remain in the complete cases.

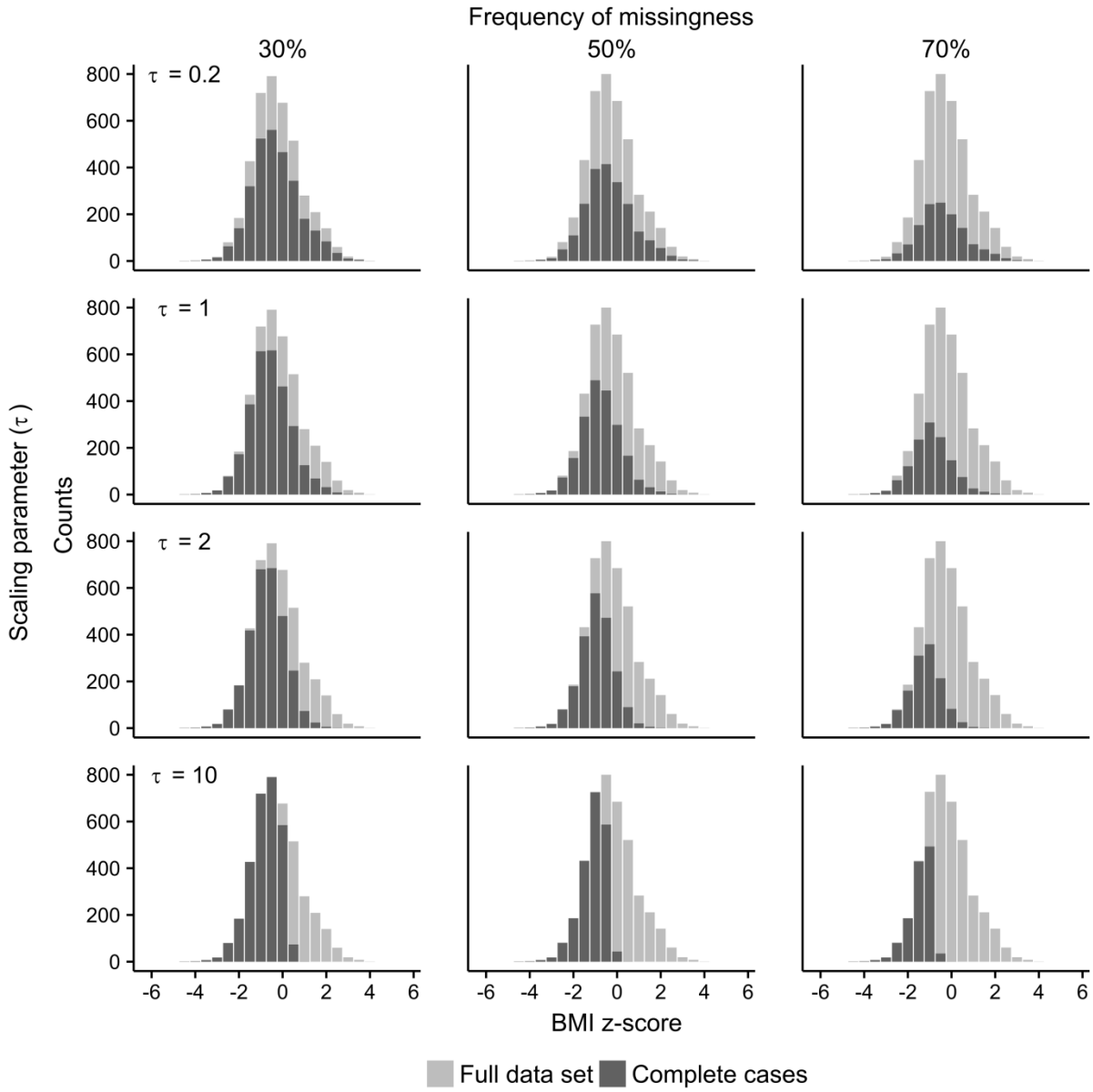

Figure 4.3 Histograms comparing the distribution of zBMI in the full data set (light grey) and in the complete cases (grey) for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$

Analysis models

To demonstrate how different modelling choices affect our conclusions from a regression analysis, we considered three nested linear regression models of increasing complexity. These three models are

called the *simple*, *intermediate* and *complex* analysis model and are specified as follows. The *simple analysis model* contained only main effects:

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot PA + \beta_7 \cdot AVM + \epsilon$$

The *complex analysis model* includes main effects, as well as, all quadratic and cubic terms:

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot zBMI^2 + \beta_7 \cdot zBMI^3 + \beta_8 \cdot PA + \beta_9 \cdot PA^2 + \beta_{10} \cdot PA^3 + \beta_{11} \cdot AVM + \beta_{12} \cdot AVM^2 + \beta_{13} \cdot AVM^3 + \epsilon$$

The *intermediate analysis model* is obtained by subjecting the *complex analysis model* to automatic model selection on the full data set with the Bayesian information criterion (BIC) as decision rule using the R-package glmulti (Calcagno and de Mazancourt, 2010) resulting in:

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot age + \beta_3 \cdot education_{medium} + \beta_4 \cdot education_{low} + \beta_5 \cdot zBMI^2 + \beta_6 \cdot AVM + \epsilon$$

Multiple imputation and inverse probability weighting

In addition to CCA, we also carried out multiple imputation (MI) (Schafer, 1999; Sterne et al., 2009) and inverse probability weighting (IPW) (Li et al., 2013) for each simulated data set. Note that, for our particular simulation setting and for a correctly specified analysis model, neither MI nor IPW should add anything to the CCA in the particular situation where zHOMA is MAR given zBMI. However, for illustrative purpose, we did not use this knowledge but rather performed the analysis from the perspective of a naive researcher who has to make assumptions about the missing mechanisms, and may choose to use either MI or IPW to account for the missing data.

The imputation model for zHOMA was a model fitted by ordinary least squares (OLS) regression with covariates as specified below. For each of the 300 data sets with missing values we created 5 imputed data sets. For each model, estimates were pooled across the imputed data sets using Rubin's rules (Rubin, 1987; Van Buuren and Groothuis-Oudshoorn, 2011). In accordance with *the simple, intermediate, and complex analysis model* outlined above, we used congenial models (Van Buuren, 2018) to impute missing values including as additional predictors migration background and country (Spratt et al., 2010; Sterne et al., 2009; Van Buuren et al., 1999). This procedure corresponds to the

situation where one believes in the MAR assumption given an extended set of covariates not all of which are used in the analysis model.

The *simple imputation model*:

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot PA + \beta_7 \cdot AVM + \beta_8 \cdot migrant_{partly} + \beta_9 \cdot migrant_{full} + \beta_{10} \cdot migrant_{unknown} + \beta_{11} \cdot country + \epsilon$$

The *complex imputation model:*

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot zBMI^2 + \beta_7 \cdot zBMI^3 + \beta_8 \cdot PA + \beta_9 \cdot PA^2 + \beta_{10} \cdot PA^3 + \beta_{11} \cdot AVM + \beta_{12} \cdot AVM^2 + \beta_{13} \cdot AVM^3 + \beta_{14} \cdot migrant_{partly} + \beta_{15} \cdot migrant_{full} + \beta_{16} \cdot migrant_{unknown} + \beta_{17} \cdot country + \epsilon$$

The *intermediate imputation model:*

$$zHOMA = \beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot age + \beta_3 \cdot education_{medium} + \beta_4 \cdot education_{low} + \beta_5 \cdot zBMI^2 + \beta_6 \cdot AVM + \beta_7 \cdot migrant_{partly} + \beta_8 \cdot migrant_{full} + \beta_9 \cdot migrant_{unknown} + \beta_{10} \cdot country + \epsilon$$

To construct the weights for IPW, we fitted a logistic regression model (Fox, 2016) with the missingness indicator as the response variable and with covariates as described below for each simulated data set; the analysis model was then fitted using the inverse of the probabilities being a complete case as weights (Lumley, 2004).

The *weighting model:*

$$P(complete\ case = 1) = \frac{exp(\beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot PA + \beta_7 \cdot AVM)}{1 + exp(\beta_0 + \beta_1 \cdot zBMI + \beta_2 \cdot sex + \beta_3 \cdot age + \beta_4 \cdot education_{medium} + \beta_5 \cdot education_{low} + \beta_6 \cdot PA + \beta_7 \cdot AVM)}$$

Results

The *simple*, *intermediate* and *complex analysis models* were fitted to the full data set to provide "true" estimates of regression coefficients that served as benchmarks to assess the impact of missing values, under the different missingness mechanisms, on CCA and the two missing data methods. The full data analysis (Table 4.1, *simple analysis model*) suggests, for instance, that a higher zBMI predicted a higher

expected zHOMA, whereas older children had a lower expected zHOMA. Results were roughly in agreement with the analysis of (Peplies et al., 2016).

In the following, we will mostly focus on the estimation of the coefficient of zBMI, β_1, as a measure of the association between the primary exposure, BMI, and the outcome, HOMA. Turning to the analyses with missing data, we first conducted an exploratory visual check for apparent non-linearities by inspecting selected scatterplots, for the complete cases, of zBMI and zHOMA for each frequency of missingness and τ (Figure 4.4). Locally weighted scatterplot smoothing (LOESS) (red lines in Figure 4.4) suggested a quadratic bivariate association between zBMI and zHOMA. If this non-linearity cannot be removed by including the other covariates then the *simple analysis model* with its linear effect of zBMI does not correctly describe the data.

Table 4.1 Linear regression of the simple, intermediate, and complex analysis model with zHOMA as dependent variable

	Simple	Intermediate	Complex
	$\hat{\beta}$ and 95% CI	$\hat{\beta}$ and 95% CI	$\hat{\beta}$ and 95% CI
zBMI	0.375 (0.347, 0.402)	0.350 (0.322, 0.378)	0.376 (0.334, 0.418)
Physical activity (h/d)	-0.126 (-0.217, -0.034)		-1.216 (-1.886, -0.547)
Sex			
Male	Reference		Reference
Female	-0.058 (-0.119, 0.004)		-0.040 (-0.101, 0.021)
Mean-centered age	-0.045 (-0.064, -0.027)	-0.052 (-0.070, -0.034)	-0.051 (-0.069, -0.032)
Educational level			
High education	Reference	Reference	Reference
Medium education	0.142 (0.079, 0.204)	0.133 (0.072, 0.195)	0.127 (0.065, 0.188)
Low education	0.331 (0.188, 0.474)	0.277 (0.135, 0.419)	0.286 (0.144, 0.429)
Audio-visual media time (h/d)	0.064 (0.033, 0.095)	0.067 (0.037, 0.098)	0.097 (0.059, 0.134)
$zBMI^2$		0.073 (0.056, 0.089)	0.073 (0.055, 0.090)
$zBMI^3$			-0.007 (-0.015, 0.002)
Physical activity $(h/d)^2$			1.239 (0.459, 2.018)
Physical activity $(h/d)^3$			-0.363 (-0.627, -0.098)
Audio-visual media time $(h/d)^2$			-0.035 (-0.067, 0.003)
Audio-visual media time $(h/d)^3$			0.003 (-0.004, 0.009)
N	4138	4138	4138

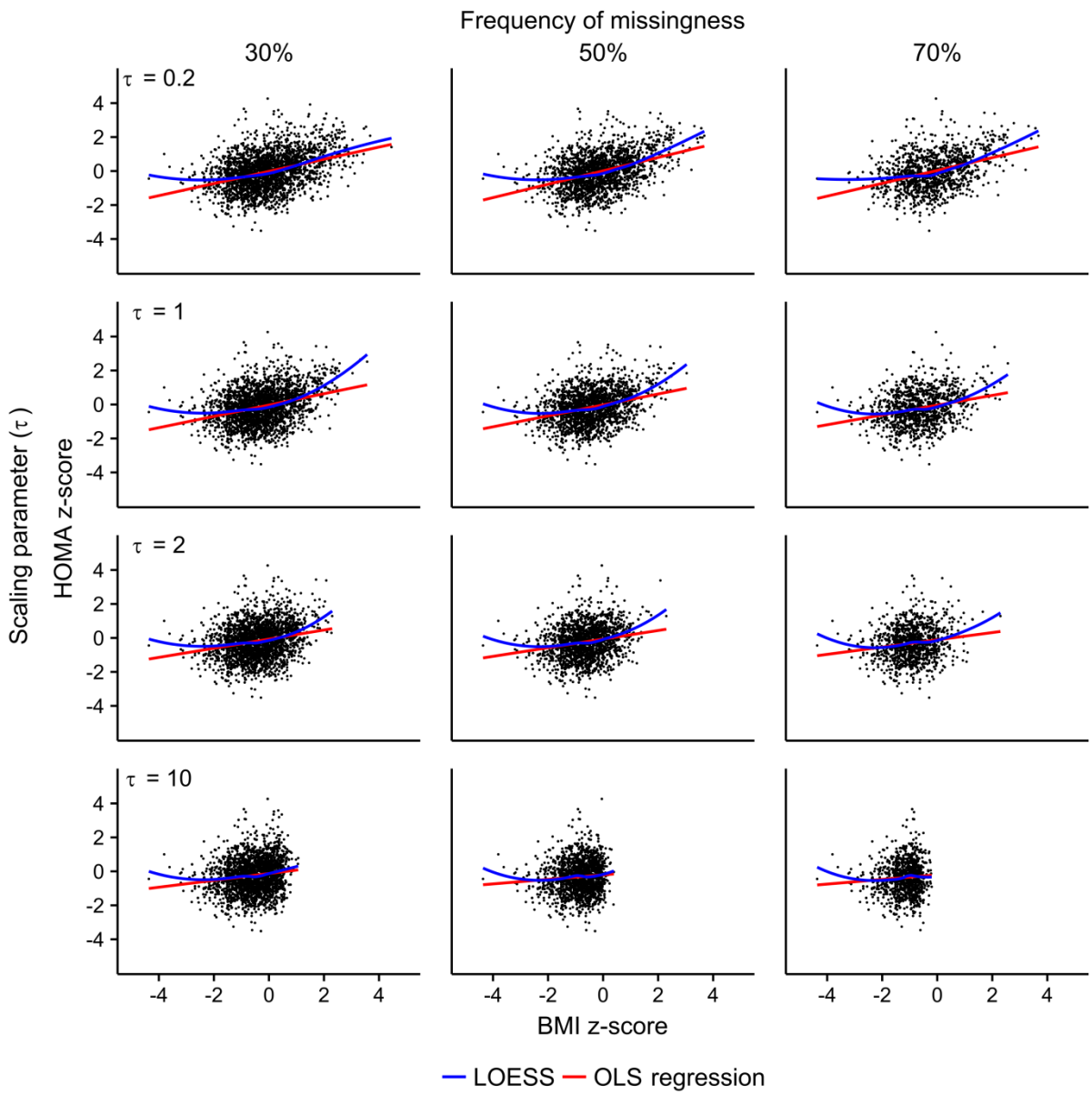

Figure 4.4 Scatterplots of zHOMA and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameters $\tau \in \{0.2, 1, 2, 10\}$. OLS-Regression (blue line) and a locally weighted scatterplotsmoothing (LOESS, red line)

Figure 4.5 shows boxplots of the distribution of CCA, MI and IPW estimates of the regression coefficient separately for the *simple, intermediate*, and *complex analysis model*. These illustrate how the different missingness mechanisms and modeling choices affected estimation of the zBMI coefficient. Figure 4.5 (Top Row, Column 1) shows the results of the CCA obtained under the *simple analysis model*. In Figure 4.5, top row, column 1, focusing on the scaling parameter of τ = 0.2 (light grey boxes) we see that, with the simple analysis model, a higher frequency of missing values lead to increased bias in the estimated regression coefficients of zBMI. A comparison across different values of τ shows that the magnitude of bias was more strongly affected by τ than by the frequency of missingness. For example, even with

a 30% frequency of missingness, a scaling parameter of τ = 10 (the most extreme threshold missingness mechanism) produced highly biased regression coefficients, whereas the bias was much smaller for τ = 0.2 (the smoother, more continuous missingness mechanism) even with 70% missing values.

As expected, under the simple analysis model a similar bias pattern was observed for MI as for the CCA. IPW was able to compensate for the presumably misspecified analysis model when the scaling parameter was $\tau \leq 1$ (Figure 4.5, top right). The *intermediate analysis* model, however, performed fairly well (Figure 4.5, middle row), except when τ = 10. In addition, the *intermediate analysis model* tended to perform overall better than the *complex analysis model* (Figure 4.5, bottom row). Although IPW exhibited the least bias on average for all analysis models, the variance of the estimates was extremely high for scaling parameters of $\tau \geq 2$. The estimated coefficients of other predictors, including age, low educational level, and audio-visual media time were still biased for the *intermediate analysis model* (see Additional file 1).

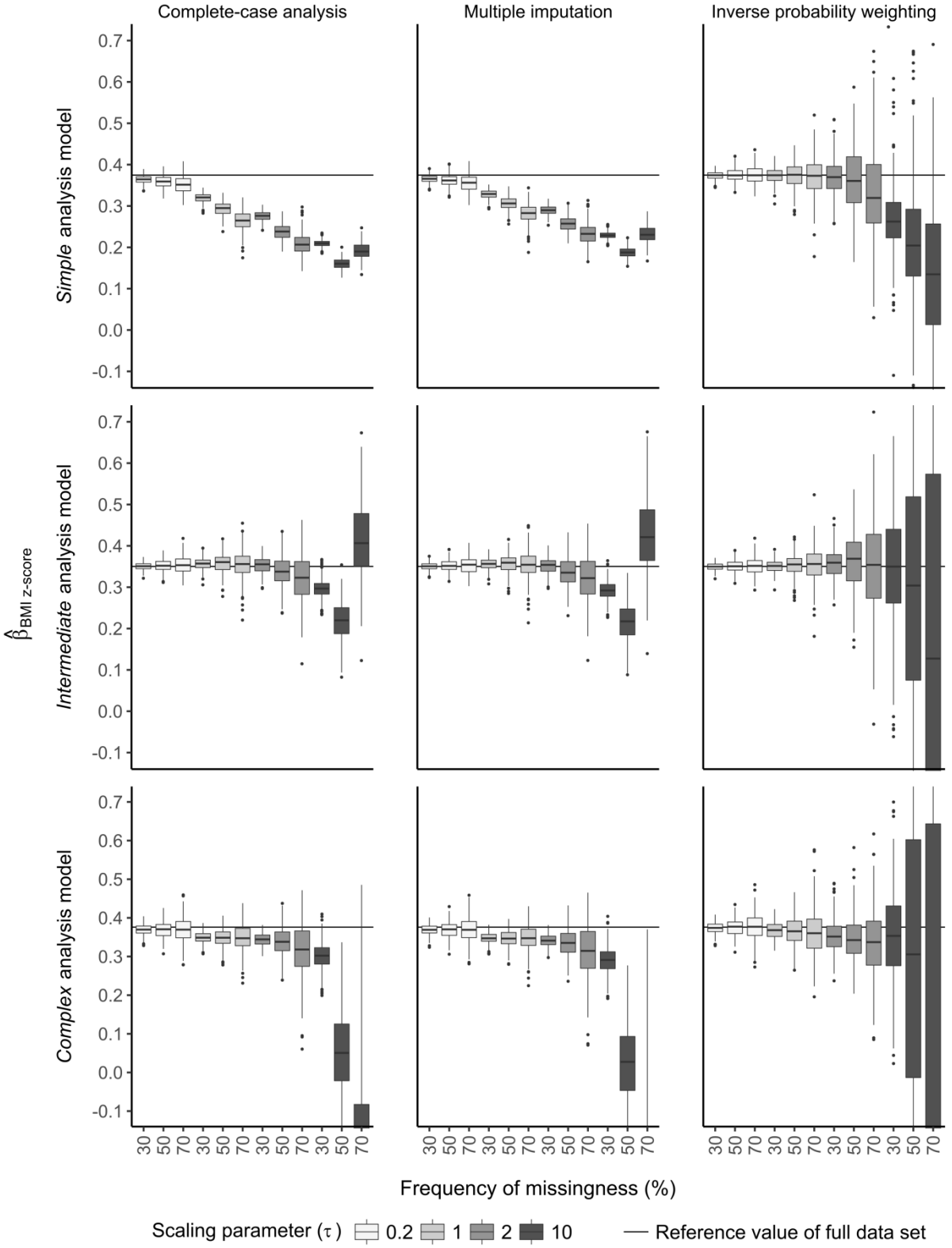

Figure 4.5 Boxplots of estimated regression coefficients of zBMI with zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are grouped by

different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axisTo verify whether model diagnostics would have revealed an apparent misspecification of the *simple analysis model*, we further inspected selected residual plots (Figure 4.6). For each frequency of missingness and each value of τ, these plots show a clear non-linear relation between residuals and zBMI indicating a misspecification of the functional form of zBMI in the simple analysis model (see Additional file 1 for residual plots of the *intermediate analysis model* and *complex analysis model*).

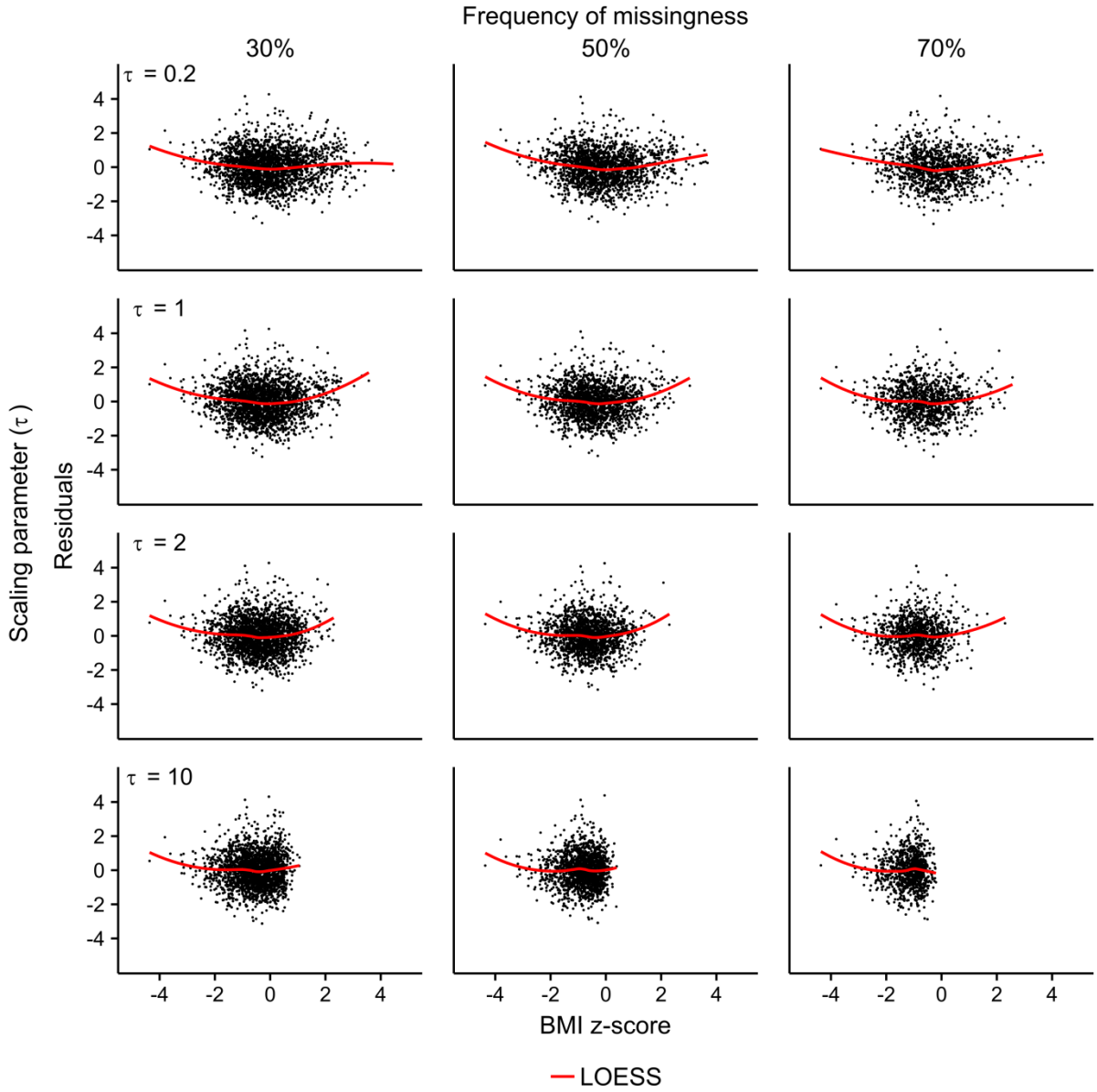

Figure 4.6 Residual plots of the *simple analysis model* and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

Secondary analysis with simulated zHOMA values

To confirm that the bias seen above was the result of a misspecified model rather than other factors, we conducted a secondary simulation study. We used the same covariate data as above, but artificially generated the response variable zHOMA from a now known outcome model; this we chose to be the *complex analysis model*, with parameter values as in Table 4.1, with added noise. We call these artificial data the *simulated zHOMA* (see Additional file 1). Since the *complex model* is the "true" generating model in this scenario, we would expect it to perform better than both the *simple* and the *intermediate model*.

Figure 4.7 shows selected scatterplots of zBMI and, in this case, simulated zHOMA for each combination of frequency of missingness and value of τ (cf. Figure 4.4). As expected, a curve fitted to the data (red line) revealed a non-linear bivariate association for all scatterplots.

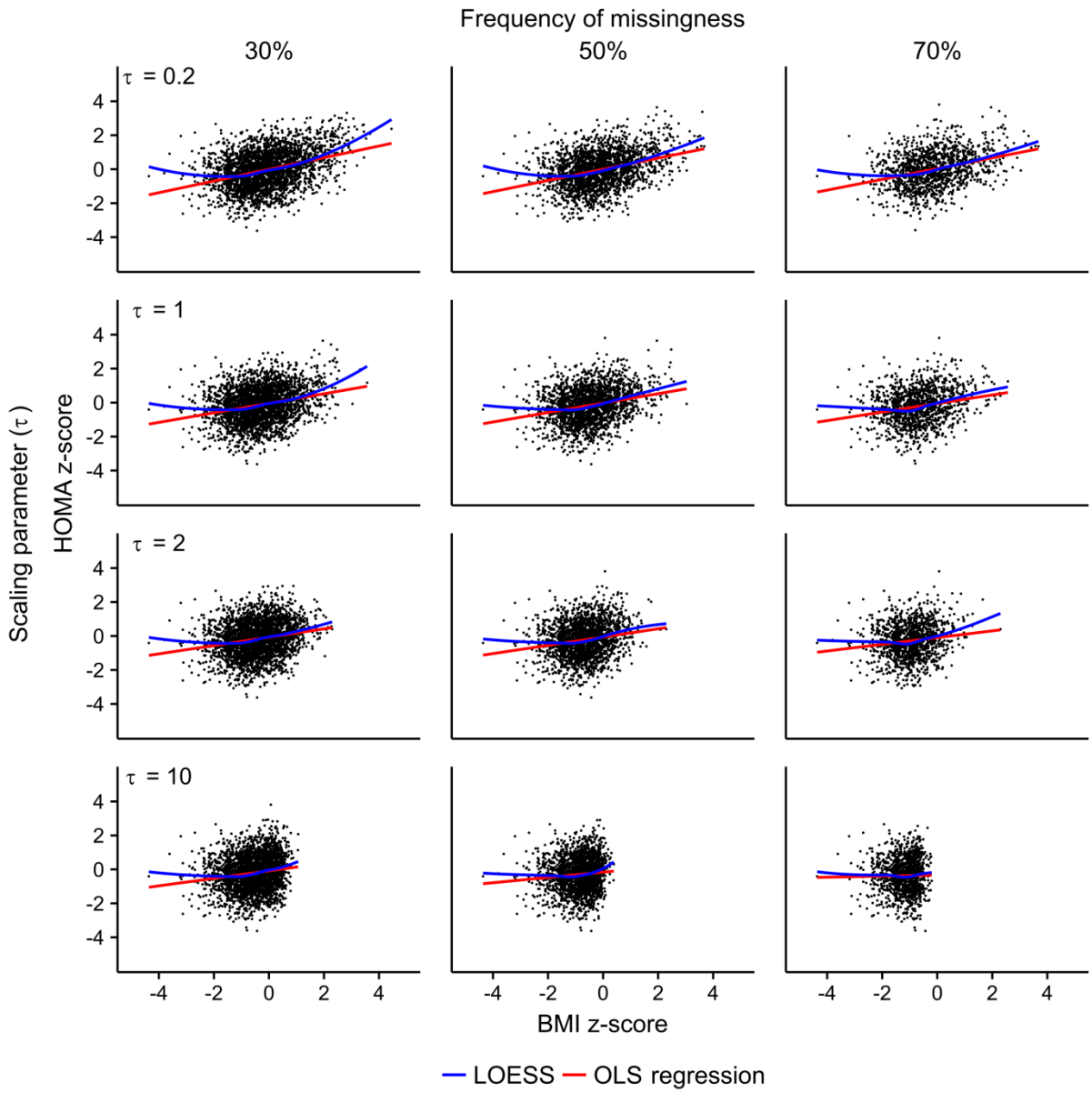

Figure 4.7 Scatterplots of simulated zHOMA and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. OLS-Regression (blue line) and a locally weighted scatterplot smoothing (LOESS, red line)

The results with the simulated zHOMA show that the regression coefficients of zBMI estimated by a CCA and by MI under the *simple analysis model* were increasingly biased with increasing τ and frequency of missingness (Figure 4.8, top row, Columns 1 to 2). IPW corrected the bias up to τ = 0.2 (Figure 4.8, top row, Column 3). We also observed a bias pattern of CCA and MI for the *intermediate analysis model*, while IPW corrected the bias up to τ = 10 (Figure 4.8, middle row, Columns 1 to 3). Finally, as expected, the estimated regression coefficients of zBMI of the *complex analysis model* were consistent using CCA, MI or IPW except for τ = 10 (Figure 4.8, bottom row, Columns 1 to 3).

The estimated coefficients of other predictors, including age, low educational level, and audio-visual media time were still biased, but to a lesser extent, for the *complex analysis model* (see Additional file 1).

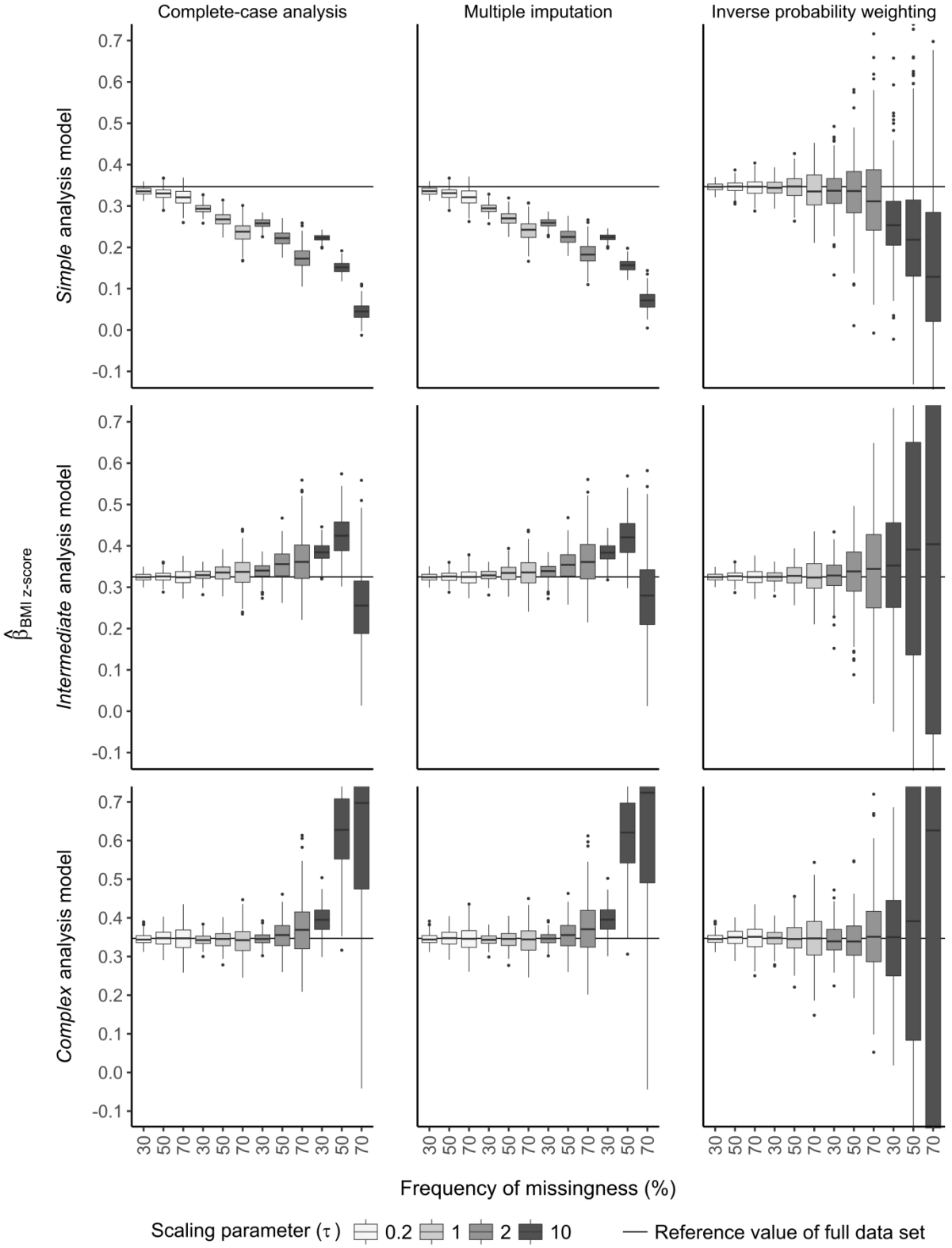

Figure 4.8 Boxplots of zBMI estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1 for complex analysis

model). Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

Like Figure 4.6, Figure 4.9 shows selected scatterplots of the residuals of the *simple analysis model* with simulated zHOMA as dependent variable and zBMI for each frequency of missingness and τ. Figure 4.9 again clearly shows non-linear relationships (red line) indicating that the functional form of zBMI is misspecified in the simple analysis model.

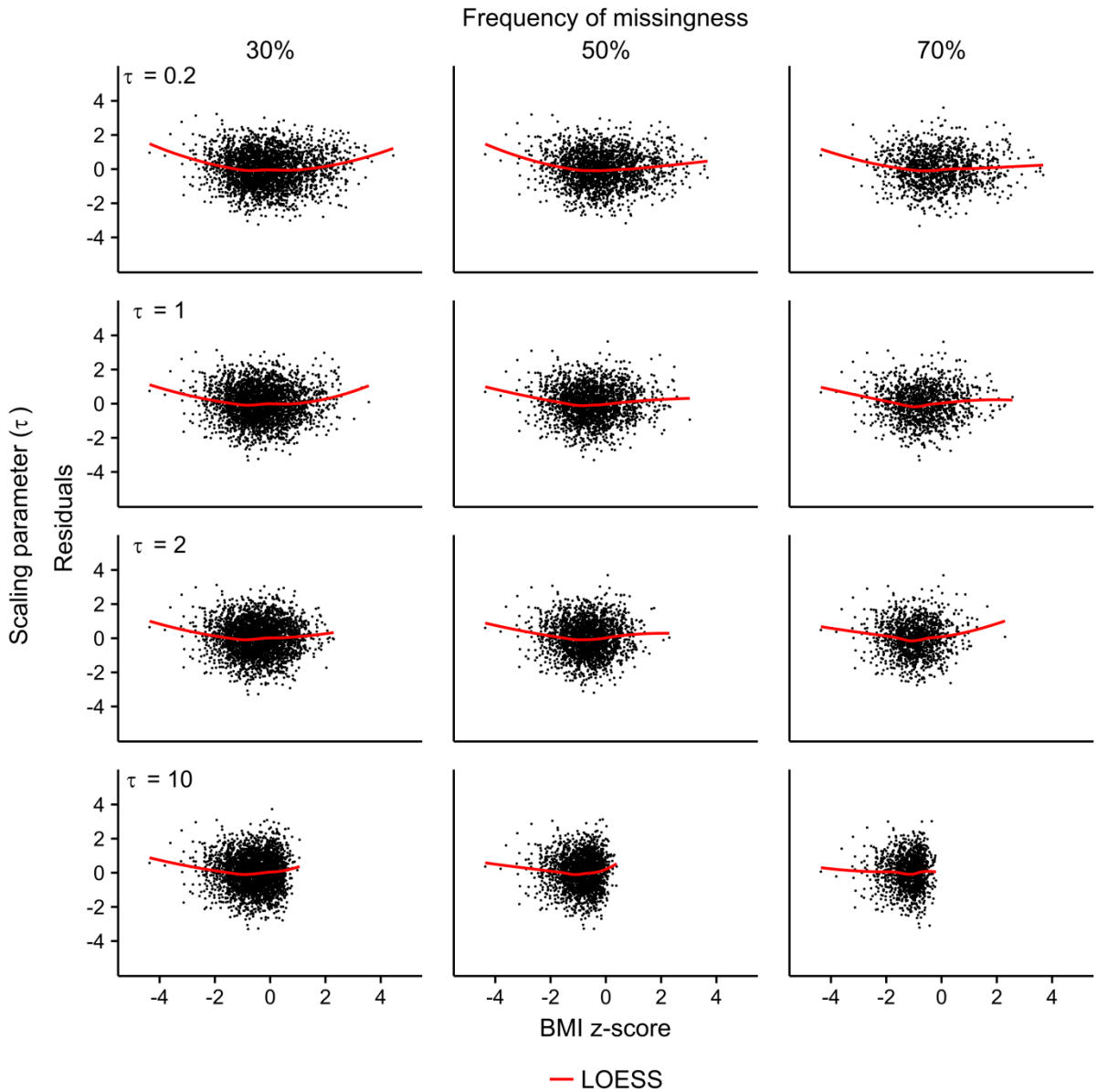

Figure 4.9 Residual plots for the *simple analysis model* using simulated zHOMA as dependent variable and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

Discussion

Our results illustrate three key points: (1) Model misspecification can lead to considerable bias when data contain missing values, even in a scenario where a CCA is known to be consistent, in principle; this bias equally affects MI, and to a lesser extent (if the weighting model is correct), IPW (although at the cost of precision). (2) In our example, basic model diagnostics, based on the complete cases, such as inspecting scatter- and residual plots for non-linearities, were sufficient to alert us to the misspecification of the simple analysis model with regard to the functional form of the exposure; this was weaker but detectable even for the most extreme missingness mechanisms. (3) Even with a correctly specified model it is hard to compensate for extreme dependence of missingness on the exposure, presumably because a complex (non-linear) model is more difficult to extrapolate into a region of exposure values where all outcome values are effectively missing.

The linear main-effects-only analysis model is usually chosen as the typical default model for analyzing data such as those in our motivating example with little attention paid to more complex models or model diagnostics, especially under the added complication of incomplete data. Thus, our findings are relevant, since popular heuristics guide researchers towards explaining phenomena with the simplest hypothesis possible (cf. Ockham's razor), possibly leading them to systematically prefer simple analysis models over complex ones. Furthermore, complex models are often difficult to interpret and may bear the danger of overfitting. However, our results suggest that the unchecked avoidance of complex models, combined with incomplete data, can lead to seriously biased results and conclusions.

In general, artificial data used in simulation studies are valuable as they allow creating scenarios with specific properties tailored to the research question. Depending on the aim of the simulation, artificial data may perfectly fit or not fit model assumptions. At the same time, however, it is often hard to determine whether artificial data possess all the relevant properties of the empirical scenario they are meant to mimic. A strength of this paper is that we used empirical data in a simulation study highlighting real-world challenges, like nonlinearities and dependencies among the covariates, that are far more ambiguous than textbook examples. To further verify our results, we added a secondary simulation study where data on the outcome were simulated using a model closely inspired by the real data, but the explanatory variables and their correlation structure was still exactly as in the empirical data. While these additional results broadly supported our primary findings, they also showed that the coefficients of other variables associated with zBMI, like age and educational level, could be affected by bias when the model misspecified the functional form of zBMI or lacks an interaction.

The missingness mechanism considered in this study was a simplification with respect to several points. In practice, missing values are unlikely to dependent on a single variable; moreover, in our simulation, the only missing values occurred in the outcome variable and their missingness depended

on a fully measured exposure, which justifies a CCA. Additionally, we exaggerated the dependence between zBMI and missingness for demonstrative purpose; for large scale parameters it was as good as impossible to still observe the outcome for large zBMI values and one cannot expect reliable inference in such situations. Furthermore, alternative missingness mechanisms might follow, e.g., a u-shaped functional form, since it is plausible that obese as well as underweight children have a higher probability of drop-out. Throughout, we compared CCA with two standard methods for dealing with missing values, MI and IPW. In our particular setting, MI could not be expected to improve on CCA with regard to bias because the only incomplete variable was the outcome, and the misspecified analysis model was also used as the imputation model. MI was also not expected to improve efficiency because the additional predictors were not actually informative, which would again not be known in practice. However, it is reassuring to see that MI was neither worse nor less efficient than CCA. Further, for MI the specification of the imputation model(s) is another issue that we did not touch upon here as our imputation model was almost identical to the analysis model except for additional predictors. A comprehensive overview of different methods for determining suitable imputation models is given in Nguyen et al. (2017), including standard regression diagnostics as well as, for example, cross-validation. MI will give unbiased results if the data are MAR and the imputation model is correctly specified. Thus if we had used the correct imputation model – i.e. used the simple analysis model, but the true (complex) imputation model, we would not expect bias – however, this would not be an approach that would be taken in practice.

In contrast, IPW mostly performed well regarding bias except for very large τ values. In the latter case IPW is very unstable and biased because when the dependence of missingness on zBMI was very strong there were no observed HOMA values at all for zBMI over a certain value and the weights became extremely large. The otherwise good performance of IPW even with a misspecified analysis model was due to using a correctly specified missingness model which ensured an accurate re-weighting of the incomplete sample. In practice, the true missingness mechanism is not known, which could be a further source of misspecification and induce its own bias resulting in a worse performance of inverse weighting. It is our impression, though, that there is greater awareness for the potential bias due to a misspecified weighting model, in contrast to the problem of misspecified analysis model.

For easier comparisons, we used the same three given models throughout, where the intermediate model was obtained by model selection on the full data. In practice, one would instead need to build the analysis model with the incomplete data. However, this is a difficult task, especially when it is unclear whether the assumptions for a CCA to be valid are fulfilled, as in our settings; then imputation and model selection may need to be iterated in a suitable way to arrive at an analysis model. There appears to be no generally agreed approach but some first ideas have been put forward by White et al. (2011). To investigate the issue, we supplemented our simulation study with one scenario where a

basic model selection procedure, based on the complete cases, was carried out for each simulated data set (Figure 4.12 in the Additional file 1); the results were in line with the previous findings.

Conclusions

In summary, we suggest that researchers carefully consider their choice of analysis model; misspecification will obviously always pose a risk of some bias, but while this may be small with fully observed data and a moderately misspecified model, missing data may seriously amplify the bias. Model building should only be based on the complete cases if the missing mechanism can reasonably be assumed to be such that CCA is in principle consistent. Otherwise more advanced techniques are called for as suggested by White et al. (2011). Moreover, further research into sensitivity analyses and model diagnostics with incomplete data is still needed.

Additional file 1

Matching

We used nearest neighborhood matching (Ho et al., 2011) with Mahalanobis distance to increase the sample size of the data for the simulation (Figure 4.10). Thus, for 2081 drop-outs at follow-up (T1) with complete baseline data (T0), we searched for matches in 2942 children with complete baseline data that participated at follow-up. In the matching, we allowed a child that participated at follow-up to match with multiple drop-outs at follow-up. The dependent variable in the matching model was drop-out at follow-up. The matching included the six variables used in the main analysis (z-scores of body mass index (zBMI), physical activity (PA), educational level, audio-visual media time (AVM), children's age, and sex of child; for details see Section "Outcomes and exposures" of the main manuscript) and the following additional variables: a score of study compliance, constructed separately for children and parents by counting the number of key examination modules they completed at baseline and at the first follow-up. The place of birth of parents served to define the migration background (full migrant: both parents foreign-born; partly migrant: one parent foreign-born; not migrant: otherwise), and country.

The *Matching model:*

$$P(drop-out_{T1}=1)=\frac{exp(\beta_0+\beta_1\cdot x_1+\beta_2\cdot x_2+\beta_3\cdot x_3+\cdots+\beta_{19}\cdot x_{19})}{1+exp(\beta_0+\beta_1\cdot x_1+\beta_2\cdot x_2+\beta_3\cdot x_3+\cdots+\beta_{19}\cdot x_{19})}$$

x_1 = $zBMI_{T0}$

x_2 = children's age_{T0}

x_3 = children's study $compliance_{T0}$

x_4 = parent's study compliance$_{T0}$

x_5 = partly migrant$_{T0}$

x_6 = full migrant$_{T0}$

x_7 = unknown migration background$_{T0}$

x_8 = medium educational level$_{T0}$

x_9 = low educational level$_{T0}$

x_{10} = female children$_{T0}$

x_{11} = Estonia$_{T0}$

x_{12} = Cyprus$_{T0}$

x_{13} = Belgium$_{T0}$

x_{14} = Sweden$_{T0}$

x_{15} = Germany$_{T0}$

x_{16} = Hungary$_{T0}$

x_{17} = Spain$_{T0}$

x_{18} = phyisical activity$_{T0}$

x_{19} = audio-visual media time$_{T0}$

Of the 2081 drop-outs at follow up, the matching resulted in 1196 children that were similar to children with complete baseline data who participated at follow-up. The final fully observed data set for the simulation study consisted of 2942 children with complete baseline data and the 1196 duplicates of children identified by matching resulting in 4138 complete cases.

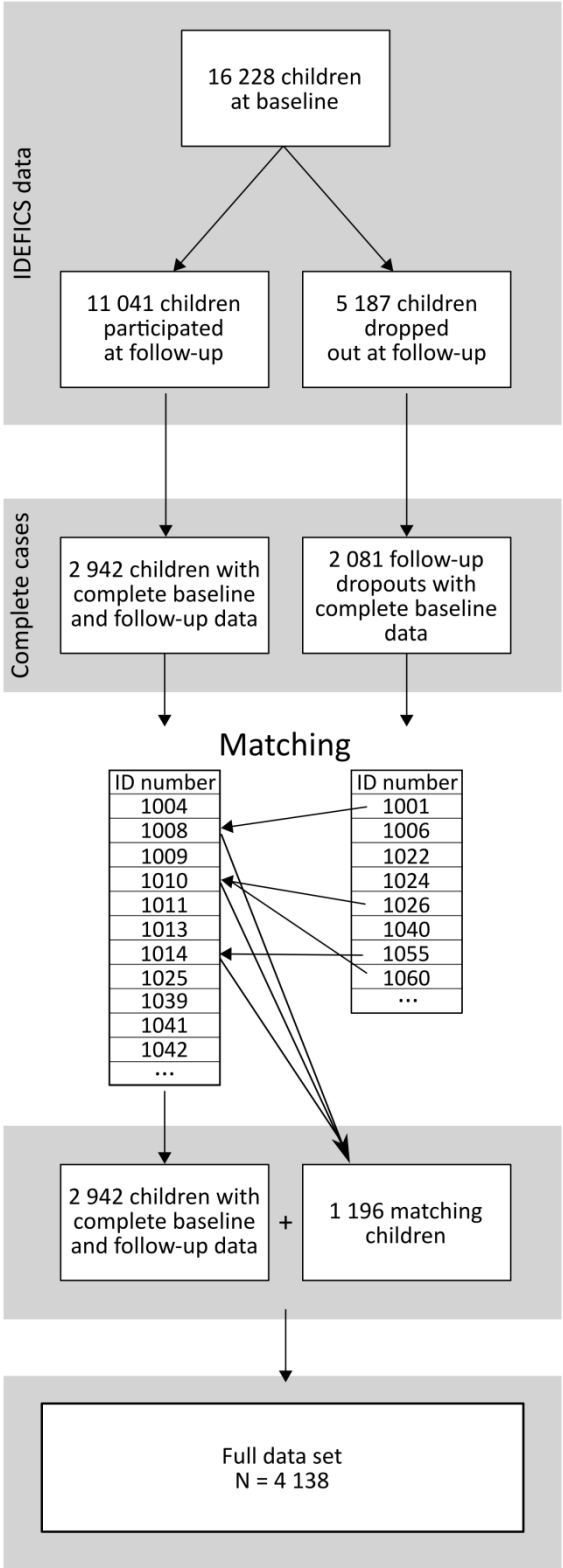

Figure 4.10 Matching procedure to create the full data set

Adding noise to predicted zHOMA values

For the repetition of simulation with artificial zHOMA values, we added noise to predicted zHOMA values that was obtained by randomly drawing from a normal distribution with a mean 0 and the estimated residual standard deviation (sd = 0.97) of the *complex analysis model*. Since the noise is very effective in masking the underlying non-linear association (cf. Figure 4.7), an example of zHOMA values without and with noise is given in Figure 4.11.

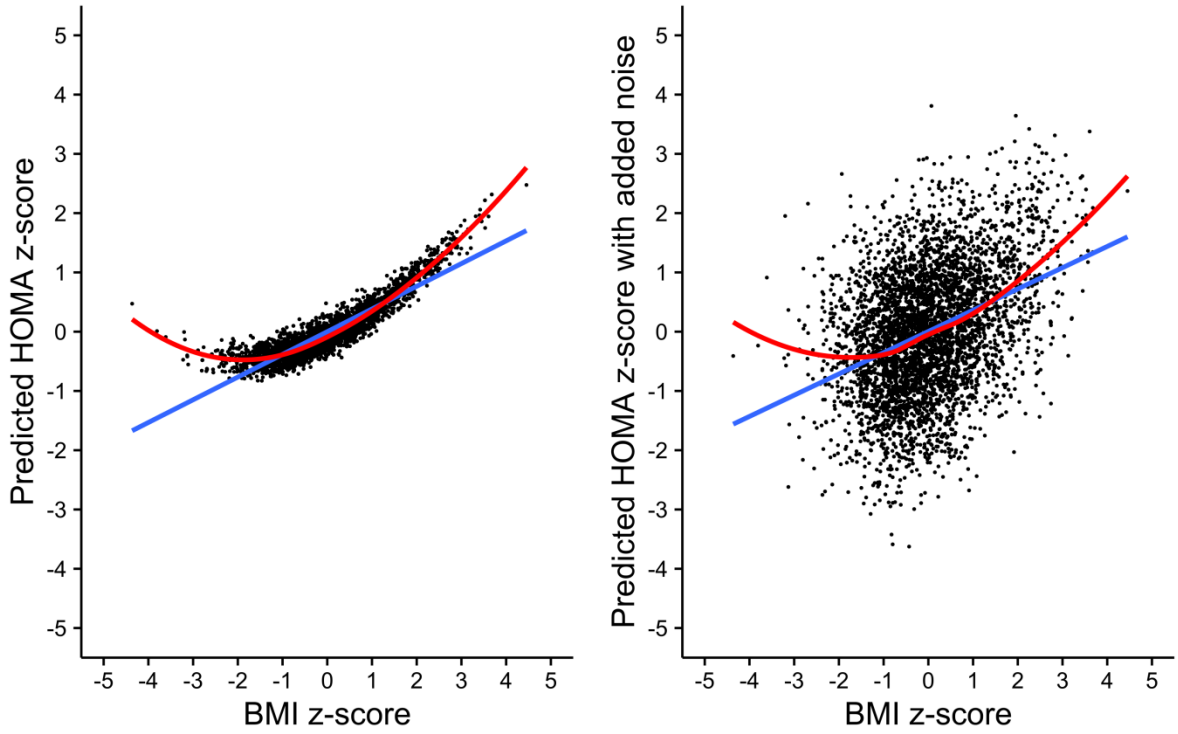

Figure 4.11 Scatterplots of BMI z-score and predicted HOMA z-score without (left panel) and with added noise (right panel). OLS-Regression (blue line) and locally weighted scatterplot smoothing (LOESS, red line)

Sensitivity analysis of model selection on complete cases

For the main analysis, the intermediate analysis model was selected by performing model selection using the Bayesian information criterion (BIC) once on the full data and then fitted to each of the 300 simulated data, so that the same set of estimated parameters was available for every data set simplifying the comparisons. Selecting the model separately for each of the 300 simulated data sets would more accurately reflect the added uncertainty due to model selection; but each of the 300 selected model would also likely differ in the number parameters included. To demonstrate that the outcomes of our analysis and our conclusions remain the same for the latter approach, Figure 4.12 depicts the results corresponding to the leftmost panel in the middle row of Figure 4.5 in the main text.

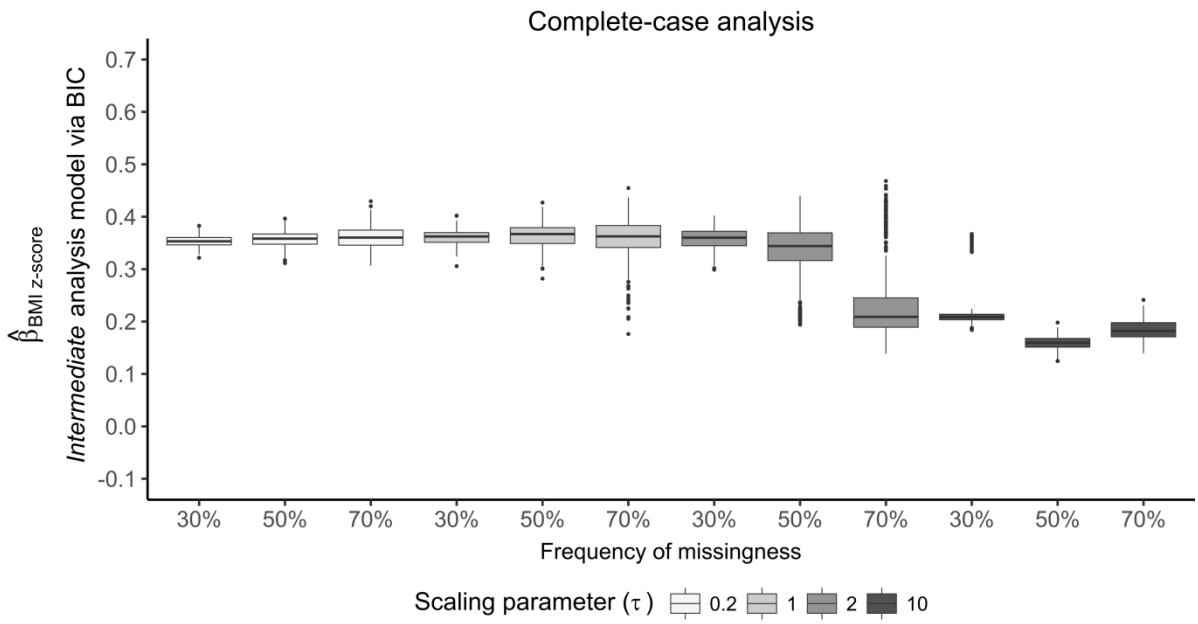

Figure 4.12 Boxplots of estimated regression coefficients of zBMI with zHOMA as the dependent variable with simulated missingness depending on zBMI. Complete-case analyses using a model selected via the Bayesian information criterion (BIC) on the data sets with deleted observations. Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shade, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis. The Figure corresponds to the results in Figure 4.5, middle row, Column 1, in the main text

Bias of estimates of coefficients for other predictors

Complete-case analysis using the *simple, intermediate and complex analysis model* also biased estimates of age of child, low educational level, and audio-visual media time (h/d) (Figure 4.13 to Figure 4.16; estimates corresponding to Figure 4.5 in the manuscript). These variables were less biased in the simulation study with simulated zHOMA (Figure 4.17 to Figure 4.21; estimates corresponding to Figure 4.8 in the manuscript).

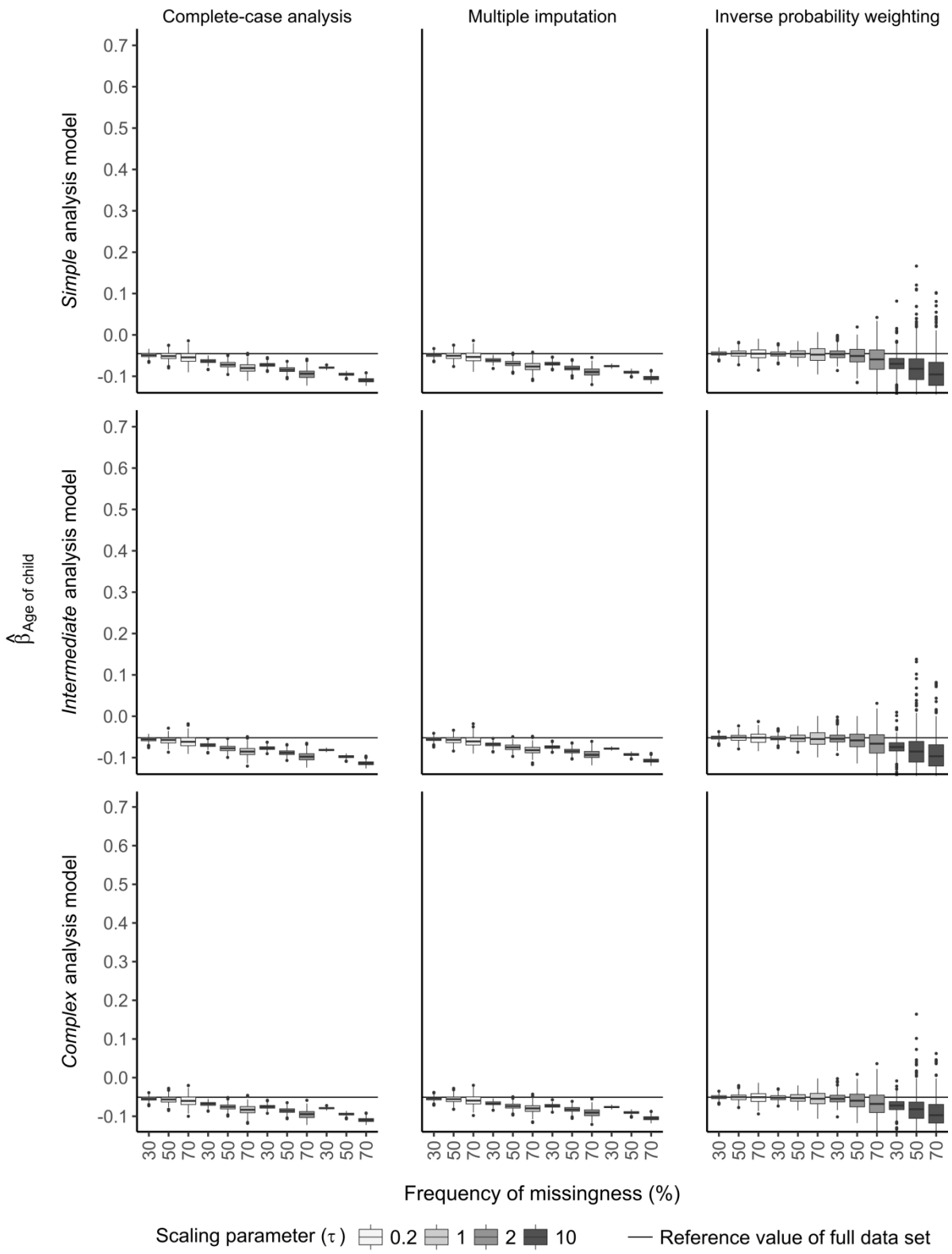

Figure 4.13 Boxplots of age of child estimates of regression models with zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are grouped by

different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

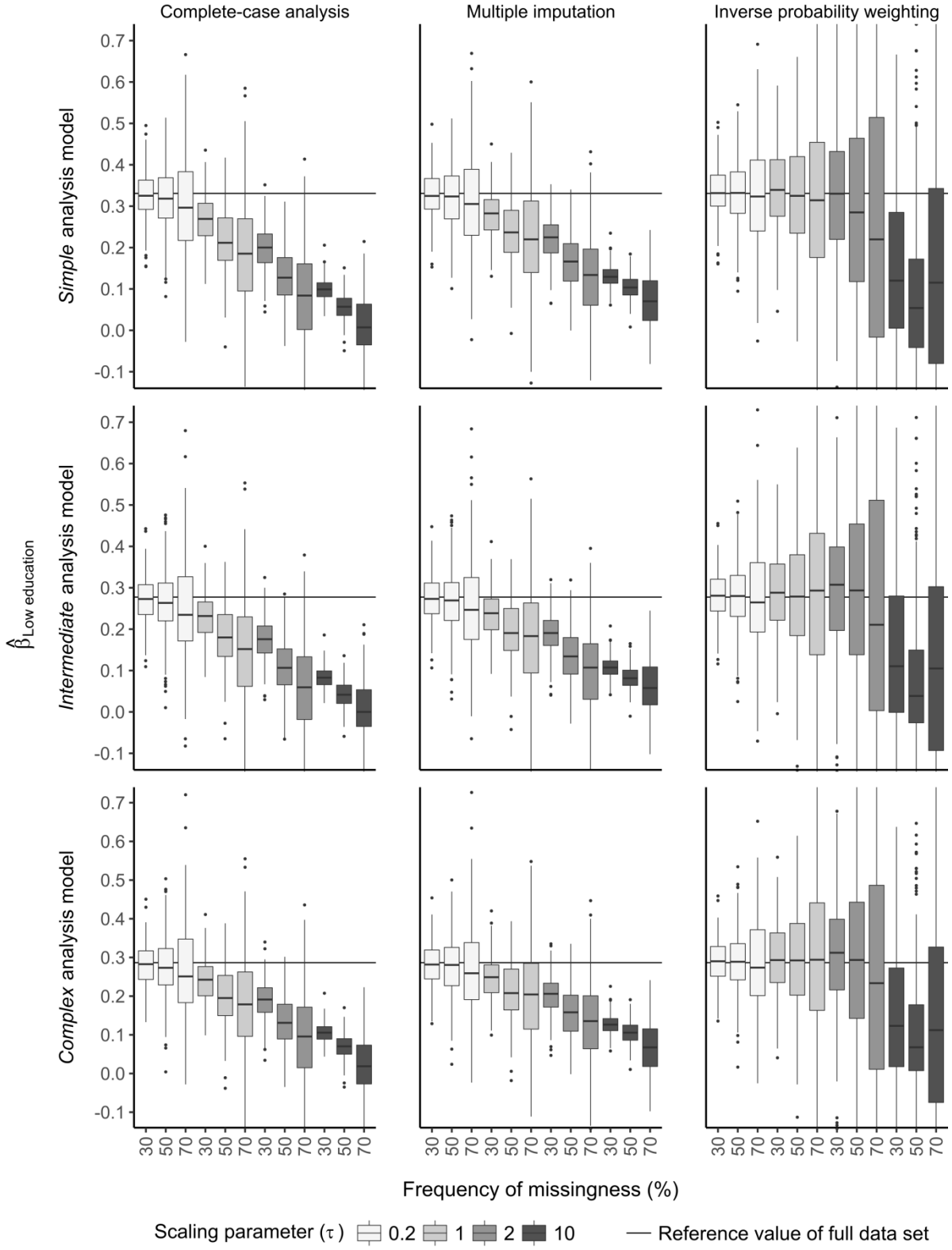

Figure 4.14 Boxplots of low educational level estimates of regression models with zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are

grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

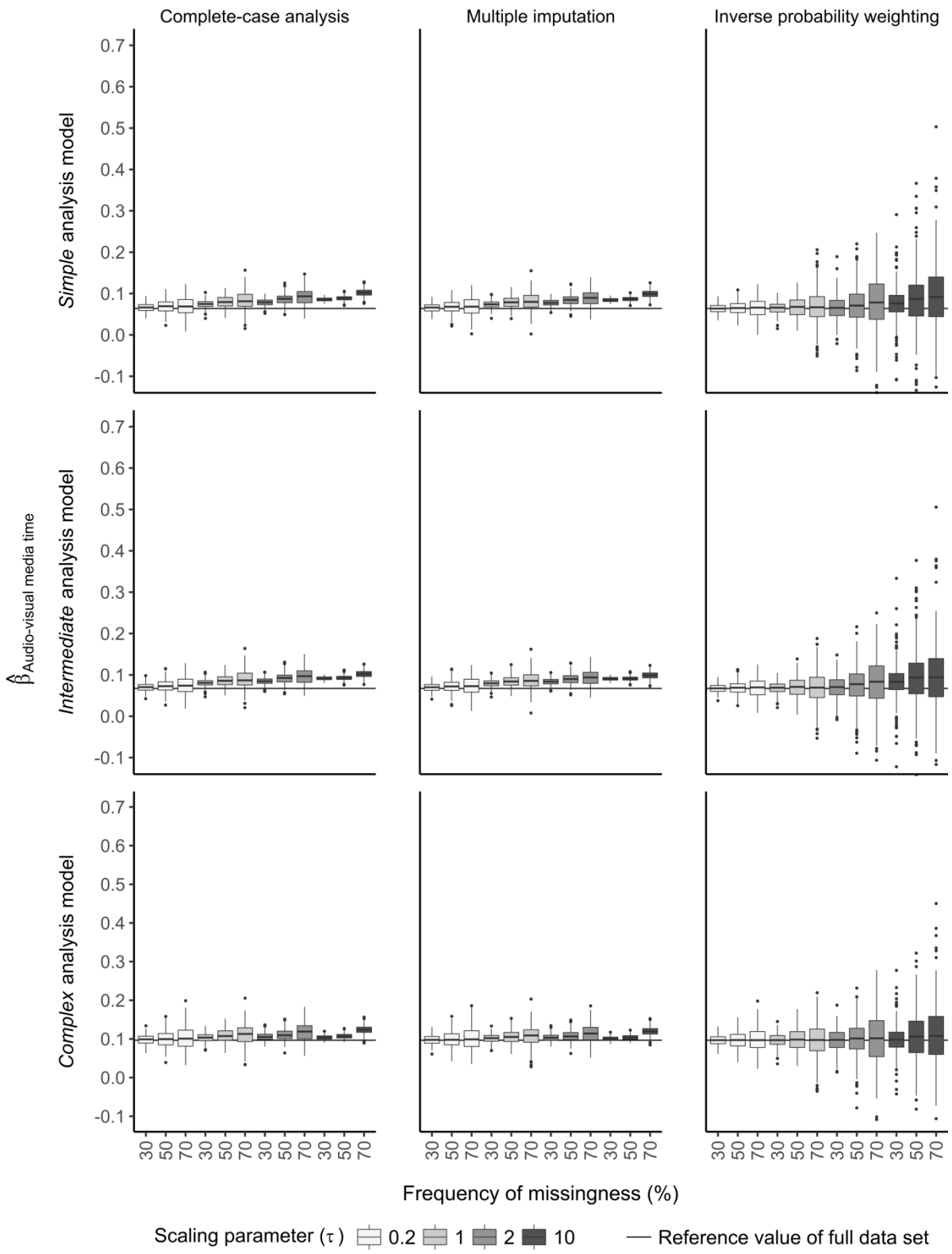

Figure 4.15 Boxplots of audio-visual media time (h/d) estimates of regression models with zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table

4.1). Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

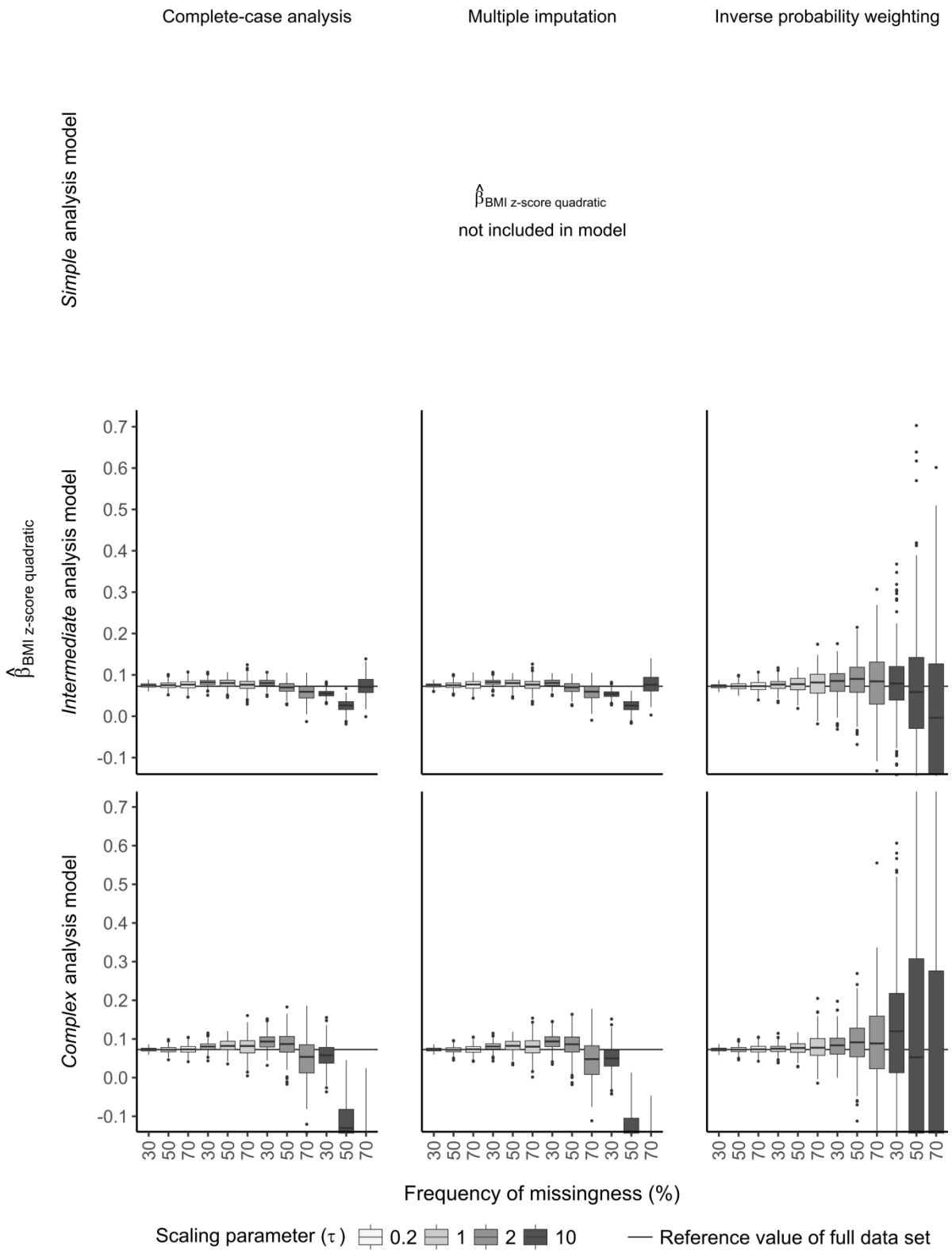

Figure 4.16 Boxplots of BMI (quadratic) estimates of regression models with zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are

grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

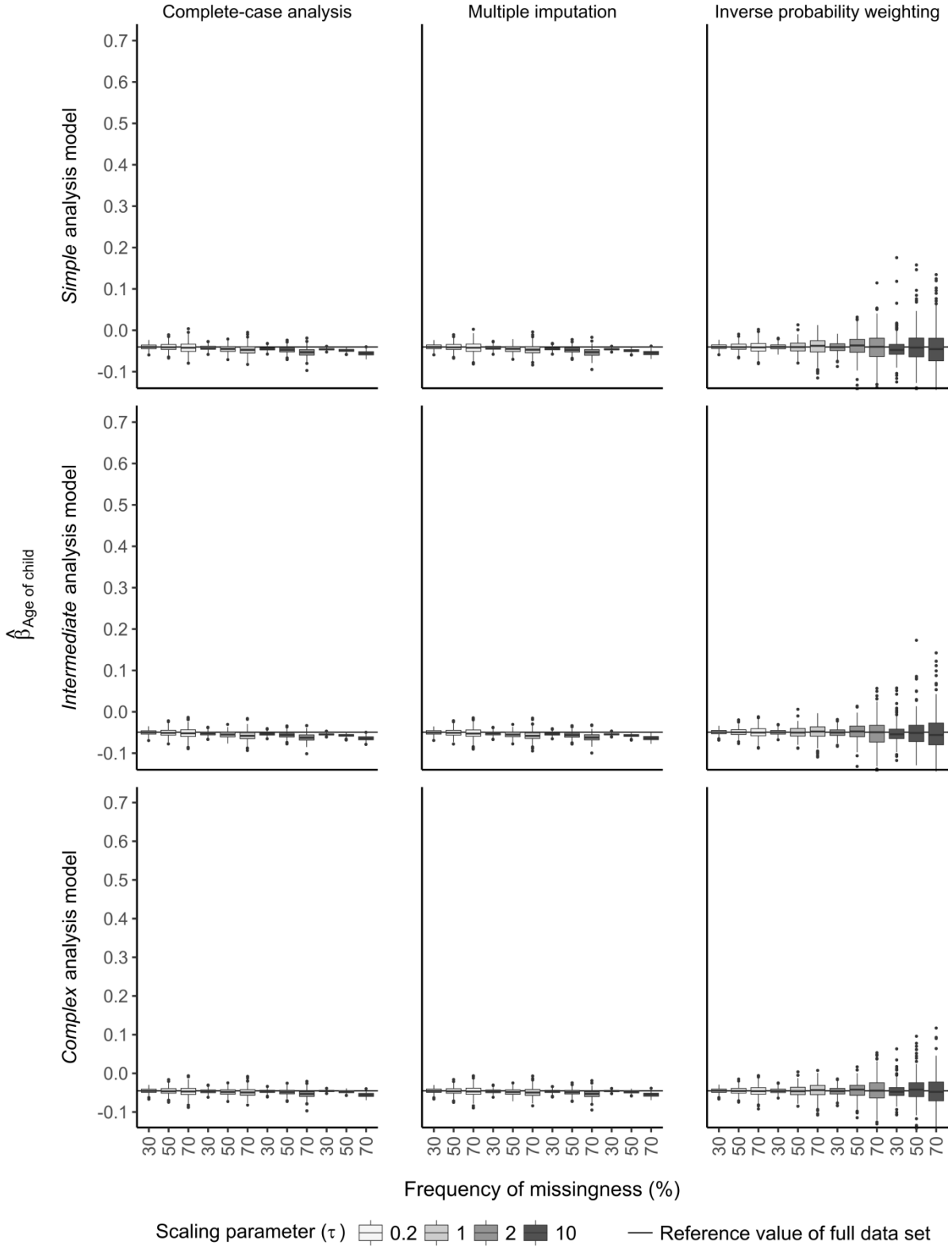

Figure 4.17 Boxplots of age of child estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are

grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

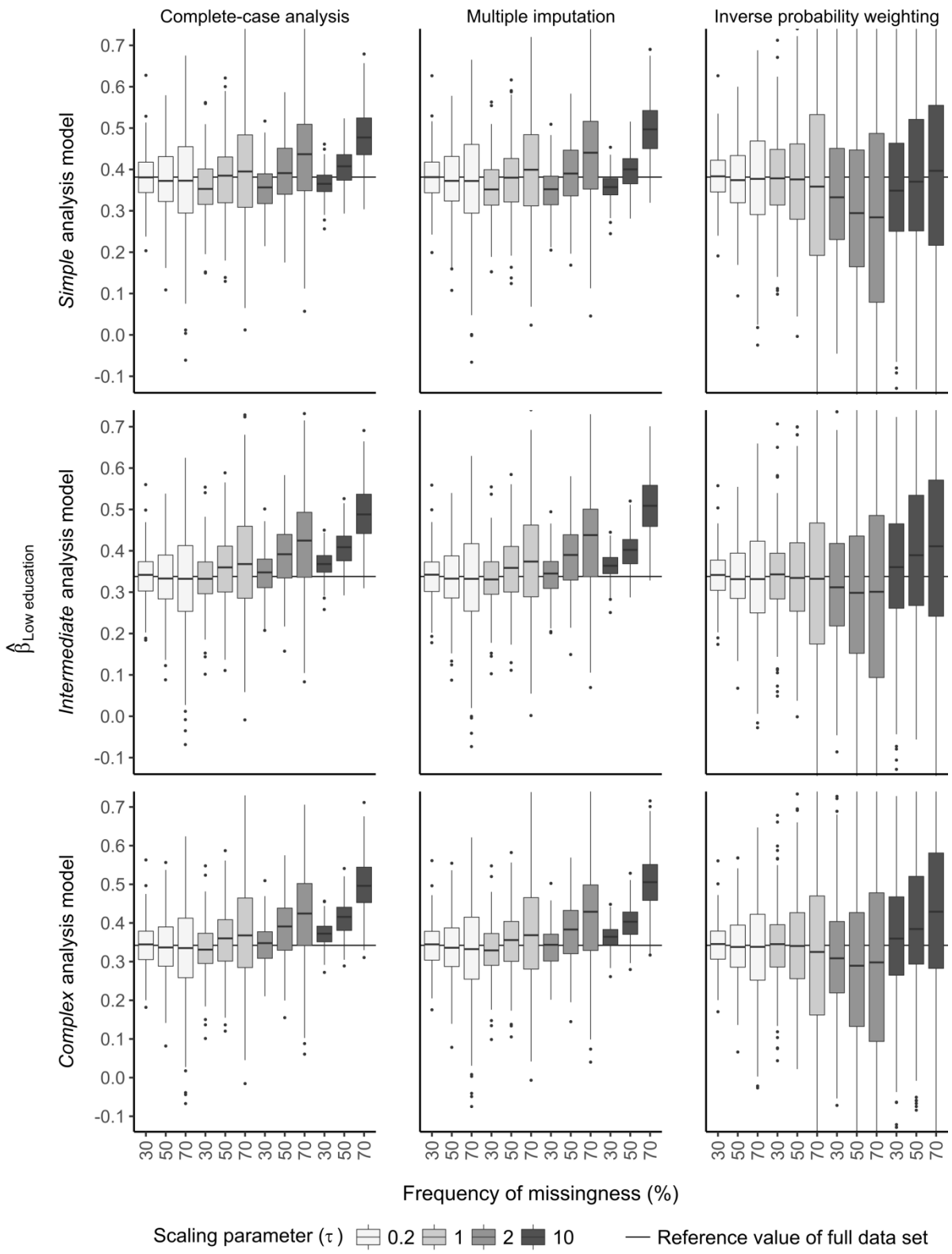

Figure 4.18 Boxplots of low educational level estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table

4.1). Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

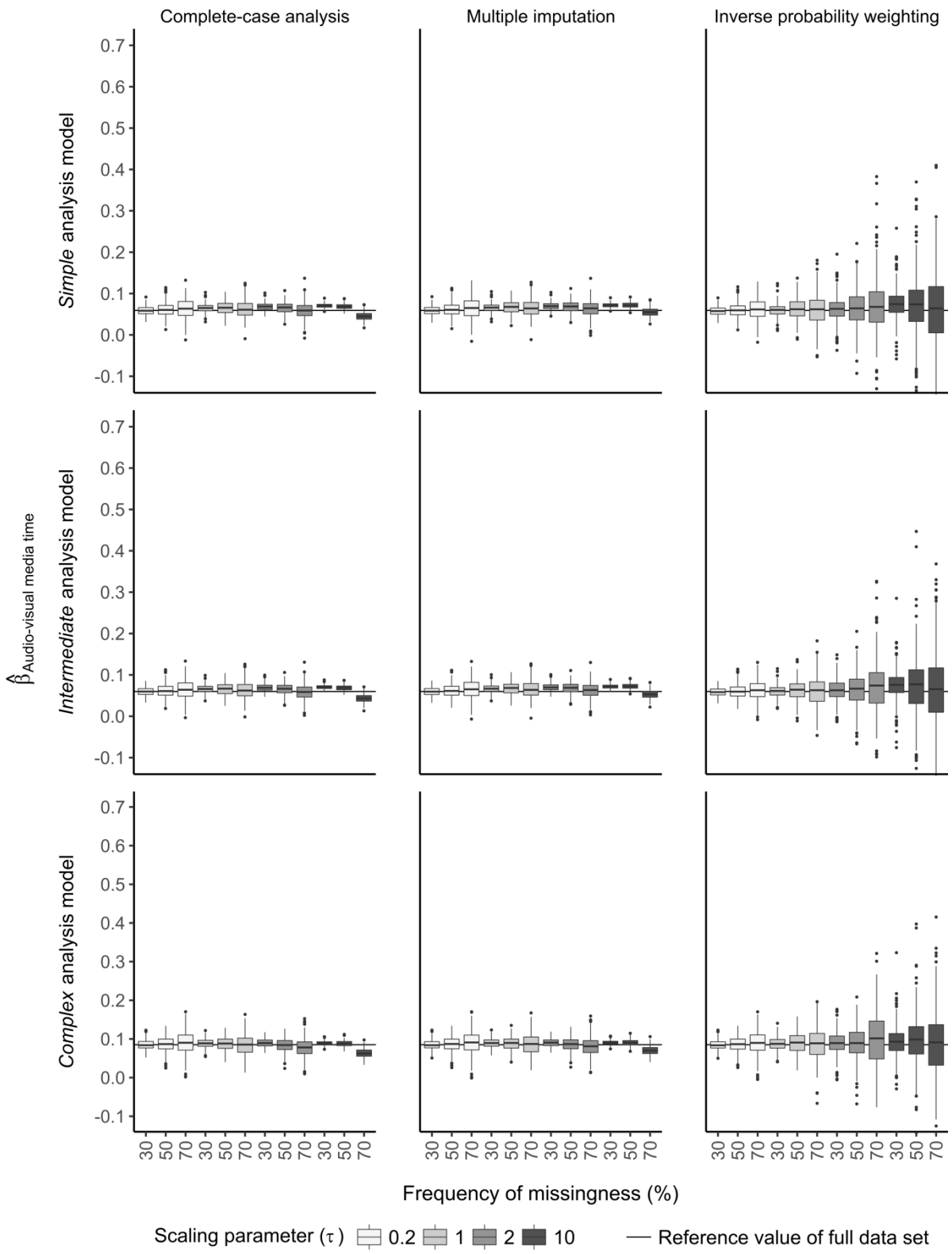

Figure 4.19 Boxplots of audio-visual media time (h/d) estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table

4.1). Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

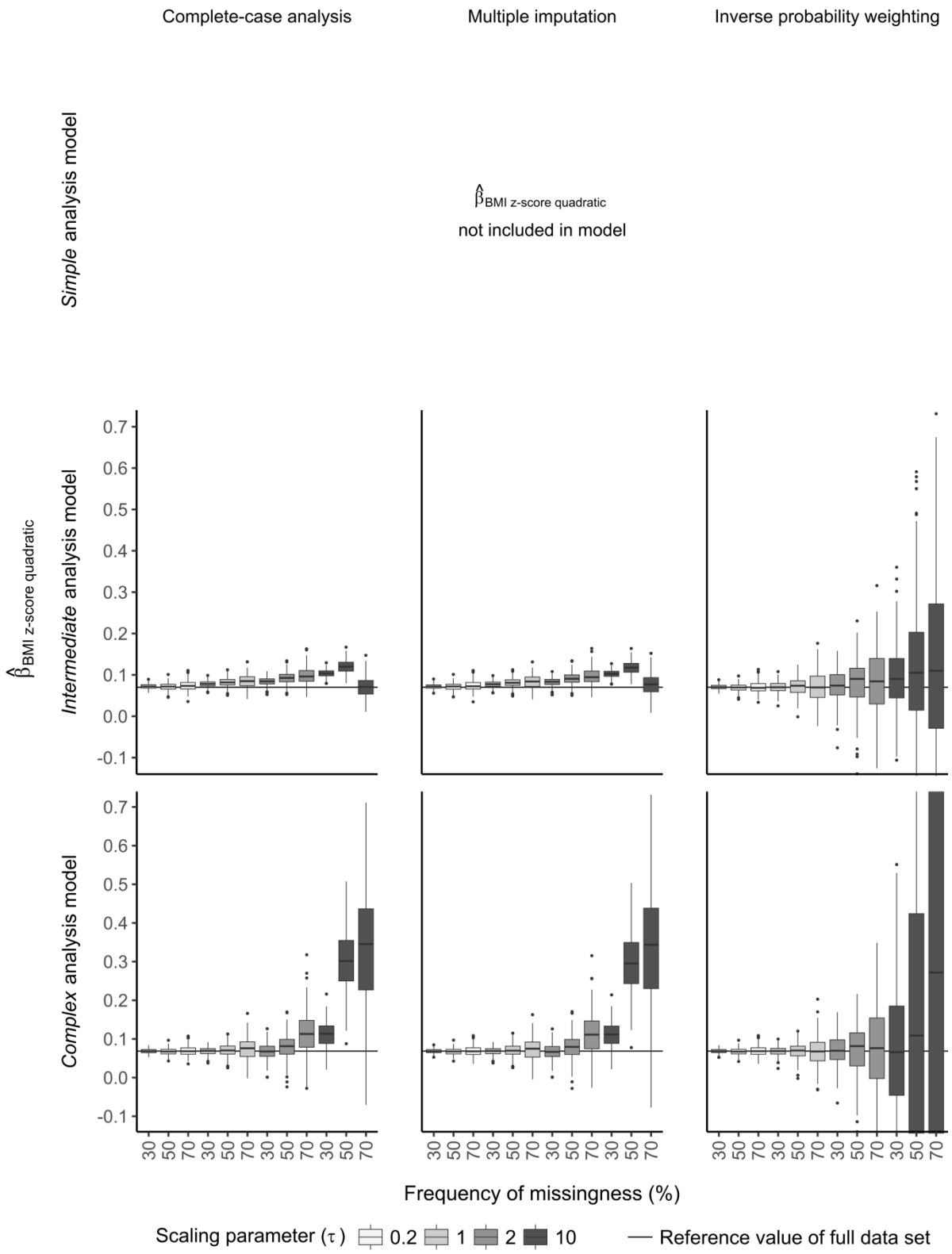

Figure 4.20 Boxplots of BMI (quadratic) estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table

4.1). Boxplots are grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

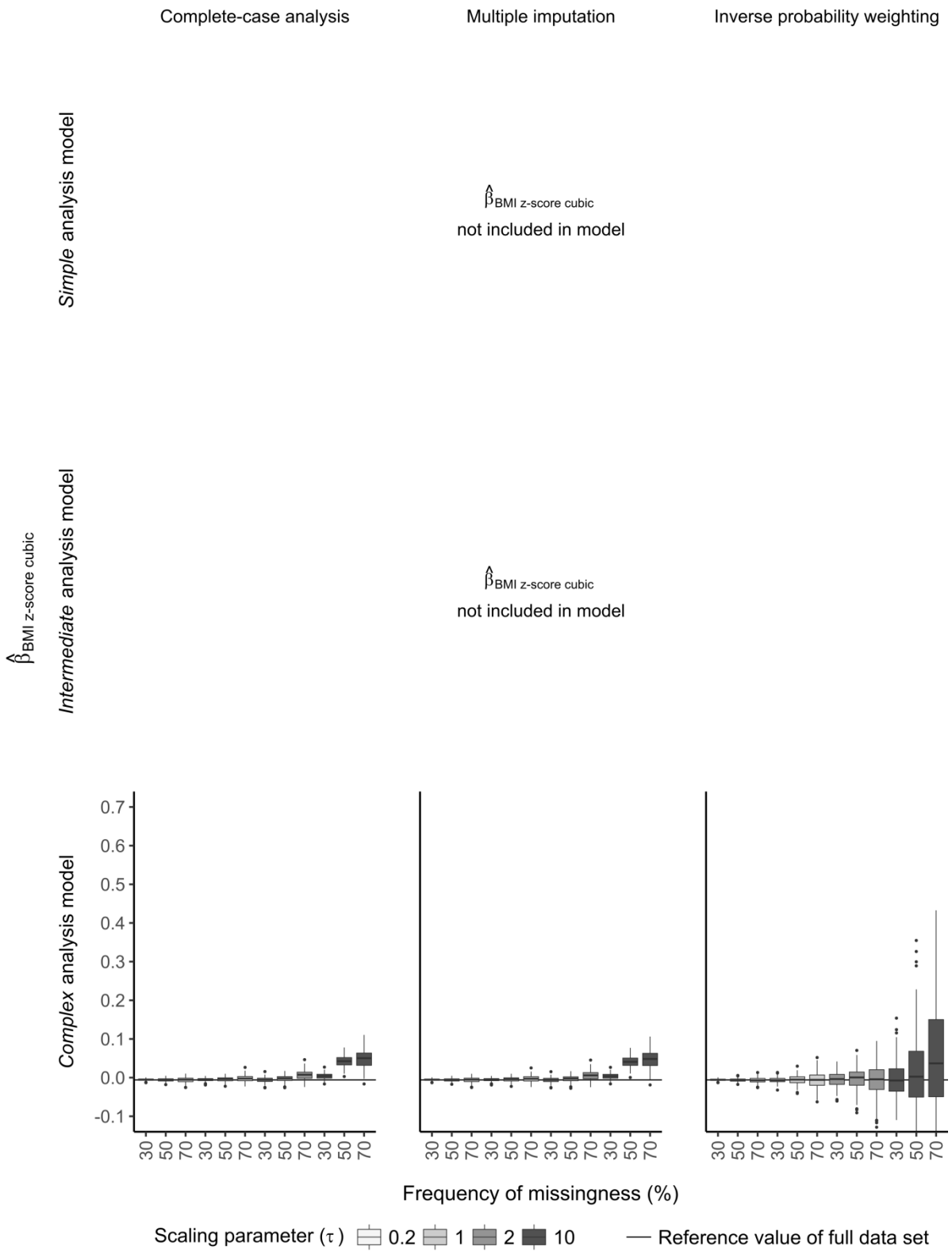

Figure 4.21 Boxplots of BMI (cubic) estimates of regression models with simulated zHOMA as the dependent variable and with simulated missingness depending on zBMI. Columns depict different missing data methods (from left to right: CCA, MI IPW), rows depict different analysis models (from top to bottom: simple analysis model, intermediate analysis model, complex analysis model). In each panel the horizontal solid line represents the estimate in the full data set (cf. Table 4.1). Boxplots are

grouped by different values of $\tau \in \{0.2, 1, 2, 10\}$, indicated by grey shading, and by different frequencies of missingness of 30%, 50% or 70%, arranged in ascending order on the x-axis.

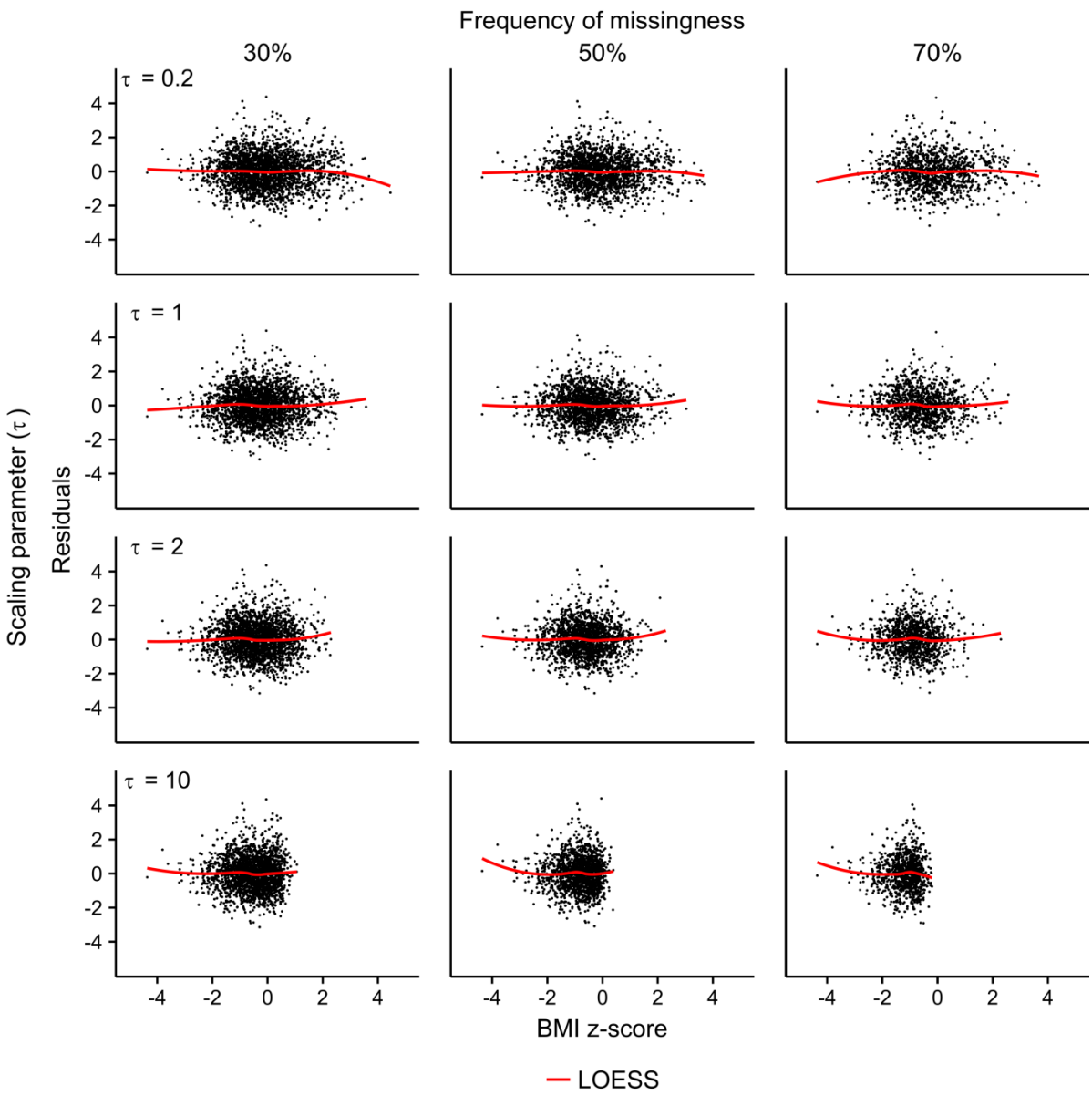

Figure 4.22 Residual plots for the *intermdiate analysis model* using zHOMA as dependent variable and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

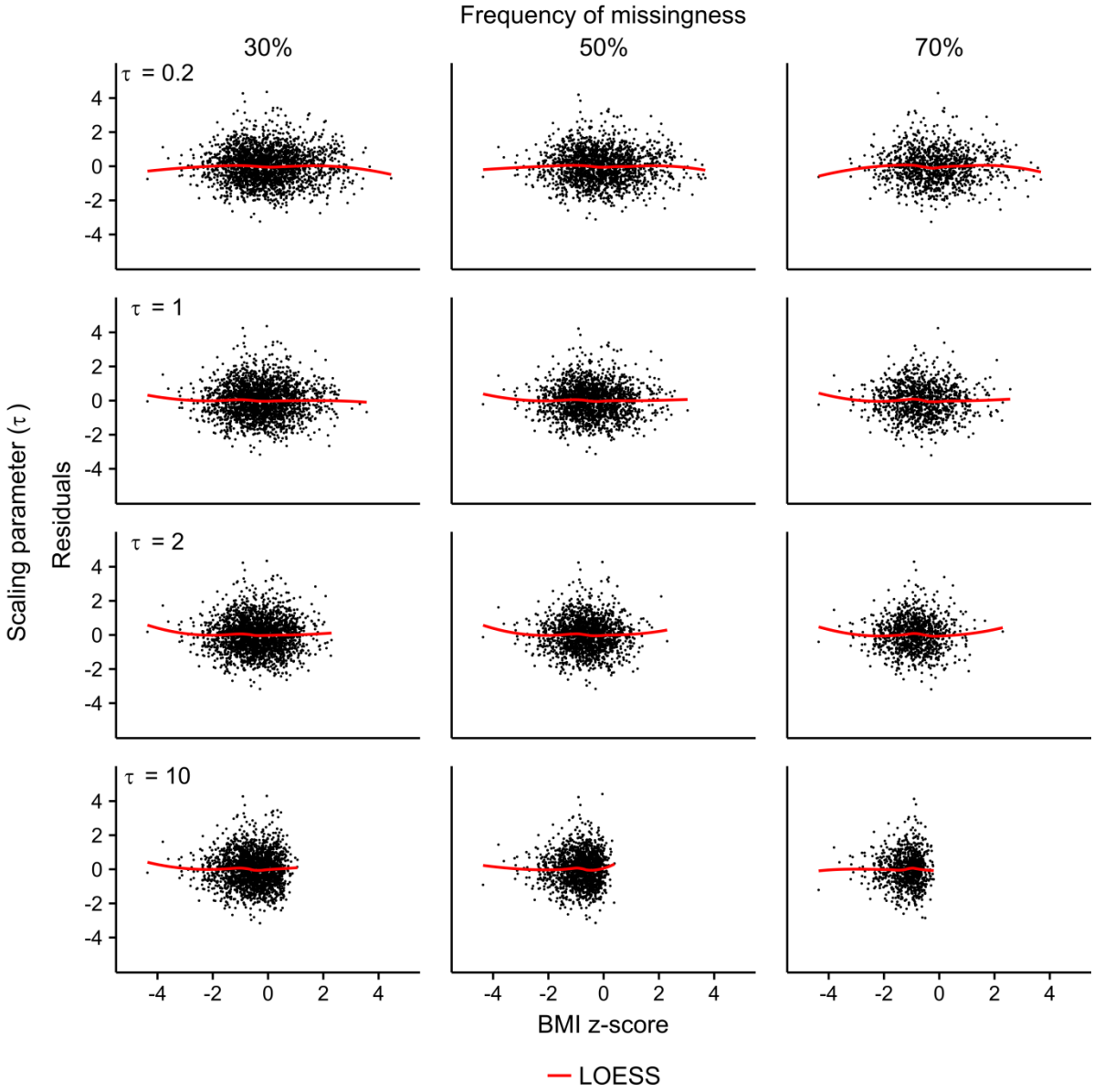

Figure 4.23 Residual plots for the *complex analysis model* using zHOMA as dependent variable and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

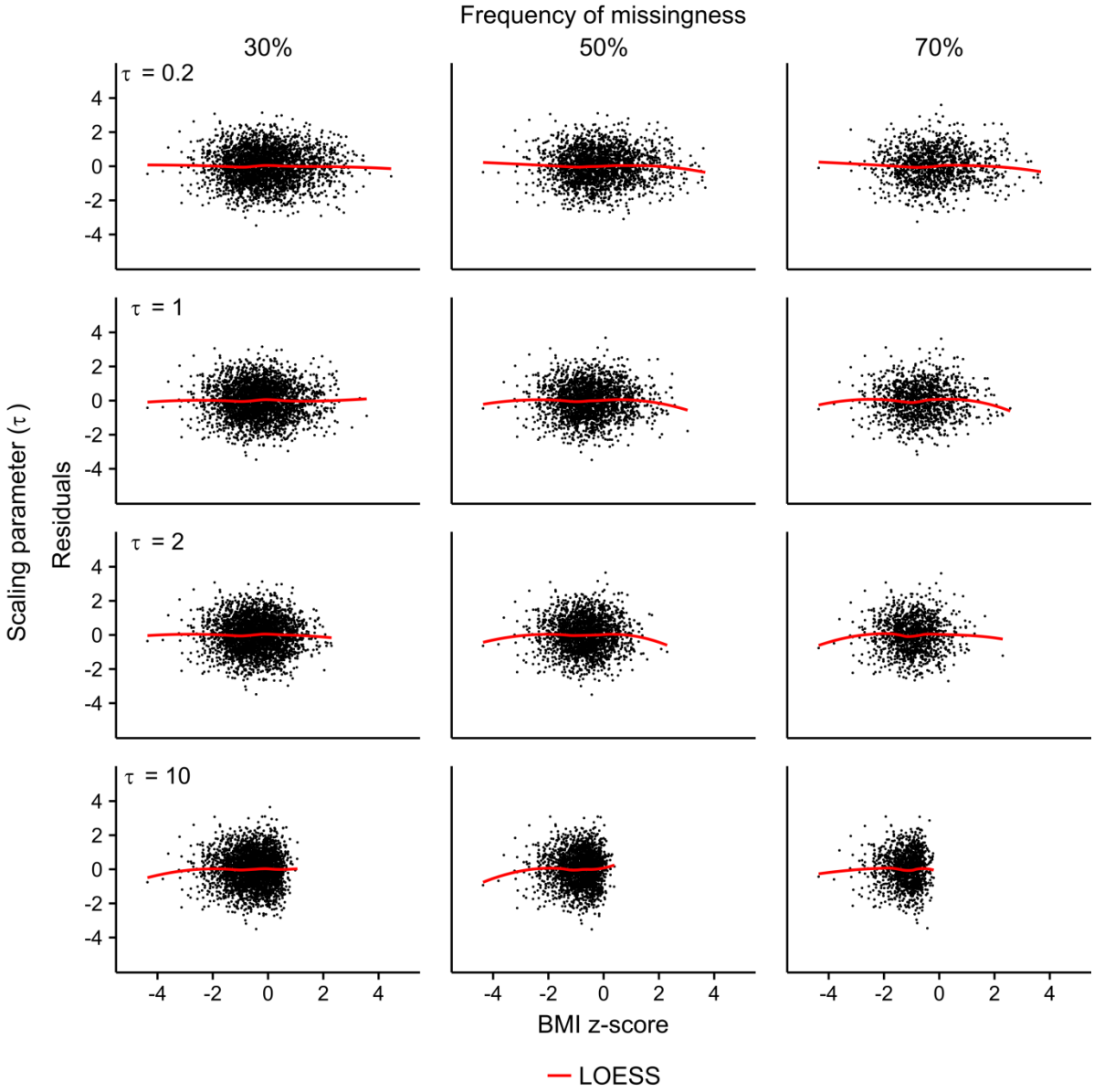

Figure 4.24 Residual plots for the *intermdiate analysis model* using simulated zHOMA as dependent variable and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

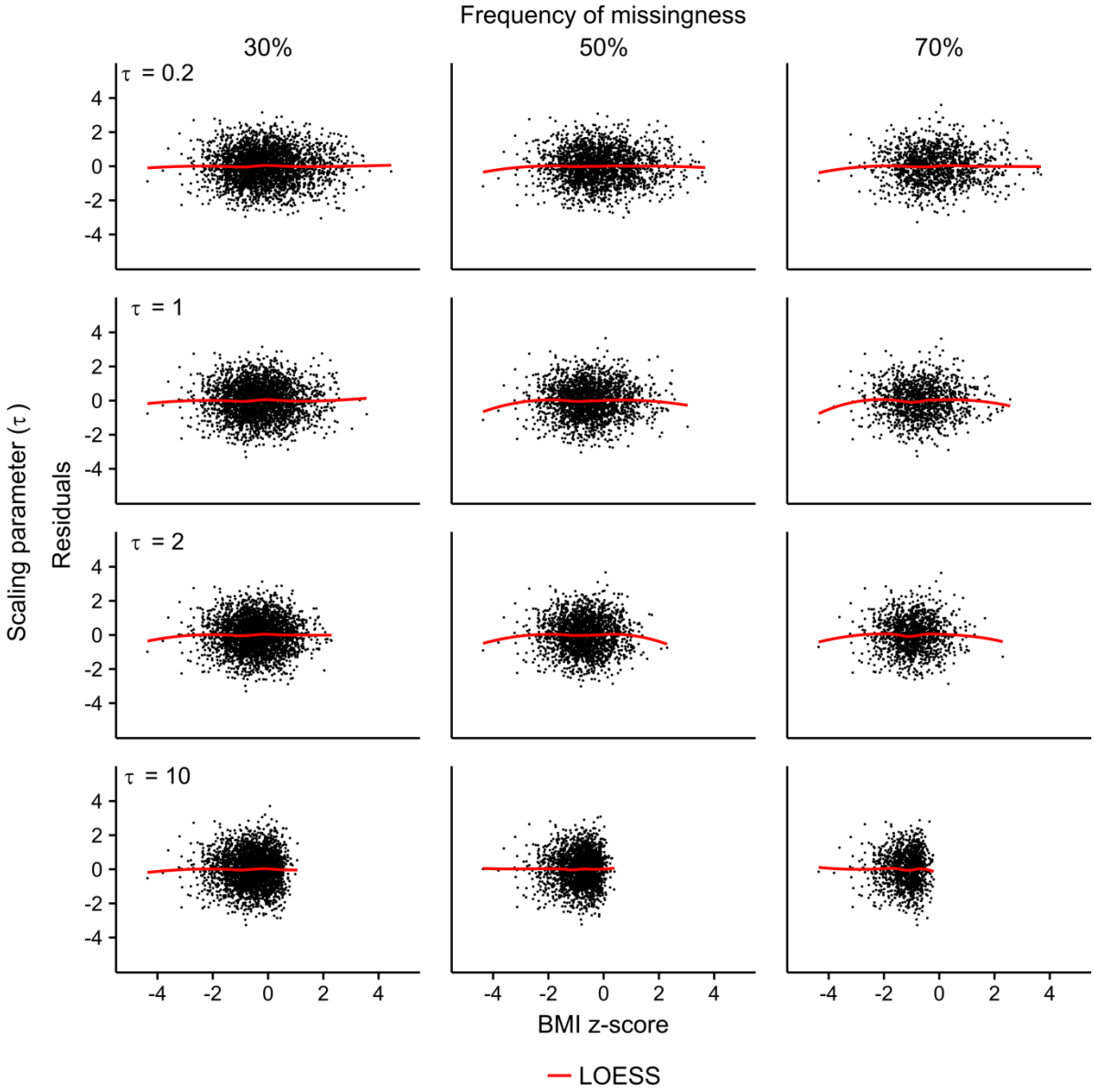

Figure 4.25 Residual plots for the *complex analysis model* using simulated zHOMA as dependent variable and zBMI for each frequency of missingness (30%, 50% or 70%) and scaling parameter $\tau \in \{0.2, 1, 2, 10\}$. Locally weighted scatterplot smoothing (LOESS, red line)

5 Study invitations with envelopes made from recycled paper do not increase likelihood of active responses or study participation in the German National Cohort

Malte Langeheine, Hermann Pohlabeln, Wolfgang Ahrens, Kathrin Günther, Stefan Rach[4]

Abstract

Objective: We conducted a trial embedded within the German National Cohort comparing the responses to study invitations sent in recycled envelopes of grey color vs. envelopes of white color. We analyzed paradata for reactions to the invitation letters by potential subjects, the duration between mailing date of the invitation and active responses, and study participation.

Results: Grey envelopes only slightly increased the chance of active responses (OR 1.16, 95% CI: 0.83, 1.62) to the invitation letter. Potential study subjects with German nationality (OR 3.75, 95% CI: 2.07, 7.66) and age groups above 50 years (50-59: OR 1.78, 95% CI: 1.19, 2.69; 60-69: OR 2.25, 95% CI: 1.48, 3.43) were more likely to actively respond to the invitation letter. The duration between mailing date of the invitation and active response was not associated with envelope color, sex, nationality, or age. Our trial replicates previous observations that the color of the envelope of a study invitation does not influence the likelihood of an active response or study participation.

Introduction

In recent years, the problem of decreasing response in population-based research has received considerable attention (Galea and Tracy, 2007; Groves, 2006; La Verda and Teta, 2006; Morton et al., 2006; Stang, 2003) and although its implications are still a matter of debate (Groves et al., 2006; Jöckel and Stang, 2013; Lacey and Savage, 2016; Nohr et al., 2006; Rothman et al., 2013; Stang and Jöckel, 2004), there seems to be a consensus that a higher response is generally preferable (Edwards et al., 2009; Lacey and Savage, 2016). A systematic review by Edwards and colleagues (Edwards et al., 2009) reported that already some low-level characteristics of the delivery (e.g., recorded or first-class delivery, hand-written addresses) can increase the response to a mailed survey. The color of envelope (brown vs. white) did not influence the response (Edwards et al., 2009) but may do so in other cultural contexts. We compared the response to grey vs. white envelopes that we used for invitations to a large cohort study in Germany. Grey envelopes are commonly used by German official authorities and we

[4] Langeheine M, Pohlabeln H, Ahrens W, Günther K, Rach S. Study invitations with envelopes made from recycled paper do not increase likelihood of active responses or study participation in the German National Cohort. BMC Research Notes. 2019;12(1):468. doi:10.1186/s13104-019-4510-y.

assumed that a more official character might influence the recipient's attitude towards the contents of the letter. The response might also be influenced the by fact that grey envelopes apparently are made from recycled paper, whereas the paper source is not obvious for white envelopes.

In this trial, embedded in the German National Cohort (GNC, German: NAKO Gesundheitsstudie German National Cohort, 2014), we investigated whether the envelope color of the first invitation influenced the probability of a reply to the invitation, the delay between mailing date and replies, and, finally, the probability of study participation.

Methods

The GNC is a cohort study investigating the causes for the development of major chronic diseases. The baseline examinations are conducted from 2014 to 2019. In 18 regional study centers across Germany, a random sample of the general population including a total of 100,000 women and 100,000 men aged 20–69 years will be examined. Potential study participants are randomly drawn from the regional registries of residents and corresponding contact details are provided to the respective study center. The baseline assessments include an extensive interview and self-completion questionnaires, a wide range of medical examinations and the collection of various biomaterials. Detailed information can be found elsewhere (German National Cohort, 2014).

The recruitment protocol of the GNC includes an invitation letter, followed by up to three reminder letters separated by waiting periods of 14 days. The invitation letter asks potential participants to either return a pre-stamped response letter, e-mail, or to call the study center using a toll-free telephone number. For potential study subjects with known phone numbers, the invitation letter is followed by phone calls, and afterwards up to three reminder letters. The recruitment is controlled by MODYS (Reineke et al., 2018), a dedicated software that schedules recruitment tasks and electronically documents all paradata, that is, detailed data about the recruitment process (e.g., events, attempted and successful contacts with potential subjects).

This trial was conducted in the Bremen study center of the GNC which will recruit a total of 10,000 cohort participants. The trial was restricted to potential subjects without known phone numbers to prevent phone calls of the study center during the waiting period.

For this trial, we planned to send out invitation letters with 1,925 white and 1,925 grey envelopes during 8 consecutive weeks between February and April 2017. With this sample size and an assumed response of approximately 11%, a response change of ± 3 percentage points can be detected with a power of 0.80. In each week letters were sent out on two days (usually Monday and Tuesday, according to the normal mailing schedule of the study) with only white envelopes used on one day and only grey ones on the other. Colors were randomly assigned to weekdays prior to the trial. The number of letters sent out per day varied between 225 and 250. Due to human error white envelopes were sent out on

a “grey day” once, resulting in a final sample size of 2,174 white and 1,595 grey letters. The remaining recruitment adhered to the general recruitment protocol outlined above. We analyzed paradata for reactions by potential subjects during the first waiting period of 14 days (responses by mail, phone, e-mail, or personal contact in the study center) and derived the outcome ‘active response’ (0: not responded vs. 1: responded). For subjects actively responding within the first waiting period, we calculated the duration in days between the mailing of the invitation letter and their response. Whether or not subjects eventually participated in the GNC baseline examination defined the second outcome ‘participation’ (0: not participated vs. 1: participated). In our analyses, we included the variables sex (female vs. male), nationality (German vs. non-German), and age (categories: 20-29, 30-39, 40-49, 50-59, and 60-69 years), as provided by the registry of residents. To adjust for potential differences between weekdays, data from the four months preceding the trial was used to calculate pre-trial baselines for the likelihood of active responses and study participation separately for each weekday letters were sent out. Likewise, pre-trial baselines were calculated for the mean duration to respond to the invitation letter.

Subjects were excluded from further analyses if invitation letters were returned as undeliverable (i.e., subject moved or address turned out to be incorrect; N = 103) or if the paradata included recruitment events before the trial started (e.g., previous invitations sent to wrong addresses; N = 143) or phone calls initiated by the study center (N = 4). Furthermore, one subject was excluded because of missing data on nationality. The resulting analysis group consisted of 3,518 subjects to whom letters with 2,022 white and 1,496 grey envelopes were sent (Figure 5.1).

To estimate associations with the outcomes active response and participation we used logistic regression models adjusted for pre-trial likelihood of active responses and study participation to calculate odds ratios (ORs) and 95% confidence intervals (CIs). For the study participation model, 71 cases were excluded from the analysis (22 cases did not complete recruitment and 49 were not eligible, not capable, or deceased), reducing the analysis group to 3,447 cases. We assessed the association between type of envelope and duration (ORs and 95% confidence intervals) to respond to the invitation letter with a linear regression model adjusted for pre-trial duration to respond to the invitation letter.

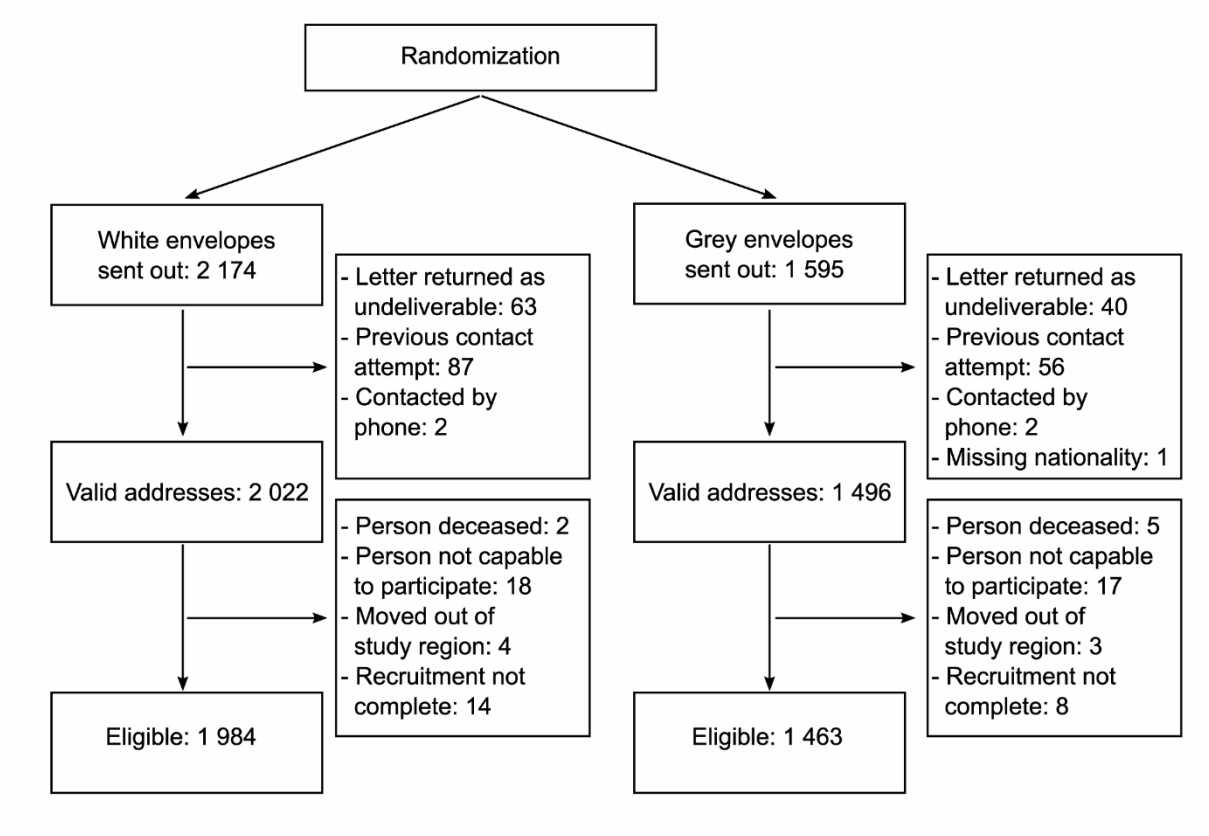

Figure 5.1 Consort flow chart

Table 5.1 Response to Invitation and Participation in the Bremen Study Center of the German National Cohort

	Active response (N = 3,518)					Participation (N = 3,447[d])				
	No		Yes[a]		OR[b] (95% CI)	No		Yes		OR[c] (95% CI)
	n	%	n	%		n	%	n	%	
Envelope color										
White	1,929	95.4	93	4.6	1	1,766	89.0	218	11.0	1
Grey	1,418	94.8	78	5.2	1.16 (0.83, 1.62)	1,308	89.4	155	10.6	0.95 (0.76, 1.19)
Sex[d]										
Female	1,108	93.8	73	6.2	1	1,003	86.2	161	13.8	1
Male	2,239	95.8	98	4.2	0.73 (0.53, 1.02)	2,071	90.7	212	9.3	0.71 (0.57, 0.90)
Nationality[d]										
Non-German	690	98.6	10	1.4	1	644	96.4	24	3.6	1
German	2,657	94.3	161	5.7	3.75 (2.07, 7.66)	2,430	87.4	349	12.6	3.46 (2.31, 5.43)
Age										
20-29	339	96.0	14	4.0	1.41 (0.73, 2.56)	319	93.3	23	6.7	0.89 (0.54, 1.41)
30-39	182	96.8	6	3.2	1.15 (0.43, 2.57)	177	95.7	8	4.3	0.58 (0.26, 1.15)
40-49	1,321	96.9	43	3.2	1	1,233	91.9	109	8.1	1
50-59	887	94.1	56	5.9	1.78 (1.19, 2.69)	784	85.0	138	15.0	1.83 (1.40, 2.40)
60-69	618	92.2	52	7.8	2.25 (1.48, 3.43)	561	85.5	95	14.5	1.67 (1.24, 2.25)
N	3,347		171			3,074		373		

Abbreviations: OR, odds ratio; CI, confidence interval
a Includes subjects who replied by letter, phone call, e-mail or visited the study center in person
b Adjusted for pre-trial likelihood of active responses to invitation letter stratified by weekdays
c Adjusted for pre-trial study participation stratified by weekdays
d Subjects excluded from analysis (N=71): recruitment not complete (N = 22), subject not eligible, not capable, or deceased (N = 49)

Results

Only 171 subjects responded actively to the invitation letter while 373 eventually participated in the study. Grey envelopes only slightly increased the chance of active responses (OR 1.16, 95% CI: 0.83, 1.62, Table 5.1). In contrast to non-Germans, potential study subjects with German nationality (OR 3.75, 95% CI: 2.07, 7.66) and, compared to the age group 40-49, age groups above 50 years (50-59: OR 1.78, 95% CI: 1.19, 2.69; 60-69: OR 2.25, 95% CI: 1.48, 3.43) were more likely to actively respond to the invitation letter and also eventually to participate in the Bremen GNC study. Male and female subjects did not differ in their likelihood to actively respond (OR 0.73, 95% CI: 0.53, 1.02) but males were less likely to eventually participate (OR 0.71, 95% CI: 0.57, 0.90). For both outcomes, we checked whether envelope color interacted with age, sex, or nationality, but no meaningful interactions were found (results not shown).

Table 5.2 Duration between mailing date of the invitation and active response

	mean	Sd	ß[ab] (95% CI)
Envelope color			
White	7.5	3.1	0
Grey	7.4	3.2	0.65 (-0.38, 1.68)
Sex[d]			
Female	7.4	3.0	0
Male	7.4	3.3	0.06 (-0.92, 1.04)
Nationality[d]			
Non-German	8.2	3.2	0
German	7.4	3.2	-0.56 (-2.59, 1.47)
Age			
20-29	7.4	2.9	-0.53 (-2.45, 1.39)
30-39	6.8	2.8	-0.66 (-3.4, 2.08)
40-49	7.5	2.8	0
50-59	8.0	3.6	0.53 (-0.71, 1.77)
60-69	6.8	3.1	-0.74 (-2.03, 0.54)
N	171		

Abbreviations: CI, confidence interval

a Includes subjects who replied by letter, phone call, e-mail or visited the study center in person

b Adjusted for pre-trial likelihood of active responses to invitation letter stratified by weekdays

The duration between mailing date of the invitation and active response was not associated with envelope color, sex, nationality, or age (mean duration white envelopes 7.5 days vs. grey envelopes 7.4 days, Table 5.2).

Discussion

Our trial replicates the observation by Edwards et al. (Edwards et al., 2009) that the color of the envelope of a study invitation does not significantly influence the likelihood of an active response or study participation. An update of the two meta-analyses relevant to our study (Analyses 20.1. and 20.2. in Edwards et al. (2009)) resulted in only slightly decreased odds ratios and no changes to the authors' original conclusions (Figure 5.2 and Figure 5.3). Furthermore, our data confirm previous reports from the pretest of the GNC indicating that subjects with a foreign background are less likely to participate (Reiss et al., 2014; Winkler et al., 2014). It should be noted, that the execution of this trial was eased by the utilization of the MODYS software for recruitment, in which all measures of interest were routinely recorded. We would therefore advocate for the routine collection of paradata that would greatly facilitate the assessment of new trials or periodic replications of previous trials testing the effects of low-level or technical characteristics of recruitment schemes. In addition to dedicated

software, however, collecting detailed paradata routinely of course also requires extra effort and diligence from the recruiting personnel, but once available, they offer opportunities for new insights on the recruitment process that would not be available without (Langeheine et al., 2017; Reineke et al., 2018).

Limitations

It is not clear, however, whether the low response observed here generalizes to GNC as a whole since this trial is based only on a small sample from only one study center. Additionally, the sample in this trial excluded subjects with known phone numbers and phone contacts are known to have a positive effect on the response (Stang et al., 2005; Winkler et al., 2014).

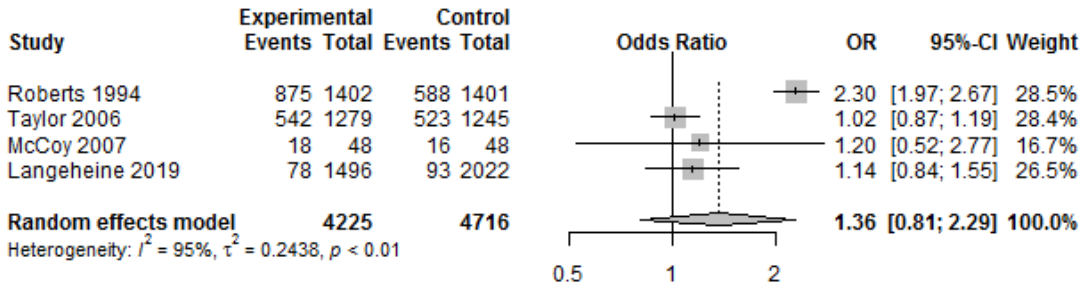

Figure 5.2 Comparison of non-white vs. white envelope, outcome first response

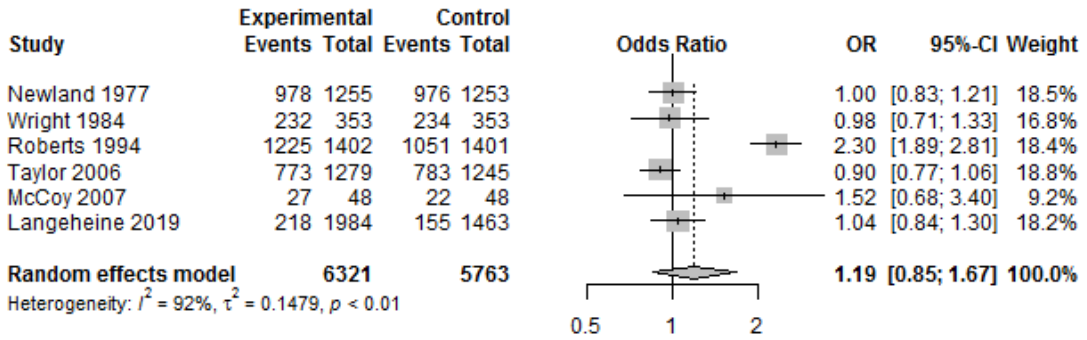

Figure 5.3 Comparison of non-white vs. white envelope, outcome final response

6 Why were there three? – Determinants of the presence of an intimate partner during face-to-face interviews

Richard Preetz, Malte Langeheine[5]

Abstract

This study analyses determinants of the presence of an intimate partner during face-to-face interviews. Based on theoretical assumptions about opportunity structure, social control, social support, and companionship, we investigated partner presence using data from the first wave of the German Family Panel (pairfam). Descriptive results revealed that an intimate partner was present in every seventh interview. Multivariate results using separate logistic regression models for the presence of the female (n = 3,272) and the male partner (n = 2,348) revealed that the opportunity structure, such as the couple's living arrangements or their employment status, had the greatest influence on the presence of both female and male partners. Gender differences existed for social control, social support and companionship. The results suggest that partner presence could most easily be prevented by taking into account the opportunity structure.

Introduction

Face-to-face interviews are influenced by aspects of both interviewer (Cannell et al., 1977) and respondent (Billiet and Davidov, 2008) as well as by characteristics of the interview situation (Johnson, 2014) that may result in systematically over- or underestimated measures of survey variables (Krumpal, 2013). A special influence on the interview situation is the presence of a third person during the interview. Earlier research found that the third person most likely to be present during an interview is the intimate partner (Aquilino, 1993; Hartmann, 1994; Lander, 2000). Previous research examining the presence of the intimate partner detected differential measurement outcomes of partnership related questions, such as marital dissolution, self-reported levels of marital conflict (Aquilino, 1993) and partnership satisfaction (Mohr, 1986).

Determinants influencing the presence of the intimate partner that are discussed in the sparse literature include structural (Reuband, 1992) as well as motivational reasons or personality characteristics of the respondent or the partner (Hartmann, 1994; Lander, 2000; Reuband, 1987). Empirical research found that non-working status was associated with a higher chance of partner presence during the interview and socioeconomic status was inversely associated with partner presence (Aquilino, 1993). Sex differences between the interviewer and the respondent were associated with a higher chance of partner presence when a woman was interviewed by a man, and

[5] Preetz R, Langeheine M. Why were there three? Determinants of the presence of an intimate partner during face-to-face interviews. Survey Methods: Insights from the Field. 2017. doi:https://doi.org/10.13094/SMIF-2017-00007.

this was interpreted as social control (Hartmann, 1994). Although the intimate partner is most likely to be the third person present during the interview, an examination of relationship quality and presence of the intimate partner is lacking in the literature. However, the presence of the intimate partner during the interview is highly likely to bias the results when both partners are concerned about potential survey questions; hence the respondent tends towards socially motivated misreporting (Aquilino et al., 2000; Krumpal, 2013). The intimate partner could, however, also assist the respondent emotionally or provide important information during an interview. Partner presence may thus also have positive effects (Reuband, 1984). Both are to be expected in the German Family Panel (pairfam) (Huinink et al., 2011) because the major research issues of pairfam are couple dynamics and partnership stability. Pairfam addresses the development of partnership relations, including the quality of the relationship, issues of labour division, the internal power distribution and the stability of the relationship in a multi-actor design, in which information on both the respondent and the partner are available.

The aim of the present study is to extend the existing literature on the presence of the intimate partner and investigate the association between characteristics of the intimate relationship and the presence of the intimate partner. This paper does not investigate the potential influences of partner presence on the response behaviour of the respondent. The paper starts with brief theoretical considerations, introduces data and method, presents the results and discusses the findings in the conclusion.

Theoretical background and hypotheses

The opportunity structure should influence the partner's presence. The employment status or living arrangements are crucial for partner presence (Hartmann, 1994). The chance that the partner will be present should be more likely if both partners live in the same household. In contrast, living in separate households should be a higher barrier for partner presence because it is more difficult to schedule a joint appointment. The partner might also be present because he is not employed and has time to spend with the respondent.

H1: The partner is more likely to be present if both partners live in the same household.

H2: The partner is more likely to be present if the partner is not employed.

In line with social control theory, people try to gain control over situations to influence these in a beneficial way, i.e. seek desired events and avoid undesirable ones (Thompson, 1981). Accordingly, the partner should be more likely to be present if the partner is keen to know the respondent's answers in an interview, so as to influence the response behaviour in a beneficial way. A more positive interpretation of this potential motive is curiosity. But it is reasonable to assume that control

motivation is a characteristic of mistrust and thus of a disturbed relationship. Control motivation should therefore be more likely if the partner is dominant. Another reason for a higher social control motivation might be the sex constellation between the respondent and the interviewer (Hartmann, 1994). Especially in situations in which the respondent and the interviewer are of opposite sex the partner might be jealous and curious what will happen during the interview. Furthermore, jealousy could be a reason for a higher control motivation if the respondent is very attractive (Guerrero et al., 2004; Lander, 2000).

H3: The higher the partner's control motivation, the higher the chance of the partner's presence during the interview.

In line with the social support theory, a further assumption is that the interview situation might be regarded as a stressful event for the respondent, in particular, if a respondent is being interviewed for the first time (Lakey and Cohen, 2000). The respondent might feel uncomfortable with this unknown situation. Therefore the partner might be more likely to be present to support the respondent during the interview. This should occur in partnerships that are characterised by strong reciprocal support. Furthermore, the presence of the partner as a source of support might be desired if the respondent has low self-worth and is uncertain how to answer survey questions. (Lander, 2000) mentioned that a higher educational level of the partner could be an indicator of social support. Thus, a higher educational level of the partner should increase the chance of his presence.

H4: The higher the partner´s social support, the higher the chance that the partner will be present during the interview.
H5: The lower the respondent´s self-worth, the higher the chance that the partner will be present.
H6: A higher educational level of the partner increases the chance of partner presence.

In contrast to the partner's control motivation, companionship might influence the partner's presence during the interview (Aquilino, 1993). A higher level of trust and intimacy within the relationship is likely to enhance the partner's presence. Partner presence might be more likely if the partners share their secrets and their leisure time activities. Companionship is expected to be a characteristic of good relationships.

H7: The greater the intimacy between the two partners, the higher the chance that the partner will be present during the interview.

Data and method

Pairfam enables extended empirical research on issues of couples and family development with a focus on couple dynamics and partnership stability, childbearing, parenting and child development, and intergenerational relationships (Huinink et al., 2011). Although pairfam's main focus is couples and family development, the paradata (information about the process of data collection) allow methodological questions to be investigated, such as the presence of the intimate partner during the interview. The annual survey is based on a cohort-sequence design and started in 2008 with 12,402 randomly selected respondents from the three birth cohorts 1991-93, 1981-83, and 1971-73. Population registers were used as the sampling frame.

To assess determinants of partner presence during the interview, we use data from the first wave of the pairfam study collected in 2008/09. The overall response rate for the first wave was 36.9%. This relatively low response rate is common for German surveys and does not result in a nonresponse bias (Huinink et al., 2011). Our analysis includes all three birth cohorts and is restricted to persons with an intimate heterosexual relationship. In pairfam, respondents are interviewed with a Computer-Assisted Personal Interview (CAPI). Sensitive questions or questions about relationship quality which might be regarded as sensitive in the presence of other household members are asked by Computer-Assisted Self-Administered Interview (CASI). The interviewer´s laptop is handed over to the respondent and the respondent fills out the questions autonomously (Huinink et al., 2011).

Operationalisation

The binary dependent variable partner presence during the interview (partner is present: 1; partner is not present: 0) is based on paradata recorded by the interviewer after each successful interview. The interviewer had to specify whether the partner, children or other persons were present during the interview or not.

Respondent´s attractiveness (very unattractive: 1; very attractive: 7) is based on the interviewer´s assessment. All other information including information about the partner´s employment status and educational level were provided by the respondent.

Opportunity structure is indicated by the respondent´s and the respondent´s partner employment status (respondent/partner was not employed (including unemployment, retirement and parental leave): 1; full- or part-time employment: 0) as well as the cohabitation status (partners live in the same household: 1; partners live in separate households: 0).

Dominance of the partner, the interviewer-respondent sex constellation and the above-mentioned attractiveness of the respondent were considered as indicators of control motivation. Dominance within the respondent-partner dyad (low: 1; high: 5) was measured via an adapted and shortened version of the Network of Relationships Inventory (NRI) (Furman and Buhrmester, 1985). The

respondent reported how often his/her partner gets his/her way when they can´t agree on something and how often their partner makes them do things in his/her way. These questions were answered in CASI mode. The interviewer-respondent sex constellation is coded 1 if the respondent and interviewer have the opposite sex and 0 for the same sex.

Social support is operationalised via dyadic coping within the respondent-partner dyad and the self-worth of the respondent. The measurement of dyadic coping (low: 1; high: 5) is based on the supportive dyadic coping of the partner scales from the Dyadic Coping Questionnaire (FDCT-N) (Bodenmann, 2000). The respondent answered several items about how often and in which way the partner supports him or her in stressful situations. The respondent's self-worth (low: 1; high: 5) was measured via three items from an adaptation of Rosenberg's Self-Worth scale (Rosenberg, 1965). Items for dyadic coping and self-worth were retrieved via CASI mode. The respondent´s and their partner´s educational level were measured via the International Standard Classification of Education (ISCED) (Schneider, 2008). A variable comparing the educational level of the two partners was included in the analysis. This variable indicates whether the respondent and the partner have the same educational level or whether the partner or the respondent is more highly educated.

Companionship was operationalised by the measure of intimacy within the respondent-partner dyad (low: 1; high: 5) and was also collected via CASI mode. The respondent answered items from the Intimacy Network of Relationships Inventory (NRI) (Furman and Buhrmester, 1985) on how often he shares his feelings and secrets with his partner.

Relationship duration in months, respondent´s age, partner´s age, and birth cohort were included as control variables in the analysis. Table 6.1 lists detailed information about the scales and items used. Table 6.2 provides descriptive statistics of the independent variables.

Table 6.1 Scales and operationalisation of the independent variables

Variable	Questionnaire	Operationalisation
Cohabitation	Cohabitation status	Both partners live in the same household vs. both partners live in separate households
Respondent non-working	Respondent´s employment status	Respondent non-working vs. respondent working
Partner non-working	Partner´s employment status	Partner non-working vs. partner working
Dominance (CASI)	How often does [name of current partner] get his/her way when you can't agree on something?	1: Never 5: Always
Interviewer-respondent sex constellation	Constellation of interviewer sex and respondent sex	Opposite sex vs. same sex
Attractiveness	How attractive do you find the respondent?	1: very unattractive 7: very attractive
Dyadic coping (CASI)	[name of partner] lets me know that he/she understands me.	1: Never 5: Always
Self-worth (CASI)	Sometimes I believe that I'm worthless.	1: Not at all 5: Absolutely
Education	Respondent´s education and partner´s education (ISCED)	1:Respondent more educated 2: Partner more educated 3: Same education
Intimacy (CASI)	How often do you tell [name of current partner] what you're thinking?	1: Never 5: Always
Relationship duration		Months
Respondent´s age		Years
Partner´s age		Years
Birth cohort		1: 1991-1993 2: 1981-1983 3: 1971-1973

Statistical analysis

Descriptive statistics were calculated to reveal the prevalence of partner presence in the first wave of pairfam. As social support and control motivation may have different effects, depending on the sex of the partner (Lander, 2000), separate logistic regression models for the presence of the male partner (n = 3,272) and presence of the female partner (n = 2,348) with average marginal effects (AMEs) were computed to allow the stepwise inclusion of variables and the comparison of coefficients across models (Long and Freese, 2014). As interviews were nested within interviewers (minimum interviews per interviewer was one, maximum was 85) and rates of partner presence differ by interviewer, we accounted for a potential clustering with robust standard errors using interviewers as clusters.

Table 6.2 Descriptive statistics of the independent variables

	Percent	Mean	SD
Living in the same household (1 = yes; 0 = no)	77.30		
Respondent not employed (1 = yes; 0 = no)	19.34		
Partner not employed (1 = yes; 0 = no)	13.56		
Dominance of female partner (1-5)		3.21	.614
Dominance of male partner (1-5)		2.92	.650
Interview sex constellation (1 = opposite; 0 = same)	49.34		
Attractiveness of the male respondent (1-7)		5.44	1.302
Attractiveness of the female respondent (1-7)		5.61	1.392
Dyadic coping of male partner (1-5)		4.21	.701
Dyadic coping of female partner (1-5)		4.28	.622
Self-worth of male respondent (1-5)		4.27	.661
Self-worth of female respondent (1-5)		4.08	.782
Educational level of couples			
Woman higher educated	23.26		
Man higher educated	30.71		
Both have same education	46.03		
Intimacy (1-5)		3.87	.774
Relationship duration (months)		95.48	72.11
Couple's age			
Woman older	17.05		
Man older	72.74		
Both have same age	10.21		
Birth cohort			
1991-1993	6.98		
1981-1983	41.60		
1971-1973	51.42		
N	5620		

Results

Descriptive Results

Table 6.3 shows descriptive results regarding partner presence as a function of birth cohort, living arrangement and respondent's sex. In 14.88% of interviews, an intimate partner was present. Stratified by the three birth cohorts, partner presence was 3.83% for the 1991-1993 birth cohort, 15.74% for the 1981-1983 birth cohort and 15.67% for the 1971-1973 birth cohort (Chi^2 = 40.63, p = 0.00). A higher degree of institutionalisation of the relationship was associated with a higher percentage of partner presence (living in separate households = 5.02% vs. cohabitation = 15.81% vs. marriage = 18.65%; Chi^2 = 132.65, p = 0.00). Partner presence was higher for male respondents than for female respondents (19.25% vs. 11.74%, Chi^2 = 60.95, p = 0.00).

Table 6.3 Percentage of partner presence stratified by birth cohort, living arrangement and respondent's sex

Respondents with partner	14.88
Birth cohort	
1991-1993	3.83
1981-1983	15.74
1971-1973	15.67
Living arrangement	
Living in separate households	5.02
Cohabitation	15.81
Marriage	18.65
Respondent's sex	
Male (female partner present)	19.25
Female (male partner present)	11.74
N	5620

Source: pairfam wave 1 2008/2009; calculations by the authors

Multivariate Results

Table 6.4 shows stepwise logistic regression models for the probability of the presence of the female partner during the interview of a male respondent. Model 1a includes only the opportunity structure

and the control variables. If both partners lived in the same household the probability of the presence of the female partner was 17.2% higher than for couples living in separate households. Furthermore, the chance that the female partner would be present during the interview increased by 15.5% if the male respondent was unemployed. If the female partner was unemployed the chance increased as well. In Model 2a variables for social control were added. The results show a significant positive effect for an opposite interviewer-respondent sex constellation: female partners were more often present (AME = 5.5%) during the interview if a male respondent was interviewed by a woman. Surprisingly, the more attractive the male respondent was, the lower the chance was that the female partner was present. Differences in the educational level were not associated with partner presence.

In Model 3a variables for social support were added. The lower the self-worth of the male respondent, the higher was the probability of the presence of the female partner. Intimacy added in Model 4a had a significant positive effect of 3.1% on the presence of a female partner. A female partner was more often present during the interview if the intimacy between the male respondent and the female partner was high. Nagelkerkes R^2 slightly increases from Model 1a to Model 4a (0.11 vs. 0.13).

Table 6.4 Logistic regression with AMEs (with logits in brackets) for the presence of a female partner during an interview of a male respondent

	Presence of female partner							
	Model 1a		Model 2a		Model 3a		Model 4a	
	AME	Logit	AME	Logit	AME	Logit	AME	Logit
Opportunity structure								
Living in the same household								
No	ref		ref		ref		ref	
Yes	.172*	(1.652)	.173***	(1.671)	.174***	(1.685)	.174***	(1.686)
Male respondent not employed								
Employed	ref		ref		ref		ref	
Not employed	.155*	(.887)	.147***	(.855)	.143***	(.836)	0.141***	(.831)
Female partner not employed								
Employed	ref		ref		ref		ref	
Not employed	.083*	(.529)	.082***	(.530)	.082***	(.531)	.081***	(.527)
Social Control								
Dominance of female partner			-.009	(-.062)	-.011	(-.076)	-.014	(-.097)
Interviewer-respondent sex constellation								
Male respondent – male interviewer			ref		ref		ref	ref
Male respondent – female interviewer			.055*	(.375)	.055*	(.379)	.055*	(.382)
Attractiveness of male respondent			-.018*	(-.127)	-.018*	(-.122)	-.017*	(-.117)
Social support								
Dyadic coping of female partner					.020	(.137)	.004	(.025)
Self-worth of male respondent					-.022+	(-.154)	-.025*	(-.177)
Educational level of respondent and partner								
Same educational level					ref		ref	
Male respondent has higher education					.003	(.024)	.003	(.020)
Female partner has higher education					-.003	(-.021)	-.004	(-.026)
Companionship								
Intimacy							.031**	(.219)
Control variables								
Relationship duration	-.001	(-.001)	-.001	(-.001)	-.001	(-.001)	-.001	(-.000)
Age of respondent and partner								
Same age	ref		ref		ref		ref	
Male respondent older	-.024	(-.172)	-.026	(-.190)	-.026	(-.190)	-.025	(-.185)
Female partner older	.009	(.058)	.005	(.037)	.005	(.035)	.005	(.034)
Birth cohort								
1991-1993	ref		ref		ref		ref	
1981-1983	.132*	(1.328)	.132*	(1.124)	.131*	(1.334)	.135*	(1.141)
1971-1973	.128*	(1.296)	.130*	(1.335)	.129*	(1.317)	.133*	(1.394)
N	2,348		2,348		2,348		2,348	
Nagelkerke R^2	.110		.123		.126		.131	

Source: pairfam wave 1 2008/2009; calculations by the authors; average marginal effects (AME); logit coefficients in parenthesis; robust standard errors for interviewer ID; ***p<0.001; **p<0.01; *p<0.05; +p<0.1

Table 6.5 Logistic regression with AMEs (with logits in brackets) for the presence of a male partner during an interview of a female respondent

	Presence of male partner							
	Model 1b		Model 2b		Model 3b		Model 4b	
	AME	Logit	AME	Logit	AME	Logit	AME	Logit
Opportunity structure								
Living in the same household								
No	ref		ref		ref		ref	
Yes	.079***	(.987)	.079***	(.982)	.081**	(1.018)	.081*	(1.020)
Female respondent not employed								
Employed	ref		ref		ref		ref	
Not employed	.058***	(.540)	.057***	(.527)	.055**	(.516)	0.055	(.514)
Male partner not employed								
Employed	ref		ref		ref		ref	
Not employed	.060**	(.512)	.059*	(.502)	.056*	(.485)	0.056	(.482)
Social Control								
Dominance of male partner			.004	(.042)	.004	(.042)	.0040	(.042)
Interviewer-respondent sex constellation								
Female respondent – female interviewer			ref		ref		ref	ref
Female respondent – male interviewer			.011	(.108)	.011	(.113)	.011	(.115)
Attractiveness female respondent			-.003	(-.030)	-.002	(-.020)	-.002	(-.020)
Social support								
Dyadic coping of male partner					.020*	(.195)	.021*	(.212)
Self-worth of female respondent					-.018*	(-.181)	-.018*	(-.177)
Educational level of respondent and partner								
Same educational level					ref		ref	
Female respondent has higher education					.001	(.013)	.002	(.015)
Male partner has higher education					.009	(.094)	.010	(.094)
Companionship								
Intimacy							-.003	(-.035)
Control variables								
Relationship duration	.001	(.001)	.001	(001)	.001	(.001)	.001	(.001)
Age of respondent and partner								
Same age	ref		ref		ref		ref	
Female respondent is older	.002	(.024)	.002	(.018)	.002	(.022)	.002	(.024)
Male partner is older	.001	(.011)	.001	(.006)	.001	(.003)	.001	(.002)
Birth cohort								
1991-1993	ref		ref		ref		ref	
1981-1983	.038	(.357)	.038	(.358)	.039	(.368)	.038	(.362)
1971-1973	-.008	(-	-.008	(-.086)	-.006	(-.073)	-.007	(-.083)
N	3,272		3,272		3,272		3,272	
Nagelkerke R^2	.057		.058		.064		.064	

Source: pairfam wave 1 2008/2009; calculations by the authors; average marginal effects (AME); logit coefficients in parenthesis; robust standard errors for interviewer ID; ***p<0.001; **p<0.01; *p<0.05; +p<0.1

Table 6.5 shows the results for the presence of the male partner. In comparison to the presence of the female partner, the AMEs for the opportunity structure variables have the same sign but are smaller in size. In contrast to the presence of the female partner, the analysis did not reveal any significant AMEs for social control. In particular, if a female respondent was interviewed by a male interviewer the chance that the male partner would be present did not increase (Model 2b). The lower the self-worth of the female respondent, the higher was the chance of the male partner´s presence. Contrary to the presence of the female partner, dyadic coping (Model 3b) was positively associated with the presence of the male partner. Compared with the presence of the female partner Model 4b does not reveal an effect of intimacy on the presence of the female partner. Nagelkerkes R^2 increased marginally from Model 1b to Model 4b (0.057 vs. 0.064).

Discussion and Conclusion

This paper extends the existing literature on the presence of an intimate partner during an interview. In contrast to most previous studies, the focus was on the causes of the partner's presence and not on the potential influences on the response behaviour of the respondent. While previous studies on partner presence focused on structural reasons, we mainly investigated the association between characteristics of relationship dynamics and personality characteristics on the one hand and partner presence on the other hand. Descriptive findings revealed that in one of seven interviews the intimate partner was present. Furthermore, female partners were more often present than male partners.

Our multivariate results show that the opportunity structure was the most important determinant of partner presence during the interview (H1, H2). If the respondent or the partner was not employed, the probability of the partner's presence increased. These results seem plausible for unemployed partners, but are surprising for unemployed respondents. Apparently, unemployed respondents did not choose specific time slots for the interview in which their partner was definitely not at home.

In the case of social control, the constellation of a male respondent and a female interviewer was associated with an increased presence of the female partner during the interview. One possible explanation is the control motivation of the female partner as a consequence of jealousy. Psychological studies have revealed that women are more affected by emotional jealousy than men, which might explain the absence of significant effects for men (Edlund et al., 2006). Furthermore, the respondent's attractiveness was only associated with the presence of the female partner. However, the respondent's attractiveness did not have the expected negative effect on female presence. Nevertheless, the mechanism of jealousy and attractiveness is unclear due to an important limitation of pairfam: there is no information on the attractiveness of the interviewer as a potential rival that would help to uncover the mechanism between attractiveness and jealousy (Buunk and Dijkstra, 2004). Thus, the hypothesis of social control (H3) was only partially confirmed.

In line with the social support theory, the presence of the female and the male partner was associated with the respondent's self-worth (H5). One possible explanation for this association might be that if the self-worth is low, an interview might raise the respondent's stress level. In this situation, the partner could act as a balance due to his/her support and could lower the respondent's stress level. In this analysis, intimacy was only associated with the presence of the female partner, which partly confirmed Hypothesis 7.

A limitation of pairfam is that no information exists on whether the respondent initiated the presence of the partner or whether it was the partner himself/herself who initiated his/her presence. As Reuband (1987) showed, the partner himself/herself initiates his presence in only 22% of interviews. In all other cases, the presence of the partner is initiated by the respondent himself/herself, or jointly by the respondent and the partner, which should especially occur when the respondent is looking for social support. Furthermore, the paradata provided by pairfam do not include any information about the length of the partner's presence during the interview. In addition, partner presence might bias the information on couple dynamics and partnership stability, so that the measurement error induced by the presence of the partner might explain the association between the explanatory variables used in this analysis and partner presence. But in pairfam all variables related to couple dynamics and partnership stability are obtained via CASI mode. Although CASI should minimize the bias induced by the partner's presence, this is not fully guaranteed (Lavrakas, 2008).

Factors (opportunity structure and social support) associated with the presence of the female partner might reflect traditional role models. Female partners in general might stay at home more often and also might be the most knowledgeable respondent in terms of specific partnership related questions, such as relationship duration, and be able to provide support during the interview.

In summary, the results illustrate that partner presence is mainly related to the opportunity structure. Intimacy and social support are also associated with partner presence, but it is not expected that these will induce a bias. As it has been reported in the literature that the partner's presence may result in socially motivated misreporting (Aquilino, 1993; Aquilino et al., 2000; Hartmann, 1994; Tourangeau and Yan, 2007), it might be easiest to prevent partner presence by taking into account the opportunity structure.

Interviewers could emphasize the importance to conduct the interview without a third person present when making the interview appointment. In pairfam, however, this is not the case because the chance that further respondents like the intimate partner participate in pairfam is higher when they are present during the interview. This benefits pairfam's multi-actor design. But for surveys that collect data from only one respondent, it might be effective to prevent third person presence if the interviewer emphasizes the importance before the interview is conducted.

Analysing the reasons for partner presence in a cross-sectional design is a first step. Based on these results future research could investigate the development of the frequency of partner presence over subsequent waves. It would be interesting to analyse whether social control and social support matter in subsequent waves, since control motivation might decrease over time after the partner was present during the interview. The need for social support might decrease as well, as learning effects should occur if the interview situation has been repeated several times.

7 General discussion

Population-based cohort studies will always be prone to some form of error (Bethlehem et al., 2011). This thesis investigates aspects related to the Total Study Error framework, in particular those concerning participation in population-based cohort studies. The following chapter begins by summarizing the main findings and discussing them in the context of previous research. Thereafter, selected methodological considerations related to Chapters 2, 3, and 5 are discussed.

7.1 Main findings

In line with previous research, Chapter 2 revealed that IDEFICS/I.Family study participants who were less healthy and disadvantaged were more likely to drop-out (Behr et al., 2005; Howe et al., 2013; Lange et al., 2014; Melton et al., 1993; Vinther-Larsen et al., 2010). The analysis of selection effects on cross-sectional exposure-outcome associations however revealed few effects on point estimates of the association between childhood overweight and social position, as well as the association between childhood overweight and adherence to key behaviors of a healthy lifestyle when restricting the full sample at baseline (T0) to participants of the first follow-up examination (T1) or the second follow-up examination (T3). This again is in line with previous research. Despite a potential selection bias due to nonresponse or attrition, previous research reported biased prevalences due to but concluded that associations remained unbiased (Behr et al., 2005; Howe et al., 2013). This also included studies using external data like registry data to obtain information on non-participants (Carter et al., 2012; Greene et al., 2011; Harald et al., 2007; Nohr et al., 2006; Søgaard et al., 2004; Van Loon, 2003).

In Chapter 3, we examined the effect of extended recruitment efforts on participation and therefore categorized individuals into late and early respondents. The concept of early and late respondents has a “long and rich history” (Sigman et al., 2014, p. 652) in study research, dating back to the late 1940s (Baur, 1947). Due to time constraints and administrative costs of studies, the majority of research on early and late respondents limited their analysis to excluding late respondents and observing how this affected the demographic composition of a sample. Fewer studies investigated whether a prolonged recruitment effort is penalized at follow-up (Haring et al., 2009). Our analyses using the German IDEFICS/I.Family data from the baseline and follow-up examination augments the existing literature as we explored the interaction between recruitment efforts at baseline and follow-up. The findings revealed a complex relationship between recruitment effort and attrition, whereby attrition was mainly driven by individuals who were late respondents (late recruitment) at baseline and early respondents (early recruitment) at follow-up. The risk of these individuals to dropout at follow-up was more than 1.5 times higher than of the other groups. Furthermore, drop-out codes taken from the MODYS paradata revealed that late respondents at baseline and early respondents at follow-up were characterized by a higher chance of being coded “no-contact”. Additional analysis suggested that invalid phone numbers at follow-up led to a higher chance of dropping-out.

In Chapter 4, we illustrated with a simulation how different modeling choices can affect our conclusions even when a complete-case analysis is in principle valid. We based our study on empirical data from the IDEFICS study and simulate only the missingness mechanism assuming different association strengths and different frequencies of missingness. In each scenario, we investigatedd the performance of three different analysis models using complete-case analysis as well as multiple imputation and inverse probability weighting as methods to correct for missing data. Model misspecification induced considerable bias when data contain missing values, even in a scenario where an ideal complete-case analysis is known to be consistent. This bias equally affected multiple imputation, and to a lesser extent, at the cost of precision, inverse probability weighting which requires a correct weighting model. In our example, basic model diagnostics were sufficient to alert us to the misspecification of the simple analysis model with regard to the functional form of the exposure; this was detectable even for the most extreme missingness mechanisms.

The trial conducted in Chapter 5 replicated previous observations (Edwards et al., 2009) that sociodemographic variables rather than the color of the envelope of a study invitation (grey color vs. envelopes of white color) influence the likelihood of an active response or study participation. Furthermore, the time span between the mailing date of the invitation and active response was not associated with envelope color, sex, nationality, or age. Our update of the two meta-analyses relevant to our study [Analyses 20.1. and 20.2. in Edwards et al. (2009)] using the results presented in Chapter 5 resulted in only slightly decreased odds ratios and no changes to the authors' original conclusions.

Chapter 6 confirmed previous findings to the effect that the opportunity structure, that is, employment status and living arrangement, is the most important correlate of partner presence during the interview (Lander, 2000). In contrast, characteristics of the intimate relationship, which were previously not investigated in detail, proved to be a less important associate with partner presence.

7.2 Methodological considerations

In this section we discuss further methodological considerations related to Chapters 2 to 5, and refer to additional analyses and tables that were not included in the journal articles due to space restrictions or editorial decisions. Broadly summarized, the topics presented here cover statistical issues, limitations of the MODYS paradata, and the possibilities to investigate nonresponse/attrition.

7.2.1 Multilevel structure in IDEFICS/I.Family

The IDEFICS/I.Family data exhibit a clustered structure in which children (siblings) are hierarchically nested in families, schools/Kindergarten, communities and, finally, within countries. Hence, not accounting for the hierarchical data structure is likely to violate the assumption of statistical independence. The method generally used to deal with clustered data structures is multilevel modelling (Snijders and Bosker, 1999).

In Chapter 3, we used kindergarten/school-level as the second level in the multilevel models. This affected the association between parents' education and being categorized as late at baseline, resulting in confidence intervals including one (see Table 3.1). This, however, was not surprising given the German school system. In Germany, parents cannot freely choose primary schools for their children, but rather have to select a primary school within their school districts. This results in a certain level of socio-economic segregation in primary schools caused by residential socio-economic segregation (Voigtlander et al., 2010). In the second-level model, the social position of parents was therefore likely to cluster on the kindergarten/school-level and mask or obscure the influence of socioeconomic variables at the first level.

7.2.2 Sensitivity analysis of non-independence for siblings

In both Chapter 2 and Chapter 3, families with more than one child were included in the analysis. As families are units with a strong relationship, measures of different children from a single family are likely to be correlated, for example, measures from a single respondent over time, or measures within a country as pointed out in the foregoing section. We therefore checked the data for the number of siblings and carried out sensitivity analyses where we excluded single children from the analysis. In Chapter 2, we observed a small variation in the overall percentage of siblings across time points (percentage of siblings at baseline (T0) 11%; at follow-up (T1) 12%). The percentage of children within each country was also relatively stable across baseline (T0) and follow-up (T1). However, between countries, the percentage of siblings strongly varied at baseline and follow-up. For example, at baseline (T0), the percentage ranged from 4% in Estonia to 20% in Sweden, and this pattern was similar at follow-up (T1) (Table 7.1).

Table 7.1 The number and percentage of families with more than one child participating in IDEFICS/I.Family for each country separately for baseline and follow-up

	Baseline (T0)				Follow-up (T1)			
	No siblings		Siblings		No siblings		Siblings	
Spain	1 303	0.88	177	0.12	998	0.86	164	0.14
Hungary	2 300	0.92	196	0.08	879	0.88	115	0.12
Germany	1 720	0.86	288	0.14	985	0.86	167	0.14
Cyprus	2 011	0.95	100	0.05	1 535	0.96	61	0.04
Estonia	1 590	0.96	60	0.04	1 240	0.96	46	0.04
Belgium	1 590	0.84	294	0.16	1 009	0.85	184	0.15
Italy	1 927	0.86	314	0.14	1 304	0.85	237	0.15
Sweden	1 391	0.80	357	0.20	1 103	0.79	287	0.21
All	13 832	0.89	1 786	0.11	9 053	0.88	1 261	0.12

To assess whether measures of different children from a single family impacted the results, we first selected one child from each family. We then calculated a random intercept logistic regression model. We repeated this procedure n = 100 times to obtain mean odds ratios and corresponding confidence intervals for each predictor for each sample, using country as a second-level variable. The highest observed difference between the logistic regression including all the children from each family and the mean odds ratios for the random sample with one child from each family was small (Table 7.2). Furthermore, the confidence intervals for the mean odds ratios were narrow, indicating no variance of the 100 estimates using a single child from each family for each estimate. The result seemed to be plausible as the overall percentage of siblings in the data was rather small. As the odds ratios of the logistic regression model including all children from each family and the mean odds ratios for the 100 samples of children did not differ substantially, we decided not to account for clustering within the families.

Table 7.2 Odds ratios with 99% confidence intervals for cohort attrition in IDEFEICS/I.Family compared to mean odds ratios with 99% confidence intervals of n = 100 random samples with only one child of each family included in the analysis

	ORa	99% CI		Mean OR	99% CI	
Time						
First follow-up (T1)	ref.			ref.		
Second follow-up (T3)	2.62	2.32	2.96	2.63	2.63	2.63
Sex of child[b]						
Male	ref.			ref.		
Female	0.99	0.93	1.07	1.00	1.00	1.00
Age child (years)[c]	1.05	1.02	1.07	1.04	1.04	1.04
Weight status child[c]						
Normal weight	ref.			ref.		
Overweight	1.17	1.05	1.29	1.16	1.16	1.16
Obese	1.18	1.03	1.36	1.22	1.22	1.22
Compliance score of child[c]	0.84	0.81	0.87	0.84	0.84	0.84
Compliance score parent(s)[c]	0.93	0.88	0.98	0.93	0.93	0.93
Mother's age (years)[c]	0.98	0.97	0.99	0.98	0.98	0.98
Weight status parents[c]						
No parent overweight	ref.			ref.		
At least one parent overweight	1.05	0.96	1.14	1.04	1.04	1.04
Missing	1.23	1.08	1.40	1.21	1.21	1.21
Migration background[b]						
No	ref.			ref.		
Partly	1.13	1.00	1.28	1.14	1.14	1.14
Full	1.41	1.21	1.63	1.40	1.40	1.40
Missing	1.36	0.96	1.91	1.40	1.39	1.40
Educational level[b]						
Low education	1.49	1.27	1.74	1.45	1.45	1.45
Medium education	1.19	1.10	1.29	1.20	1.20	1.20
High education	ref.			ref.		
Missing	1.12	0.77	1.62	1.17	1.17	1.17
Number of adults in household[c]						
One	ref.					
Two	0.89	0.78	1.02	0.91	0.91	0.91
Three	0.98	0.82	1.18	1.00	1.00	1.00
Four or more	1.02	0.80	1.30	1.02	1.02	1.03
Missing	1.02	0.72	1.44	1.01	1.01	1.01
Siblings aged < 18 years[c]						
Yes	0.90	0.83	0.99	0.91	0.91	0.91
No	ref.			ref.		
Missing	0.94	0.67	1.34	0.95	0.95	0.96
Region[b]						
Intervention	ref.			ref.		
Control	1.05	0.98	1.13	1.08	1.08	1.08
N	25 932			100 (repetitions)		

Abbreviations: OR, odds ratio; CI, confidence interval; ref., reference category

[a] Adjusted for country.

[b] Time invariant variable using information from baseline (T0).

[c] Time variant variable using information from baseline (T0) and first follow-up (T1).

Table 7.3 Odds ratios (with 95% CIs) for attrition at the follow-up as predicted by a model with interaction compared to odds ratios (with 95% CIs) including only the first child of each family

	OR	95% CI		OR	95% CI	
Grouped baseline/follow-up recruitment effort						
Early baseline recruitment & early follow-up recruitment	ref.			ref.		
Early baseline recruitment & late follow-up recruitment	1.13	0.81	1.58	1.29	0.90	1.84
Late baseline recruitment & early follow-up recruitment	1.65	1.19	2.28	1.70	1.21	2.39
Late baseline recruitment & late follow-up recruitment	1.11	0.81	1.52	1.12	0.80	1.57
Telephone phone numbers at baseline						
One landline number	ref.			ref.		
Two landline numbers	1.01	0.49	2.09	1.04	0.47	2.32
One mobile number	1.51	1.01	2.27	1.57	1.02	2.42
Landline and mobile	0.76	0.54	1.06	0.75	0.52	1.07
Sex child (baseline)						
Male	ref.			ref.		
Female	1.14	0.91	1.42	1.15	0.91	1.45
Age child at baseline (years)[a]	1.22	1.11	1.33	1.24	1.12	1.37
Weight status child (baseline)						
Under weight	0.74	0.51	1.08	0.74	0.49	1.12
Normal weight	ref.			ref.		
Overweight	1.29	0.95	1.75	1.18	0.86	1.63
Mother's age at baseline (years) [a]	0.96	0.94	0.98	0.96	0.94	0.99
Weight status parents (baseline)						
No parent overweight	ref.			ref.		
At least one parent overweight	1.01	0.76	1.34	1.06	0.78	1.43
Missing	1.30	0.86	1.96	1.38	0.89	2.14
Migration background (baseline)						
No	ref.			ref.		
Partly	1.04	0.72	1.50	1.14	0.77	1.68
Full	1.02	0.75	1.38	1.09	0.78	1.51
Missing	1.84	0.33	10.35	1.88	0.33	10.69
Income level (baseline)						
Low income	ref.			ref.		
Medium income	0.83	0.62	1.10	0.80	0.59	1.07
High income	0.95	0.6	1.52	0.94	0.57	1.54
Missing	1.53	1.02	2.29	1.56	1.01	2.39
Educational level (baseline)						
Low education	ref.			ref.		
Medium education	0.85	0.63	1.15	0.87	0.63	1.19
High education	0.60	0.39	0.90	0.61	0.39	0.95
Missing	1.26	0.51	3.15	1.83	0.66	5.07
Number of adults in household (baseline)						
One	ref.			ref.		
Two	0.95	0.66	1.38	1.09	0.74	1.62
Three	1.05	0.58	1.90	1.18	0.63	2.22
Four or more	1.68	0.68	4.18	1.58	0.61	4.07
Missing	1.18	0.31	4.46	1.28	0.34	4.85
Siblings aged <18 years (baseline)						
Yes	ref.			ref.		
No	1.28	0.98	1.67	1.22	0.92	1.61
Missing	1.62	0.38	6.94	1.52	0.35	6.60
Moved						
Stayer	ref.			ref.		
Mover within study region	1.36	0.85	2.18	1.37	0.84	2.25
Mover outside study region	2.98	0.95	9.37	2.58	0.80	8.34
Missing	4.49	0.41	49.1	4.18	0.38	46.26
All	1 691			1 501		

Odds ratios are adjusted for sex child (baseline), age child at baseline (years), weight status child (baseline), mother's age at baseline (years), weight status parents (baseline), migration background (baseline), income level (baseline), educational level (baseline), number of adults in household (baseline), siblings aged <18 years (baseline)

[a] Odds ratios correspond to a 1-year increase in age

In Chapter 3, 89% of the families in the data had one child only. From the 1 501 families included in the analysis, 150 families had two eligible children and 17 had three or more eligible children. The resulting mean children size for families was 1.13. We selected the first child of each family and compared the results of multilevel models with and without siblings, using kindergarten/school-level as the second level. Although we observed some deviations of the odds ratios, for example for partly migrants and movers outside (Table 7.3), the results for the analysis with and without siblings did not differ substantially. Hence, as in Chapter 2, we decided not to account for clustering within the families in Chapter 3.

7.2.3 Dealing with missing data

Throughout Chapter 2 and Chapter 3, we used the missing-indicator method (Groenwold et al., 2012) to handle missing covariate data. This is a popular and simple to implement method, whereby missing observations are assigned a fixed value and a dummy variable indicating whether a value for a variable is missing is added to the multivariable model (Groenwold et al., 2012). By doing so, statistical power is maintained as each subject can be included in the analysis regardless of whether some values are missing or not, and all the information that is available is retained. The method has however been criticized as it is likely to result in biased estimates due to the fact that the estimated association of exposure and outcome is the weighted average of two associations, namely, the association estimated within the subgroup of subjects with complete data and that within the subgroup of subjects with incomplete data. As exposures are likely to be related in nonrandomized studies, this relation may not be adequately adjusted for within the subgroup of subjects with incomplete data (Groenwold et al., 2012).

To verify whether the missing-indicator method affected estimates, we carried out sensitivity analyses for the main results of Chapter 2 and Chapter 3, and compared the estimates obtained using the missing-indicator method with those obtained using complete-case analysis (Table 7.5 and

Table ***7.7***). The latter is the standard regression model method to handle missing data and entails the exclusion of subjects with a single missing in a covariate from the analysis (Allison, 2007). To facilitate comparison, we compared confidence intervals attained using the missing-indicator method to those attained using complete-case analysis.

Table 7.4 The number and percentage of missing data in the analysis of cohort attrition carried out in Chapter 2

	Missing	%	Not missing	%
Cohort attrition	0	0.0	25 932	100
Time	0	0.0	25 932	100
Sex of child	0	0.0	25 932	100
Age child (years)	0	0.0	25 932	100
Weight status child	0	0.0	25 932	100
Compliance score of child	0	0.0	25 932	100
Compliance score of parent(s)	0	0.0	25 932	100
Mother's age (years)	0	0.0	25 932	100
Weight status of parents	2 800	10.8	23 132	89.2
Migration background	596	2.3	25 336	97.7
Educational level	523	2.0	25 409	98.0
Number of adults in household	2 725	10.5	23 207	89.5
Siblings aged < 18 years	2 425	9.4	23 507	90.6
Region	0	0.0	25 932	100
Country	0	0.0	25 932	100

As illustrated by the descriptive results in Table 7.4, the proportion of missing data on the covariates in the final sample in Chapter 2 was relatively small. We used the missing-indicator method on the covariates weight status of parents (10.8% missing data), migration background (2.3% missing data), educational level (2% missing data), number of adults in household (10.5% missing data), and Siblings aged < 18 years (9.4% missing data). Using complete-case analysis, the sample size was reduced from 25 932 to 20 503. The results however remained quite stable as the differences between the estimates of the model using the missing-indicator method and that using complete-case analysis were small, except for the estimate Second follow-up (T3) (Table 7.5). Furthermore, the confidence intervals of both methods did overlap. For example, the confidence interval for obese children using the missing-indicator method ranged from 1.03 to 1.36, and that for complete-case analysis ranged from 1.05 to 1.45.

Table 7.5 Odds ratios with 99% confidence intervals for cohort attrition in IDEFEICS/I.Family using the missing-indicator method to handle missing data, compared to odds ratios with 99% confidence intervals using complete-case analysis

	Missing-indicator method			Complete-case analysis		
	OR[a]	(99% CI)		OR[a]	(99% CI)	
Time						
First follow-up (T1)	ref.			ref.		
Second follow-up (T3)	2.62	2.32	2.96	2.78	2.43	3.18
Sex of child[b]						
Male	ref.			ref.		
Female	0.99	0.93	1.07	1.01	0.93	1.09
Age child (years)[c]	1.05	1.02	1.07	1.05	1.02	1.07
Weight status child[c]						
Normal weight	ref.			ref.		
Overweight	1.17	1.05	1.29	1.17	1.04	1.32
Obese	1.18	1.03	1.36	1.24	1.05	1.45
Compliance score of child[c]	0.84	0.81	0.87	0.83	0.80	0.87
Compliance score parent(s)[c]	0.93	0.88	0.98	0.94	0.88	1.00
Mother's age (years)[c]	0.98	0.97	0.99	0.98	0.97	0.99
Weight status parents[c]						
No parent overweight	ref.			ref.		
At least one parent overweight	1.05	0.96	1.14	1.04	0.95	1.14
Missing	1.23	1.08	1.40			
Migration background[b]						
No	ref.			ref.		
Partly	1.13	1.00	1.28	1.12	0.97	1.29
Full	1.41	1.21	1.63	1.48	1.25	1.76
Missing	1.36	0.96	1.91			
Educational level[b]						
Low education	ref.			ref.		
Medium education	1.19	1.10	1.29	1.18	1.08	1.29
High education	1.49	1.27	1.74	1.39	1.16	1.67
Missing	1.12	0.77	1.62			
Number of adults in household[c]						
One	ref.			ref.		
Two	0.89	0.78	1.02	0.87	0.74	1.03
Three	0.98	0.82	1.18	0.92	0.75	1.14
Four or more	1.02	0.80	1.30	1.00	0.76	1.31
Missing	1.02	0.72	1.44			
Siblings aged < 18 years[c]						
Yes	ref.			ref.		
No	0.90	0.83	0.99	0.89	0.80	0.98
Missing	0.94	0.67	1.34			
Region[b]						
Intervention	ref.			ref.		
Control	1.05	0.98	1.13	1.02	0.94	1.10
N	25 932			20 503		

Abbreviations: OR, odds ratio; CI, confidence interval; ref., reference category

[a] Adjusted for country.

[b] Time invariant variable using information from baseline (T0).

[c] Time variant variable using information from baseline (T0) and first follow-up (T1).

In Chapter 3 the following covariates had missing data (Table 7.6): weight status of parents (T0) (13.1% missing data), migration background (T0) (0.5% missing data), income level (T0) (8.8% missing data), educational level (T0) (1.5% missing data), number of adults in household (T0) (3.3% missing data), siblings aged < 18 years (T0) (2.6% missing data), and moved (0.3% missing data).

Table 7.6 The number an percentage of missing data in the analysis of cohort attrition carried out in Chapter 3

	Missing	%	Not missing	%
Cohort attrition	0	0.0	1 691	100
Grouped T0/T1 recruitment effort	0	0.0	1 691	100
Telephone phone numbers at T0	0	0.0	1 691	100
Sex child (T0)	0	0.0	1 691	100
Age child (years) (T0)	0	0.0	1 691	100
Weight status child (T0)	0	0.0	1 691	100
Mother's age (years) (T0)	0	0.0	1 691	100
Weight status parents (T0)	221	13.1	1 470	86.9
Migration background (T0)	9	0.5	1 682	99.5
Income level (T0)	149	8.8	1 542	91.2
Educational level (T0)	26	1.5	1 665	98.5
Number of Adults in Household (T0)	55	3.3	1 636	96.7
Siblings aged < 18 years (T0)	44	2.6	1 647	97.4
Moved	5	0.3	1 686	99.7

Missing data reduced the sample size from 1 691 to 1 297. The following estimates of the model using complete-case analysis deviated from the model using the missing-indicator method: two landline numbers, underweight, full migration background, high education, three adults in household, more than four adults in household, and mover within as well as mover outside (

Table ***7.7***). Although we observed these deviations the confidence intervals of the two analysis methods generally overlapped.

Within the context of the IDEFICS/I.Family study that focused on overweight and obesity (Ahrens et al., 2017), it seemed reasonable that the estimate of underweight decreased towards 1 and in light of the literature on attrition and full migration background increased (Frank et al., 2018; Winkler et al., 2014). The same applies for the tendency observed in the model with complete-case analysis, whereby families with two landline phone numbers had a lower chance of attrition compared to families with one landline phone number. Although the results obtained using complete-case analysis slightly differed from those obtained using the missing-indicator method, they were in line with the results discussed in Chapter 3, where a higher chance of cohort attrition was reported to be associated with the group late baseline recruitment and early follow-up and baseline participants with only one known mobile phone number. Families with high education had a lower chance of cohort attrition, and the

age of the child was positively associated with cohort attrition, while the mother's age was inversely associated.

Table 7.7 Odds ratios (with 95% CIs) for attrition at the follow-up (T1) as predicted by a model with interaction using the missing-indicator method to handle missing data, compared to a model using complete-case analysis

	Missing-indicator method			Complete-case analysis		
	OR	95% CI		OR	95% CI	
Grouped baseline/follow-up recruitment effort						
Early baseline recruitment & early follow-up recruitment	ref.			ref.		
Early baseline recruitment & late follow-up recruitment	1.13	0.81	1.58	1.08	0.73	1.58
Late baseline recruitment & early follow-up recruitment	1.65	1.19	2.28	1.74	1.21	2.50
Late baseline recruitment & late follow-up recruitment	1.11	0.81	1.52	1.07	0.74	1.54
Telephone phone numbers at baseline						
One landline number	ref.			ref.		
Two landline numbers	1.01	0.49	2.09	0.82	0.37	1.84
One mobile number	1.51	1.01	2.27	1.52	0.90	2.56
Landline and mobile	0.76	0.54	1.06	0.71	0.48	1.05
Sex child (baseline)						
Male	ref.			ref.		
Female	1.14	0.91	1.42	1.22	0.94	1.57
Age child at baseline (years)[a]	1.22	1.11	1.33	1.25	1.12	1.39
Weight status child (baseline)						
Under weight	0.74	0.51	1.08	0.92	0.60	1.43
Normal weight	ref.			ref.		
Overweight	1.29	0.95	1.75	1.27	0.89	1.82
Mother's age at baseline (years) [a]	0.96	0.94	0.98	0.97	0.94	0.99
Weight status of parents (baseline)						
No parent overweight	ref.			ref.		
At least one parent overweight	1.01	0.76	1.34	1.07	0.79	1.45
Missing	1.30	0.86	1.96			
Migration background (baseline)						
No	ref.			ref.		
Partly	1.04	0.72	1.50	1.09	0.71	1.66
Full	1.02	0.75	1.38	1.19	0.83	1.69
Missing	1.84	0.33	10.36			
Income level (baseline)						
Low income	ref.			ref.		
Medium income	0.83	0.62	1.10	0.82	0.60	1.12
High income	0.95	0.60	1.52	0.94	0.57	1.57
Missing	1.53	1.02	2.29			
Educational level (baseline)						
Low education	ref.			ref.		
Medium education	0.85	0.63	1.15	0.93	0.65	1.33
High education	0.60	0.39	0.90	0.71	0.44	1.14
Missing	1.26	0.50	3.15			
Number of adults in household (baseline)						
One	ref.			ref.		
Two	0.95	0.66	1.38	0.88	0.55	1.42
Three	1.05	0.58	1.90	0.89	0.43	1.84
Four or more	1.69	0.68	4.18	1.41	0.51	3.89
Missing	1.18	0.31	4.45			

	Missing-indicator method			Complete-case analysis		
	OR	95% CI		OR	95% CI	
Siblings aged <18 years (baseline)						
Yes	ref.			ref.		
No	1.28	0.98	1.67	1.26	0.92	1.73
Missing	1.62	0.38	6.96			
Moved						
Stayer	ref.			ref.		
Mover within study region	1.36	0.85	2.18	1.56	0.89	2.71
Mover outside study region	2.98	0.95	9.38	2.57	0.75	8.85
Missing	4.50	0.41	49.26			
All	1 691			1 297		

Odds ratios adjusted for sex child (baseline), age child at baseline (years), weight status child (baseline), mother's age at baseline (years), weight status parents (baseline), migration background (baseline), income level (baseline), educational level (baseline), number of adults in household (baseline), siblings aged <18 years (baseline)

[a] Odds ratios correspond to a 1-year increase in age

7.2.4 Compliance score

In Chapter 2 we constructed a separate score of study compliance for children and parents by counting the number of key examination modules they completed at baseline and at first follow-up. The compliance score for children included the examination modules blood pressure, bioelectrical impedance analysis (fasting state), waist-to-hip ratio, skinfold thickness (subscapularis and triceps), venous or capillary blood (fasting state), morning urine, and saliva. The compliance score for parents included the availability of the general questionnaire, food frequency questionnaire, medical history, and 24-h dietary recall. For all examination modules we constructed a binary variable that contained the information whether a given module was available, coded as 1, and not available, coded as 0. The final score was constructed by calculating the sum of all binary variables for each case in the IDEFICS/I.Family data separately for the examination modules of children and parents. Using the sum of the raw scores has the advantage that it is simple to construct and interpret. We then constructed an instrument based on the assumption that a higher score was correlated with the unobserved latent variable 'study compliance' that was measured via manifest items, that is participation in the different examination modules. Adding up the raw scores however neglects the fact that that the examination modules were not likely to require the same level of overcoming to participate (Boone, 2016). In the IDEFICS/I.Family baseline examination, for example, a higher percentage of children provided the waist-to-hip ratio compared to venous or capillary blood (fasting state). Thus, the sums of the raw scores for different IDEFICS/I.Family children are not accurate comparisons of the study compliance as they ignore the differences in difficulty across the examinations modules. This means that the assumption of a linear relationship between the sum of the raw scores and the study compliance is not necessarily appropriate (Boone, 2016; Geiser and Eid, 2010).

To increase the precision of an instrument, such as the compliance score, there are several statistical methods that can be applied. For instance, the item-response-theory enables the analysis of responses

to psychological tests, for example the mathematical competence of pupils or customer satisfaction (Strobl, 2012). In the construction phase of such tests, the Rasch-model (Rasch, 1960) can be used to take unequal difficulties into account. Tests are constructed, iteratively modified by reformulating questions or excluding single items, and finally applied in real live (Strobl, 2012).

For our purposes, we retrospectively used the information that was available to construct an ad-hoc score of study compliance. As illustrated in the contingency table of the compliance scores of children and parents (see Table 7.8) the compliance score for 76% of the study subjects (n = 19 706) was above 5 for children and above 2 for parents. This indicates that once children and their parents decided to participate in IDEFICS/I.Family, they tended to take part in the complete study protocol rather than in selected modules. The Rasch-model is hence redundant in this particular setting. Nevertheless, the fact that we did not construct the compliance scores under statistical guidance and that study enrollment was followed by participation in the majority of study modules limits the validity of the compliance scores we constructed.

Table 7.8 Contingency table of the parent's and children's compliance scores in IDEFICS/I.Family (see Chapter 2)

	Score parents				
Score children	1	2	3	4	N
0	6	62	13	15	96
1	25	125	50	24	224
2	20	45	49	57	171
3	30	51	64	101	246
4	48	112	146	304	610
5	144	369	699	1 160	2 372
6	162	912	2 129	3 413	6 616
7	158	1 275	4 253	9 911	15 597
N	593	2 951	7 403	14 985	25 932

Although it is reasonable to assume that higher study compliance is associated with lower attrition, from a theoretical perspective, a measure of study compliance captures the willingness to participate before the actual participation, whereas the experience of having participated should reframe the willingness to participate at follow-up (Esser, 1999). From the perspective of a study practitioner, it could be useful to quantify the experiences of having taken part at previous time points of a cohort study (Lynn, 2014), either with regard to particular dimensions of interest (e.g. study personnel, duration of the examination, or personal interest in the investigation) or with the aim to design a set of items that can be summarized into a single measure, for example of customer satisfaction or service quality. Experiences are likely to be crucial for participation in follow-ups, in particular in epidemiological studies in which study subjects are invited to a study center and exposed to invasive methods of data collection, such as the collection of biomaterials.

According to De Battisti et al. (2005), quality service applications are defined by the quality and the satisfaction of the customer. Even when the quality of a service is on a constant level, the satisfaction of consumers with the quality may differ because of habit or cultural differences. Therefore, service quality can be divided into attribute factors, such as properties of the quality giver, and person factors, such as properties of the user of a service, which both determine the answers given to a questionnaire on service quality. The Rasch-model essentially captures two features: (1) The difficulty of an item, for example, a construct that ought to measure intelligence may include different tests with increasing difficulty. (2) The ability of a person, which is the likelihood of a single person to solve a certain test. De Battisti et al. (2005) argue that the Rasch-model can also be applied to construct a measure of service quality on a statistical basis, as attribute factors of a service (the quality) can be regarded as difficulty, and the person factors (satisfaction) can be regarded as the ability. For example, Ferrari and Salini (2008) used the Eurobarometer study data to investigate consumer satisfaction ex-post, using Likert-scaled items (Overall, what do you think of 'XXX' service that you use?) with the categories very good, fairly good, fairly bad, and very bad.

It could indeed be fruitful to have a measure of service quality for epidemiological cohort studies. Potential problematic issues that may arise when using such an instrument are social desirable answers, therefore questionnaires regarding the experience of participation could be filled in in an anonymous setting at the end of the examination so as to prevent socially desirable answers (participant in room without presence of study personnel). Although letting participants fill in such a questionnaire at home via a link to an online questionnaire or app on a smartphone could be appealing to save time, the participation in such an instrument outside the setting of the examination might be low, thereby compromising the idea of such an instrument. In cases where participants have to fill in an instrument at home, this could be complemented by a similar assessment of the interview situation by the study personnel after each successful interview/examination. The reliability of interviewer assessments may however be limited as the interviewers will be rating the quality of the examination/interview directly related to their own work, and may therefore overestimate its success.

7.2.5 Country specific paradata

As pointed out by Bristle et al. (2018), although paradata have been used in the literature at the individual or interviewer level, cross-national research using paradata is scarce. This is probably due to the fact that the respective paradata are not available or that they are not harmonized across countries, which then limits comparisons (Bristle et al., 2018). Such information would however be valuable (Bristle et al., 2018), in particular for the investigation of factors associated with locating, contacting and obtaining co-operation with study subjects (Lepkowski and Couper, 2002)

In Chapter 2, where we investigated attrition across eight different countries participating in IDEFICS/I.Family, we observed that some predictors of attrition deviated from the overall pattern of a pooled estimate (female children in Belgium, medium educational level in Italy, control region in Belgium). Unfortunately, we were not able to explain these deviations as an electronic documentation system to document the recruitment, such as MODYS, was only used in Germany, and detailed paradata were hence not available for the other countries. Although paradata were implicitly included in our analysis as we used country as a second-level variable in the multilevel model, a major benefit of having paradata is that they would have offered the possibility to assess the consequences of different country-specific recruitment strategies on attrition. In IDEFICS/I.Family, recruitment strategies differed across countries, in particular with respect to contact modes. Families were recruited via schools, mail, e-mail, phone, and home visits. For example, at the second follow-up examination (T3), some countries recruited study subjects primarily via schools and made additional phone calls or sent mails/e-mails (Italy, Estonia, Belgium depending on study region, Sweden, Hungary). In Cyprus on the other hand, recruitment was carried out via the phone, and SMS messages were used as reminders for appointments. In Germany, recruitment included mailed invitations, phone calls, and where phone numbers were not available, home visits. In Spain, recruitment started via mail but, due to low participation, the recruitment protocol was adjusted and school events were organized, where the I.Family project was introduced and then appointments with the families made (Wolters, 2014).

Detailed paradata, including a recruitment history for each study subject within each country, would have enabled the quantification of the recruitment effort as well as the inclusion of relocations of subjects into the investigation of attrition across countries. The latter is very relevant as relocation is known to be highly associated with attrition (Behr et al., 2005; Lynn and Lugtig, 2017; Watson and Wooden, 2009). Additionally, it would have been fruitful to validate the findings of Chapter 3 by investigating the interaction between recruitment effort at baseline and follow-up for different countries.

7.2.6 Documentation of phone numbers

In Chapter 3, we observed that the dropout code 'no-contact' was assigned to participants belonging to the late baseline recruitment & early follow-up recruitment group more frequently than expected. Thus, our aim was to further investigate whether the phone number given at IDEFICS/I.Family baseline turned out to be invalid at follow-up as this would have explained the observed interaction between recruitment effort at baseline and follow-up. The available MODYS paradata included area-codes or mobile phone number primaries at baseline and follow-up, categorized as private, business and mobile. Due to data protection regulations, full numbers were not provided as they would have enabled the direct identification of study subjects. A first challenge regarding the data was that private

landline and private mobile numbers were not stored in separate categories, but together under the category “private numbers”. As we also wanted to investigate the association between the type of phone number and 'no-contact', we distinguished the phone numbers into landline and mobile phone numbers. The paradata enabled us to determine whether a respondent with a documented landline or mobile phone number at baseline had a documented landline or mobile phone number at follow-up. The information value of such data is however limited, as the level of statement was not related to the phone numbers themselves, but to the participants. We in fact did not investigate whether a particular phone number at baseline turned out to be invalid at follow-up, as the history of the validity of phone numbers for each potential study participant was not available.

Furthermore, we were not able to distinguish the source of the phone numbers as they could have been provided by the participants via the informed consent forms, or searched for in public telephone registers. The phone numbers acquired via the latter method presumably have a higher likelihood of invalidity compared to those collected from participants directly, leading to higher levels of “no contact”. Finally, as area-codes and mobile number primaries were provided after completion of the field-phase, it is possible that some of the phone number changes that occurred were not recorded.

7.2.7 Free-text field

MODYS allows users to type additional information for a given event into a free-text field. For smaller amounts of data it is possible to classify the text manually, for bigger amounts, a form of text analysis can be applied (Grimmer and Stewart, 2013). As stated, in the foregoing section, our aim in Chapter 3 was to investigate whether a given phone number at IDEFICS/I.Family baseline turned out to be invalid at follow-up. Indeed, the paradata included free-text fields with statements describing the reasons of no contact, for example “The mobile number you have called is currently not available”. The inconsistency of documentation in free-text fields however poses a great challenge. For instance, in the German IDEFICS/I.Family study, follow-up (T1) paradata revealed that if a contact attempt was documented as ‘no contact’ and the contact description as ‘nobody answering’, then only 14% of the free-text fields contained at least an entry, regardless of the content of the information. Hence, in 86% of the cases, it is not clear whether a phone number was valid, or an invalid number was not documented. Thus, as long it is not mandatory for fieldworkers to provide additional information in the free text field, the free-text field will not be a reliable/valid source of information to investigate reasons for ‘no-contact’. Furthermore, as has already been pointed out, in cases where subjects provided more than one phone number, MODYS paradata do not allow the linkage of a particular free-text field to a specific phone number.

In order to be able to investigate reasons of ‘no-contact’ in more detail, it might be advisable to add a documentation field within MODYS for the documentation of the validity of a given phone number,

such as a drop-down menu with predefined operationalizations of 'no-contact' and also including the category "unknown".

7.2.8 Weekdays in envelope experiment

In Chapter 5, where we investigated whether the color of the envelope of a study invitation (grey color vs. envelopes of white color) influenced the likelihood of an active response or study participation, we sent out the mailings on Mondays and Tuesdays. This might raise the concern that randomization was not conducted on an individual basis, but instead, 'clusters' of participants were randomized. It could further be argued, that intra-cluster correlation ought to have been taken into account in the sample size estimation and in the analysis. However, as participants for the mail invitations were randomly drawn from a larger random sample provided by the registry office and randomly assigned to envelope color (weekday), we assumed that there would be no clustering of characteristic or traits within these groups that would make them different from other groups. The only difference to the other groups would hence only be the envelope color which was our outcome variable, and the day of the mailing, which was randomly assigned. To the best of our knowledge, our procedure resulted in the random distribution of all variables known to us (name, sex, age, nationality as supplied by the registry office) or unknown to us across all groups.

7.3 Conclusion

Despite the long history of research on nonresponse and attrition, missing data is still a matter of debate. Paradata are valuable information for the investigation of nonresponse and attrition in cases where they capture all the information required to answer a given research question. At first sight, it may seem that the documentation of paradata itself requires considerable effort, but dedicated documentation software such as MODYS actually reduce the effort to be invested in conducting a study as they can also be used as a tool to organize the field phase efficiently. However, from the perspective of investigating nonresponse and attrition, it seems reasonable to collect more paradata than required for the design of the documentation software because these data may be valuable for further analysis on nonresponse and attrition. In particular, it seems reasonable to collect paradata specifically for scientific purposes. This means that in the planning phase of a study, it should already be defined how nonresponse and attrition should be investigated using paradata. Having specific research questions in mind will also help to define the necessary paradata required to obtain insightful answers. As stated in the guidelines of good epidemiological practice, for example from the German Society for Epidemiology, "[t]he planning of every epidemiological study requires explicit and operationalisable questions, which should be formulated specifically and precisely." (Hoffmann et al., 2019, p. 303)

Software for the collection of paradata, such as MODYS, is primarily designed by software engineers, eventually tested by users, for example field staff, and then iteratively adapted accordingly to ensure

a smooth workflow during the field phase. The software requirements, especially for studies recruiting study subjects via the phone, are a clear software interface and fast and intuitive handling. Thus, since its first version in the mid-1990s, the design of MODYS was mainly orientated towards the purpose of organizing the field phase. Further, the documented data was not collected for scientific purpose, but was primarily used to carry out automatized actions of the program. However, in cases where systems for collecting paradata are programmed in-house in a research institute, such as MODYS in BIPS, the fact that it is possible to design the paradata for scientific purpose prior to the study is a great strength that potentially increases the quality of the collected paradata. This is important as paradata are not only a 'by-product', but provide considerable scientific benefit, for example in the field of nonresponse analysis. For instance, the analysis carried out in this thesis using MODYS paradata resulted in the improved documentation of master data changes (in German so called Stammdatenänderungen) to capture changes in telephone numbers.

7.3.1 Suggestions for future research

Although dedicated paradata are valuable information for the analysis of nonresponse and attrition they cannot solve all problems related to the analysis of nonresponse and attrition, especially as they are not likely to include information on key study variables. Just as many empirical problems can be subdivided into a set of smaller units of problems, paradata often capture a particular aspect that explains nonresponse and attrition to a certain extent. In this thesis we used data from various sources to investigate nonresponse and attrition, including collected data from study subjects, MODYS paradata capturing the recruitment history, as well as information from the sampling frame containing sociodemographic information on potential study subjects and study participants. In order to be able to explain a higher degree of variation in nonresponse and attrition, future research should include data from more sources into a single analysis. Using data from different sources will benefit the research as it will enable the different data sources to be tested against each other regarding confounding. It will also enable a more detailed investigation of potential selection bias.

As previous research demonstrated, adding individual-level data on key variables from population registers to the data collected via a population-based study would be particularly beneficial (Ludvigsson et al., 2009). Although a unique identifier such as a personal identity number for multiple registers does not exist in Germany, it is still theoretically possible to combine individual level data from multiple registers using record linkage (March et al., 2018). Record linkage “refers to the task of extracting record information from various input data sources and combining them in such a way that each output record corresponds a distinct real-world entity” (Gruenheid, 2019, see paragraph 'Definitions').

The potential data sources for record linkage for Germany are statutory health insurances, social security data, pension schemes, register of residents, and census data. According to March and

colleagues (March et al., 2018), although record linkage is increasingly being used in Germany, there are significant methodological and legal challenges involved. For instance, for prospective studies, linkage via a direct identifier, for example via the social security number or multiple direct identifiers such as name and birthdate, requires informed consent. Previous research reported differences in the respondents' willingness to consent to this (Sakshaug et al., 2012), leading to effects similar to nonresponse and attrition, and stressed the role of interviewer characteristics (Korbmacher and Schroeder, 2013) as well as the benefit of appropriate wording of the consent form (Sakshaug and Kreuter, 2014). Where direct identifiers are not available, for example in a retrospective analysis, linkage can be obtained using indirect identifiers, for example, age, sex, hospital admission date and time (Maier et al., 2015). This, however, has to be done in compliance with data protection laws (March et al., 2018). The availability of data sources, comparability of data sources, and the legal requirements pose a further burden to carry out nonresponse analysis, particularly for multicenter cohort study carried out in different countries.

Additional individual-level data could further be complemented by spatial data on the aggregate level (Schweers et al., 2016). For Chapter 3, we searched for secondary data such as socioeconomic data on aggregate level, to investigate nonresponse at the German IDEFICS/I.Family baseline. Governmental statistical offices publishing data for administrative divisions are a potential source of spatial data, part of which captures dimensions of social deprivation. Research has shown that regional deprivation is associated with epidemiological endpoints such as cancer, ischemic heart disease, and subjective health assessment (Hofmeister et al., 2016). Furthermore, previous research linked governmental data to dimensions of social deprivation. These dimensions can be used as single measures or can be combined into indices of multiple deprivation that capture the extent of deprivation for spatial units (Noble et al., 2006). For Germany, Hofmeister et al. (2016) reported the existence of indices of multiple deprivation for the different administrative divisions (Kreise) and municipalities (Gemeinden), as well as for the city-states Hamburg, Bremen, and Berlin using the German Index of Multiple Deprivation (GIMD) (Hofmeister et al., 2016). The GMID includes the dimensions (so-called domains) income, employment, and education, income of municipality, social capital, environment, and security. For example, the indicator for the domain employment is the number of unemployed of a spatial unit, and for social capital, voter turnout for the national parliament elections, in percent.

In Germany, up-to-date data, for administrative districts or municipalities, including street directories, are publicly available via the Federal Statistical Office. The street directories could hence be used to link potential study subjects to spatial units and the data could then be included in nonresponse analyses. However, within municipalities, data on a small area level may be scarce. For Chapter 3, we searched for aggregated data for urban districts of the German municipalities Delmenhorst and Wilhelmshaven, where the eligible IDEFICS/I.Family children lived. The data were available for

Wilhelmshaven, but not for Delmenhorst. It would have been helpful to have smaller units of aggregated data, for example building blocks, as it can be assumed that the correlation within smaller spatial units is higher than within larger ones. However, even if such data were to be available, access would be restricted due to privacy legislations (Hofmeister et al., 2016; Schweers et al., 2016).

8 References

Agresti, A., 2007. An Introduction to Categorical Data Analysis, 2nd ed. John Wiley & Sons, New York.

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.A., Pitsiladis, Y.P., Reisch, L., Tornaritis, M., Veidebaum, T., Pigeot, I., Idefics Consortium, 2011. The IDEFICS Cohort: Design, Characteristics and Participation in the Baseline Survey. Int J Obes (Lond) 35 Suppl 1, S3-15.

Ahrens, W., Pigeot, I., Pohlabeln, H., De Henauw, S., Lissner, L., Molnar, D., Moreno, L.A., Tornaritis, M., Veidebaum, T., Siani, A., consortium, I., 2014. Prevalence of overweight and obesity in European children below the age of 10. Int J Obes (Lond) 38 Suppl 2, S99-107.

Ahrens, W., Siani, A., Adan, R., De Henauw, S., Eiben, G., Gwozdz, W., Hebestreit, A., Hunsberger, M., Kaprio, J., Krogh, V., 2017. Cohort Profile: The Transition from Childhood to Adolescence in European Children–How I. Family Extends the IDEFICS Cohort. International Journal of Epidemiology 46, 1394-1395j.

Allison, P.D., 1985. Event History Analysis. Regression Models for Longitudinal Event Data. Sage, Beverly Hills, California.

Allison, P.D., 2007. Missing Data. SAGE Publications, Inc, Thousand Oaks, California.

Alonso, A., Segui-Gomez, M., de Irala, J., Sanchez-Villegas, A., Beunza, J.J., Martinez-Gonzalez, M.A., 2006. Predictors of Follow-Up and Assessment of Selection Bias from Dropouts Using Inverse Probability Weighting in a Cohort of University Graduates. European Journal of Epidemiology 21, 351-358.

Aquilino, W., 1993. Effects of Spouse Presence During the Interview on Survey Responses Concerning Marriage. Public Opinion Quarterly 3, 358–376.

Aquilino, W., Wright, D., Supple, A., 2000. Response effects due to bystander presence in CASI and Paper-and-Pencil Surveys of Drug Use and Alcohol Use. Substance Use & Misuse 35, 845–867.

Battaglia, M.P., Khare, M., Frankel, M.R., Murray, M.C., Buckley, P., Peritz, S., 2007. Response Rates: How have they Changed and Where are they Headed? In: Lepkowski, J.M., Tucker, C., Brick, J.M., De Leeuw, E.D., Japec, L., Lavrakas, P.J., Link, M.W., Sangster, R.L. (Eds.), Advances in Telephone Survey Methodology. John Wiley & Sons, Inc., Hoboken, pp. 529-560.

Baur, E.J., 1947. Response Bias in a Mail Survey. Public Opinion Quarterly 11, 595–600.

Behr, A., Bellgart, E., Rendtel, U., 2005. Extent and Determinants of Panel Attrition in the European Community Household. European Sociological Review 21, 489-512.

Bethlehem, J., Cobben, F., Schouten, B., 2011. Handbook of Nonresponse in Household Surveys. John Wiley & Sons, Hoboken, New Jersey.

Billiet, J., Davidov, E., 2008. Testing the Stability of an Acquiescence Style Factor Behind Two Interrelated Substantive Variables in a Panel Design. Sociological Methods & Research 36, 542–562.

Bodenmann, G., 2000. FDCT-N. Fragebogen zur Erfassung des dyadischen Copings als generelle Tendenz. Institut für Familienforschung und -beratung,, Fribourg.

Boone, W.J., 2016. Rasch Analysis for Instrument Development: Why, When, and How? CBE Life Sci Educ 15, rm4.

Brilleman, S.L., Pachana, N.A., Dobson, A.J., 2010. The Impact of Attrition on the Representativeness of Cohort Studies of Older People. BMC Medical Research Methodology 10.

Bristle, J., Celidoni, M., Dal Bianco, C., Weber, G., 2018. The Contributions of Paradata and Features of Respondents, Interviewers and Survey Agencies to Panel Co-operation in the Survey of Health, Ageing and Retirement in Europe. Journal of the Royal Statistical Society: Series A (Statistics in Society) 182, 3-35.

Buunk, B., Dijkstra, P., 2004. Gender Differences in Rival Characteristics that Evoke Jealousy in Response to Emotional versus Sexual Infidelity. Personal Relationships 11, 395-408.

Calcagno, V., de Mazancourt, C., 2010. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. Journal of Statistical Software 34, 1-29.

Cannell, C., Oksenberg, L., Converse, J., 1977. Striving for Response Accuracy: Experiments in New Interviewing Techniques. Journal of Marketing Research 14, 306-315.

Carter, K.N., Imlach-Gunasekara, F., McKenzie, S.K., Blakely, T., 2012. Differential Loss of Participants Does not Necessarily Cause Selection Bias. Australian and New Zealand Journal of Public Health 36, 218-222.

Cohen, S.B., Machlin, S.R., Branscome, J.M., 2000. Patterns of survey attrition and reluctant response in the 1996 medical expenditure panel survey. Health Services and Outcomes Research Methodology 1, 131-148.

Cole, S.R., Platt, R.W., Schisterman, E.F., Chu, H., Westreich, D., Richardson, D., Poole, C., 2010. Illustrating Bias Due to Conditioning on a Collider. Int J Epidemiol 39, 417-420.

Cole, T.J., Lobstein, T., 2012. Extended International (IOTF) Body Mass Index Cut-offs for Thinness, Overweight and Obesity. Pediatric Obesity 7, 284-294.

Corry, N.H., Williams, C.S., Battaglia, M., McMaster, H.S., Stander, V.A., 2017. Assessing and Adjusting for Non-response in the Millennium Cohort Family Study. BMC Medical Research Methodology 17.

Curtin, R., Presser, S., Singer, E., 2005. Changes in Telephone Survey Nonresponse over the Past Quarter Century. Public Opinion Quarterly 69, 87-98.

Daniel, R.M., Kenward, M.G., Cousens, S.N., De Stavola, B.L., 2011. Using Causal Diagrams to Guide Analysis in Missing Data Problems. Statistical Methods in Medical Research 21, 243-256.

De Battisti, F., Nicolini, G., Salini, S., 2005. The Rasch Model to Measure the Service Quality. The ICFAI Journal of Services Marketing, 58-80.

de Leeuw, D., de Heer, W., 2002. Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R. (Eds.), Survey Nonresponse. Wiley, New York, pp. 41-54.

Dijkema, M.B.A., Grievink, L., Stellato, R.K., Roorda, J., van der Velden, P.G., 2005. Determinants of Response in a Longitudinal Health Study Following the Firework-disaster in Enschede, The Netherlands. European Journal of Epidemiology 20, 839-847.

Eaton, W.W., Anthony, J.C., Tepper, S., Dryman, A., 1992. Psychopathology and Attirition in the Epidemiologic Catchment Area Surveys. American Journal of Epidemiology 135, 1051-1059.

Edlund, J., Heider, J., Scherer, C., Farc, M.-M., Sagarin, B., 2006. Sex Differences in Jealousy in Response to Actual Infidelity. Evolutionary Psychology 4, 462-470.

Edwards, P.J., Roberts, I., Clarke, M.J., Diguiseppi, C., Wentz, R., Kwan, I., Cooper, R., Felix, L.M., Pratap, S., 2009. Methods to Increase Response to Postal and Electronic Questionnaires. Cochrane Database Syst Rev, MR000008.

Eerola, M., Huurre, T., Aro, H., 2005. The Problem of Attrition in a Finnish Longitudinal Survey on Depression. European Journal of Epidemiology 20, 113-120.

Ellis, R.A., Endo, C.M., Armer, J.M., 1970. The Use of Potential Nonrespondents for Studying Nonresponse Bias. The Pacific Sociological Review 13, 103-109.

Elwert, F., 2013. Graphical Causal Models. In: Morgan, S.L. (Ed.), Handbook of Causal Analysis for Social Research. Springer Science + Business Media, Dordrecht pp. 245-273.

Esser, H., 1999. Soziologie : Allgemeine Grundlagen, 3 ed. Campus-Verlag, Frankfurt/M.

Evenson, K.R., Catellier, D.J., Gill, K., Ondrak, K.S., McMurray, R.G., 2008. Calibration of Two Objective Measures of Physical Activity for Children. J Sports Sci 26, 1557-1565.

Ferrari, P.A., Salini, S., 2008. Measuring Service Quality: The Opinion of Europeans about Utilities. In: (FEEM), F.E.E.M. (Ed.), Privatisation Regulation Corporate Governance Working Papers 36758.

Filion, F.L., 1976. Exploring and Correcting for Nonresponse Bias Using Follow-Ups of Nonrespondents. The Pacific Sociological Review 19, 401-408.
Fletcher, A.C., Hunter, A.G., 2003. Strategies for Obtaining Parental Consent to Participate in Research. Family Relations 52, 216-221.
Fox, J., 2016. Applied Regression Analysis and Generalized Linear Models, 3 ed. Sage, Los Angeles.
Frank, L., Yesil-Jürgens, R., Born, S., Hoffmann, R., Santos-Hövener, C., Lampert, T., 2018. Improving the Inclusion and Participation of Children and Adolescents with a Migration Background in KiGGS Wave 2. Journal of Health Monitoring 3, 126-142.
Furman, W., Buhrmester, D., 1985. Children's Perceptions of the Personal Relationships in their Social Networks. Developmental Psychology 21, 1016–1024.
Gades, N.M., Jacobson, D.J., McGree, M.E., Lieber, M.M., Roberts, R.O., Girman, C.J., Jacobsen, S.J., 2006. Dropout in a Longitudinal, Cohort Study of Urologic Disease in Community Men. BMC Medical Research Methodology 6.
Gaertner, B., Seitz, I., Fuchs, J., Busch, M.A., Holzhausen, M., Martus, P., Scheidt-Nave, C., 2016. Baseline Participation in a Health Examination Survey of the Population 65 Years and Older: Who is Missed and Why? BMC Geriatrics 16.
Galea, S., Tracy, M., 2007. Participation Rates in Epidemiologic Studies. Ann Epidemiol, 643-653.
Geiser, C., Eid, M., 2010. Item-Response-Theorie. In: Wolf, C., Best, H. (Eds.), Handbuch der sozialwissenschaftlichen Datenanalyse. VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 311-332.
German National Cohort, C., 2014. The German National Cohort: Aims, Study Design and Organization. Eur J Epidemiol 29, 371-382.
Giorgi, R., Belot, A., Gaudart, J., Launoy, G., 2008. The Performance of Multiple Imputation for Missing Covariate Data within the Context of Regression Relative Survival Analysis. Statistics in Medicine 27, 6310-6331.
Greene, N., Greenland, S., Olsen, J., Nohr, E.A., 2011. Estimating Bias From Loss to Follow-up in the Danish National Birth Cohort. Epidemiology 22, 815-822.
Grigoletto, F., Anderson, D.W., Rocca, W.A., Meneghini, F., Morgante, L., Reggio, A., Savettieri, G., Di Perri, R., 1994. Attrition and Use of Proxy Respondents and Auxiliary Information in the Sicilian Neuroepidemiologic Study. American Journal of Epidemiology 139, 219-228.
Grimmer, J., Stewart, B.M., 2013. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis 21, 267–297.
Groenwold, R.H.H., White, I.R., Donders, A.R.T., Carpenter, J.R., Altman, D.G., Moons, K.G.M., 2012. Missing Covariate Data in Clinical Research: When and When not to Use the Missing-Indicator Method for Analysis. Canadian Medical Association 184, 1265–1269.
Groves, R.M., 2006. Nonresponse Rates and Nonresponse Bias in Household Surveys. Public Opinion Quarterly 70, 646-675.
Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta, G.P., Nelson, L., 2006. Experiments in Producing Nonresponse Bias. Public Opinion Quarterly 70, 720-736.
Groves, R.M., Fowler, F.J., Couper, M., Lepkowski, J.M., Singer, E., Tourangeau, R., 2009. Survey Methodology, 2 ed. John Wiley & Sons, Inc, Hoboken, New York.
Groves, R.M., Heeringa, S.G., 2006. Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs. Journal of the Royal Statistical Society: Series A 169, 439-457.
Groves, R.M., Lyberg, L.E., 2010. Total Survey Error. Past, Present, and Future. Public Opinion Quarterly 74, 849-879.
Gruenheid, A., 2019. Record Linkage. In: Sakr, S., Zomaya, A.Y. (Eds.), Encyclopedia of Big Data Technologies. Springer International Publishing, Cham, pp. 1409-1413.
Gubhaju, L., Banks, E., Macniven, R., Joshy, G., McNamara, B.J., Bauman, A., Eades, S.J., 2016. Factors Relating to Participation in Follow-up to the 45 and up Study in Aboriginal and non-Aboriginal Individuals. BMC Medical Research Methodology 16.

Guerrero, L.K., Spitzberg, B.H., Yoshimura, S.M., 2004. Sexual and Emotional Jealousy. In: Harvey, J., Wenzel, A., Sprecher, S. (Eds.), The Handbook of Sexuality in Close Relationships. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 311-345.

Guey, L.T., Bromet, E.J., Gluzman, S.F., Zakhozha, V., Paniotto, V., 2008. Determinants of Participation in a Longitudinal Two-stage Study of the Health Consequences of the Chornobyl Nuclear Power Plant Accident. BMC Medical Research Methodology 8.

Hall, J., Brown, V., Nicolaas, G., Lynn, P., 2013. Extended Field Efforts to Reduce the Risk of Non-response Bias: Have the Effects Changed over Time? Can Weighting Achieve the Same Effects? Bulletin de Méthodologie Sociologique 117, 5-25.

Hansen, M.H., Hurwitz, W.N., 1946. The Problem of Non-Response in Sample Surveys. Journal of the American Statistical Association 41, 517-529

Harald, K., Salomaa, V., Jousilahti, P., Koskinen, S., Vartiainen, E., 2007. Non-participation and Mortality in Different Socioeconomic Groups: The FINRISK Population Surveys in 1972-92. Journal of Epidemiology & Community Health 61, 449-454.

Haring, R., Alte, D., Volzke, H., Sauer, S., Wallaschofski, H., John, U., Schmidt, C.O., 2009. Extended Recruitment Efforts Minimize Attrition but not Necessarily Bias. J Clin Epidemiol 62, 252-260.

Hartmann, P., 1994. Interviewing when the Spouse is Present. International Journal of Public Opinion Research 6, 298–306.

Hense, S., Pohlabeln, H., Michels, N., Mårild, S., Lissner, L., Kovacs, E., Moreno, L.A., Hadjigeorgiou, C., Veidebaum, T., Iacovello, L., Pitsiladis, Y., Reisch, L., Siani, A., Ahrens, W., 2013. Determinants of Attrition to Follow-Up in a Multicentre Cohort Study in Children-Results from the IDEFICS Study. Epidemiology Research International 2013, 1-9.

Hernán, M.A., Hernández-Díaz, S., Robins, J.M., 2004. A Structural Approach to Selection Bias. Epidemiology 15, 615-625.

Higgins, J.P.T., 2003. Measuring inconsistency in meta-analyses. Bmj 327, 557-560.

Ho, D.E., Imai, K., King, G., Stuart, E.A., 2011. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. Journal of Statistical Software 42.

Hoffmann, W., Latza, U., Baumeister, S.E., Brünger, M., Buttmann-Schweiger, N., Hardt, J., Hoffmann, V., Karch, A., Richter, A., Schmidt, C.O., Schmidtmann, I., Swart, E., van den Berg, N., 2019. Guidelines and recommendations for ensuring Good Epidemiological Practice (GEP): a guideline developed by the German Society for Epidemiology. European Journal of Epidemiology 34, 301-317.

Hofmeister, C., Maier, W., Mielck, A., Stahl, L., Breckenkamp, J., Razum, O., 2016. Regionale Deprivation in Deutschland: Bundesweite Analyse des Zusammenhangs mit Mortalität unter Verwendung des ‚German Index of Multiple Deprivation (GIMD)'. Gesundheitswesen 78, 42-48.

Howe, L.D., Tilling, K., Galobardes, B., Lawlor, D.A., 2013. Loss to Follow-Up in Cohort Studies: Bias in Estimates of Socioeconomic Inequalities. Epidemiology 24, 1-9.

Hox, J.J., Moerbeek, M., van de Schoot, R., 2010. Multilevel Analysis: Techniques and Applications, 2. ed. ed. Routledge, Taylor & Francis, New York.

Huinink, J., Brüderl, J., Nauck, B., Walper, S., Castiglioni, L., Feldhaus, M., 2011. Panel Analysis of Intimate Relationships and Family Dynamics (pairfam) : Conceptual Framework and Design. Zeitschrift für Familienforschung 23, 77-101.

Jacobs, S., Stallmann, C., Pigeot, I., 2015. Verknüpfung großer Sekundär- und Registerdatenquellen mit Daten aus Kohortenstudien. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 58, 822-828.

Jöckel, K.H., Stang, A., 2013. Cohort Studies with Low Baseline Response May Not Be Generalisable to Populations with Different Exposure Distributions. Eur J Epidemiol 28, 223-227.

Johnson, T., 2014. Sources of Error in Substance Use Prevalence Surveys. International Scholarly Research Notices.

Kamtsiuris, P., Lange, M., Rosario, A.S., 2007. Der Kinder- und Jugendgesundheitssurvey (KiGGS): Stichprobendesign, Response und Nonresponse-Analyse. Robert Koch-Institut, Epidemiologie und Gesundheitsberichterstattung.

Knai, C., Lobstein, T., Darmon, N., Rutter, H., McKee, M., 2012. Socioeconomic patterning of childhood overweight status in Europe. Int J Environ Res Public Health 9, 1472-1489.

Koch, A., Blohm, M., 2015. Nonresponse Bias. SDM Survey Guidelines. GESIS – Leibniz-Institut für Sozialwissenschaften, Mannheim.

Konstabel, K., Veidebaum, T., Verbestel, V., Moreno, L.A., Bammann, K., Tornaritis, M., Eiben, G., Molnar, D., Siani, A., Sprengeler, O., Wirsik, N., Ahrens, W., Pitsiladis, Y., consortium, I., 2014. Objectively Measured Physical Activity in European Children: The IDEFICS Study. Int J Obes (Lond) 38 Suppl 2, S135-143.

Korbmacher, J.M., Schroeder, M., 2013. Consent when Linking Survey Data with Administrative Records: The Role of the Interviewer. Survey Research Methods 7, 115-131.

Kovacs, E., Hunsberger, M., Reisch, L., Gwozdz, W., Eiben, G., De Bourdeaudhuij, I., Russo, P., Veidebaum, T., Hadjigeorgiou, C., Sieri, S., Moreno, L.A., Pigeot, I., Ahrens, W., Pohlabeln, H., Molnar, D., consortium, I., 2015. Adherence to combined lifestyle factors and their contribution to obesity in the IDEFICS study. Obes Rev 16 Suppl 2, 138-150.

Kreienbrock, L., Pigeot, I., Ahrens, W., 2012. Epidemiologische Methoden, 5. Aufl. ed. Spektrum Akademischer Verlag, Heidelberg.

Kreuter, F., Casas-Cordero, C., 2010. Paradata.

Krumpal, I., 2013. Determinants of Social Desirability Bias in Sensitive Surveys: a Literature Review. Quality & Quantity 47, 2025–2047.

Kuskowska-Wolk, A., Rössner, S., 2009. The “True” Prevalence of Obesity: A Comparison of Objective Weight and Height Measures versus Self-Reported and Calibrated Data. Scandinavian Journal of Primary Health Care 7, 79-82.

La Verda, N., Teta, M.J., 2006. Re: “Reporting Participation in Epidemiologic Studies: A Survey of Practice”. American Journal of Epidemiology 164, 292-292.

Lacey, J.V., Jr., Savage, K.E., 2016. 50 % Response Rates: Half-Empty, or Half-Full? Cancer Causes Control 27, 805-808.

Lakey, B., Cohen, S., 2000. Social Support Theory and Measurement. In: Cohen, S., Brittney, L., Gottlieb, B. (Eds.), Social Support Measurement and Intervention. A Guide for Health and Social Scientists. Oxford University Press, Oxford, New York, pp. 29-52.

Lander, B., 2000. Anwesenheitseffekte im Wandel. Eine Sekundäranalyse zur Anwesenheit des Partners im Interview anhand des ALLBUS 1980 bis 1998. Zeitschrift für Soziologie 29, 227–238.

Lange, M., Butschalowsky, H.G., Jentsch, F., Kuhnert, R., Schaffrath Rosario, A., Schlaud, M., Kamtsiuris, P., Group, K.S., 2014. Die erste KiGGS-Folgebefragung (KiGGS Welle 1). Studiendurchführung, Stichprobendesign und Response. Bundesgesundheitsblatt 57, 747–761.

Langeheine, M., Pohlabeln, H., Ahrens, W., Rach, S., 2017. Consequences of an Extended Recruitment on Participation in the Follow-Up of a Child Study: Results from the German IDEFICS Cohort. Paediatric and Perinatal Epidemiology 31, 76-86.

Lavrakas, P., 2008. Encyclopedia of Survey Research Methods. SAGE, Thousand Oaks.

Lepkowski, J.M., Couper, M.P., 2002. Nonresponse in the Second Wave of Longitudinal Household Surveys. In: Groves, R.M., Dillman, D.A., Eltinge, J.L., Little, R.J.A. (Eds.), Survey Nonresponse. John Wiley, New York, pp. 257-272.

Li, L., Shen, C., Li, X., Robins, J.M., 2013. On Weighting Approaches for Missing Data. Stat Methods Med Res 22, 14-30.

Lin, I.-F., Schaeffer, N.C., 1995. Using Survey Participants to Estimate the Impact of Nonparticipation. Public Opinion Quarterly 59, 236-258.

Little, R.J.A., Rubin, D.B., 2002. Statistical Analysis with Missing Data. John Wiley & Sons, Inc, Hoboken, New Jersey.

Long, D.L., Howard, G., Long, D.M., Judd, S., Manly, J.J., McClure, L.A., Wadley, V.G., Safford, M.M., Katz, R., Glymour, M.M., 2019. An Investigation of Selection Bias in Estimating Racial Disparity in Stroke Risk Factors: The REGARDS Study. American Journal of Epidemiology 188, 587-597.

Long, J., Freese, J., 2014. Regression Models for Categorical Dependent Variables using Stata. Stata Press, College Station, Tex.

Ludvigsson, J.F., Otterblad-Olausson, P., Pettersson, B.U., Ekbom, A., 2009. The Swedish Personal Identity Number: Possibilities and Pitfalls in Healthcare and Medical Research. Eur J Epidemiol 24, 659-667.

Lugtig, P., 2014. Panel Attrition: Separating Stayers, Fast Attriters, Gradual Attriters and Lurkers. Sociological Methods and Research 43, 699-723.

Lumley, T., 2004. Analysis of Complex Survey Samples. Journal of Statistical Software 9, 1-19.

Lynn, P., 2009. Methods for Longitudinal Surveys. In: Lynn, P. (Ed.), Methodology of Longitudinal Surveys. Wiley, Hoboken, pp. 1-20.

Lynn, P., 2014. Longer Interviews May Not Affect Subsequent Survey Participation Propensity. Public Opinion Quarterly 78, 500-509.

Lynn, P., Lugtig, P.J., 2017. Total Survey Error for Longitudinal Surveys. In: Biemer, P.P., de Leeuw, E.D., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L.E., Tucker, N.C., West, B.T. (Eds.), Total Survey Error in Practice. John Wiley & Sons, Hoboken, New York, pp. 279-298.

Maclennan, B., Kypri, K., Langley, J., Room, R., 2012. Non-response bias in a community survey of drinking, alcohol-related experiences and public opinion on alcohol policy. Drug Alcohol Depend 126, 189-194.

Maier, B., Wagner, K., Behrens, S., Bruch, L., Busse, R., Schmidt, D., Schühlen, H., Thieme, R., Theres, H., 2015. Deterministisches Record Linkage mit indirekten Identifikatoren: Daten des Berliner Herzinfarktregisters und der AOK Nordost zum Herzinfarkt. Das Gesundheitswesen 77, e15-e19.

March, S., Antoni, M., Kieschke, J., Kollhorst, B., Maier, B., Müller, G., Sariyar, M., Schulz, M., Enno, S., Zeidler, J., Hoffmann, F., 2018. Quo vadis Datenlinkage in Deutschland? Eine erste Bestandsaufnahme. Das Gesundheitswesen 57, e20-e31.

Martikainen, P., Laaksonen, M., Piha, K., Lallukka, T., 2016. Does Survey Non-response Bias the Association Between Occupational Social Class and Health? Scandinavian Journal of Public Health 35, 212-215.

Martin, J., Tilling, K., Hubbard, L., Stergiakouli, E., Thapar, A., Smith, G.D., O'Donovan, M.C., Zammit, S., 2016. Association of Genetic Risk for Schizophrenia With Nonparticipation Over Time in a Population-Based Cohort Study. American Journal of Epidemiology 183, 1149-1158.

May, A.M., Adema, L.E., Romaguera, D., Vergnaud, A.C., Agudo, A., Ekelund, U., Steffen, A., Orfanos, P., Slimani, N., Rinaldi, S., Mouw, T., Rohrmann, S., Hermann, S., Boeing, H., Bergmann, M.M., Jakobsen, M.U., Overvad, K., Wareham, N.J., Gonzalez, C., Tjonneland, A., Halkjaer, J., Key, T.J., Spencer, E.A., Hellstrom, V., Manjer, J., Hedblad, B., Lund, E., Braaten, T., Clavel-Chapelon, F., Boutron-Ruault, M.C., Rodriguez, L., Sanchez, M.J., Dorronsoro, M., Barricarte, A., Huerta, J.M., Naska, A., Trichopoulou, A., Palli, D., Pala, V., Norat, T., Mattiello, A., Tumino, R., van der Daphne, A., Bueno-de-Mesquita, H.B., Riboli, E., Peeters, P.H.M., 2012. Determinants of Non-Response to a Second Assessment of Lifestyle Factors and Body Weight in the EPIC-PANACEA Study. BMC Medical Research Methodology 12.

Mein, G., Johal, S., Grant, R.L., Seale, C., Ashcroft, R., Tinker, A., 2012. Predictors of Two Forms of Attrition in a Longitudinal Health Study Involving Ageing Participants: An Analysis based on the Whitehall II Study. BMC Medical Research Methodology 12.

Melton, L.J., 3rd, Dyck, P.J., Karnes, J.L., O'Brien, P.C., Service, F.J., 1993. Non-response Bias in Studies of Diabetic Complications: The Rochester Diabetic Neuropathy Study. J Clin Epidemiol 46, 341-348.

Miller, A.B., Goff Jr., D.C., Bammann, K., Wild, P., 2014. Cohort Studies. In: Ahrens, W., Pigeot, I. (Eds.), Handbook of Epidemiology. Springer Science +Business Media New, New York, pp. 260-291.

Mohr, H., 1986. Dritte beim Interview. Ergebnisse zu Indikatoren aus dem Bereich Ehe und Partnerschaft mit den Daten des "Wohlfahrtssurvey 1984". ZA Information 19, 52-71.

Morton, L.M., Cahill, J., Hartge, P., 2006. Reporting Participation in Epidemiologic Studies: A Survey of Practice. American Journal of Epidemiology 163, 197-203.

Nederhof, E., Jörg, F., Raven, D., Veenstra, R., Verhulst, F.C., Ormel, J., Oldehinkel, A.J., 2012. Benefits of extensive recruitment effort persist during follow-ups and are consistent across age group and survey method. The TRAILS study. BMC Medical Research Methodology 93, 3-15.

Nguyen, C.D., Carlin, J.B., Lee, K.J., 2017. Model Checking in Multiple Imputation: An Overview and Case Study. Emerging Themes in Epidemiology 14, 8.

Noble, M., Wright, G., Smith, G., Dibben, C., 2006. Measuring Multiple Deprivation at the Small-area Level. Environment and Planning A38, 169–185.

Nohr, E.A., Frydenberg, M., Henriksen, T.B., Olsen, J., 2006. Does Low Participation in Cohort Studies Induce Bias? Epidemiology 17, 413-418.

Pearce, N., 2012. Classification of Epidemiological Study Designs. Int J Epidemiol 41, 393-397.

Pearl, J., 1995. Causal Diagrams for Empirical Research Biometrika 82, 669–710.

Pearl, J., Glymour, M., Jewell, N.P., 2016. Causal Inference in Statistics : A Primer. Wiley, Chichester, West Sussex, United Kingdom.

Peplies, J., Bornhorst, C., Gunther, K., Fraterman, A., Russo, P., Veidebaum, T., Tornaritis, M., De Henauw, S., Marild, S., Molnar, D., Moreno, L.A., Ahrens, W., Idefics consortium, 2016. Longitudinal Associations of Lifestyle Factors and Weight Status with Insulin Resistance (HOMA-IR) in Preadolescent Children: The Large Prospective Cohort Study IDEFICS. Int J Behav Nutr Phys Act 13, 97.

Pohlabeln, H., Rach, S., De Henauw, S., Eiben, G., Gwozdz, W., Hadjigeorgiou, C., Molnár, D., Moreno, L.A., Russo, P., Veidebaum, T., Pigeot, I., 2017. Further evidence for the role of pregnancy-induced hypertension and other early life influences in the development of ADHD: results from the IDEFICS study. European Child & Adolescent Psychiatry 26, 957-967.

Powers, J., Tavener, M., Graves, A., Loxton, D., 2015. Loss to follow-up was used to estimate bias in a longitudinal study: a new approach. Journal of Clinical Epidemiology 68, 870-876.

Rasch, G., 1960. Probabilistic Models for Some Intelligence and Attainment Tests. Nissen & Lydicke, Kopenhagen.

Regber, S., Novak, M., Eiben, G., Lissner, L., Hense, S., Sandström, T.Z., Ahrens, W., Mårild, S., 2013. Assessment of Selection Bias in a Health Survey of Children and Families - The IDEFICS Sweden-Study. BMC Public Health 13, 418-428.

Reineke, A., Pigeot, I., Ahrens, W., Rach, S., consortia obotIaIF, 2018. MODYS – A Modular Control and Documentation System for Epidemiological Studies. In: Bammann, K., Lissner, L., Pigeot, I., Ahrens, W. (Eds.), Instruments for Health Surveys in Children and Adolescents. Springer Nature, pp. 25-45.

Reineke, A., Voss, S.K., Asendorf, H., 2006. Modys - ein modulares Steuerungs- und Dokumentationssystem für epidemiologische Studien. Medizinische Dokumentation - wichtig oder nichtig? , 9. Fachtagung des Deutschen Verbandes Medizinischer Dokumentare (DVMD), Erlangen.

Reiss, K., Dragano, N., Ellert, U., Fricke, J., Greiser, K.H., Keil, T., Krist, L., Moebus, S., Pundt, N., Schlaud, M., Yesil-Jurgens, R., Zeeb, H., Zimmermann, H., Razum, O., Jockel, K.H., Becher, H., 2014. Comparing Sampling Strategies to Recruit Migrants for an Epidemiological Study. Results from a German Feasibility Study. Eur J Public Health 24, 721-726.

Reuband, K.-H., 1984. Dritte Personen beim Interview – Zuhörer, Adressaten oder Katalysatoren der Kommunikation? . In: Meulemann, H., Reuband, K.-H. (Eds.), Soziale Realität im Interview. Campus, Frankfurt, pp. 117-156.

Reuband, K.-H., 1987. Unerwünschte Dritte beim Interview. Zeitschrift für Soziologie 16, 303-308.
Reuband, K.-H., 1992. On Third Persons in the Interview Situation and their Impact on Responses. International Journal of Public Opinion Research 4, 269-274.
Riley, R.D., Higgins, J.P., Deeks, J.J., 2011. Interpretation of random effects meta-analyses. Bmj 342, d549.
Rinsky, J.L., Richardson, D.B., Wing, S., Beard, J.D., Alavanja, M., Freeman, L.E.B., Chen, H.L., Henneberger, P.K., Kamel, F., Sandler, D.P., Hoppin, J.A., 2017. Assessing the Potential for Bias From Nonresponse to a Study Follow-up Interview: An Example From the Agricultural Health Study. American Journal of Epidemiology 186, 395-404.
Rogers, W.H., 1993. Regression standard errors in clustered samples., Stata Technical Bulletin Reprints. Stata Press, College Station, TX, pp. 88-94.
Rosenberg, M., 1965. Society and the adolescent self-image. Princeton University Press,, Princeton.
Rothman, K.J., 2012. Epidemiology : An Introduction, 2 ed. Oxford Univ. Press, New York, NY.
Rothman, K.J., Gallacher, J.E., Hatch, E.E., 2013. Why Representativeness Should Be Avoided. International Journal of Epidemiology 42, 1012-1014.
Rubin, D.B., 1987. Multiple Imputation for Nonresponse in Surveys. John Wiley & Sons, Inc, Hoboken, New Jersey.
Sakshaug, J.W., Couper, M.P., Ofstedal, M.B., Weir, D.R., 2012. Linking Survey and Administrative Records: Mechanisms of Consent. Sociological Methods & Research 41, 535-569.
Sakshaug, J.W., Kreuter, F., 2014. The Effect of Benefit Wording on Consent to Link Survey and Administrative Records in a Web Survey. Public Opinion Quarterly 78, 166-176.
Schafer, J.L., 1999. Multiple Imputation: A Primer. Stat Methods Med Res 8, 3-15.
Schilpzand, E.J., Sciberras, E., Efron, D., Anderson, V., Nicholson, J.M., 2015. Improving survey response rates from parents in school-based research using a multi-level approach. PLoS One 10, e0126950.
Schimpl-Neimanns, B., Weiss, F., 2014. Zur Bereitstellung amtlicher Mikrodaten für die Wissenschaft aus sozialwissenschaftlicher Perspektive. AStA Wirtschafts- und Sozialstatistisches Archiv 8, 205–219.
Schneider, S., 2008. Applying the ISCED-97 to the German educational qualifications. In: Schneider, S. (Ed.), The International Standard Classification of Education (ISCED-97). An Evaluation of Content and Criterion Validity for 15 European Countries. MZES, Mannheim, pp. 76-102.
Schweers, S., Kinder-kurlanda, K., Müller, S., Siegers, P., 2016. Conceptualizing a Spatial Data Infrastructure for the Social Sciences: An Example from Germany. Journal of Map & Geography Libraries 12.
Siddiqi, A., Holmes-Rovner, M., 2006. Factors Associated with Patient Attrition in Clinical Trials. American Journal of Epidemiology 163, S230-S230.
Sigman, R., Lewis, T., Yount, N.D., Lee, K., 2014. Does the Length of Fielding Period Matter? Examining Response Scores of Early Versus Late Responders.
Slymen, D.J., Drew, J.A., Elder, J.P., Williams, S.J., 1996. Determinants of Non-compliance and Attrition in the Elderly. International Journal of Epidemiology 25, 411-419.
Snijders, T.A.B., Bosker, R.J., 1999. Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling. Sage.
Søgaard, A.J., Selmer, R., Bjertness, E., Thelle, D., 2004. The Oslo Health Study: The Impact of Self-selection in a Large, Population-based Survey. International Journal for Equity in Health 3.
Spiers, S., Oral, E., Fontham, E.T.H., Peters, E.S., Mohler, J.L., Bensen, J.T., Brennan, C.S., 2018. Modelling Attrition and Nonparticipation in a Longitudinal Study of Prostate Cancer. BMC Medical Research Methodology 18.
Spieß, M., 2010. Der Umgang mit fehlenden Werten. In: Wolf, C., Best, H. (Eds.), Handbuch der sozialwissenschaftlichen Datenanalyse. VS Verlag für Sozialwissenschaften, Wiesbaden, pp. 117-142.

Spratt, M., Carpenter, J., Sterne, J.A.C., Carlin, J.B., Heron, J., Henderson, J., Tilling, K., 2010. Strategies for Multiple Imputation in Longitudinal Studies. American Journal of Epidemiology 172, 478-487.

Stang, A., 2003. Nonresponse Research - An Underdeveloped Field in Epidemiology. European Journal of Epidemiology 18, 929-931.

Stang, A., Jöckel, K.H., 2004. Studies with Low Response Proportions May Be Less Biased than Studies with High Response Proportions. American Journal of Epidemiology 159, 204-210.

Stang, A., Moebus, S., Dragano, N., Beck, E.M., Mohlenkamp, S., Schmermund, A., Siegrist, J., Erbel, R., Jockel, K.H., Heinz Nixdorf Recall Study Investigation, G., 2005. Baseline Recruitment and Analyses of Nonresponse of the Heinz Nixdorf Recall Study: Identifiability of Phone Numbers as the Major Determinant of Response. Eur J Epidemiol 20, 489-496.

Sterne, J.A., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., Carpenter, J.R., 2009. Multiple Imputation for Missing Data in Epidemiological and Clinical Research: Potential and Pitfalls. Bmj 338, b2393.

Strobl, C., 2012. Das Rasch-Modell : eine verständliche Einführung für Studium und Praxis, 2. erw. Aufl. ed. Hampp, München [u.a.].

Stuart, A.L., Pasco, J.A., Brennan-Olsen, S.L., Berk, M., Betson, A.G., Bennett, K.E., Timney, E.N., Williams, L.J., 2019. Sample Selection and Reasons for Non-participation in the PRedictors and Outcomes of incident FRACtures (PROFRAC) Study. J Public Health Res 8, 1475.

Studer, J., Baggio, S., Mohler-Kuo, M., Dermota, P., Gaume, J., Bertholet, N., Daeppen, J.B., Gmel, G., 2013. Examining Non-Response Bias in Substance Use Research. Are Late Respondents Proxies for Non-Respondents? Drug and Alcohol Dependence 132, 316-323.

Suchman, E.A., McCandless, B., 1940. Who Answers Questionnaires? Journal of Applied Psychology 24, 758-769.

Thompson, S., 1981. Will it Hurt Less if I Can Control it? A Complex Answer to a Simple Question. Psychological Bulletin 90, 89-101.

Tourangeau, R., Yan, T., 2007. Sensitive Questions in Surveys. Psychological Bulletin 133, 859-883.

UNESCO, 2012. International Standard Classification of Education ISCED 2011. UNESCO Institute for Statistics, Montreal, Canada.

Van Buuren, S., 2018. Flexible Imputation of Missing Data, 2 ed. CRC Press LLC, Boca Raton, FL.

Van Buuren, S., Boshuizen, H.C., Knook, D.L., 1999. Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis. Statistics in Medicine 18, 681-694.

Van Buuren, S., Groothuis-Oudshoorn, K., 2011. mice: Multivariate Imputation by Chained Equations in R. Journal of Statistical Software 45, 1-67.

Van Loon, A., 2003. Survey Non-response in the Netherlands Effects on Prevalence Estimates and Associations. Annals of Epidemiology 13, 105-110.

Vinther-Larsen, M., Riegels, M., Rod, M.H., Schiøtz, M., Curits, T., Grønbæk, M., 2010. The Danish Youth Cohort: Characteristics of Participants and Non-participants and Determinants of Attrition. Scandinavian Journal of Public Health 38, 648–656.

Voigtlander, S., Berger, U., Razum, O., 2010. The Impact of Regional and Neighbourhood Deprivation on Physical Health in Germany: A Multilevel Study. BMC Public Health 10, 403.

Wallborn, T., Grafe, N., Quante, M., Geserick, M., Casprzig, N., Ludwig, J., Hiemisch, A., Kiess, W., 2013. Never Plan a Population Based Cohort Study on a Rainy Friday after a Sunny Thursday. European Journal of Epidemiology 28, 285-286.

Wang, J., Tsang, W.W., Marsaglia, G., 2003. Evaluating Kolmogorov's distribution. Journal of Statistical Software 8, 1-4.

Watson, D., 2003. Sample Attrition between Waves 1 and 5 in the European Community Household Panel. European Sociological Review 19, 361-378.

Watson, N., Wooden, M., 2009. Identifying Factors Affecting Longitudinal Survey Response. In: Lynn, P. (Ed.), Methodology of Longitudinal Surveys. John Wiley and Sons, Chichester, pp. 157-181.

White, I.R., Royston, P., Wood, A.M., 2011. Multiple Imputation Using Chained Equations: Issues and Guidance for Practice. Statistics in Medicine 30, 377-399.

Winkler, V., Leitzmann, M., Obi, N., Ahrens, W., Edinger, T., Giani, G., Linseisen, J., Loffler, M., Michels, K., Nothlings, U., Schipf, S., Kluttig, A., Wichmann, H.E., Hoffmann, B., Jockel, K.H., Becher, H., 2014. Response in Individuals With and Without Foreign Background and Application to the National Cohort in Germany: Which Factors Have an Effect? Int J Public Health 59, 555-563.

Wolfenden, L., Kypri, K., Freund, M., Hodder, R., 2009. Obtaining active parental consent for school-based research: a guide for researchers. Aust N Z J Public Health 33, 270-275.

Wolke, D., Waylen, A., Samara, M., Steer, C., Goodman, R., Ford, T., Lamberts, K., 2009. Selective Drop-Out in Longitudinal Studies and Non-Biased Prediction of Behaviour Disorders. British Journal of Psychiatry 195, 249-256.

Wolters, M., 2014. D2.2.2.1 Compilation of Short Local Field Reports on Number of Participants, Non-Response and Survey Roll-Out of Follow-Up Survey by Each Survey Centre - unpublished.