Universität Hamburg

Institut für kriminologische Sozialforschung

Nonresponse to Items on Self-Reported Delinquency
A Review and Evaluation of Missing Data Techniques

Masterarbeit
im Studiengang
Internationale Kriminologie

Vorgelegt von Patrick Präg

Erstgutachter: Prof. Dr. Peter Wetzels

Zweitgutachter: Dr. Dirk Enzmann

Oktober 2007

Table of contents

1. Introduction	1
2. Validity problems in research on crime and delinquency	3
Problems of official crime statistics	3
The development of the self-report method for measuring delinquency	4
Problems of the self-report methodology	5
Explanations for misreporting and nonresponse	6
Over- and underreporting	7
Types of nonresponse	8
Summary	10
3. Types of missing data	11
4. Conventional missing data techniques	15
Listwise deletion	15
Pairwise deletion	17
Mean substitution	17
Regression estimation/mean substitution by subgroups	18
Mean substitution with dummy variable adjustment	19
Ipsative mean imputation	20
Summary	21
5. Modern missing data techniques	22
Expectation Maximization	22
Full-Information Maximum Likelihood	24
Multiple Imputation	25
Summary	28
6. Data and methods	29
Analytical strategy	29
Data	29
Summary	33
7. Analysis of nonresponse	34
Descriptive univariate and bivariate analysis	34
Univariate analysis of the Russian sample	34
Univariate analysis of the German sample	37
Summary and discussion of univariate results	40
Bivariate analysis of both samples	41

Multivariate analysis	43
Data and method.....	44
Results	46
Summary	49
8. Simulation of missing data technique performance under various conditions.....	50
Data used in the simulation	50
Simulated scenarios	50
Strategies and techniques used in the simulation	52
Results	55
Values missing at random on independent variables	55
Values missing at random on independent variables and the dependent variable	61
Values not missing at random on independent variables	66
Values not missing at random on independent variables and the dependent variable	70
Discussion and summary.....	73
9. Using Multiple Imputation to analyze the data set.....	75
10. Conclusion.....	77
11. References	79
Erklärung gemäß § 14 (8) Fakultätsmasterordnung	88

1. Introduction

Missing values are a common problem in any quantitative social science. They complicate data analysis and can be a threat to the validity of study results. Traditional approaches have several shortcomings, such as that they waste data, diminish statistical power, and can yield biased estimates of parameters and standard errors. However, in the past thirty years, remarkable methodological advances have been achieved in the way missing data can be handled (Little and Rubin 2002; Rubin 1976; 1987; Schafer 1997). These have been begun to be used more frequently rather recently, since computational time is not a problem anymore and using these more sophisticated algorithms is not an issue anymore.

The present study focuses on the application of missing data techniques in self-reported delinquency (SRD) research. Although in research on crime and delinquency missing observations are at least as common as in any other social science, not much research has been conducted on the application of modern missing data techniques in research on delinquency (Brame and Paternoster 2002; Raab, Purdon, and Bruckner 2006). One should think that testing these techniques is especially attractive in a field where some respondents might have felt the need to refuse disclosing certain acts they have committed, and thus the analysis of missing values could be very closely related to the actual field of study.

The awareness of the problem of less than optimal missing data techniques has spread in the field of social science research over the past few years and this process is likely to continue. Wilkinson and the APA Task Force on Statistical Inference (1999) have claimed in their guidelines for psychology journals that traditional missing data techniques such as pairwise or listwise deletion are ‘among the worst methods available for practical applications.’ West (2001) has stated that there will be a major focus on missing data techniques in methodological research in this decade. Some assume that missing data techniques will be used routinely in the future (Azar 2002). Peugh and Enders (2004) claim that in education and applied psychology journals missing data were far more often reported in 2003 than in 1999 and attribute this to the risen awareness among researchers for the missing data problem. The yearly number of publications related to the modern missing data technique Multiple Imputation has grown formidably in the last years (Van Buuren 2005).

Acknowledgements. I thank Dirk Enzmann, Barbara Beham, Katharina Schulte, and Roland Verwiebe for their continuous help and encouragement.

Furthermore, modern missing data techniques will be more readily available. Acock (2005) expects that all major statistical software packages will incorporate some sort of multiple imputation routine within the next few years. Horton and Kleinman (2007) give detailed information about which software applications are currently supporting modern missing data techniques.

The present study aims at providing a comprehensive and non-technical review of conventional and modern missing data techniques. It will give an introduction into problems of missing data and assess the performance and applicability of modern missing data techniques when used for solving missing data problems in crime research. In order to do so, a secondary data analysis of a SRD survey among high school students in Germany and Russia will be conducted. The German part of the sample has been analyzed in Wilmers et al. (2002), an analysis of the Russian part and a comparison of both parts can be found in Siegmunt (2005).

This study is set out as follows. The next chapter discusses the problem of validity in crime research to explain the practical relevance of this topic. Chapter 3 gives some definitions necessary to understand the methodological discussion about missing data and ways of handling them. Chapter 4 then reviews conventional missing data techniques and discusses their benefits and drawbacks. Chapter 5 provides an overview about the state-of-the-art missing data techniques, namely Expectation Maximization, Full-Information Maximum Likelihood, and Multiple Imputation. Their advantages and shortcomings are reviewed. Chapter 6 introduces the data set to be analyzed and gives an outline of the analysis to come. Chapter 7 assesses the extent of and possible reasons for nonresponse in order to account for possible bias created by missing values. In chapter 8 the performance of various conventional and modern missing data techniques is simulated under different conditions. The findings of chapters 7 and 8 are used in chapter 9 to establish prevalence rates that account for nonresponse. Chapter 10 concludes this study with a short summary and a discussion of the findings.

2. Validity problems in research on crime and delinquency

The validity of measures is a crucial issue in research on delinquency and crime (Farrington et al. 1996) and on any social problem in general (Best 2001; 2004). This chapter will review two of the main sources of information in the field, official statistics and self-reports. It is set out as follows. The first section offers a short review of the development of the self-report method. Flaws of official statistics are discussed and it is shown that the self-report method marked a major advance in the study of delinquent behavior. The following section explores a number of threats to the validity of the self-report method, namely overreporting, underreporting, and unit nonresponse. Another validity problem, the central issue of this thesis, is extensively discussed in the last section: The problem of item nonresponse and missing data. A concluding section summarizes the main points of this chapter.

Problems of official crime statistics

Although the self-report method for research on delinquency nowadays is well-established, it has come in use rather recently. Early studies on crime and delinquency (e.g., Quetelet 1831) have mostly relied on data generated by official institutions, such as the police, courts, and prisons. These data sources are known to have serious limitations with regard to their validity. First of all, criminal acts have to be discovered in order to find their ways into official statistics. Some crimes remain undiscovered since their victims did not recognize them as criminal acts (e.g., in cases of fraud). In cases of drug trade or drug possession for instance, there are no direct victims. Research has shown that only a fraction of all police-registered crimes were discovered by the police, the majority was reported by citizens (Coleman and Moynihan 1996: ch. 2). Increasing the productivity of the police force (O'Brien 1985) or the size of the police force yields a higher number of reported crimes (Levitt 1998). Whenever the police decide to crack down on until-then hidden delinquency, this can also have a serious impact on official crime statistics (Maguire 2002). A second limitation of official statistics is that most crimes have to be reported to the police in order to get registered. However, not many victims are willing to do so. The decision whether or not to report a crime is highly cognitive and patterned (Skogan 1976), and yields systematical bias in official records (Myers 1980). Victimization surveys show that citizens' willingness to report crimes to the police largely depends on the type of crime. The results of the 2000 *International Crime Victims Survey* state that while nearly every stolen car is reported to the police, the share of reported sexual assaults for instance is on average only 28 per cent (Van Kesteren, Mayhew, and Nieuwbeerta 2000:

ch. 4). Crimes reported to the police have to be recorded there by the police. Even in Germany, where the police adhere to the principle of legality much stricter than in other countries, such as England (Jasch 2004), the police can exercise a certain degree of discretion in their recording behavior of crimes (Kury 2001; Schäfer 1999). Police discretion in crime recording is another potential source of bias, especially since the police is a part of the political system of a state and can become an object of political pressure, as illustrated by Seidman and Couzen (1974).

As discussed previously, official crime rates are subject to several biasing factors, which might even interact in incomprehensible ways. Already in 1963, Kitsuse and Cicourel proposed to interpret official crime rates merely as reflections of institutional processes that produce the statistics rather than reflections of actual behavior. For the study of criminality within a society alternative sources of knowledge are needed.

The development of the self-report method for measuring delinquency

Thornberry and Krohn (2000) state that Sutherland with his ground-breaking work on white-collar crime is credited for giving the impulse towards finding a more valid way of measuring delinquency. Sutherland (1940) challenged the prevailing notion that delinquent behavior was mainly limited to the underclass, although this notion had been backed empirically by official statistics many times before. Sutherland claimed that the correlation between criminal behavior and low socio-economic status was a measurement artifact caused by the fact that official statistics were biased towards members of the underclass.

Sutherland's work fell within a period of time where major changes were happening with respect to the way society, media, and politics perceived and dealt with delinquency. Maguire (2002) claims that a 'data explosion' took place in the second half of the twentieth century and links this to a heightened concern over crime in the public, a more prominent role of crime in the political discourse, the rise of risk management, economization of public services, crime prevention, and a growth in attention to crime victims. These developments spawned the need for new sources of information on the extent of delinquency and thus caused this 'data explosion': new types of knowledge about crime evolved, which were based on self-report and victimization surveys. These approaches attempt to avoid the uncertainties of crime reporting and police recording behavior and to generate a more complete and valid picture of crime.

Short and Nye (1958) pioneered the self-reported delinquency method with their collection of self-report data from high school students and students who attended ‘training schools’ (i.e. correctional facilities for juveniles). They focused not only on the relationship between social class and delinquency (not finding any substantial associations; Short, Nye, and Olson 1958), but also on methodological issues such as reflecting on the validity of self-reports (Short and Nye 1957) and developing a scale for measuring delinquent behavior (Nye and Short 1957). In the context of scale construction in SRD research it should be noted that one should not expect an SRD scale to be internally consistent (Huizinga and Elliott 1986): Criminal behavior is not a one-dimensional domain, and it cannot be expected that the reporting of several criminal acts is highly interrelated, i.e. just because someone steal cars does not mean that they also beat their spouse and do not pay their taxes.

Although there are some early studies using the SRD method in the 1940s (e.g., Porterfield 1943), Short and Nye were the first to pay attention to methodological issues in a systematic manner (Thornberry and Krohn 2000; 2003). Once grounded on a sound methodological foundation, self-report studies on delinquency gained popularity. Another strong argument for their use is their twofold objective: Self-reports are not only capable of estimating prevalence and incidence rates of delinquent behavior within a population (just as official statistics), they also allow for testing theories on criminal behavior and its correlates (Junger-Tas and Marshall 1999).

In the 1960s and 1970s, the SRD methodology was further improved and SRD studies became common for developing and testing etiological theories of delinquent behavior (Coleman and Moynihan 1996: ch. 3; Thornberry and Krohn 2003). Large national and cross-national research undertakings such as the National Youth Survey (NYS; Elliott, Huizinga and Ageton 1985) or the International Self-Report Delinquency Study (ISRD; Junger-Tas, Terlouw, and Klein 1994) were conducted. Kleck, Tark, and Bellows (2006) state that in contemporary crime research, survey methods are far more popular than the use of official data.

Problems of the self-report methodology

The advantages of SRD data over official statistics are obvious. Crimes reported in a survey have not necessarily been discovered by the police or the victims, the victims or witnesses do not have to have reported them to the police, and it does not make a difference whether and

how crimes were recorded by the police. Furthermore, the information on delinquency can be analyzed in regard to other information collected in the survey. This means that actual theories attempting to explain delinquent behavior can be tested with SRD data.

Of course, the advantages of this method come at a certain price. Self-report studies do not only require that respondents are able to answer questions correctly (e.g., to understand the questions and to remember correctly), they also must be willing to do so. In order to enable respondents to answer correctly, rather technical issues such as wording of questions, scale construction, and controlling memory effects have to be considered. These topics are extensively reviewed by Junger-Tas and Marshall (1999), Thornberry and Krohn (2000; 2003), and Tourangeau and McNeeley (2003) specifically for SRD studies. A general introduction into these topics is given by Schaeffer and Presser (2003) and Tourangeau, Rips, and Rasinski (2000). These matters will not be discussed here. The willingness to answer questions on one's own delinquent behavior is at the center of this research, and therefore I will focus on misreporting and especially nonresponse in the following sections.

Explanations for misreporting and nonresponse

The main reason for the occurrence of the problems of nonresponse and misreporting is rather obvious. SRD items have to be considered as sensitive questions that probe for information about socially undesirable behavior, respondents may perceive them as intrusive. In addition, respondents might perceive a risk associated with the possible disclosure of their reports to third parties (Tourangeau, Rips, and Rasinski 2000: ch. 9). According to Tourangeau and his colleagues, respondents' concern about social desirability is determined both by situational aspects and by respondents' personality traits. The dependence on situational aspects becomes visible when similar questions yield different results with regard to their mode of administration and the context in which they were asked. Tourangeau and Smith (1996) for instance observed a higher rate of reported sensitive behavior (use of illegal drugs and number of sex partners) when the mode of data collection was self-administered. Oberwittler and Naplava (2002) found a higher prevalence rate of reported delinquent behavior among juveniles when respondents were questioned in the classroom with a self-administered, written questionnaire rather than in a face-to-face interview at their home. Social desirability can also partially be attributed to respondents' personality characteristics. Some respondents may have a stronger need for social approval than other respondents, causing them to edit their responses in a way

that is socially desired. A classic way of assessing this personality characteristic is the Marlowe-Crowne social desirability scale (Crowne and Marlowe 1960).

However, the concept of social desirability is somewhat problematic since it can lead to paradoxical outcomes when the distinction between respondents' striving for a socially accepted lifestyle and respondents' tendency to answer questions in a way that they deem socially accepted is not accounted for in the instrument that is supposed to measure respondents' social desirability (Enzmann 1995: appendix 5). When this is the case, respondents scoring high on social desirability do report criminal behavior (since they act like they are expected to and respond honestly), while other respondents achieving lower or the same social desirability scores misreport or do not report criminal behavior.

Watkins and Melde (2007) report that nonresponse to single questions is related to respondents' self-control, in particular impulsivity. Impulsive persons are less likely to answer all questions of a questionnaire, since they tend to lack the patience necessary to answer several hundred questions in a row.

When confronted with questions of a sensitive nature, some respondents employ different strategies to cope with them. Sensitive behaviors are misreported systematically. Tourangeau, Rips, and Rasinski (2000: ch. 9) distinguish between two different types of misreporting, namely overreporting and underreporting, and two types of nonresponse, unit nonresponse and item nonresponse.

Over- and underreporting

Typically, respondents underreport socially undesirable behaviors and deny having unpopular or extreme attitudes and overreport the socially desirable ones. Other than criminal behavior, Tourangeau and colleagues (2000) list a number of other commonly underreported activities and data, such as illicit drug use, abortion, smoking, energy consumption, income, or the number of sexual partners (the latter only when the respondent is female). Examples for overreported activities and data are voting, seat belt use, church attendance, and the number of sexual partners (the latter only when the respondent is male). For some of these commonly misreported behaviors such as smoking or drug use, the accuracy can be improved by employing 'bogus pipeline' procedures (cf. Jones and Sigall 1971). Respondents are told that false answers can somehow be detected (regardless of whether this is actually possible). In other

cases, external sources of information can be tapped, such as reliable official records (for comparisons on an aggregate level) or other external validation data (for comparisons on the individual level). But as pointed out at the beginning of this chapter, official records of crime are not highly reliable and thus may not allow such comparison on the aggregate level. In addition, external validation at the individual level is rather difficult and threatened by other methodological problems such as matching (Miller and Groves 1985) or the type of attrition that Köllisch and Oberwittler (2004) for example encountered: About 30 per cent of juveniles of the realized sample refused to have their records (of police contacts) checked, and for roughly another 10 per cent the parents refused to have their offsprings' records checked. Apart from that, there is an unquantifiable chance that the official records are not accurate (i.e. police contacts that actually took place were not recorded by the police). Another approach of external validation data at the individual level is third party data collection. This approach has for instance been applied by Farrington et al. (1996) and is frequently used in clinical child research, where not only children complete questionnaires, but data is also collected from other informants such as parents or teachers. However, research has shown that there frequently exists a large discrepancy between information from different sources (Achenbach, McConaughy, and Howell 1987; De Los Reyes and Kazdin 2004; Eisner and Ribeaud 2007), thus it is questionable whether this approach really is a good source of cross-validation information, especially when the persons on whom research is done have an interest in hiding the behavior that is probed for.

Types of nonresponse

Unit nonresponse occurs when sampled respondents do not participate in a survey at all. It does not pose a problem when samples are large and unit nonresponse is random. Common reasons for nonresponse are refusals, being unreachable, or language problems. A dropping response rate in surveys of all kinds has become a serious issue in the social sciences during the past few decades (Schnell 1997). A large body of literature on this topic emerged (e.g., Goyder 1987; Tourangeau 2004), but it falls beyond the scope of this study. This does not mean that the data used in this study are immune to the problem of unit nonresponse. When surveys are administered in school classes, unit nonresponse can pose a serious threat to the validity of the survey if the nonresponse is due to truancy. Being truant can be a start for a delinquent career and is a serious risk factor for current and future criminal behavior (Huizinga and Jakob-Chien 1998; Loeber and Farrington 2001; Loeber and LeBlanc 1990; Robins and Ratcliffe 1980; Wilmers et al. 2002: ch. 15). Therefore, when a large share of the unit

nonresponse is caused systematically by truancy and not by neutral reasons such as absence due to students' illness, survey results can be seriously biased. For a discussion of the extent of unit nonresponse that occurred when gathering data for this study, see chapter 6.

Partial nonresponse is frequent phenomenon in longitudinal research. Parts of the sample are not available for subsequent interviews, which is referred to as panel mortality or attrition. This phenomenon poses a serious risk for longitudinal research, and hence there is a large body of research on this problem, e.g., Farrington et al. (1990) or Hart, Rennison, and Gibson (2005) on attrition due to respondent fatigue in the *National Crime Victimization Survey*. Although it is not extensively discussed in the literature, partial nonresponse can also occur in the context of cross-sectional surveys, either in the form of break-offs during telephone surveys (De Leeuw, Hox, and Huisman 2003), or in other types of survey such as those conducted in a classroom setting. Respondents may experience fatigue when faced with a long and monotone questionnaire and stop filling out the questionnaire, or they cannot cope with the amount of questions within a limited time span, and thus turn in a questionnaire that is only partially completed, with answers frequently missing towards the end of the questionnaire.

Item nonresponse occurs when single questions in an interview are not answered or entire parts of a questionnaire are not completed. Just as with unit nonresponse, it does not pose a serious problem when samples are large and nonresponse is random. However, drawing from what is known about misreporting, systematic patterns of nonresponse are quite frequent. Apart from problems understanding the question or retrieving the requested information, item nonresponse can also be caused by unwillingness to give an accurate answer and lying. Junger-Tas and Marshall (1999) suggest conducting face-to-face interviews in order to lower rates of partial nonresponse. Despite the fact that face-to-face interviews are very costly, one can still not tell whether the information elicited this way is accurate. Apart from that, item nonresponse can have severe consequences from a data-analytic standpoint. The way how missing values are handled can have a large impact on the results found when analyzing the data. When handled not appropriately, item nonresponse can yield biased sample estimates and increase the risk of type II errors.

Summary

To conclude, we have seen that the self-report method in crime research offers a vast potential for gathering information that is not included in official crime statistics, which have serious validity problems. However, the self-report method also has validity problems of its own. Respondents can misreport requested information or can refuse to answer. Assuming that not much can be done about respondents who willingly misreport information, this study will proceed with item nonresponse due to respondents who refuse to answer single questions. The next chapter will familiarize the reader with assumptions about missing values from a data analyst's perspective and chapter 4 will review conventional missing data techniques that are inappropriate for most situations.

3. Types of missing data

Even when values are missing, a lot can be said about them. The present chapter reviews different types of missing data and underlying assumptions that are presented in the literature. These assumptions intend to guide researchers to see missing values not as mere blanks in their data set, but to choose appropriate analytical strategies in order to achieve precise and unbiased results.

In a seminal paper, Rubin (1976) proposed an initial taxonomy of different missing data mechanisms. They can be seen as probabilistic explanations for why data are missing and represent assumptions under which a particular missing data technique will provide optimal performance. The performance of the various missing data techniques depends crucially on these assumptions (Little 1992). However, it is important to note that all three mechanisms are not mutually exclusive. They can all be present in a single data set (Yuan and Bentler 2003). Schafer and Graham (2002) developed the following graphical representation of the three mechanisms that is reproduced here and that will be referred to throughout this chapter.

Figure 3.1. Graphical representation of data (a) missing completely at random, (b) missing at random, and (c) not missing at random. X stands for completely observed variables, Y stands for partially missing variables, Z stands for the component of the causes of missingness unrelated to X and Y, and R stands for the missingness. (Source: Schafer and Graham 2002: fig. 2)

Data are called *Missing Completely at Random* (MCAR), or, in Rubin's (1976) initial terminology, 'observed at random,' when the probability that causes a value on a variable Y to be missing is independent of the underlying value of Y and also independent of the values of all other variables, such as X (Figure 3.1 a). Although missing values are not necessarily distributed randomly, there is no systematic relationship of the underlying values that are missing to any data values. Hence, when data are MCAR, the missing values can be seen as a random sample of the hypothetically complete data set.

Although the assumptions of MCAR are rather restrictive, there are cases when they are tenable, e.g., values missing by design. In multiple matrix sampling for example, test items and examinees are randomly sampled from their respective populations. For each respondent, a randomly chosen sample of all questions is selected and asked. This procedure has the advantage of much shorter questionnaires, which in turn can lead to lower unit nonresponse, less drop out due to fatigue in the course of an interview, less lapses in respondents' concentration, and more efficient data collection (De Leeuw, Hox, and Huisman 2003). However, this procedure can lead to data sets with large fractions of (intentionally) missing values that can not be analyzed with traditional missing data techniques (Mislevy et al. 1992). The uncertainty introduced by the imputation procedure diminishes the statistical power of the analysis (compared to data gathered in a conventionally designed study with the same sample size).

An example for data being MCAR in a crime research context could be a survey among younger children, where only a random sample of questions of an extensive questionnaire is asked in order to avoid respondents' fatigue and to collect data more efficiently. Apart from data being MCAR due to the design of the study, it is hard to think of examples in crime research where item nonresponse could be completely at random. An example could be a case where the 'sealed envelope technique'* (as described in Becker 2006) is employed, and a number of envelopes is lost. The fact that this example is rather constructed illustrates that MCAR usually is a rather unrealistic assumption in research on crime and delinquency.

Empirical evidence to support the MCAR assumption that missing values on a variable Y are unrelated to the values of all other variables included in the data set can be gained by using two-sample t-tests. When the means of each observed variable differ significantly between groups defined by whether they have missing data on another variable or not, this suggests that the data are not MCAR. In data sets with a large number of variables, this procedure requires a growing number of tests and leads to multiple-comparison problems. Little (1988) developed a global multivariate test of the MCAR assumption to avoid this problem. However, of course, the assumption that a value on X is missing due to the underlying value cannot be tested.

* In a face-to-face interview, the interviewer hands a short self-administered questionnaire with sensitive questions to the respondent who is to hand back the completed questionnaire to the interviewer in a sealed envelope.

Data are called *Missing at Random* (MAR) when the probability of a value on a given variable Y to be missing is independent of the value of Y itself, but depends on the values of other variables, such as X (Figure 3.1 b). Graham, Cumsille, and Elek-Fisk (2003) have proposed that this mechanism should rather be called ‘missing conditionally random,’ in order to distinguish it from the MCAR mechanism more easily. Graham and Donaldson (1993) suggested calling the mechanism ‘accessible’ missingness, since the use of the term ‘random’ in the phrase MAR can be misleading. The MAR assumption is satisfied when the probability of a missing value on a variable Y depends on the respondent’s value on another variable such as X, but within each category of X, the probability of the missing value on Y does not differ. This means that data on Y is missing randomly within subgroups. This assumption is less restrictive than the MCAR assumption and more realistic for the data collected and analyzed in social sciences (Acock 2005).

To give an example in crime research, the probability of missing data on a given SRD item depends on the respondent’s gender, but for both male and female respondents the probability of a missing value on the SRD item is unrelated to the delinquent acts that they were questioned about. As can be easily seen from this example, the MAR assumption can not be tested since the values that are missing are unknown. The only way to record the missing values would be to re-approach the nonrespondents and obtain the missing values from them (Glynn, Laird, and Rubin 1993), and/or to find out about the reasons of nonresponse (Huisman, Krol, and Van Sonderen 1998; Stoop 2004). Usually this is not an option when nonresponse is only partial and the mode of data collection was anonymous or when a secondary data analysis is being conducted.* However, the more auxiliary variables (i.e. variables related to missingness) can be identified and the stronger they are related to missingness, the more plausible is the assumption that data are MAR (Collins, Schafer, and Kam 2001; Schafer and Olsen 1998).

The third mechanism is called *Not Missing at Random* (NMAR), ‘non-ignorable’ (e.g., King et al. 2001), or ‘missing not at random’ (e.g., Schafer and Graham 2002). In this case, the probability that a value on a variable Y is missing may depend on the actual value of Y or that the mechanism that causes the values to be missing can not be found (Figure 3.1 c). This mechanism is usually seen as the least desirable missing data mechanism since it complicates treatment of the missing values and serious bias may occur when data are not treated properly.

* When data collection from nonrespondents is possible in terms of feasibility and cost, it should be done. Of course, when all missing values can be obtained, there is no more need to test the MAR assumption (or to use any missing data technique at all).

An example for this assumption in crime research can be easily found. If respondents who have committed a larger number of delinquent acts are less likely to report these acts, then data is called NMAR.

To conclude, there are three available missing data mechanisms. Although there is no way to distinguish reliably between two of them, i.e. when data being MAR or NMAR, Little's (1988) MCAR test yields some evidence on whether data are MCAR, but it cannot do so with complete certainty. The MAR assumption can be considered to be fulfilled approximatively, the larger the number of auxiliary variables is and the stronger these are correlated with the fact that values are missing. Detecting a mechanism for the missing values is important for the further treatment of the missing data and should be tackled with great care.

4. Conventional missing data techniques

The missing data techniques that are most prevalent in social science research in the past decades (King et al. 2001; Peugh and Enders 2004; Roth 1994) are frequently referred to as ‘traditional’ (Acock 2005; Peugh and Enders 2004), ‘old’ (Schafer and Graham 2002), ‘unacceptable’ (Graham, Cumsille, and Elek-Fisk 2003), ‘ad hoc’ (Jamshidian 2004), ‘naïve’ (Little and Schenker 1995), ‘classical’ (Lüdtke et al. 2007), or ‘conventional’ (Allison 2002). These methods mainly focus on removing missing values and editing data (Schafer and Graham 2002) in order to conduct standard statistical analyses. This chapter provides an overview of the most common techniques, including listwise deletion, pairwise deletion, mean substitution, conditional mean substitution, mean substitution with dummy adjustment, and ipsative mean substitution. Their disadvantages and advantages are discussed and recommendations in the literature are summarized.

Listwise deletion

The procedure known as ‘listwise deletion’ is by far the most common method used in social sciences (Acock 2005; Graham and Schafer 2002; King et al. 2001; Roth 1994). Only complete cases are taken into consideration for analysis, while all incomplete cases are merely discarded. Listwise deletion thus is easy to perform, and there is no special computation necessary and it can be used for any type of statistical analysis.

Listwise deletion is regarded to be a conservative strategy since no data is ‘made up’ (Acock 2005). It is often accepted when data are MCAR (Allison 2002; Little 1992), since the remaining complete cases are a random subsample from the initial sample. Nevertheless, there is an important caveat to be mentioned. Listwise deletion should be used with caution when the sample size is small and the fraction of cases with missing values is relatively large. Discarding the incomplete cases can result in a large loss of data, especially if one considers multivariate models where fractions of missing values can cumulate. Allison (2002: p. 2) illustrates this problem: A data matrix with 20 variables and 1,000 cases can result in a net sample of 360, when each of the 20 variables independently has missing values on 5 per cent of the cases. According to a review of articles in influential political science journals by King et al. (2001), a loss of 50 per cent of cases occurs frequently after listwise deletion. In these cases the power of statistical tests can be severely diminished and type II errors may be the consequence. This means the alternative hypothesis is rejected (instead of the null hypothesis) when

the alternative hypothesis is in fact true (Cohen 1992). King et al. (2001) even state that bias caused by omitting control variables (and retaining more cases) is often preferable to the bias induced by loss of power due to loss of cases. However, when power is not an issue and the MCAR assumption holds sufficiently well, listwise deletion can be a viable missing data technique (Little and Rubin 2002: ch. 3; Allison 2002: ch. 3).

When data are not MCAR, listwise deletion mainly yields biased estimates (although there are exceptions to this rule), since the complete cases are not representative of the population and even point estimators are likely to be biased. Wirtz (2004) claims that using listwise deletion when data are not MCAR usually creates more homogeneous samples with an attenuated variance, which in turn diminishes the size of correlations among the retained cases.

But as mentioned earlier, there are certain conditions when listwise deletion performs well if data are not MCAR. When missingness on independent variables in regression analysis does not depend on the values of the dependent variable, unbiased estimates can be expected (Allison 2002: ch. 3; Little and Rubin 2002: ch. 3). For instance, when predicting the incidence of criminal behavior from scores on scales measuring self-control and attitude towards violence, regression analysis works well when missingness on the predictor variables is interrelated between those variables, but is independent of the incidence of criminal behavior. Graham and Donaldson (1993) report that listwise deletion even yields unbiased estimates when missingness on the dependent variable is MAR and depends on the values of the predictor variables. For logistic regression models where missingness in the predictor variables depends merely on the dependent variable, Vach (1994: ch. 4) states that after listwise deletion only the estimated intercept is biased, while slope coefficients are accurate. However, although listwise deletion can yield unbiased parameter estimates under the aforementioned conditions, one should be cautious because power can become an issue when the number of deleted cases is too large (Graham, Cumsille, and Elek-Fisk 2003; Little 1992).

Listwise deletion might be a viable way of dealing with incomplete data when the fraction of missing values is small, power is not an issue, and, of course, when the MCAR assumption holds (Acock 2005; Graham, Cumsille, and Elek-Fisk 2003; Jamishidian 2004; Little 1992; Roth 1994; Wirtz 2004). Apart from that, listwise deletion can be used for explorative purposes (Lüdtke et al. 2007) or in regression analysis, when the missingness in the predictor variables does not depend on the values of the dependent variable. Allison (2002) even explic-

itly recommends listwise deletion for regression analyses when the probability of missing data on independent variables does not depend on the values of the dependent variable.

Pairwise deletion

Pairwise deletion is a less popular strategy in social sciences (Acock 2005, Roth 1994). It is also called ‘available-case analysis’ in the literature and refers to the deletion of all observations that have missing values on any pair of variables analysed, regardless of whether they have missing values on other variables. This means that a covariance matrix is calculated including all available cases that have values on any pair of variables. Any subsequent analysis will be based on this covariance matrix. A large amount of information can be retained in comparison to the use of listwise deletion (Acock 2005; Roth 1994; Wirtz 2004).

Nevertheless, pairwise deletion has a number of undesirable features. Covariances in one covariance matrix might be based on completely different parts of a sample, which in turn results in a covariance matrix that is not positive definite, and thus makes it impossible to conduct common statistical analyses such as linear regression. This can even be the case when data are MCAR. In addition, if covariances in multivariate analysis are based on different, but overlapping subsets of the initial sample, the population about which inferences are to be drawn becomes unclear. It is not easy to determine the correct number of degrees of freedom since the correct number of cases a model is based on cannot be determined (Acock 2005; Lüdtke et al. 2007). Furthermore, pairwise deletion can only be used for analytic strategies based on covariance matrices and it relies on the MCAR assumption to yield unbiased parameter estimates.

While Roth (1994) recommends using pairwise deletion when data are MCAR, others recommend to refrain from using it generally (Allison 2002; Acock 2005; Graham, Cumsille and Elek-Fisk 2003; Lüdtke et al. 2007).

Mean substitution

Mean substitution, also known as ‘unconditional mean imputation,’ is a simple and appealing technique that is frequently used in many different variations (Jinn and Sedransk 1989). It is based on the assumption that the best estimate for a missing value a respondent has on a given variable is the mean of the present observations for that variable. For example, when some respondents fail to report the number of their delinquent peers, the average number of delin-

quent peers of all respondents who answered the question replaces the missing values. After replacing the missing values by means, the data matrix is no longer incomplete, and any statistical analysis can be carried out.

Mean substitution has several disadvantages. Although the mean of a distribution does not change by replacing missing values with the mean, this procedure changes the shape of the distribution. Furthermore, the variance of a distribution will be reduced since more cases have the exact same value, which in turn reduces the covariance with other variables (e.g., Peugh and Enders 2004). When multiple variables have different shares of missing values, their variances are attenuated at different degrees, thus while most relationships will appear to be weaker, some will be artificially increased (Acock 2005). In addition, sample size is overestimated when using mean substitution, therefore the risk of type I error increases. In the case of mean substitution the mechanism that causes the data to be missing can be neglected, since mean substitution tends to bias all parameters except for the mean, it does not matter whether data are MCAR, MAR or NMAR (Peugh and Enders 2004).

Due to its numerous and serious drawbacks, mean substitution is generally not recommended in the literature.

Regression estimation/mean substitution by subgroups

Regression estimation, or as it is also called, ‘conditional mean imputation,’ is a variation of mean substitution. Instead of plugging in the same mean for every respondent, the mean for each unit is estimated by a regression equation using the complete cases (e.g., Buck 1960). For instance, when confronted with missing answers on the question how many delinquent peers a respondent has, a researcher would come up with predictors for the number of delinquent peers, such as gender or socio-economic status. With these auxiliary variables the researcher would estimate the number of delinquent peers for each respondent by conducting a regression analysis for the complete cases, and then substitute missing values with the estimates. According to Allison (2002), regression estimation produces unbiased estimates when data are MCAR, sample size is large, and estimation is merely based on independent variables, and not on the dependent variable. When the latter condition is not fulfilled, correlations can be massively overestimated since all plugged-in values lie on one regression line.

But even when all missing values could be substituted perfectly in a sense that they would preserve all correlations between all variables and the distribution of each variable, similar problems as with simple mean substitution remain. Subsequent statistical analyses yield too small standard errors since standard analyses do not account for the uncertainty caused by missing values (Allison 2002: ch. 3; Jamishidian 2004; Lüdtke et al. 2007; Schafer and Olson 1998). After all, the substituted values are nothing more than estimates of the unknown missing values. Schafer and Schenker (2000) suggested solving this problem by using techniques to correct the size of standard errors. However, Lüdtke et al. (2007) claim that such techniques are not practical to use when conducting multivariate analyses. And even when the size of standard errors is correct, the caveat mentioned in the paragraph above still applies.

While some researchers advocate the use of regression estimation under certain conditions (Allison 2002; Acock 2005; Roth 1994), the majority of scholars, such as Graham, Cumsille and Elek-Fisk (2003) and Lüdtke et al. (2007) do not recommend mean substitution and regression-based single imputation.

Mean substitution with dummy variable adjustment

Another variation of mean substitution that is intuitively appealing yet potentially harmful is mean substitution with dummy adjustment. This procedure is also known as ‘indicator variable adjustment’ or the ‘missing-indicator method.’ Its popularization is widely credited to Cohen and Cohen’s (1983) classical textbook. The basic idea is strikingly simple: Missing values on a predictor variable in a linear regression model are simply replaced by the overall mean. Then a dummy variable is added to the model having the value 1 when a missing value has been replaced by the mean, and being 0 when a value has not been missing. The coefficient of the dummy variable should then be interpreted as the difference in the predicted values between the group of respondents with missing data on a given variable and the group of respondents without missing values. Schafer and Graham (2002) point out that this strategy is merely a redefinition of the problem and not a solution, since this strategy yields biased estimates even when data are MCAR (Allison 2002; Jones 1996). Acock (2005) further admonishes that—as with the other mean substitution-based techniques—power is generally overestimated, and that multiple indicator variables will lead to a multicollinearity problem when the presence of missing values on one variable is related to missing values on other variables. The use of dummy variable adjustment, however, can be appropriate, when the missing values are non-existent (Allison 2002), for instance, when prison experiences are used among other vari-

ables to predict future criminal behavior. Respondents that have never been to prison certainly cannot provide answers to questions aiming at this topic, but a dummy variable in a regression model indicating whether a respondent has not been in jail and therefore did not give an answer would yield unbiased estimates.

Ipsative mean imputation

Ipsative mean imputation (Schafer and Graham 2002) or person mean imputation (Huisman 2000) is a variation of mean substitution that is widely used when calculating values for scales. Scales are used for measuring latent properties that cannot be measured by a single question, but have to be measured using multiple items. Ipsative mean imputation now means that whenever a respondent fails to answer all items that form a scale, the value of the scale will not be set missing, but instead all available items are averaged. Schafer and Graham (2002) admonish the lack of knowledge and even awareness about the consequences of this procedure among researchers. However, in simulation studies by Huisman (2000) and Roth, Switzer, and Switzer (1999), other simple methods, such as hot-deck imputation, do not perform clearly superior to ipsative mean imputation. The reason that this type of mean substitution performs better than the general type is that the mean being substituted is the mean of the nonmissing variables, not that of nonmissing cases. Thus, the missing values are not replaced by the same mean, but by the mean of each individual case. As a consequence, variance of variables with substituted values is attenuated substantially less (Graham, Cumsille and Elek-Fisk 2003). However, it leads to an overestimation of the internal consistency of a scale, especially since scales based on fewer items are supposed to have a lower reliability (Huisman 2000). A more sophisticated variation of ipsative mean imputation can even be used when difficulties of the scale items differ (Sijtsma and Van der Ark 2003), but this technique is under suspicion of artificially deflating Cronbach's alpha (Enzmann 2005). Schafer and Graham (2002) state that a prerequisite to using person mean imputation are that reliability of a scale should be high ($\alpha > .70$) in the first place and that each scale to be averaged should be one-dimensional. However, Schafer and Graham (2002) recommend that more principled approaches such as multiple imputation (to be reviewed in the next chapter) should be preferred when dealing with scales. One last caveat: person mean imputation and related techniques assume that missing values substituted this way are parts of scales, and as laid out in Chapter 2, this assumption does not hold when criminal behavior is being measured.

Summary

Many conventional missing data techniques have potentially serious shortcomings, mainly concerning power and biased parameter estimates. However, there are some remarkable exceptions, namely the applicability of listwise deletion in regression analysis when missingness on predictor variables is not related to the dependent variable even when data are not MCAR and the good performance of ipsative mean imputation in comparison to other simple methods. The next chapter reviews modern and more promising missing data techniques.

5. Modern missing data techniques

This chapter introduces modern alternatives for working with missing values to the classical and partly inappropriate techniques discussed in the last chapter: Multiple imputation (MI) and two maximum likelihood approaches, the Expectation Maximization (EM) algorithm and Full-Information Maximum Likelihood (FIML) estimation.

According to Schafer and Olsen (1998), the formalization of the EM algorithm at the end of the 1970's was a symptom of the fundamental shift that occurred in the way incomplete data were seen. Rather than being regarded as a nuisance of which one was to 'get rid' (i.e., case deletion and substitution), incomplete data began to be seen as a source of variability to be averaged over, namely by more advanced missing data techniques.

All techniques described in this section are known to perform well as long as the MAR assumption is tenable. Instead of editing data or attempting to recover the missing observations, they attempt to make valid and efficient inferences (Schafer and Graham 2002). In order to assure valid inference, the correct degree of randomness has to be injected into the imputed values and this uncertainty has to be regarded when standard errors are calculated.

Expectation Maximization

The Expectation Maximization (Dempster, Laird, and Rubin 1977) algorithm is a modern missing data technique that produces asymptotically unbiased point estimates of means, variances, and correlations when data are MCAR or MAR. In some implementations, e.g., the MVA module of SPSS, it is also used to impute missing values in the data set.

EM is an algorithm for finding the maximum-likelihood estimates of the parameters of an underlying distribution given the data and it is a variation of regression imputation. The logic behind the EM algorithm is easy to understand. If one knew the missing values, then one could estimate the parameters of the data model. And: If one had information about the parameters of the data model, then one could predict the missing values. This recursive logic suggests an iterative approach as it can be found in the EM algorithm (Schafer and Olsen 1998): The values are imputed iteratively until subsequent imputations yield sufficiently alike parameter estimates. EM consists of two different steps, the E(xpectation) step imputes the

missing values using regression imputation, and the M(aximization) step calculates means and covariances.

In the E-step, all missing values are replaced by their regression estimates, this means that for every variable with missing values a regression model is calculated with the said variable as dependent variable and all other variables as predictors. It is not important which traditional missing data technique is used for this step (Peugh and Enders 2004), thus listwise deletion is a viable option. The missing values are replaced with their estimates plus a residual term which makes up for the loss of residual variability that occurs when regression estimation is used. For instance, this residual term can comprise regression residuals (Allison 2002: ch. 4).

In the M-step, means and covariances are calculated by maximizing the likelihood of the parameters given the data, based on the data set imputed in the E-step. In the subsequent E-step, new imputations are created based on the parameter estimates from the M-step, which means in practice that the estimated covariances are used for the regression models predicting the missing values. The algorithm converges (i.e. this sequence of E and M-steps) stops, when after several cycles of E and M-steps the parameter estimates do not substantially differ.

EM is said to converge rather quickly, but it will take longer the more variables there are (Acock 2005) and the larger the fraction of missing data is (Little and Rubin 2002: ch. 8). Schafer and Olsen (1998) recommend inspecting the loglikelihood function in order to monitor the convergence process and switching the starting values to check whether the loglikelihood function does have a single maximum, this is especially helpful when the sample size is small, the fraction of missing values is high, and the number of parameters to be estimated is large. Slow convergence is a warning sign that the inferences are highly dependent on the assumptions made about the missing values and one might think about omitting variables with very high rates of missingness from the model in order to improve convergence and the quality of the estimates (Schafer 1997: ch. 3).

Apart from maximum-likelihood estimated means and covariances for the incomplete data set, the EM algorithm creates a single rectangular data set yielding these parameters. The means and covariances can be used for further analyses, e.g., regression analysis etc. Graham, Cumsille, and Elek-Fisk (2003) recommend using it as a default for exploratory factor analysis. However, the use of the imputed data set is deemed inappropriate by some (e.g., Fichman

and Cummings 2003). Furthermore, Von Hippel (2004) warns of the SPSS implementation of EM: The data set thus created has no random variation incorporated into the imputed values, hence parameter estimates based on this data set will probably be biased. When planning to create EM imputed data sets, Acock (2005) therefore recommends using the program 'Em-cov' by Graham (1995), where residual variance can be added to the imputed values.

In sum, EM is a powerful algorithm that can yield an unbiased and efficient variance-covariance matrix whenever data are MAR and inferential analysis is not desired (Acock 2005; Allison 2002: ch. 4; Lüdtke et al. 2007). Wirtz (2004) claims that EM can even be robust against moderate departures from MAR. When the parameter matrix obtained from EM is used for calculating standard errors, these will generally be deflated. Roth (1994) admonishes that the material explaining the ML-based EM algorithm is so complex that he counts it as an explicit disadvantage of EM. And furthermore, EM is based on formal distributional assumptions, i.e. the observed data are a random sample drawn from a population with a multivariate normal distribution (Allison 2002: ch. 4). Although Allison also states that EM/ML can be robust against departures from this assumption, the assumption of multivariate normality is frequently questionable when analyzing data in researching crime and delinquency.

Full-Information Maximum Likelihood

The EM algorithm yields maximum likelihood estimates when performing rather simple analyses such as a multiple-regression analysis. However, for more complex analyses such as structural equation models with latent variables, a correlation matrix obtained from EM should not be used (Graham, Cumsille, and Elek-Fisk 2003). Instead Full-Information Maximum Likelihood should be used. The technique is available in programs such as Mplus, Amos, HLM, or Lisrel.

FIML is a model-based missing data procedure (Graham, Cumsille, and Elek-Fisk 2003), it does not actually impute missing values, but rather uses all information present to calculate maximum-likelihood estimates of parameters. Both missing data and data analysis are handled at the same time. Maximum-likelihood is a general estimation procedure that is also used for data sets without missing values. The underlying logic is that parameter estimates should maximize the probability of observing what actually has been observed. To achieve this, both observed data and assumptions about the distribution of the variables in the population have to be present. In an iterative process it is then tested, for which population parameters the data

are the most likely to have occurred, i.e. the likelihood function of the parameter has to be maximized (Allison 2002: ch. 4). Maximizing the likelihood is still possible even when not every case has complete information (Peugh and Enders 2004). FIML estimation makes it unnecessary to edit the data or to discard incomplete cases before running an analysis and during the estimation process even information from incomplete cases is utilized. It also accounts for the uncertainty induced by the missing values by increasing the standard errors (Allison 2002: ch. 4).

Nonetheless, there are disadvantages to this technique. Just as EM, it is based on the assumption of multivariate normality, and a simulation study has shown that FIML tends to give biased results when data are nonnormal (Enders 2001). Specialized software is needed to conduct the analysis and Lüdtke et al. (2007) admonish that the use of auxiliary variables, i.e. variables that can be mechanisms for missingness, is cumbersome. FIML relies on the MAR assumption, but Roth (1994) recommends it for use with NMAR data, even when the fraction of missing values is large.

Allison (2002: ch. 4) states that FIML is especially suitable when the sample size is large. FIML is recommended in particular for MAR data when linear models (Allison 2002: ch. 4) or, more specifically, structural equation models (Lüdtke et al. 2007) are estimated. Allison (2002: ch. 5) admonishes that for other applications of FIML such as ordered logistic regression, either theory or software is unavailable, which complicates data analysis substantially.

Multiple Imputation

Multiple Imputation is, just like EM, a data-based missing data technique (Graham, Cumsille, and Elek-Fisk 2003). It seeks to find solutions to the problems many conventional missing data techniques have: Many approaches cannot deal with data that are not MCAR, some attenuate variance and under- or overstate covariance, and most do not account for the uncertainty created by the lack of information.

The general idea of multiple imputation is easy to understand: First, missing values are imputed in the data using an appropriate imputation model that incorporates random variation. Second, this procedure is repeated M times, thus creating M imputed data sets. Third, the desired analysis in each of the imputed data sets is carried out using the standard methods for

complete data sets. Fourth, combine the resulting estimators using the rules stated by Rubin (1987: ch. 3).

A number of algorithms have been proposed to accomplish the first step. Schafer (1997) proposed an algorithm called Data Augmentation (DA). Just as EM, DA is a two-step algorithm. It starts off with the parameters obtained from EM. In the first i-step, DA imputes the data set based on this set of parameters, and in the next p-step new parameters are drawn from a Bayesian posterior distribution comprising the observed and imputed data. The next step is again an i-step where the data are imputed using the new parameters (Schafer and Olsen 1998). The more steps are carried out, the less the drawn parameters (and the imputed data sets) depend on one another, thus resembling random draws from one population (Graham, Cumsille, and Elek-Fisk 2003). After a number of steps, the algorithm converges in a sense that the imputed data sets do not depend on each other. Then an imputed data set can be saved, the DA algorithm starts until again convergence is reached and another data set can be saved. This procedure is repeated until the desired number of imputed data sets is obtained. As attractive as this method seems, it still requires a multivariate distribution to be specified for all the variables, but there is some evidence that MI also performs well when this assumption is not tenable (Allison 2002: ch. 6).

An alternative approach is Multiple Imputations by Chained Equations (MICE), which was popularized by Van Buuren, Boshuizen, and Knook (1999). Here, for each single variable with missing values a separate, appropriate regression model is chosen and variability is created by using draws from the posterior predictive distribution (Royston 2005b).

Van Buuren, Boshuizen, and Knook (1999) give recommendations which variables should be included in model that should impute missing values. Those variables that will be in the analytical model, factors known to be associated with missingness and factors that explain a considerable amount of variance for the target variables. Collins, Schafer, and Cam (2001) have shown that it can be beneficial to include theoretically irrelevant variables into MI models as well, as long as they are correlated to the variables containing missing data. According to the simulation results of Collins and colleagues, the liberal use of auxiliary variables reduces the possibility of omitting relevant causes of missingness by mistake and is able to reduce the bias created by missing data mechanisms that cannot be completely included in the missing data model.

Allison (2002: ch. 6) states that it is essential to include the dependent variable of an analytical model into the imputation model. When single imputation that draws on the dependent variable is used for substituting missing values, regression coefficients can be inflated. However, due to the random component in the imputation process, this tendency is neutralized.

In the second step, the preceding step is repeated a number of times, thus multiple imputed data sets are created. There is some discussion on the necessary number of imputations. For normal applications, a number of five imputed data sets performs well and gives highly efficient estimates when the MAR assumption is tenable, many authors state (Allison 2002: ch. 5; Rubin 1987: ch. 4; Schafer 1997: ch. 6). They all agree that increasing the number of imputations to ten or 20 usually improves efficiency only marginally. However, Raab, Purdon, and Bruckner (2006) recommend using 20 to 50 imputed data sets to improve estimates of standard errors, Royston (2004) also recommends using rather large numbers of imputed data sets.

In the third step of MI, the desired analysis is conducted with each of the data sets created in the second step. What sounds like a tedious task, can in fact be a tedious task when Schafer's highly endorsed program 'Norm' (1997; 1999; cf. Graham, Cumsille, and Elek-Fisk 2003 or Schafer and Olsen 1998 for detailed instructions) is used. Some standard statistical software packages such as SAS and Stata, however, provide routines that simplify working with multiply imputed data sets. Nonetheless, when planning to use Norm, it is recommendable to use the corresponding R routine provided by Schafer. Also available for R is an implementation of MI using chained equations by Van Buuren and Oudshoorn (2000), and a package by Honaker, King and Blackwell (2007) called 'Amelia II', that is based on an algorithm similar to DA called Sampling Importance/Resampling (SIR). Current versions of SAS also offer a DA implementation of MI. Acock (2005) recommends the use of Royston's program 'ice' (2004; 2005a; 2005b), which implements the use of MI by chained equations for Stata.

In the fourth step, parameter estimates are pooled across the various imputed data sets using the formulas provided by Rubin (1987). Parameter estimates are simply the mean of all respective estimates. Standard errors are calculated by averaging the respective standard errors and adding the variance between all respective standard errors. When all imputed data sets yield very similar results, the thus obtained standard errors are rather small. When they differ, this uncertainty is incorporated into the parameter estimates by adding the variance of the

standard errors across imputations. As discussed below, there is some uncertainty on how many imputations are necessary to obtain accurate standard errors, but usually there is nothing wrong with choosing the largest number that is computationally feasible.

To conclude, MI has some drawbacks, such as that some implementations require the assumption that data are multivariate normal. Furthermore, MI is not necessarily robust against departures from MAR. The number of imputations needed to achieve unbiased standard errors is still disputed. Although when the number is large, the standard errors are likely to be accurate.

Summary

Modern missing data techniques all require MAR data to perform well. Expectation maximization is a maximum-likelihood algorithm yielding a variance-covariance matrix that is suitable for descriptive analyses, since usually no standard errors are provided. Full-Information Maximum Likelihood is a technique also based on the assumption of multivariate normality. It yields approximately unbiased and efficient parameter estimates and accurate standard errors. It tends to require special software and is not readily available for all applications. Multiple Imputation also yields good results when conducted correctly. Furthermore, implementations of MI are somewhat more widespread and not each implementation is based on the assumption of multivariate normality.

6. Data and methods

This chapter introduces the data set to be used in the analysis—a self-report survey among Russian and German high school student— and gives an outline of the analytical strategy for the following chapters.

Analytical strategy

In order to analyze possible bias caused by nonresponse in the present data set, the extent of nonresponse is studied with a thorough, descriptive univariate and bivariate analysis. Biased estimates can occur when those respondents especially prone to commit delinquent acts are more likely to not respond and only complete cases are used during the analysis. In order to test whether factors influencing nonresponse found in the bivariate analysis have a significant effect even when the influence of other, potentially confounding variables is controlled for, a multiple logistic regression model is estimated. By doing so, the need for accounting for the nonresponse during the analysis is to be assessed. In a following step, the complete cases of the present data set are used to simulate the performance of several missing data techniques under various conditions. In the last step of the analysis, an exemplary analysis of the original data set is carried out to estimate the prevalence rates of delinquent behavior using missing data techniques that performed well in the simulation. All methods will be explained in greater detail in the respective chapters.

Data

This study draws on data collected in the year 2000 among ethnic German and Russian high school students of about 14 years of age. Data were gathered in classroom surveys in one Russian, one East German, and three West German cities. The survey in German cities was conducted by the Criminological Research Institute of Lower Saxony (KFN). The survey in a Russian city was the result of a cooperation by both the KFN and the FH Merseburg and was organized by the Volgograd State Pedagogical University.

The data used in this study have been analyzed in previous publications. The German part of the sample was used in Willmers et al. (2002) and Enzmann, Brettfeld, and Wetzels (2003), and both the German and the Russian sample were used in Enzmann and Siegmunt (2005) and Siegmunt (2005). Therefore, for more detailed information see Willmers et al. (2002: ch. 2 and

appendix 1) concerning the German part of the sample, and Siegmunt (2005: ch. 4) for in-depth information on the Russian part.

The survey was designed to provide quantitative information about juvenile delinquency and the social conditions of juveniles' lives. The aims of the project were to establish the prevalence and incidence of both victimization and delinquent behavior among the juvenile respondents as well as possible correlates of these and to allow for a comparison between Russian and German juveniles.

Data were gathered by means of pencil-and-paper questionnaires of about 25 pages in length. The Russian and German questionnaires differed in the topics covered and sometimes in response categories for identical questions, but were sufficiently similar in the core parts to be used in this analysis. Among the topics covered in the survey are self-reported delinquency (using an adaptation of Lösel 1975), victimization, self-control (using an adaptation of Grasmick et al. 1993), relations to peers, attitudes towards violence (cf. Siegmunt 2005: ch. 8; Wilmers et al. 2002: ch. 9), fear of crime, empathy, school climate, substance abuse, and parenting styles.

Russian school authorities deemed some questions in the Russian questionnaire to be provocative and incomprehensible and particularly demanded that one question was removed from the questionnaire. The questionnaire was changed as requested. However, in two instances instructions for skip patterns became erroneous: Respondents were asked to skip questions that they in fact should have answered. In one instance respondents who stated that they had never been victims of armed assault were asked to skip the question on whether they had fallen victim to unarmed assault (Siegmunt 2005: ch. 6). In another instance respondents who stated that they spend their leisure time with a group of peers were instructed not to answer the question on how often they meet. Those respondents stating that they do not spend their leisure time with a group of peers were directed not only to ignore said question on the frequency that they meet, but to ignore a whole section of questions. Unfortunately this section included not only questions on the nature of their relationship to a group of peers, but also Grasmick et al.'s (1993) self-control scale. Thus, errors in the questionnaire inflated nonresponse in the Russian sample, namely in regard to victimization of unarmed assault and self-control.

The aim of the sample design was to obtain a representative sample of ninth-grade students^{*}. The sampling frame consisted of all ninth-grade classes in the West German cities Hamburg, Hanover, and Munich; the East German city Leipzig, and the Russian city Volgograd. Sampling strata were created based on the school types and a random sample was drawn.

The survey was conducted in Germany in the first three months of 2000 and in Russia in April and May of the same year. It was administered in classroom settings by trained interviewers. They instructed students how to fill out the questionnaires and informed them that participation was voluntary and that the data gathered would be anonymized. Students had two lessons time to complete the questionnaire. Teachers were present during the interviews, but could not see the completed questionnaires.

Tab. 6.1: Sample coverage

	German sample	Russian sample
Gross sample	14,526	2,765
Ineligibility/refusals by teachers	-3,316	-457
Adjusted sample	11,210	2,308
Absent students	-1,011	-438
Present students	10,199	1,870
Refusals by students or parents	-267	-103
Respondents	9,932	1,767
Unusable questionnaires	-131	-20
Realized sample	9,801	1,747
Response rate based on ad- justed sample	87.4%	75.7%

Source: Wilmers et al. (2002: appendix 1) for the German sample and Siegmunt (2005: ch. 4) for the Russian sample.

For the German sample, Wilmers et al. (2002: ch. 2) claim that the coverage rate based on the adjusted sample can be seen as very good (87.4 per cent). Most coverage problems were associated with rather marginal school types such as special education schools, which had to be removed from the sample altogether (and are not included in Table 6.1) due to the undercoverage. However, Wilmers et al. (2002: appendix 1) state that this non-participation can be considered mostly neutral. Non-participation in these school types was mostly due to teachers who refused to have class time used for interviews since they claimed that class time was too scarce already. About ten per cent of students were absent on the day of the interview, 267 did

^{*} Both in the Russian and German school system, ninth-grade students are about 14 years of age.

not participate due to refusals (either by themselves or by their parents), and 131 questionnaires were excluded from the analysis since they were not filled out correctly, e.g., respondents who obviously had severe language problems or who had filled in random patterns. The overall response rate is 67.5 per cent of the gross sample and 96.1 per cent of the students who attended school on the day of the interview. For the analysis to come, respondents that are not ethnic Germans, 24.6 per cent of the realized sample, will be removed from the sample.

For the Russian sample, Siegmunt (2005: ch. 4) states that the reasons for non-participation of entire classes are mostly independent of the topics covered in the survey, e.g. one school was shut down, but in five cases no reasons were given for non-participation. A rather large proportion of students (19 per cent, compared to 9 per cent in the German sample) was absent on the day the survey was conducted in their classes. Another 103 students did not participate in the survey due to refusals. 20 questionnaires had to be excluded. This yields overall response rates of 63.2 per cent of the gross sample, 75.7 per cent of the adjusted sample, and 93.4 per cent of all students present on the day of the interview. Siegmunt (2005: ch. 4) states that this coverage rate may be called very good as well.

The sample to be used in the analysis is comprised of 7,231 ethnic German respondents and 1,747 Russian respondents. Respondents range from 13 to 18 years of age, however, the vast majority in Germany (96.4 per cent) ranges between 14 and 16 years and in Russia (98.2 per cent) between 14 and 15 years. German students are about half a year older than their Russian counterparts ($M_{\text{German}} = 15.1$; $SD_{\text{German}} = .7$; $M_{\text{Russian}} = 14.6$; $SD_{\text{Russian}} = .5$).

In the German sample, 50.3 per cent of the respondents are male and 49.7 per cent female, and in the Russian sample 45.3 per cent of the respondents are male and 54.7 per cent are female.

In order to allow a comparison of the different school systems, the dichotomy proposed by Siegmunt (2005: ch. 4) was used. Russian students attending high schools such as *lyceums* or *gymnasiums* are combined into one category denoting a higher educational level, while students attending other school forms such as *obscheobrazovatelnye* schools or schools offering intensive studies in certain subjects are put together into a second category of lower educational level. German students attending a *gymnasium* are categorized as having a higher edu-

cational level, while students attending other school forms are categorized as having a lower educational level. According to this categorization, 60.6 per cent of the German students are on the lower level of education, just as 66.8 per cent of the Russian students.

Summary

In the year 2000 a classroom survey on self-reported delinquency and topics issues was conducted among Russian and German high school students of about 14 years of age. Respondents were selected in a stratified random sampling process. The overall response rate was high in both countries. For some items in the Russian sample, however, item nonresponse was inflated due to errors in the questionnaire. Furthermore, German respondents are on average six months older than the Russian students and male students are somewhat underrepresented in the Russian sample. Also, the rate of students on a lower educational level is somewhat lower in the German sample.

In order to analyze nonresponse in this data set, a variety of techniques will be used: A simple descriptive analysis will be followed by a multivariate analysis that allows for controlling additional variables. Then a simulation is conducted to assess performance of various missing data techniques under various conditions will be conducted. Furthermore, missing data techniques will be used to analyze the data set, correcting for the nonresponse.

7. Analysis of nonresponse

In this chapter, the overall extent of nonresponse in the sample is studied in order to account for possible bias created by missing values. For instance, when respondents who exhibit many risk factors of delinquent behavior are less likely to complete the survey, it is possible that the extent of delinquency of the population will be seriously underestimated when only complete cases are analyzed. In order to study nonresponse in this survey, first a univariate and bivariate analysis will be conducted and then a multivariate analysis, to verify and focus the findings of the earlier analysis.

Descriptive univariate and bivariate analysis

In order to assess the extent and development of nonresponse during the process of completing the questionnaire, all variables in the data set were arranged in the order they occurred in the questionnaire. Since the questionnaires used in Russia and Germany differ markedly, especially in the order of questions, this had to be done separately for each sample.

Each variable was displayed on an axis in the order it had been placed in the questionnaire, then the rate of missing values of each variable was plotted, resulting in the graphs displayed as Figures 7.1 and 7.2. A few variables that only applied to few respondents due to filters were omitted from these plots.

Univariate analysis of the Russian sample

The Russian sample comprises 1,747 respondents. For the first 150 items, the graph shows a normal pattern, it slightly oscillates near the Y-axis, meaning that there is hardly any nonresponse, and probably not in a systematic manner.

Figure 7.1: Rate of missing values in the Russian sample in the process of completing the questionnaire. X-axis displays the sequences of times in the questionnaire, Y-axis the share of missing values (0 per cent to 40 per cent).

The short ridge between variables 165 and 169 is due to the mistakes in the questionnaire mentioned in chapter 6: Respondents stating that they have never been victims of armed assault are not only asked to skip questions probing for details about an armed assault, erroneously they are also asked to skip the question about unarmed assault. Variables 187 to 189, the next peak of the graph, contain the information about the respondents' last grades in Russian, math, and sports. This could be explained by the assumption that students consider this topic to be sensitive and thus are reluctant to give information about this.

At the following peak, the variables (206 to 208) contain information about the respondents' living situations. Students are asked whether they have to share the room they are sleeping in. Response options are 'In our apartment I have got a room of my own' or 'I share my room with ____ siblings/step-siblings and ____ other persons'. The variable that refers to the number of other persons the respondent is living with in their room has 401 missing values, while the variable referring to the number of siblings has only 58 missing values. This could be explained by the assumption that a number of respondents, when faced with a numeric open-end

question do not fill in a value when the answer they intend to give is actually zero, since these questions do not apply to them^{*}. Furthermore, respondents were not explicitly instructed to fill in zeros when faced with such questions. A fact that backs this assumption is that the average number of reported siblings is 1.2 (SD = .68), while the average number of other persons is only .44 (SD = .79). The same problem assumption could explain the large rates of missingness for the next two variables, which are also based on a numeric open-ended question: 'How many older and how many younger siblings or step-siblings do you have? I have ____ younger (step-) siblings and ____ older (step-) siblings.' Both variables have 347 or 324 missing values respectively. The assumption that respondents did not refuse to answer, but meant to say zero seems more plausible than other possible assumptions such as the information probed for is considered sensitive by the respondent or the respondent does not know the requested information.

The few smaller peaks (variables 212 to 214 and 222) relate to questions about religious beliefs, namely one's own and parents' religious affiliation. They could be related to the fact that in Russia, which has a large share of the population being irreligious and a history that features periods where religious groups were suppressed, children might not know exactly about the information probed for or might consider the information as sensitive.

The sequence between variables 228 and 255 where the graph oscillates is due to the fact that respondents were faced with two parallel matrix questions. They were asked not only about their own opinion on statements describing Russian society, they were also to report their parents' opinion on the said statement. The latter question has frequently been left unanswered, yielding the peculiar pattern. It should be noted that the graph in this sequence is already artificially smoothened by not counting missing values from respondents who do not live with both of their parents.

The high peak at variable 325 is again due to an error in the questionnaire. As already mentioned, due to incorrect jump instructions the question before variable 325 all respondents were asked to skip this question. However, it is interesting to see that only roughly 30 per cent of respondents adhered to that instruction.

^{*} Mittag et al. (2005) provide evidence for a similar phenomenon they named 'checklist effect': Confronted with a sequence of multiple choice questions that have only 'yes' and 'no' as response options, some respondents merely check off the items that apply to them, leaving blank the items where the correct answer would have been 'no'.

The high ridge between 340 and 364 is Grasmick et al.'s (1993) self-control scale, which has also been left uncompleted by a large share of the sample because of the incorrect branching instruction just mentioned.

The sequence just after the ridge, variables 365 to 396, comprises the items self-reported delinquency and drug usage. In this sequence there are two relative peaks, the 12-month incidence of truancy (292 missing values) and the 12-month incidence of driving a motor vehicle without a license (251 missing values).

Towards the end of the questionnaire, the slope of the missing-value rate becomes steeper, the last items in the questionnaire have about 16 per cent of all possible values missing.

Univariate analysis of the German sample

Figure 7.2: Rate of missing values in the German sample in the process of completing the questionnaire. X-axis displays the sequences of times in the questionnaire, Y-axis the share of missing values (0 per cent to 40 per cent).

The graph for the German questionnaire has a number of peaks at the beginning: 350 of the total 7238 respondents fail to report the number of their true friends of their age group and

1303 respondents do not report the number of true adult friends who are not family members. A likely explanation for those two peaks is the just mentioned behavior of not filling in zeros but rather leaving the answering section blank.

These two items with very high rates of nonresponse are followed by a sequence of variables with moderately high nonresponse. The corresponding questions are about respondents' recreational activities and respondents are faced with a two-dimensional table of very detailed response options comprising '___ hours' (per week), or if less than weekly, '___ times' (per month), and 'I never do that'. It can be assumed that a less complex way to arrange the response options would have yielded a lower nonresponse rate.

The next large peak is variable 34, which contains information on whether respondents feel safe at the local youth center. 1251 respondents did not provide an answer to this question, however, further analysis has shown that 1140 of those nonrespondents have stated that they never visit a youth center. In this case, the high nonresponse rate can be explained with the missing 'don't know' response option.

The smaller peaks from variables 62 to 67 and 77 to 80 are likely to be related to the nonexistence of the 'don't know' response option as well. Respondents are asked what other persons or groups of persons, such as parents, teachers, best female and male friends, etc. think about violent behavior and whether their parents, youth group leaders, or coaches are role models for them. Respondents who do not attend youth groups or do not engage in organized athletic activities are more likely not to be able to map their prospective answer onto the given response categories and thus refuse to answer.

Another relatively minor peak is variable 173, which is a branching question for respondents probing for crimes that have occurred to them within the city limits of their home town. Response options are somewhat unclear, i.e. the option 'None of this has happened to me in [my home town]' is placed on the right hand side of all other possible responses, which could have curbed the response rate for this question, especially since earlier questions on victimization have lower nonresponse rates.

Just as other numeric open-ended questions, the questions on the number of times students have changed schools and have repeated school years (variables 177 and 178) have high rates

of missing values. It can be assumed that respondents perceive this information to be sensitive, but the rate of nonresponse is not higher than for other numeric open-ended questions. However, none of this possible information applies to the high share of missing responses on students' physical education grades in their last report card (variable 176). The rate is three times as high (244 missing values) as for the grades in German and mathematics (75 and 81 missing values) and further analysis shows that this phenomenon occurs in all cities and across all school types, therefore the lack of a 'don't know' response option for students of schools where no physical education classes are offered (or where they are not graded) probably is not the problem in this case.

Some variables (269 to 273) related to truancy also have relatively high shares of missing answers. The items probe for the opinions of other people than the respondents, i. e. their classmates, teachers, parents. These do not have the 'don't know' response option, which could explain the size of the share of missing data.

Furthermore, a number of respondents do not report their parents' ages (variables 291 and 292), and the problem already mentioned regarding the number of people respondents share their room with and the number of siblings occurs in the German sample as well (variables 297, 299, and 300).

The items on self-reported delinquency range from 371 to 393. Relative peaks are the 12-month incidence rate for using the public transport system without buying a ticket and—just as in the Russian sample—driving a motor vehicle without a license.

Towards the end of the questionnaire there is another group of variables with a relatively high nonresponse rate (variable 418 to 424). It is a numeric open-ended question, here respondents are to report the number of their friends having committed the offenses listed within the last 12 months. Again it seems probable that many respondents intending to report that they have no friends who have committed a certain offense have not written down zeros. This sequence of questions is followed by questions on religious affiliation (variables 425 to 430), and a relatively high nonresponse rate is observed here too, probably due to a perceived sensitivity of the information probed for.

The last items of the questionnaire have around 7 per cent of their values missing.

Summary and discussion of univariate results

When comparing results from both samples, it is obvious that open-ended numeric questions clearly pose a problem, perhaps not as much for the respondent as for the data analyst. This type of question should always be accompanied with the instruction to fill in zeros when zero is a valid answer or with an additional check box labeled for instance ‘I have no older (step-) siblings’. High rates of nonresponse also occur when response options are arranged in a complex fashion, e.g. when respondents are asked to give more than one answer in a single line or when response options are not all neatly lined up right next to each other but rather arranged in a two-dimensional table.

Problems also occur when asking respondents to report their last grades, yielding high rates of nonresponse in Russia for all classes and a high rate in Germany only for physical education. Another topic where respondents are reluctant to give information is religion: Questions on religion tend to have relatively large shares of nonresponse both in German and the Russian sample. In addition, nonresponse also occurs when respondents are asked about the opinions of others. Questions about respondents’ own delinquent behavior, however, do not result in overly high rates of nonresponse. Some respondents are reluctant to report 12-month incidence rates for some minor offenses, fare evasion (for Germans), truancy (for Russians), and driving a car without a driver’s license (for both).

Another difference between the German and the Russian sample is obvious: The extent of partial nonresponse is considerably larger among the Russian respondents. However, any attempt to explain differences in reporting behavior of German and Russian students is highly speculative. The questionnaires used in this survey were not identical and possible external factors influencing their behavior were not recorded. For instance, it is not known to what extent assistance and motivation by the interviewers who administered the survey differed in the two countries. The difference could also be caused by the different age structure in the two samples, i.e. Germans are on average half a year older. The erroneous filtering instructions in the Russian questionnaire might also have created a cognitive dissonance among the respondents, resulting in earlier fatigue and fewer completely filled-out questionnaires.

Due to the fact that attention deficits are known predictors of delinquent behavior (Farrington 2007; Hawkins et al. 1998), the nonresponse poses a threat to data validity since many re-

spondents at risk will not even see the items on self-reported delinquency located towards the end of the questionnaire.

Bivariate analysis of both samples

In order to gain additional information on factors that influence the amount of missing values the graphs are plotted again for different subsamples. Two factors to be taken into account are educational level and gender.

As a dichotomous variable that is based on the school type a respondent attends, educational level is only a rough approximation for intelligence or cognitive ability. Furthermore, it is known that low intelligence is not only a risk-factor for delinquent behavior (Farrington 2007; Hawkins et al. 1998; Lipsey and Derzon 1998), it is also known that a lack of cognitive skills has an influence on the response process in surveys (Holbrook et al. 2000; Narayan and Krosnick 1996).

Gender is a major predictor of delinquent behavior insofar that crimes are disproportionately committed by males (e.g., Heidensohn and Gelstorphe 2007; Junger-Tas, Ribeaud, and Cruyff 2004; Simpson and Herz 1999; Steffensmeier et al. 2006). Explanations offered are biological reasons, differences in socialization, cognitive abilities, and differences in exposition to risk factors and to social control (Bennett, Farrington, and Huesmann 2005; Simpson and Herz 1999). It can be expected that the boys in the sample report more crimes.

Gender and educational level are mostly unconfounded, since both boys and girls are to be found at both educational levels.

Figure 7.3 shows that the slope for the boys' graph is steeper than the girls' graph and that the graph for lower educational level is somewhat steeper than the graph for higher educational level. The same pattern occurs in the German sample (Figure 7.4), but the effects of the two factors are much weaker.

Figure 7.3: Rate of missing values in the process of completing the questionnaire: Subsamples of the Russian sample. Upper left graph: boys; upper right graph: girls; lower left graph: higher educational level; lower right graph: higher educational level.

Figure 7.4: Rate of missing values in the process of completing the questionnaire: Subsamples of the German sample. Upper left graph: boys; upper right graph: girls; lower left graph: higher educational level; lower right graph: higher educational level. X-axis displays the sequences of times in the questionnaire, Y-axis the share of missing values (0 per cent to 40 per cent).

It is obvious that at least in the Russian sample, nonresponse causes a threat to validity since those in the sample who are at risk, i.e. boys and those at a lower educational level have a larger rate of nonresponse. Therefore, omitting respondents with missing values on central variables such as self-reported delinquency is likely to yield biased estimates.

In order to test whether this association does exist when additional variables are taken into account, a multivariate analysis of nonresponse will be conducted in the next section.

Multivariate analysis

The multivariate analysis will test whether variables which are usually positively correlated to criminal behavior are also positively correlated to nonresponse on questions probing for delinquent acts. To do so, a dummy variable indicating whether the questionnaire section on self-reported delinquency was completed, will serve as the outcome.

It is hypothesized that respondents exhibiting known risk factors of criminal behavior are more likely to refuse an answer to the SRD items. Apart from the already discussed risk factors gender and educational level, additional factors will be introduced.

One risk factor to be tested is the respondent's socio-economic status. Although the notion of the 'dangerous classes' (Frégier 1840) has been surfacing frequently in discussions about the causes of crimes already since the nineteenth century (e.g., Agnew 1992; Merton 1938), the link between social inequality and criminal behavior is still heavily disputed (e. g., Katz 1988: ch. 12; Paul 2006; Tittle, Villemez, and Smith 1978). In their synthesis of research, however, Hawkins et al. (1998) claim that poverty is a predictor of youth violence, which is in line with Farrington (2007).

Acceptance of violence is correlated with delinquent behavior (Baier et al. 2006; Siegmunt 2005: ch. 8; Wetzels et al. 2001: ch. 12; Wilmers et al. 2002: ch. 9). It is assumed that juveniles with a positive attitude towards the use of violence are more prone to commit delinquent acts. From a social learning perspective, it is plausible that juveniles who use violence successfully to reach their goals develop a positive attitude towards it. Based on this attitude, juveniles continue using violence as a means to achieve their aims, thus reinforcing their acceptance of violence.

Research has shown that empathy can be an important predictor for criminal behavior (Farrington 2007; Hosser and Beckurts 2005; Jolliffe and Farrington 2004). It is assumed that persons who are able to recognize and feel the emotion of other persons are less likely to victimize someone. According to Jolliffe and Farrington (2004), this is especially the case for cognitive empathy (as opposed to emotional empathy).

Data and method

Listwise deletion results in an analytic sample of 7,056 German and 1,710 Russian respondents, i.e. 2.5 per cent of cases were lost in the German sample and 2.1 per cent in the Russian.

The method used to test these hypotheses is binary logistic regression (e. g., Long 1997: ch. 3). Binary logistic regression is a generalized linear model which is to be used when the outcome variable is dichotomous, e.g., takes on merely the values 0 or 1. Predictor variables can be of any type. In binary logistic regression, the logit of the dependent variable (i.e. the natural log of the odds of the outcome occurring or not) is modelled as a linear function of the independent variables. To ease interpretation of the results, the logit coefficients are transformed to odds ratios using the exponential function. Odds ratios above the value of 1 denote that there are positive odds for the outcome variable to occur. Odds ratios of about 1 indicate that a predictor has no influence whether the dependent variable occurs or not. Accordingly, odds ratios below 1 denote that the odds for the outcome variable to occur are negative.

As already indicated, the missing data technique to be used here will be listwise deletion. From an earlier chapter we know that listwise deletion performs well in logistic regression even when there is nonrandom missingness on the independent variables (as long as it is confined to the independent variables).

Dependent variable will be a *nonresponse indicator for the SRD items*. This indicator has the value 1 when the respondent has not provided an answer for the items on whether or how often they have ever committed one of the following acts: Beating and injuring another person; breaking into a car in order to steal something out of it; threatening another person with a weapon; intentionally destroying a window, phone booth, street light, or something similar; stealing a bicycle, moped, or any other vehicle; extorting protection money from someone for not beating this person; robbing something from someone; driving a motor vehicle without a license. When all answers were given, the response indicator will have the value 0.

Predictor variables will be the previously used variables *gender* (1 = male; 0 = female) and *educational level* (1 = lower level; 0 = higher level). Respondents' *socio-economic status* is also controlled for by a dummy variable. The variable has the value 1 when the respondent has reported that at least one of their parents is unemployed or the family is living on welfare, and otherwise the value 0. This indicator is similar to the one used by Siegmunt (2005: ch. 5).

Furthermore, two subscales of a scale measuring respondents' *attitudes towards violence* (Siegmunt 2005: ch. 8; Wilmers et al. 2002: ch. 9) will also be used as predictors. One subscale measures approval of violence as a means of entertainment (with three items such as 'Without violence, everything would be much more boring'), the other subscale measures approval of violence in a normative context with six items, e.g. 'An eye for an eye and a tooth for a tooth—that's how life is.' Participants responded to the items using a 4-point Likert scale ranging from 'strongly agree' (1) to 'strongly disagree' (4). In order to retain more cases after listwise deletion, missing values on scale items were replaced with the respondent's scale mean (ipsative mean imputation) as long as two thirds of the scale items are present. Both subscales have sufficiently high Cronbach's (1951) alphas, for normative approval = .82 and for entertainment approval = .77 in the German sample, and for normative approval = .72 and for entertainment approval = .68 in the Russian sample. In order to ease comparison of parameter estimates, the scales were linearly transformed to a range of 0 to 1, where 0 means the respondent rejects violent behavior and 1 means the respondent highly approves of violence.

In order to inspect the impact of respondents' *empathy* on their reporting behavior, two subscales of a scale measuring empathy (Davis 1980) are used (Wilmers et al 2002: ch. 12 and appendix 2). One subscale that consists of five items measures perspective-taking (e.g., 'When I'm upset at someone, I usually try to "put myself in his shoes" for a while.'). and a second subscale (comprising eight items) measures empathic concern (such as in the item, 'I often have tender, concerned feelings for people less fortunate than me.'). Response options in the questionnaire were similar to the options for the attitude towards violence scale, just as the linear transformations conducted with them. Again, ipsative mean imputation was used for computing the scale values when at least two thirds of answers were present. Internal consistency of the scales is sufficient, for empathic concern Cronbach's alpha equals .83 in Ger-

many and .76 in Russia, while for perspective-taking alpha is .78 in Germany and .70 in Russia.

Results

Table 7.1: Descriptive statistics

	Germany		Russia	
	M	SD	M	SD
Nonresponse rate	0.10	0.29	0.15	0.36
Male respondent	0.50	0.50	0.45	0.50
Age	15.08	0.70	14.56	0.53
Low educational level	0.60	0.49	0.66	0.47
Low socio-economic status	0.12	0.33	0.31	0.46
<i>Attitude towards violence</i>				
Normative purpose	0.34	0.22	0.46	0.20
Entertainment purpose	0.19	0.21	0.12	0.18
<i>Empathy</i>				
Empathic concern	0.67	0.19	0.76	0.15
Perspective-taking	0.58	0.20	0.64	0.18
N	7056		1710	

Although a number of cases were lost due to the listwise deletion procedure, the proportions of sample characteristics described in chapter 6 were mostly preserved.* The share of girls in the Russian sample is slightly lower than the corresponding share in the German sample. The Russian students are half a year younger than their German counterparts, and fewer of them attend schools classified as being on a higher educational level. As was expected, the share of poor respondents is larger in the Russian sample. On the scales measuring attitude towards violence; Russians score higher on the scale for approval of violence for normative purposes, while their German counterparts score higher on the scale concerned with approval due to entertainment purposes. This phenomenon is described in more detail by Enzmann and Siegmunt (2005). Overall, levels of approval of violence are rather low. Both for empathic concern and perspective-taking, Russian respondents score higher than Germans. This difference can at least partially be attributed to the fact that the share of female respondents is slightly larger in the Russian sample.

* It should be noted that these rather small shares of lost cases are artificially low due to the special nature of the outcome variable. If the same model was used to predict criminal behavior instead of predicting response to the questions, at least 10 per cent of cases in the German sample and 15 per cent in the Russian sample would have been discarded after listwise deletion.

Table 7.2: Logistic regression, dependent variable: nonresponse indicator

	Germany			Russia		
	Odds ratio	Lower CI limit (95 %)	Upper CI limit (95 %)	Odds ratio	Lower CI limit (95 %)	Upper CI limit (95 %)
Constant	0.04 ***			0.04 ***		
Male respondent	1.77 ***	1.46	2.14	2.77 ***	2.04	3.77
Low educational level	2.05 ***	1.69	2.48	1.56 ***	1.14	2.13
Low socio-economic status	0.69 **	0.53	0.90	1.04	0.77	1.39
<i>Attitude towards violence</i>						
Normative purpose	3.01 ***	1.77	5.14	3.56 **	1.66	7.62
Entertainment purpose	1.47	.88	2.44	1.09	0.49	2.44
<i>Empathy</i>						
Empathic concern	0.72	0.42	1.25	0.46	0.16	1.33
Perspective-taking	0.81	0.50	1.31	2.25	0.97	5.22
Model Chi ² (7)		280.33 ***			105.15 ***	
R ² (Cox and Snell)		0.04			0.06	
R ² (Nagelkerke)		0.08			0.11	

* $p < .05$, ** $p < .01$, *** $p < .001$, two-tailed tests

As it was expected, gender has a massive impact on the probability of nonresponse. Being a male in the Russian sample increases the odds of not answering all questions by 177 per cent ($p < .001$). In the German sample, the effect is slightly weaker, male gender increases the odds by 77 per cent ($p < .001$).

The influence of the educational level does not disappear when control variables are taken into account. Both in Germany and Russia the educational level is highly predictive for nonresponse ($p < .001$). The controversial predictor socio-economic status does have an impact on nonresponse, but only in the German sample. However, the direction of the effect is not as

expected: Having at least one unemployed parent and/or living on welfare in fact decreases the odds of not answering all questions by 31 per cent^{*}.

One of the two scales measuring attitudes towards violence has a significant positive effect on the outcome variable ($p < .001$ in Germany and $p < .01$ in the Russian sample). Approval of violent behavior due to normative reasons has odds ratios which are among the highest in both models. However, it should be noted that the four scales in the model are not dichotomies like the other variables used, but vary in between 0 and 1. Approval of violence for entertainment purposes has no significant effect on reporting behavior, just as the scales measuring respondents' empathy: Neither empathic concern nor perspective-taking have a significant effect.[†]

The overall fit of the models is significant ($p < .001$), however, the pseudo- R^2 measures show that the strength of association between predictor variables and the outcome is rather low.

The multivariate analysis was not able to confirm all hypothesized associations in the data. However, the predictors identified in the bivariate analysis—gender and educational level—were significant even when other variables were controlled for. Among the attitudinal risk factors commonly associated with delinquent behavior, approval of violence due to normative purposes increases the probability of nonresponse both in the German and the Russian sample. Cognitive as well as affective empathy and approval of violence for entertainment purposes do not have an effect on the reporting behavior. In Germany, however, for low socio-economic status a significant positive effect on the probability of response was found. This finding cannot be explained by the theoretical assumptions made.

^{*} This effect remains significant even when an interaction term for poverty and low educational level is included: After centering the predictors in order to reduce multicollinearity, the odds ratio of the interaction term itself does not significantly differ from 1, the odds ratio of low socio-economic status does not change and remains significant.

[†] An additional model estimated for the German sample that controlled for a self-control scale (Grasmick et al. 1993; Siegmunt 2005: ch. 8; Wilmers et al. 2002: ch. 11) showed that low self-control is a major predictor for nonresponse (transformed like the other scales, low self-control has an odds ratio of 2.9). This is in line with the findings of Watkins and Melde (2007). However, a similar analysis cannot be conducted for the Russian sample when listwise deletion is used. The self-control scale in the Russian questionnaire was left blank by a large share of the Russian respondents and it was placed near the end of the questionnaire. Nonresponse on this scale is highly related to both the values on the dependent as well as the independent variables. Respondents who have a value for self-control differ systematically from those who do not.

Summary

Every analysis has shown that nonresponse is a problem in the current data set, especially for the Russian data set. Questions towards the end of the questionnaire were left unanswered by up to 16 per cent in the Russian sample and 7 per cent in the German sample. Item nonresponse was frequently associated with certain question formats. Educational level and gender have effects on the likeliness of nonresponse. Focusing on SRD items, the multivariate analysis confirmed this finding. The multivariate analysis further showed that additional risk factors for delinquent behavior increase the probability of nonresponse. Therefore the complete cases cannot be regarded as random subsamples of all cases: When analyzing complete cases only, the extent of delinquency will probably be underestimated.

8. Simulation of missing data technique performance under various conditions

In order to see how the various missing data techniques perform in comparison, a simulation of different scenarios was conducted. This section describes the samples that were prepared for the simulation, presents several strategies and techniques that will be used to analyze the samples, and reports the findings.

Data used in the simulation

To obtain a working sample without any missing values, all cases with values on a number of variables of interest were retained. Cases with missing values on the said variables were discarded. Variables considered to be interesting in this context were the variables used in chapter 7 (prevalence rates of self-reported delinquency, nationality, gender, age, educational level, empathy, self-control, attitude towards violence, socio-economic status), also prevalence rates of victimization, and whether the respondent has a steady group of peers. This listwise deletion process yielded a sample of 7,490 complete cases. The complete cases will serve as a standard for comparisons with the results achieved by the various strategies and techniques. Using a backward stepwise logistic regression procedure, a model predicting the prevalence of simple assault was estimated among the variables in the sample (e.g., Table 8.4). All predictors have highly significant effects on the dependent variable ($p < .001$). This model will serve as a benchmark for comparison.

Although the models displayed in the following are binary logistic regression models just like the model in the last chapter, in this chapter logit coefficients are displayed, and not odds ratios. Logit coefficients are the natural logs of the odds ratios. These are not as easy to interpret as odds ratios: A one-unit increase in an independent variable is associated with a change in the log odds of the dependent variable that equals the logit coefficient. However, they ease comparison of the size of effects since they can range between plus and minus infinity, with a logit coefficient of 0 indicating no effect.

Simulated scenarios

Several different scenarios of missing data were created. In one variant missing values are confined to the independent variables, in the other missing values occur both on the independent variables and the dependent variable. Furthermore, a scenario was created where data are

MAR, and a scenario where data are NMAR. Each of these four combinations was implemented for different shares of missing data.

Two independent variables were chosen to have missing values: self-control and empathic concern. In the complete sample, the self-control score follows the normal distribution very closely, while the empathic concern score has a slight negative skew. For the MAR assumption, values for self-control were eligible for deletion, when at least one of the following, somewhat arbitrary conditions was met: respondent attends a school of lower educational level and respondent has a low socio-economic status. In order to delete values for empathic concern at random, the following conditions were used: Again respondents attending schools of lower educational level were eligible for deletion, as well as respondents having a score of more than .37 (about one standard deviation above the mean) on the approval of violence (entertainment purposes) scale. For the NMAR assumption, the low self-control score was deleted disproportionately more often if it ranged above .6 (about one standard deviation above the mean). The empathic concern score was deleted disproportionately more often when it had a value lower than .5 (about one standard deviation below the mean).

In order to create data sets that have missing values both on the independent variables and the dependent variable, the mechanisms described above were used, and in addition, values on the dependent variable—prevalence of simple assault—were deleted at random, when at least one of the following conditions was met: respondent attends a school of lower educational level, respondent has a low socio-economic status, and respondent has a score of more than .37 on the approval of violence (entertainment purposes) scale. In order to create NMAR data, respondents who admitted committing simple assault had their values deleted disproportionately more often.

To modify the amounts of missing values, all above-named conditions were combined with variables containing random draws from a uniform distribution. Only if the random value exceeded a certain threshold, the values were deleted. These thresholds were chosen so that in one data set each variable has a small share of missing values (around 5 per cent), in another data set each variable has a moderate share (around 15 per cent), and finally a data set where each variable has a large share of missing values (around 25 per cent).

This adds up to 13 different data sets, i.e. the complete data set and twelve modified ones.

Table 8.1 shows what effects the deletion process had on the means and standard errors of the variables on which values were deleted. The MAR deletion routine changes little about the means of empathic concern and low self-control. While empathic concern slightly rises, low self-control decreases a little. The effect on the prevalence of simple assault is more markedly. The prevalence rate decreases by up to 11.6 per cent. The NMAR deletion routines have a stronger impact on the data: Empathic concern rises up to 5 per cent, low self-control decreases 4.9 per cent, and the prevalence rate decreases by up to 67.1 per cent.

Strategies and techniques used in the simulation

To analyze these data sets, a range of conventional and modern approaches and strategies were chosen.

Listwise deletion was chosen since it is a very established technique that is frequently as the default method and it is said to perform better than it is often thought (cf. chapter 4).

Two ad hoc approaches were also chosen. The first is mean substitution for subgroups. A reasonable guess of two main predictors in cross-national comparisons of crime and delinquency would be gender and nationality. Thus, for cases with missing values on self-control and empathic concern means obtained from the complete cases according to their group affiliation were plugged in. Since it is not possible to substitute the mean for the binary variable, predicted values obtained from a logistic regression of simple assault on gender and nationality were substituted instead. This technique is an example of a ‘quick and dirty’ approach, where missing values are substituted based on a reasonable guess.

The second ad hoc approach is regression substitution. Missing values for empathic concern and low self-control were estimated by regressing these variables on the variables used for deciding whether to delete their values. Low self-control was regressed on lower educational level and lower socio-economic status in an OLS regression model, and empathic concern was likewise regressed on lower educational level and the approval of violence (entertainment purposes) scale. Missing values on the simple assault variable were substituted by predictions obtained from logistic regression models where the simple assault variable was regressed on lower educational level, lower socio-economic status, and the approval of violence (entertainment purposes) scale. This technique demonstrates how well single imputation without

Table 8.1: Changes in estimated means and standard errors after deletion of values

	Empathic concern					Low self-control					Prevalence rate simple assault				
	M	%	S.E.	%	N	M	%	S.E.	%	N	M	%	S.E.	%	N
<i>Full sample</i>	0.69		0.0021		7,490	0.45		0.0016		7,490	0.21		0.0047		7,490
Missing at random															
-5%	0.69	0.2%	0.0022	2.2%	7,128	0.45	-0.2%	0.0016	2.0%	7,147	0.21	-2.2%	0.0048	1.9%	7,097
-15%	0.69	0.8%	0.0023	6.1%	6,401	0.45	-0.2%	0.0017	7.7%	6,399	0.20	-7.3%	0.0050	5.3%	6,388
-25%	0.70	1.4%	0.0024	12.0%	5,592	0.45	-0.7%	0.0018	14.1%	5,610	0.19	-11.6%	0.0052	10.1%	5,633
<i>Missing not at random</i>															
-5%	0.71	2.4%	0.0020	-4.3%	7,067	0.44	-2.3%	0.0015	-2.8%	7,094	0.17	-18.4%	0.0045	-4.7%	7,057
-15%	0.71	3.5%	0.0021	-3.5%	6,376	0.43	-3.6%	0.0016	-0.1%	6,356	0.12	-41.9%	0.0041	-13.1%	6,401
-25%	0.72	5.0%	0.0020	-4.7%	5,684	0.43	-4.9%	0.0016	1.4%	5,596	0.07	-67.1%	0.0034	-28.6%	5,693

adding residual variance can perform: The imputed values are based from the exact variables which have been used to delete the missing values. This approach will not be used for the data sets where missing values are not missing at random, since the mechanism responsible for the missing values cannot be modeled using regression prediction.

The last four approaches are all based on multiple imputation by chained equations, as implemented by Royston (2004; 2005a; 2005b) in the 'ice' module for Stata. For each data set with missing values, 15 imputed data sets were created. While the appropriate number of imputations is still disputed (cf. ch. 5), the number chosen equals what was computationally feasible.

The 'ice' module sequentially imputes each variable with missing values using regression models appropriate for the type of variable which is being imputed. Random variance is added during the imputation process. Parameters obtained were combined using Rubin's (1987) rules. In order to assess model fit, the R^2 values of each imputed dataset were averaged (as suggested by Acock 2005).

MI is used in a number of different approaches. The first approach used draws on the resentment of some researchers to impute values on an outcome variable (e.g., Cohen and Cohen 1983: ch. 7). Therefore, for the first multiple imputation approach used, the dependent variable is not included in the imputation model. When data sets are analyzed where there are missing values on the dependent variable, the respective cases are discarded before imputation, i.e. the approach is combined with listwise deletion. Included in the imputation model are all other variables used in the model to be analyzed and the variables that were used for creating the missing values.

A second approach based on MI is very similar to the two latter approaches. Here the imputation model merely consists of the variables used in the model to be analyzed (including the dependent variable) and the variables used for creating the missing values when data are MAR. On the one hand, this approach is based on an ideal choice of variables, on the other hand it is a restrictive approach that does not use as much information as possible.

The third MI-based approach uses an inclusive imputation model. It draws on nearly all variables in the data set, such as age, the other prevalence variables both on victimization and

self-reported delinquency. However, it does not include the variable indicating educational level, which is in part responsible for the missing values in the MAR data set. This model was chosen to assess how well MI performs when conditions are suboptimal and a researcher trying to impute missing values is slightly off with their assumptions about the missing data.

The last MI-based approach also uses an inclusive imputation model, but this approach is a worst-case scenario. This approach uses all variables in the data set for the imputation model, except for all variables that were used for deleting values when data are NMAR.

For the data sets where values are missing non-randomly, only an approach based on an inclusive imputation strategy drawing on all variables contained in the data set is used.

Results

Values missing at random on independent variables

All methods and strategies except for listwise deletion yield fairly accurate point estimates for the means of the samples (Table 8.2). Even when 25 per cent of the values are missing at random, no method gives estimates that are off by more than 1 per cent. Mean substitution for subgroups and regression imputation result in means that are hardly biased when the share of missing data is small. Estimates obtained from the different MI approaches are fairly accurate, though the approach that excludes the dependent variable from the imputation model performs worse than the others. Listwise deletion, however, does not only result in biased estimates for the variables imputed in the other approaches, it also biases all other means. For instance, when 25 per cent of the values of two variables are missing, the prevalence rate of simple assault is biased towards zero by 18.9 per cent.

In terms of standard errors, listwise deletion again yields rather undesirable results (Table 8.3). Due to the loss of cases, all standard errors increase considerably, e. g. for the data set with the largest amount of missing data each standard error is 24.7 per cent inflated. Results obtained by mean substitution and regression imputation are also alarming. Here standard errors do not become larger, but instead precision is overestimated. Although estimates are virtually based on less information than in the complete model, standard errors are up to 14.4 per cent smaller when 25 per cent of cases have missing values on two independent variables. MI-based approaches generally yield larger standard errors for the imputed variables in order to account for the uncertainty induced by the missing information. Nonetheless, they are

Table 8.2: Estimated means and magnitude of their errors as a percentage of the estimate for the complete sample when data are MAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M [%]
	M	%	M	%	M	%	M	%	M	%	M	%	
<i>Complete sample</i>	0.87		0.48		0.34		0.69		0.45		0.21		
<i>Listwise deletion</i>													
(5) N = 6,810	0.87	0.5%	0.48	-0.4%	0.34	-1.6%	0.69	0.4%	0.45	-0.4%	0.20	-3.8%	1.2%
(15) N = 5,547	0.88	1.2%	0.47	-1.7%	0.33	-5.4%	0.70	0.9%	0.44	-1.5%	0.19	-11.0%	3.6%
(25) N = 4,406	0.89	2.0%	0.45	-5.0%	0.31	-10.4%	0.70	1.8%	0.44	-2.8%	0.17	-18.9%	6.8%
<i>Mean substitution</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.2%	0.45	-0.2%	0.21	0.0%	0.1%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.6%	0.45	-0.2%	0.21	0.0%	0.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	1.0%	0.45	-0.7%	0.21	0.0%	0.3%
<i>Regression estimation</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	-0.1%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.3%	0.45	0.2%	0.21	0.0%	0.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.6%	0.45	0.1%	0.21	0.0%	0.1%
<i>MI with dependent variable excluded</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.1%	0.45	0.0%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.4%	0.45	0.1%	0.21	0.0%	0.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.7%	0.45	0.0%	0.21	0.0%	0.1%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	-0.1%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.3%	0.45	0.2%	0.21	0.0%	0.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.5%	0.45	0.0%	0.21	0.0%	0.1%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.0%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.1%	0.45	0.1%	0.21	0.0%	0.0%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.3%	0.45	0.1%	0.21	0.0%	0.1%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.1%	0.45	0.0%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.1%	0.45	0.2%	0.21	0.0%	0.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.2%	0.21	0.0%	0.0%

(5): 5%, (15): 15%, (25): 25 % of values missing at random both on empathic concern and low self-control

Table 8.3: Estimated standard errors and magnitude of their errors as a percentage of the estimate for the complete sample when data are MAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.0039		0.0058		0.0025		0.0021		0.0016		0.0047		
<i>Listwise deletion</i>													
(5) N = 6,810	0.0040	3.5%	0.0061	4.9%	0.0026	4.3%	0.0022	4.4%	0.0016	4.2%	0.0049	3.4%	4.1%
(15) N = 5,547	0.0044	12.2%	0.0067	16.1%	0.0028	13.0%	0.0024	13.7%	0.0018	13.8%	0.0052	11.2%	13.3%
(25) N = 4,406	0.0048	22.6%	0.0075	30.0%	0.0031	23.0%	0.0027	25.7%	0.0020	26.2%	0.0057	20.3%	24.7%
<i>Mean substitution</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-2.2%	0.0015	-2.6%	0.0047	0.0%	0.8%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-7.8%	0.0014	-7.9%	0.0047	0.0%	2.6%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0018	-13.6%	0.0013	-14.3%	0.0047	0.0%	4.7%
<i>Regression estimation</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-1.9%	0.0015	-2.6%	0.0047	0.0%	0.7%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-6.0%	0.0014	-7.9%	0.0047	0.0%	2.3%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0019	-10.0%	0.0013	-14.4%	0.0047	0.0%	4.1%
<i>MI with dependent variable excluded</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	1.4%	0.0016	1.3%	0.0047	0.0%	0.4%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.9%	0.0017	7.9%	0.0047	0.0%	2.1%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0024	12.7%	0.0019	19.8%	0.0047	0.0%	5.4%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	2.8%	0.0016	1.7%	0.0047	0.0%	0.8%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.7%	0.0017	8.0%	0.0047	0.0%	2.1%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0024	12.2%	0.0019	18.8%	0.0047	0.0%	5.2%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	1.0%	0.0016	1.4%	0.0047	0.0%	0.4%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	2.0%	0.0017	7.8%	0.0047	0.0%	1.6%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0023	8.6%	0.0018	12.0%	0.0047	0.0%	3.4%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	0.7%	0.0016	3.0%	0.0047	0.0%	0.6%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	2.6%	0.0016	3.1%	0.0047	0.0%	1.0%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	2.7%	0.0017	11.1%	0.0047	0.0%	2.3%

(5): 5%, (15): 15%, (25): 25 % of values missing at random both on empathic concern and low self-control

in all instances smaller than those obtained from a complete-case analysis. For the data set with a rather large amount of missing values, they can increase up to 19.8 per cent. This is less than the increase caused by listwise deletion. And again, listwise deletion yields increased standard errors for all variables, while the increase for the MI approaches is confined to the variables with missing values. Excluding the dependent variable from the imputation model has less negative effects in terms of standard errors. Furthermore, it pays off to use an inclusive imputation strategy. The approaches using larger imputation models have smaller standard errors than those restricted to the exact variables related to missingness.

The various strategies have rather little impact on the model fit (Table 8.4). Even when a large amount of values is missing, McFadden's R^2 is never off by more than 3.5 per cent. For the single imputation approaches, R^2 is understated the larger the fraction of missing data is. For the multiple imputation approach excluding the dependent variable from the imputation model, this is also the case. However, when the information from the dependent variable is used for the imputation, the opposite is the case, R^2 is overstated, but generally less than for the latter approaches. Findings for listwise deletion are less consistent, e. g., for the data set with few missing values, R^2 is overstated, although when many values are missing, R^2 is slightly understated.

For the estimates of logit coefficients, the MI approaches that include the dependent variable in the imputation model produce the least overall bias. Even when a large amount of values is missing, the average deviation from the coefficients obtained in the analysis of the complete sample does not exceed 5.7 per cent. Occasionally, gross misestimations occur even when only a moderate amount of data is missing, e. g., the impact of empathic concern is overestimated up to 17.9 percent. The MI approach that excludes the dependent variable tends to perform worse compared to the other approaches based on MI. The single imputation strategies mostly result in a rather large bias, mostly confined to the imputed variables, but not when the amount of missing values is large. However, when few values are missing, mean substitution for subgroups performs reasonably well. The overall bias for listwise deletion is slightly larger than for the other approaches. But while for MI approaches bias is mostly confined to the imputed predictors, listwise deletion yields biased estimates even for completely observed predictors, e. g., when only a moderate amount of data is missing, the effect of gender on the outcome variable is overestimated by ten per cent.'

Table 8.4: Estimates of McFaddens R^2 and logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are MAR on independent variables. Logistic regression. dependent variable: simple assault.

	Model fit		German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	R^2	%	b	%	b	%	b	%	b	%	b	%	b	%	
<i>Complete sample</i>	0.197		0.85		0.90		2.96		-0.78		3.32		-4.83		
<i>Listwise deletion</i>															
(5) N = 6,810	0.202	2.2%	0.84	-1.4%	0.95	5.6%	2.90	-2.2%	-0.78	0.7%	3.54	6.7%	-4.97	2.8%	3.2%
(15) N = 5,547	0.200	1.6%	0.81	-4.7%	0.99	10.3%	2.90	-2.2%	-0.81	3.6%	3.58	7.8%	-5.03	4.0%	5.4%
(25) N = 4,406	0.197	-0.2%	0.78	-8.0%	1.08	20.4%	2.88	-2.8%	-0.68	-12.3%	3.65	9.8%	-5.20	7.6%	10.2%
<i>Mean substitution</i>															
(5) N = 7,490	0.196	-0.6%	0.87	2.4%	0.88	-2.3%	3.03	2.2%	-0.80	2.8%	3.30	-0.6%	-4.83	0.0%	1.7%
(15) N = 7,490	0.193	-2.2%	0.89	4.7%	0.84	-6.0%	3.21	8.3%	-0.84	8.2%	3.16	-4.7%	-4.80	-0.7%	5.4%
(25) N = 7,490	0.190	-3.5%	0.95	12.0%	0.84	-6.3%	3.40	14.8%	-0.72	-7.7%	3.16	-4.8%	-4.99	3.2%	8.1%
<i>Regression estimation</i>															
(5) N = 7,490	0.197	-0.3%	0.86	1.3%	0.88	-1.6%	2.99	1.0%	-0.85	8.7%	3.33	0.3%	-4.80	-0.7%	2.3%
(15) N = 7,490	0.194	-1.5%	0.86	1.6%	0.86	-3.7%	3.10	4.5%	-0.98	25.4%	3.20	-3.6%	-4.68	-3.1%	7.0%
(25) N = 7,490	0.193	-2.1%	0.89	4.6%	0.86	-3.8%	3.19	7.6%	-1.05	35.2%	3.24	-2.3%	-4.70	-2.7%	9.4%
<i>MI with dependent variable excluded</i>															
(5) N = 7,490	0.196	-0.6%	0.86	1.9%	0.89	-0.9%	3.01	1.6%	-0.76	-1.8%	3.21	-3.4%	-4.82	-0.4%	1.7%
(15) N = 7,490	0.193	-2.4%	0.87	2.7%	0.87	-3.3%	3.16	6.5%	-0.81	3.9%	2.79	-15.9%	-4.64	-4.0%	6.0%
(25) N = 7,490	0.190	-3.5%	0.90	6.1%	0.88	-2.3%	3.27	10.4%	-0.72	-8.0%	2.60	-21.6%	-4.68	-3.1%	8.6%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>															
(5) N = 7,490	0.198	0.5%	0.85	0.6%	0.90	-0.3%	2.92	-1.4%	-0.78	0.3%	3.43	3.3%	-4.87	0.8%	1.1%
(15) N = 7,490	0.200	1.2%	0.84	-1.5%	0.88	-1.9%	2.87	-3.0%	-0.90	15.9%	3.49	5.1%	-4.78	-1.0%	4.7%
(25) N = 7,490	0.201	1.8%	0.86	1.7%	0.91	1.7%	2.85	-3.7%	-0.73	-5.7%	3.71	11.6%	-5.03	4.1%	4.8%
<i>MI, inclusive strategy, most mechanism variables included</i>															
(5) N = 7,490	0.198	0.4%	0.86	0.8%	0.89	-0.5%	2.93	-1.1%	-0.78	0.8%	3.40	2.5%	-4.86	0.6%	1.1%
(15) N = 7,490	0.200	1.4%	0.84	-0.9%	0.88	-2.2%	2.88	-3.0%	-0.88	13.3%	3.52	6.0%	-4.81	-0.4%	4.3%
(25) N = 7,490	0.202	2.6%	0.84	-1.3%	0.89	-1.0%	2.81	-5.1%	-0.89	14.6%	3.71	11.6%	-4.88	0.9%	5.7%
<i>MI, inclusive strategy, no mechanism variables included</i>															
(5) N = 7,490	0.198	0.6%	0.85	0.5%	0.89	-0.6%	2.92	-1.5%	-0.79	1.9%	3.43	3.3%	-4.86	0.6%	1.4%
(15) N = 7,490	0.200	1.5%	0.84	-1.3%	0.88	-2.4%	2.87	-3.0%	-0.92	17.9%	3.52	6.1%	-4.79	-1.0%	5.3%
(25) N = 7,490	0.203	2.8%	0.85	-0.3%	0.90	-0.2%	2.83	-4.7%	-0.86	10.4%	3.77	13.5%	-4.94	2.3%	5.2%

(5): 5%, (15): 15%, (25): 25 % of values missing at random both on empathic concern and low self-control

Table 8.5: Estimates of standard errors for logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are MAR on independent variables. Logistic regression, dependent variable: simple assault.

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.11		0.07		0.18		0.20		0.28		0.26		
<i>Listwise deletion</i>													
(5) N = 6,810	0.12	7.9%	0.08	7.1%	0.20	6.8%	0.21	7.0%	0.30	7.4%	0.28	7.6%	7.3%
(15) N = 5,547	0.14	24.4%	0.09	23.0%	0.23	22.8%	0.24	22.8%	0.35	23.2%	0.32	23.1%	23.2%
(25) N = 4,406	0.16	46.4%	0.10	43.3%	0.26	44.2%	0.28	43.4%	0.40	42.9%	0.38	45.4%	44.3%
<i>Mean substitution</i>													
(5) N = 7,490	0.11	-0.1%	0.07	0.1%	0.18	-0.7%	0.20	2.3%	0.29	1.7%	0.26	1.0%	1.0%
(15) N = 7,490	0.11	-0.6%	0.07	1.0%	0.18	-2.8%	0.21	8.8%	0.30	5.4%	0.27	3.3%	3.6%
(25) N = 7,490	0.11	-0.7%	0.07	1.9%	0.17	-5.0%	0.23	17.7%	0.31	11.1%	0.29	9.1%	7.6%
<i>Regression estimation</i>													
(5) N = 7,490	0.11	0.0%	0.07	-0.4%	0.18	-0.3%	0.20	2.1%	0.29	2.1%	0.27	1.3%	1.0%
(15) N = 7,490	0.11	-0.5%	0.07	-1.0%	0.18	-1.0%	0.21	7.3%	0.30	6.3%	0.27	4.1%	3.4%
(25) N = 7,490	0.11	-0.5%	0.07	-1.5%	0.18	-1.5%	0.22	13.9%	0.32	13.0%	0.29	9.7%	6.7%
<i>MI with dependent variable excluded</i>													
(5) N = 7,490	0.11	0.1%	0.07	-0.4%	0.18	0.4%	0.20	1.4%	0.29	2.4%	0.27	1.5%	1.0%
(15) N = 7,490	0.11	-0.4%	0.07	-1.1%	0.19	1.0%	0.21	6.6%	0.32	12.8%	0.28	7.2%	4.9%
(25) N = 7,490	0.11	-0.4%	0.07	-1.3%	0.19	1.2%	0.22	12.5%	0.33	15.2%	0.29	10.5%	6.8%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.11	0.4%	0.07	0.4%	0.18	0.7%	0.20	2.6%	0.30	4.3%	0.27	3.1%	1.9%
(15) N = 7,490	0.11	0.9%	0.07	0.9%	0.19	2.6%	0.22	11.3%	0.31	10.2%	0.28	5.9%	5.3%
(25) N = 7,490	0.12	3.0%	0.07	4.1%	0.20	6.4%	0.26	32.7%	0.34	18.3%	0.32	23.7%	14.7%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.11	0.3%	0.07	0.3%	0.18	0.7%	0.20	2.4%	0.29	2.7%	0.27	2.1%	1.4%
(15) N = 7,490	0.11	0.9%	0.07	1.6%	0.19	3.6%	0.22	14.9%	0.31	9.6%	0.28	8.6%	6.5%
(25) N = 7,490	0.11	1.7%	0.07	2.6%	0.20	8.1%	0.22	11.8%	0.36	25.9%	0.29	10.0%	10.0%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.11	0.3%	0.07	0.4%	0.18	0.9%	0.20	3.0%	0.29	3.6%	0.27	2.8%	1.8%
(15) N = 7,490	0.11	1.1%	0.07	1.0%	0.19	3.6%	0.21	9.1%	0.32	12.3%	0.28	6.0%	5.5%
(25) N = 7,490	0.11	0.9%	0.07	2.3%	0.19	6.1%	0.22	14.8%	0.35	25.4%	0.29	11.6%	10.2%

(5): 5%, (15): 15%, (25): 25 % of values missing at random both on empathic concern and low self-control

Standard errors obtained in the logistic regression model are relatively small when single imputation strategies are employed (Table 8.5). This is an awkward finding when remembering that these approaches can yield highly biased point estimates. It also applies to the MI strategy where the dependent variable is excluded from the imputation model. The other MI approaches including the dependent variable yield rather high standard errors, however, they are still smaller than the standard errors obtained by listwise deletion. And again, with MI approaches larger standard errors occur only for the constant and the imputed predictors, while listwise deletion inflates standard errors for all estimates.

Values missing at random on independent variables and the dependent variable

When missing values are not confined to the continuous independent variables, but also occur on the binary dependent variable, larger differences in terms of estimated means (Table 8.6) become apparent for the different approaches. While for the continuous variables substantially the same things can be said as about Table 1, for the binary variable the more elaborate approaches clearly perform better. While the bias does not exceed 3.3 per cent even when a large share of values is missing, MI is used and the dependent variable is imputed as well, the ad hoc approaches fail to estimate the prevalence rate correctly. The larger the share of missing values is, the larger the bias. Listwise deletion creates a substantial underestimation of the prevalence, more than twice that of any other method.

Most standard errors for means obtained when values are also missing on the dependent variable (Table 8.7) for most approaches do not differ from those reported in Table 8.3. However, standard errors for listwise deletion are larger than those obtained by any other method, since a larger number of incomplete cases had to be discarded than in Table 8.3. The same applies to the MI approach, where the dependent variable is excluded from the imputation model. Here, cases with missing values on the dependent variable were deleted, and the thus smaller sample yields larger standard errors. For the single imputation methods, standard errors for the prevalence rate are deflated. The MI approaches that employ inclusive imputation models yield smaller standard errors than any other approach, even the MI approach based on the variables the missingness depends on. Standard errors obtained from complete-case analysis are inflated, even for the variables that do not have missing values.

When there are missing values on the dependent variable, model fit is affected (Table 8.8). For the approaches that involve listwise deletion, in general McFadden's R^2 is the more understated, the larger the fraction of missing values is, up to 11.1 per cent. This applies as well

Table 8.6: Estimated means and magnitude of their errors as a percentage of the estimate for the complete sample when data are missing MAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	M	%	M	%	M	%	M	%	M	%	M	%	
<i>Complete sample</i>	0.87		0.48		0.34		0.69		0.45		0.21		
<i>Listwise deletion</i>													
(5) N = 6,544	0.87	0.6%	0.47	-1.0%	0.34	-2.5%	0.69	0.6%	0.45	-0.6%	0.20	-5.3%	1.8%
(15) N = 4,998	0.88	1.3%	0.46	-3.5%	0.32	-8.0%	0.70	1.6%	0.44	-2.0%	0.17	-16.9%	5.6%
(25) N = 3,837	0.89	2.4%	0.44	-7.9%	0.30	-14.0%	0.71	3.0%	0.43	-3.4%	0.15	-27.2%	9.6%
<i>Mean substitution</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.2%	0.45	-0.2%	0.20	-7.3%	1.3%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.6%	0.45	-0.2%	0.17	-20.9%	3.6%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	1.0%	0.45	-0.7%	0.14	-33.5%	5.9%
<i>Regression estimation</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	-0.1%	0.20	-2.8%	0.5%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.3%	0.45	0.2%	0.19	-9.8%	1.7%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.6%	0.45	0.1%	0.18	-16.4%	2.8%
<i>MI with dependent variable excluded</i>													
(5) N = 7,097	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.3%	0.45	-0.3%	0.21	-2.2%	0.5%
(15) N = 6,388	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.5%	0.45	-0.7%	0.20	-7.3%	1.4%
(25) N = 5,633	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	1.0%	0.44	-1.3%	0.19	-11.6%	2.3%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.0%	0.21	0.0%	0.0%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.2%	0.45	0.1%	0.21	-1.3%	0.3%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.5%	0.45	0.1%	0.21	-0.4%	0.2%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.0%	0.21	-0.3%	0.1%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.1%	0.45	0.1%	0.21	-1.6%	0.3%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	-0.3%	0.45	0.0%	0.21	-2.4%	0.5%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.1%	0.45	0.0%	0.21	-0.6%	0.1%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.2%	0.21	-2.0%	0.4%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.69	0.0%	0.45	0.1%	0.20	-3.3%	0.6%

(5): 5%, (15): 15%, (25): 25 % of values missing at random on empathic concern, low self-control and prevalence simple assault

Table 8.7: Estimated standard errors and magnitude of their errors as a percentage of the estimate for the complete sample when data are MAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.0039		0.0058		0.0025		0.0021		0.0016		0.0047		
<i>Listwise deletion</i>													
(5) N = 6,544	0.0041	5.2%	0.0062	6.9%	0.0026	5.6%	0.0023	6.2%	0.0017	5.9%	0.0049	4.9%	5.8%
(15) N = 4,998	0.0046	17.7%	0.0071	22.2%	0.0029	16.2%	0.0025	18.7%	0.0019	19.1%	0.0054	14.1%	18.0%
(25) N = 3,837	0.0051	29.7%	0.0080	38.9%	0.0032	27.7%	0.0028	31.9%	0.0021	34.3%	0.0058	23.5%	31.0%
<i>Mean substitution</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-2.2%	0.0015	-2.6%	0.0046	-2.8%	1.3%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-7.8%	0.0014	-7.9%	0.0043	-8.6%	4.1%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0018	-13.6%	0.0013	-14.3%	0.0040	-14.9%	7.1%
<i>Regression estimation</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-1.9%	0.0015	-2.6%	0.0047	-1.0%	0.9%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-6.0%	0.0014	-7.9%	0.0045	-3.8%	3.0%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0019	-10.0%	0.0013	-14.4%	0.0044	-6.6%	5.2%
<i>MI with dependent variable excluded</i>													
(5) N = 7,097	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.7%	0.0016	3.5%	0.0048	1.9%	1.7%
(15) N = 6,388	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0024	11.4%	0.0018	14.2%	0.0050	5.3%	5.1%
(25) N = 5,633	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0025	18.1%	0.0019	22.1%	0.0052	10.1%	8.4%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	1.2%	0.0016	0.7%	0.0049	4.5%	1.1%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.8%	0.0017	5.4%	0.0052	10.7%	3.5%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0025	18.5%	0.0019	20.4%	0.0056	18.0%	9.5%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	1.5%	0.0016	2.2%	0.0048	2.1%	1.0%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.5%	0.0017	8.1%	0.0049	3.4%	2.7%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	4.0%	0.0017	5.5%	0.0055	15.7%	4.2%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	1.5%	0.0016	0.9%	0.0048	1.0%	0.6%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0022	3.4%	0.0018	14.5%	0.0049	4.1%	3.7%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0023	8.2%	0.0018	15.3%	0.0055	15.8%	6.5%

(5): 5%, (15): 15%, (25): 25 % of values missing at random on empathic concern, low self-control and prevalence simple assault

Table 8.8: Estimates of McFaddens R^2 and logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are MAR on independent variables and the dependent variable. Logistic regression, dependent variable: simple assault.

	Model fit		German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	R^2	%	b	%	b	%	b	%	b	%	b	%	b	%	
<i>Complete sample</i>	0.197		0.85		0.90		2.96		-0.78		3.32		-4.83		
<i>Listwise deletion</i>															
(5) N = 6,544	0.200	1.1%	0.80	-5.6%	0.98	8.7%	2.89	-2.4%	-0.80	2.6%	3.45	4.0%	-4.91	1.5%	4.1%
(15) N = 4,998	0.192	-2.8%	0.70	-17.5%	1.01	12.2%	2.84	-4.2%	-0.84	8.2%	3.53	6.4%	-4.92	1.7%	8.4%
(25) N = 3,837	0.182	-7.6%	0.68	-20.0%	1.05	17.3%	2.91	-1.7%	-0.69	-11.8%	3.40	2.5%	-5.06	4.6%	9.6%
<i>Mean substitution</i>															
(5) N = 7,490	0.177	-10.1%	0.79	-6.9%	0.90	0.7%	2.66	-10.3%	-0.74	-5.5%	3.21	-3.3%	-4.74	-2.0%	4.8%
(15) N = 7,490	0.143	-27.3%	0.75	-11.1%	0.82	-8.8%	2.37	-19.9%	-0.71	-8.8%	2.83	-14.6%	-4.57	-5.4%	11.4%
(25) N = 7,490	0.175	-11.2%	0.71	-16.6%	0.79	-11.5%	2.02	-32.0%	-0.57	-26.3%	2.81	-15.4%	-4.64	-4.0%	17.6%
<i>Regression estimation</i>															
(5) N = 7,490	0.200	1.5%	0.82	-3.8%	0.92	2.6%	3.02	1.9%	-0.90	15.8%	3.27	-1.6%	-4.78	-1.1%	4.5%
(15) N = 7,490	0.201	2.0%	0.82	-3.5%	0.88	-1.8%	3.17	6.8%	-1.20	53.6%	3.06	-7.9%	-4.64	-4.0%	12.9%
(25) N = 7,490	0.203	2.8%	0.80	-5.3%	0.83	-7.7%	3.19	7.7%	-1.63	108.8%	3.07	-7.6%	-4.44	-8.1%	24.2%
<i>MI with dependent variable excluded</i>															
(5) N = 7,097	0.195	-1.1%	0.81	-4.9%	0.94	4.5%	2.96	-0.3%	-0.82	5.8%	3.20	-3.5%	-4.74	-1.9%	3.5%
(15) N = 6,388	0.186	-5.7%	0.82	-3.3%	0.92	2.5%	3.03	2.3%	-0.82	4.8%	2.90	-12.8%	-4.67	-3.3%	4.8%
(25) N = 5,633	0.175	-11.1%	0.81	-4.2%	0.92	3.0%	3.08	4.0%	-0.70	-9.8%	2.64	-20.4%	-4.66	-3.6%	7.5%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>															
(5) N = 7,490	0.199	0.9%	0.80	-5.5%	0.93	3.8%	2.88	-2.8%	-0.87	11.4%	3.35	0.8%	-4.74	-2.0%	4.4%
(15) N = 7,490	0.197	0.0%	0.76	-10.8%	0.92	2.6%	2.80	-5.6%	-1.00	28.5%	3.32	0.0%	-4.58	-5.2%	8.8%
(25) N = 7,490	0.196	-0.7%	0.77	-8.7%	0.95	5.6%	2.79	-5.9%	-0.82	5.4%	3.39	2.1%	-4.75	-1.8%	4.9%
<i>MI, inclusive strategy, most mechanism variables included</i>															
(5) N = 7,490	0.199	0.7%	0.79	-7.2%	0.93	3.9%	2.87	-3.1%	-0.85	9.4%	3.38	1.7%	-4.75	-1.7%	4.5%
(15) N = 7,490	0.198	0.1%	0.77	-9.3%	0.91	1.0%	2.84	-4.2%	-0.93	19.4%	3.39	2.0%	-4.68	-3.1%	6.5%
(25) N = 7,490	0.198	0.6%	0.78	-7.6%	0.92	2.5%	2.81	-5.4%	-0.82	5.7%	3.54	6.5%	-4.84	0.2%	4.7%
<i>MI, inclusive strategy, no mechanism variables included</i>															
(5) N = 7,490	0.199	0.8%	0.79	-6.4%	0.94	5.1%	2.87	-3.2%	-0.81	4.5%	3.41	2.7%	-4.81	-0.6%	3.8%
(15) N = 7,490	0.200	1.2%	0.81	-4.8%	0.93	3.5%	2.84	-4.2%	-0.87	12.0%	3.47	4.6%	-4.82	-0.4%	4.9%
(25) N = 7,490	0.195	-1.4%	0.82	-3.3%	0.90	-0.1%	2.80	-5.6%	-0.82	5.0%	3.52	6.0%	-4.87	0.7%	3.4%

(5): 5%, (15): 15%, (25): 25 % of values missing at random on empathic concern, low self-control and prevalence simple assault

Table 8.9: Estimates of standard errors for logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are missing MAR on independent variables and the dependent variable. Logistic regression, dependent variable: simple assault.

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.11		0.07		0.18		0.20		0.28		0.26		
<i>Listwise deletion</i>													
(5) N = 6,544	0.12	10.4%	0.08	10.0%	0.20	10.1%	0.21	9.5%	0.31	10.3%	0.29	9.8%	10.0%
(15) N = 4,998	0.15	30.9%	0.10	32.3%	0.24	33.3%	0.26	31.5%	0.37	31.7%	0.34	31.6%	31.9%
(25) N = 3,837	0.18	59.2%	0.11	57.5%	0.30	61.6%	0.31	58.8%	0.45	57.8%	0.42	60.3%	59.2%
<i>Mean substitution</i>													
(5) N = 7,490	0.11	1.5%	0.07	2.2%	0.18	-0.6%	0.20	3.1%	0.29	2.6%	0.27	1.8%	2.0%
(15) N = 7,490	0.12	6.5%	0.08	8.1%	0.18	-0.5%	0.22	12.7%	0.31	9.2%	0.28	7.2%	7.4%
(25) N = 7,490	0.13	12.5%	0.08	15.7%	0.18	0.8%	0.25	27.3%	0.34	19.3%	0.31	18.0%	15.6%
<i>Regression estimation</i>													
(5) N = 7,490	0.11	0.5%	0.07	1.2%	0.18	0.8%	0.20	3.0%	0.29	3.1%	0.27	2.2%	1.8%
(15) N = 7,490	0.12	3.7%	0.07	4.0%	0.19	2.5%	0.22	10.8%	0.31	10.1%	0.28	7.5%	6.4%
(25) N = 7,490	0.12	7.0%	0.08	6.9%	0.19	4.1%	0.24	20.6%	0.34	19.7%	0.30	15.7%	12.3%
<i>MI with dependent variable excluded</i>													
(5) N = 7,097	0.12	3.5%	0.07	3.4%	0.19	4.9%	0.21	6.6%	0.31	8.2%	0.28	5.4%	5.3%
(15) N = 6,388	0.13	12.1%	0.08	9.6%	0.21	12.0%	0.23	15.7%	0.33	17.8%	0.30	14.4%	13.6%
(25) N = 5,633	0.14	21.5%	0.08	17.7%	0.22	20.8%	0.26	35.4%	0.36	26.5%	0.34	31.3%	25.5%
<i>MI, restricted strategy, all mechanism variables in imputation model</i>													
(5) N = 7,490	0.12	2.9%	0.08	4.7%	0.19	5.8%	0.20	4.0%	0.31	11.0%	0.28	5.8%	5.7%
(15) N = 7,490	0.12	11.3%	0.08	13.0%	0.20	11.0%	0.27	37.1%	0.32	12.1%	0.35	32.1%	19.4%
(25) N = 7,490	0.15	33.7%	0.09	19.1%	0.25	36.8%	0.27	40.9%	0.43	50.6%	0.37	42.1%	37.2%
<i>MI, inclusive strategy, most mechanism variables included</i>													
(5) N = 7,490	0.11	1.1%	0.07	3.4%	0.19	2.1%	0.22	11.2%	0.30	5.1%	0.29	10.0%	5.5%
(15) N = 7,490	0.12	11.6%	0.08	9.5%	0.21	13.7%	0.23	20.4%	0.32	14.3%	0.31	16.9%	14.4%
(25) N = 7,490	0.12	8.8%	0.08	12.5%	0.22	18.2%	0.24	21.5%	0.41	43.9%	0.34	28.3%	22.2%
<i>MI, inclusive strategy, no mechanism variables included</i>													
(5) N = 7,490	0.12	3.2%	0.08	4.4%	0.19	2.9%	0.21	8.6%	0.31	8.3%	0.28	6.6%	5.6%
(15) N = 7,490	0.12	11.5%	0.08	12.9%	0.20	8.7%	0.25	25.8%	0.36	28.2%	0.32	21.1%	18.0%
(25) N = 7,490	0.15	32.5%	0.08	15.4%	0.22	20.4%	0.30	52.9%	0.37	30.2%	0.35	33.9%	30.9%

(5): 5%, (15): 15%, (25): 25 % of values missing at random on empathic concern, low self-control and prevalence simple assault

to the mean substitution approach. Regression estimation moderately overstates model fit, up to 2.8 per cent when many values are missing. The pure MI approaches yield a rather accurate R^2 , deviations ranging from -1.4 to .7 per cent.

On average, the logit coefficients are not too extremely biased for all approaches except for the single imputation ones, even when large amounts of data are missing. For the data set with the maximum amount of missing values, listwise deletion yields logit estimates that are on average only 9.6 per cent off. Although listwise deletion rather accurately reflects the effects of the independent variables with missing values (as long as the amount of missing data is not large), it tends to misestimate the effects of some independent variables without missing values, e.g., the coefficients of gender and nationality are off by up to 20 per cent. Single imputation strategies perform well when the share of missing is low. However, regression imputation grossly overestimates the effect of empathic concern, but this problem occurs for a number of approaches, although to a lesser extent. The MI approach where cases with missing values on the dependent variable are discarded also performs reasonably well, although it fails to estimate the effect of low self-control accurately when the share of missing values is moderate or large. The MI approaches where the dependent variable is imputed also tend to have problems estimating the effect of empathic concern, however, overall bias is fairly small, especially for the inclusive strategies.

The standard errors for the model coefficients (Table 8.9) are on average all rather large, at least compared to the model without missing values on the dependent variable (Table 8.5). Listwise deletion inflates standard errors for each predictor up to 60 per cent (Table 8.9). Standard errors obtained from single imputation approaches are not deflated, but still smaller than those obtained from the pure multiple imputation approaches. It is questionable whether the single imputation standard errors are not still too small to account for the uncertainty induced by the missing data. For the multiple imputation approaches, standard errors are on average smaller than those obtained from listwise deletion, however, for the imputed independent variables standard errors can be nearly as high as those obtained with listwise deletion. Furthermore, the imputation models that use an inclusive strategy tend to have smaller standard errors.

Values not missing at random on independent variables

The means from samples with non-random missingness (Table 8.10) show that most bias is induced by listwise deletion, again not only on those variables which have missing values, but

also on variables where all values were recorded. Means obtained from MI that makes use of all information in the data set are generally only half as biased as those obtained from listwise deletion. Mean substitution performs somewhat better than listwise deletion, but worse than MI.

In terms of standard errors, non-random missingness has ominous consequences (Table 8.11). For listwise deletion, standard errors in general become larger^{*}, while precision of the means obtained from mean substitution is highly overestimated. Standard errors resulting from MI are also somewhat off, mostly deflated, but less than the ones obtained from single imputation. It is apparent that even MI is hardly able to account for the uncertainty induced when data are NMAR.

Fit for models based on non-randomly missing values on independent variables is mostly understated (Table 8.12). Mc Fadden's R^2 is greatly reduced after listwise deletion the larger the share of missing values is. Mean substitution and MI also mostly decrease model fit.

For low and moderate shares of missing values, listwise deletion yields rather unbiased logit coefficients. However, the effect of empathic concern is highly underestimated when a large share of data are missing. Mean substitution yields biased estimates independent of the amount of missing values. MI also yields biased estimates for empathic concern when a small or moderate amount of values is missing. However, it yields the best estimates overall when the share of missing data is large.

The standard errors accompanying these models obtained after listwise deletion are the largest and become larger the larger the share of missing data is (Table 8.13). Standard errors obtained after mean substitution are smaller than the ones yielded after listwise deletion and MI. MI yields standard errors that also increase with the size of the share of missing data and that are disproportionately high for the variables with missing values.

^{*} The fact that the standard errors for the variables with few missing values actually decrease after listwise deletion while they increase in the other data sets after the same procedure can be explained by the way the values were deleted. The deletion process was not completely proportional across all data sets; deleting cases for the data set with few missing values attenuated variance less than in the other data sets. Therefore, when the standard error is the ratio of standard deviation of the population and the square root of the sample size, the standard error can either decrease or increase when variance and population size are reduced disproportionately.

Table 8.10: Estimated means and magnitude of their errors as a percentage of the estimate for the complete sample when data are missing NMAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	M	%	M	%	M	%	M	%	M	%	M	%	
<i>Complete sample</i>	0.87		0.48		0.34		0.69		0.45		0.21		
<i>Listwise deletion</i>													
(5) N = 6,712	0.87	-0.2%	0.46	-4.4%	0.33	-4.8%	0.71	2.9%	0.44	-3.1%	0.19	-11.0%	4.4%
(15) N = 5,456	0.86	-0.5%	0.45	-6.1%	0.32	-7.0%	0.72	4.2%	0.43	-4.8%	0.18	-16.0%	6.4%
(25) N = 4,323	0.86	-0.6%	0.43	-9.2%	0.31	-9.3%	0.73	5.8%	0.42	-6.2%	0.17	-19.9%	8.5%
<i>Mean substitution</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	2.0%	0.44	-2.3%	0.21	0.0%	0.7%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.71	3.0%	0.43	-3.5%	0.21	0.0%	1.1%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.72	4.3%	0.43	-4.8%	0.21	0.0%	1.5%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	1.4%	0.44	-1.5%	0.21	0.0%	0.5%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	2.1%	0.44	-2.4%	0.21	0.0%	0.7%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.71	3.0%	0.44	-3.3%	0.21	0.0%	1.1%

Table 8.11: Estimated standard errors for means and magnitude of their errors as a percentage of the estimate for the complete sample when data are NMAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.0039		0.0058		0.0025		0.0021		0.0016		0.0047		
<i>Listwise deletion</i>													
(5) N = 6,712	0.0041	6.3%	0.0061	5.4%	0.0025	1.4%	0.0021	-3.0%	0.0015	-1.1%	0.0048	1.1%	3.0%
(15) N = 5,456	0.0046	18.7%	0.0067	16.7%	0.0027	9.3%	0.0022	1.8%	0.0017	6.0%	0.0052	9.7%	10.4%
(25) N = 4,323	0.0052	33.8%	0.0075	30.6%	0.0030	20.0%	0.0023	5.9%	0.0018	12.9%	0.0057	20.9%	20.7%
<i>Mean substitution</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0019	-9.2%	0.0014	-7.9%	0.0047	0.0%	2.9%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0018	-16.6%	0.0013	-15.1%	0.0047	0.0%	5.3%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0016	-25.7%	0.0012	-24.1%	0.0047	0.0%	8.3%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-3.5%	0.0015	-2.7%	0.0047	0.0%	1.0%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-4.7%	0.0016	0.1%	0.0047	0.0%	0.8%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-4.4%	0.0017	5.6%	0.0047	0.0%	1.7%

(5): 5%, (15): 15%, (25): 25 % of values not missing at random on empathic concern and low self-control

Table 8.12: Estimates of McFaddens R^2 and logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are NMAR on independent variables. Logistic regression, dependent variable: simple assault

	Model fit		German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	R^2	%	b	%	b	%	b	%	B	%	b	%	b	%	
<i>Complete sample</i>	0.197		0.85		0.90		2.96		-0.78		3.32		-4.83		
<i>Listwise deletion</i>															
(5) N = 6,712	0.182	-7.8%	0.82	-3.3%	0.93	3.7%	2.92	-1.3%	-0.78	0.7%	3.45	3.9%	-4.88	1.0%	2.3%
(15) N = 5,456	0.169	-14.6%	0.82	-3.0%	0.96	6.5%	2.94	-0.7%	-0.72	-8.1%	3.38	1.9%	-4.92	1.8%	3.7%
(25) N = 4,323	0.151	-23.4%	0.70	-17.3%	0.99	10.4%	2.80	-5.5%	-0.60	-23.0%	3.46	4.2%	-4.87	0.8%	10.2%
<i>Mean substitution</i>															
(5) N = 7,490	0.193	-2.3%	0.87	2.2%	0.87	-3.4%	3.24	9.3%	-0.92	18.0%	3.09	-7.1%	-4.68	-3.2%	7.2%
(15) N = 7,490	0.187	-5.2%	0.91	7.2%	0.86	-4.7%	3.51	18.3%	-0.89	14.0%	2.68	-19.2%	-4.62	-4.5%	11.3%
(25) N = 7,490	0.190	-3.5%	0.96	13.2%	0.87	-3.4%	3.77	27.1%	-0.64	-17.3%	2.27	-31.7%	-4.72	-2.3%	15.8%
<i>MI, inclusive strategy</i>															
(5) N = 7,490	0.198	0.3%	0.83	-2.6%	0.89	-0.8%	2.96	-0.2%	-0.92	18.7%	3.47	4.6%	-4.75	-1.8%	4.8%
(15) N = 7,490	0.195	-1.1%	0.84	-1.6%	0.89	-0.9%	3.06	3.1%	-0.90	15.9%	3.34	0.5%	-4.72	-2.4%	4.0%
(25) N = 7,490	0.192	-2.8%	0.85	0.1%	0.90	0.4%	3.19	7.5%	-0.81	4.5%	3.17	-4.5%	-4.75	-1.8%	3.1%

Table 8.13: Estimates of standard errors for logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are NMAR on independent variables. Logistic regression, dependent variable: simple assault.

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.11		0.07		0.18		0.20		0.28		0.26		
<i>Listwise deletion</i>													
(5) N = 6,712	0.12	7.7%	0.08	8.2%	0.20	10.3%	0.23	16.3%	0.32	14.3%	0.29	12.2%	11.5%
(15) N = 5,456	0.13	19.9%	0.09	20.8%	0.23	25.5%	0.27	36.9%	0.37	31.3%	0.34	29.4%	27.3%
(25) N = 4,323	0.15	33.1%	0.10	35.1%	0.26	42.8%	0.33	66.7%	0.44	56.8%	0.40	52.7%	47.8%
<i>Mean substitution</i>													
(5) N = 7,490	0.11	-0.9%	0.07	-0.9%	0.18	-3.2%	0.21	8.7%	0.30	5.3%	0.27	2.3%	3.6%
(15) N = 7,490	0.11	-1.1%	0.07	-0.7%	0.17	-5.5%	0.23	18.4%	0.32	11.3%	0.28	7.0%	7.3%
(25) N = 7,490	0.11	-1.1%	0.07	-0.8%	0.17	-7.6%	0.26	33.8%	0.35	22.7%	0.30	16.2%	13.7%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.11	-0.1%	0.07	0.3%	0.18	0.8%	0.22	12.3%	0.31	8.6%	0.28	7.0%	4.8%
(15) N = 7,490	0.11	0.6%	0.07	0.2%	0.19	3.9%	0.23	18.3%	0.34	19.1%	0.29	9.2%	8.5%
(25) N = 7,490	0.11	1.2%	0.07	1.8%	0.19	2.9%	0.28	41.9%	0.36	27.2%	0.33	24.3%	16.5%

Values not missing at random on independent variables and the dependent variable

When values are non-randomly missing on a binary variable, this obviously leads to greatly biased means (Table 8.14). After listwise deletion, the mean of the prevalence variable is underestimated by 73 per cent when a large share of data is missing. For a moderate share, it is 53.1 per cent, and even for a small share it is 27.9 per cent. Means obtained after mean substitution and MI are about just as inaccurate. For a small share of missing data, the mean of the prevalence variable is underestimated by 23.1 per cent after mean substitution and 13.4 per cent after MI. For the data set with a large share of missing data, it is -67.1 per cent and -52.8 per cent, respectively.

Standard errors for these means (Table 8.15) are for the most part similar to those displayed in Table 8.11. A noteworthy exception are the estimates for the prevalence variable. Due to the sharp decrease in variance when removing many cases that have a value of 1 indicating that they have committed simple assault, the standard error decreases as well. Therefore, precision is—at least slightly—overstated for these highly biased point estimates, even when using MI.

When data are non-randomly missing both on the dependent variable and on independent variables (Table 8.16), listwise deletion and mean substitution tend to understate model fit by up to 50.9 per cent. MI tends to overstate model fit, but to a lesser degree (9.8 per cent).

Logit coefficients all show rather harsh bias, even when the share of missing values is small. Even when only 5 per cent of the values of empathic concern are imputed using MI, the estimate is off by 19.5 per cent. Under these conditions, the bias induced by listwise deletion seems moderate: On average, each estimate is off by 5 per cent, with the highest deviations being low self-control (8.7 per cent) and the constant (8.3 per cent). Mean substitution creates major bias, up to 98.5 per cent for the logit coefficient of empathic concern, when the share of missing data is high.

Standard errors for the logit coefficients are greatly inflated (Table 8.17), especially when compared to the standard errors that accompany the coefficient estimates from the MAR data sets (Table 8). Standard errors obtained after listwise deletion are inflated by 19.2 per cent to 199.0 per cent. Mean substitution and MI yield standard errors that are relatively small, however, they are all far from being called reasonable.

Table 8.14: Estimated means and magnitude of their errors as a percentage of the estimate for the complete sample when data are missing NMAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	M	%	M	%	M	%	M	%	M	%	M	%	
<i>Complete sample</i>	0.87		0.48		0.34		0.69		0.45		0.21		
<i>Listwise deletion</i>													
(5) N = 6,357	0.86	-0.5%	0.45	-6.7%	0.32	-6.2%	0.71	3.4%	0.43	-3.7%	0.15	-27.9%	8.1%
(15) N = 4,738	0.86	-1.1%	0.42	-11.9%	0.31	-10.8%	0.73	5.4%	0.42	-6.0%	0.10	-53.1%	14.7%
(25) N = 3,398	0.86	-0.9%	0.40	-16.5%	0.30	-14.1%	0.74	7.2%	0.41	-7.9%	0.06	-73.6%	20.0%
<i>Mean substitution</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	2.0%	0.44	-2.3%	0.16	-23.1%	4.6%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.71	3.0%	0.43	-3.5%	0.10	-50.3%	9.5%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.72	4.3%	0.43	-4.8%	0.05	-75.0%	14.0%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	1.3%	0.44	-1.5%	0.18	-13.4%	2.7%
(15) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.70	2.1%	0.44	-2.4%	0.15	-31.0%	5.9%
(25) N = 7,490	0.87	0.0%	0.48	0.0%	0.34	0.0%	0.71	3.0%	0.43	-3.4%	0.10	-52.8%	9.9%

Table 8.15: Estimated standard errors for means and magnitude of their errors as a percentage of the estimate for the complete sample when data are NMAR

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Prevalence simple assault		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.0039		0.0058		0.0025		0.0021		0.0016		0.0047		
<i>Listwise deletion</i>													
(5) N = 6,357	0.0043	9.9%	0.0062	8.0%	0.0026	3.2%	0.0021	-1.2%	0.0016	1.1%	0.0045	-4.5%	4.7%
(15) N = 4,738	0.0051	29.6%	0.0072	24.3%	0.0029	14.4%	0.0023	6.3%	0.0018	12.2%	0.0043	-8.0%	15.8%
(25) N = 3,398	0.0059	52.1%	0.0084	45.6%	0.0033	30.6%	0.0025	15.9%	0.0020	25.5%	0.0039	-16.5%	31.0%
<i>Mean substitution</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0019	-9.2%	0.0014	-7.9%	0.0043	-9.6%	4.5%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0018	-16.6%	0.0013	-15.1%	0.0035	-25.0%	9.4%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0016	-25.7%	0.0012	-24.1%	0.0026	-45.2%	15.8%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0021	-2.6%	0.0015	-1.8%	0.0047	-0.2%	0.8%
(15) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0020	-6.7%	0.0015	-1.3%	0.0043	-8.0%	2.7%
(25) N = 7,490	0.0039	0.0%	0.0058	0.0%	0.0025	0.0%	0.0019	-9.5%	0.0017	9.1%	0.0047	-0.3%	3.2%

(5): 5%, (15): 15%, (25): 25 % of values not missing at random on empathic concern, low self-control and prevalence simple assault

Table 8.16: Estimates of McFaddens R^2 and logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are NMAR on independent variables and the dependent variable. Logistic regression, dependent variable: simple assault

	Model fit		German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	R^2	%	b	%	b	%	b	%	b	%	b	%	b	%	
<i>Complete sample</i>	0.197		0.85		0.90		2.96		-0.78		3.32		-4.83		
<i>Listwise deletion</i>															
(5) N = 6,357	0.186	-5.7%	0.80	-5.2%	0.91	1.2%	3.05	2.9%	-0.81	3.4%	3.61	8.7%	-5.24	8.3%	5.0%
(15) N = 4,738	0.159	-19.3%	0.57	-32.3%	0.98	9.4%	2.81	-5.1%	-0.64	-17.5%	3.95	19.0%	-5.66	17.1%	16.7%
(25) N = 3,398	0.154	-21.9%	0.88	3.8%	1.33	48.6%	2.86	-3.7%	-0.36	-53.4%	3.95	18.9%	-6.93	43.4%	28.6%
<i>Mean substitution</i>															
(5) N = 7,490	0.172	-12.9%	0.79	-7.4%	0.84	-5.9%	3.06	3.1%	-0.68	-12.4%	2.88	-13.1%	-4.96	2.7%	7.4%
(15) N = 7,490	0.122	-38.2%	0.55	-35.5%	0.77	-13.9%	2.71	-8.6%	-0.47	-39.8%	2.31	-30.5%	-4.94	2.2%	21.8%
(25) N = 7,490	0.097	-50.9%	0.76	-10.8%	0.85	-5.1%	2.59	-12.7%	-0.01	-98.5%	2.04	-38.6%	-6.07	25.6%	31.9%
<i>MI, inclusive strategy</i>															
(5) N = 7,490	0.209	6.0%	0.85	-0.3%	0.90	-0.1%	3.11	4.9%	-0.93	19.5%	3.66	10.3%	-5.15	6.6%	7.0%
(15) N = 7,490	0.205	3.8%	0.70	-17.2%	0.90	0.4%	3.04	2.5%	-0.97	24.2%	3.86	16.1%	-5.38	11.3%	12.0%
(25) N = 7,490	0.217	9.8%	1.06	24.4%	1.06	18.4%	3.44	16.2%	-0.45	-42.2%	3.82	15.0%	-6.82	41.1%	26.2%

Table 8.17: Estimates of standard errors for logit coefficients along with their magnitude of errors as a percentage of the estimate for the complete sample when data are NMAR on independent variables and the dependent variable. Logistic regression, dependent variable: simple assault.

	German respondent		Male respondent		Approval of violence: normative		Empathic concern		Low self-control		Intercept		M %
	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	S.E.	%	
<i>Complete sample</i>	0.11		0.07		0.18		0.20		0.28		0.26		
<i>Listwise deletion</i>													
(5) N = 6,357	0.13	19.2%	0.09	21.5%	0.22	22.4%	0.25	28.7%	0.36	27.7%	0.33	25.0%	24.1%
(15) N = 4,738	0.17	50.9%	0.12	65.1%	0.30	65.0%	0.36	83.7%	0.50	76.4%	0.45	73.7%	69.1%
(25) N = 3,398	0.28	153.0%	0.19	160.3%	0.46	152.5%	0.58	199.0%	0.80	181.0%	0.74	181.8%	171.3%
<i>Mean substitution</i>													
(5) N = 7,490	0.12	8.0%	0.08	9.3%	0.19	2.7%	0.23	15.7%	0.32	12.7%	0.29	9.9%	9.7%
(15) N = 7,490	0.14	21.3%	0.09	29.9%	0.21	12.0%	0.28	43.4%	0.38	34.8%	0.34	29.6%	28.5%
(25) N = 7,490	0.20	76.7%	0.13	80.8%	0.26	41.2%	0.43	118.1%	0.55	95.3%	0.50	89.9%	83.6%
<i>MI, inclusive strategy</i>													
(5) N = 7,490	0.12	10.7%	0.08	11.6%	0.20	9.5%	0.23	17.1%	0.33	16.8%	0.30	13.0%	13.1%
(15) N = 7,490	0.15	31.2%	0.10	34.7%	0.23	28.1%	0.30	51.7%	0.44	56.0%	0.40	52.2%	42.3%
(25) N = 7,490	0.21	83.6%	0.14	96.7%	0.33	81.1%	0.55	184.5%	0.58	104.2%	0.69	163.1%	118.9%

Discussion and summary

Means obtained after *listwise deletion* should generally not be interpreted as representative for all cases in the sample. They are the more biased, the larger the share of missing data is, and their standard errors are inflated. When data are NMAR, they are markedly worse than the raw means or those obtained by MI.

When used for a logistic regression model, listwise deletion tends to seriously understate model fit when the amount of missing values becomes large. Furthermore, listwise deletion yields reasonable or good coefficient estimates as long as the share of missing data is small, even when data are NMAR. However, all coefficients can be somewhat biased, even for predictors that do not have missing values. This is a disadvantage compared to MI approaches, where biased coefficient estimates tend to occur only for the variables with missing values. A similar problem occurs regarding the standard errors. While MI tends to inflate standard errors mostly for the predictors with missing values when missing values are confined to the independent variables, listwise deletion inflates the standard error of each and every predictor. Although listwise deletion performed best when there was a moderate amount of NMAR values on the independent variables, standard errors were so inflated that power could have become an issue.

In any simulated scenario, means obtained from the two *single imputation* methods are in general not more accurate than those yielded by MI, but the standard errors accompanying those means are deflated since variance is lost. In terms of means the quick-and-dirty approach failed when it was used to predict the binary dependent variable: Means are highly underestimated in this case. Model fit tended to be understated, and regression coefficients in general were estimated worse than when using other approaches. When the share of missing data was small, bias for the regression coefficients was moderate, though not optimal. When larger parts of a data set had to be imputed, results were outrageous. Single imputation such as mean substitution and regression estimation are generally not recommendable. However, when the share of missing data is small and no advanced method is at hand, one could consider using it.

All *multiple imputation* approaches (except for the one that was combined with listwise deletion) generally yielded the most accurate estimates for means when compared to the raw data

(Table 8.1). Under extreme conditions, i.e. NMAR data with many missing values, MI using the maximum amount of information performed better than the other listwise deletion, or mean imputation, and was more accurate than the mean of the raw data. Still, the MI estimate was 52.8 per cent too low. Remarkable under these conditions was that even MI standard errors can become deflated, although not as much as the ones obtained after listwise deletion.

When used with logistic regression models, all MI approaches tend to perform reasonably. Model fit is assessed rather accurate as long as not many values are NMAR. When comparing the results of the four MI approaches used for MAR data, it becomes clear that excluding the dependent variable from the imputation model has no beneficial effect for any analysis.* Furthermore, the inclusive strategies tend to perform better than the restrictive strategy that models the exact missingness mechanism, although they are based merely on approximations of the mechanism or draw from a large amount of information in the data set. The inclusive strategies tend to perform about as good as listwise deletion in terms of coefficient estimates when data are MAR, but yield smaller standard errors. When values are NMAR confined on the independent variables, MI performs clearly worse than listwise deletion when the share of missing data is small or moderate, however, when the share is large, MI is clearly superior. When data is NMAR both on independent variables and dependent variable, MI estimates are far from being accurate. However, for moderately-sized and large missing data problems, they perform somewhat better than listwise deletion.

Which recommendations can be given based on this simulation? Every tested approach is better than single-imputation approaches. When trying to estimate means for a sample, MI can improve the estimates. When planning to use listwise deletion, one should have an eye on sample size to avoid type II errors. When planning to use MI, one should use the maximum amount of information possible; when data are MAR, even mere approximations of the missing data mechanism can perform well. It is crucial to include the dependent variable into the imputation model. And when trying to tackle a missing data problem, multiple different strategies should be undertaken to assess the extent of the problem; since obviously there is no single and superior solution to all missing data problems.

* However, von Hippel (2007) shows that including the dependent variable into the imputation model and then deleting the cases with imputed values before the analysis can in fact be a viable strategy.

9. Using Multiple Imputation to analyze the data set

In this chapter, an attempt to use an advanced missing data technique for a substantial analysis of data is undertaken. Results obtained from Multiple Imputation are compared to results achieved by conventional methods.

Enzmann and Siegmunt (2005) report a remarkable finding in their analysis of the data set that is also used in this study. In their analysis of the incomplete cases they find that the prevalence rates in Germany and Russia follow a remarkable pattern. While Russian juveniles tend to commit more acts that might be called 'instrumental' in a sense of acquiring other people's property, German juveniles tend to commit the non-instrumental simple assault much more often. They interpret this finding against the background of institutional anomie theory (Messner and Rosenfeld 2006), assuming that crime is created by an imbalance of important societal institutions. The sudden shift from a communist society that was dominated by the political system to a capitalist society dominated by the economy might have led to a cynical view of individual responsibility and an increase in violent and property crime.

In order to test whether this finding holds true when a modern missing data technique is used, the variables were imputed using Multiple Imputation. Again Royston's (2004; 2005a; 2005b) 'ice' for Stata was used. In order to allow for the maximum amount of information to be used, the data set was split for Germans and Russians and missing values of the four prevalence rates were imputed. Information used comprised a large number of variables that were correlated either with the prevalence rate variables in question or with missing data indicator variables for these prevalence rate variables. Among these were the variables indicating whether respondent's father or mother had died, the prevalence rates for the other criminal acts that were probed for, prevalence rates of victimization, low self-control, attitude towards violence, whether respondent's parents were divorced or separated, self-esteem, competence to solve conflicts, attachment to parents, parents' perceived attitude towards violence, and teachers handling of violence acts among students, low socio-economic status, and for Germany: right-wing extremism. The maximum number of imputations that was computationally feasible was 10. Most authors assume this to be conservative choice.

Table 9.1: Prevalence rates and 95% confidence intervals for complete cases only and complete and imputed cases

		Complete cases			Complete and imputed cases		
		Lower CI-95	M	Upper CI-95	Lower CI-95	M	Upper CI-95
Robbery	Germany	3.82%	4.89%	5.96%	3.45%	6.08%	8.71%
	Russia	4.91%	5.45%	5.98%	5.18%	5.75%	6.32%
Extortion	Germany	1.04%	1.30%	1.57%	1.14%	1.45%	1.75%
	Russia	1.84%	2.63%	3.43%	2.16%	3.61%	5.06%
Threat w/ a weapon	Germany	2.78%	3.20%	3.61%	3.00%	3.43%	3.87%
	Russia	2.17%	3.02%	3.87%	1.82%	4.22%	6.63%
Simple assault	Germany	22.09%	23.09%	24.08%	22.79%	23.80%	24.80%
	Russia	10.49%	12.12%	13.74%	11.40%	13.33%	15.25%

When comparing the MI results to those obtained by conventional methods, it becomes clear that the values obtained from MI do not challenge Enzmann and Siegmunt's (2005) findings. All means are corrected upwards somewhat, and sometimes means in one sample become higher than the means in the other one, however, due to the uncertainty induced confidence intervals, especially in Russia, become wider. This comparison of course cannot rule out the possibility that data are nonrandomly missing. In terms of amount of information used in the imputation model, the MAR assumption should be justified.

Enzmann and Siegmunt (2005) conduct further analyses such as a cluster analysis to test whether institutional anomie theory is really applicable. It would be interesting to see if results achieved in these analyses will also hold true with the imputed data.

10. Conclusion

Missing data can pose a serious threat to validity in quantitative research on crime and delinquency, perhaps even more than in other areas of research. Conventional ways of handling missing data tend to produce biased estimates, diminish power, and waste data. Modern missing data techniques attempt to make use of all available information (even if it is incomplete for some cases) and to account for this uncertainty in the process of statistical inference. However, these techniques are based on assumptions about the missing values that can hardly be tested and are sometimes difficult to justify. But when the assumptions are tenable, the advanced missing data techniques are said to perform optimal.

The simulation has shown that using Multiple Imputation for solving missing data problems is definitely worth considering. When doing so, the maximum amount of information should be used. Single imputation strategies are not recommendable. Listwise deletion is a strategy that is worth considering, however, the loss of cases can seriously reduce statistical power. One should consider comparing the results of different missing data techniques to assess the extent of the missing data problem. However, the performance of missing data techniques under various conditions requires more systematic research. The question how missing data techniques perform when estimating incidence rates remains. But no method presented is a panacea for the problem of missing values.

Using Multiple Imputation to tackle the missing data problem in a simple analysis did not yield substantially different results than a complete-case analysis, however, it would be interesting to see results from more complex analyses.

Furthermore, this study has shown that many missing data problems can probably be solved by improving data collection. Complicated question structures seem to be responsible for many missing values, as well as open-ended numerical questions. Also, it is striking that partial nonresponse is much higher in the Russian sample, although the questionnaires in both countries had about the same length. More time for completing the questionnaire could probably have reduced bias by nonresponse more than any sophisticated missing data technique.

But improvements of the data collection process do not make the presented modern missing data techniques dispensable. When using multiple matrix sampling, questionnaires can be effectively shortened, thus reducing nonresponse by respondents with lower self-control and fewer cognitive abilities. The resulting data should be MCAR (for the most part), therefore modern missing data techniques should perform as desired in the analysis.

And when researchers know about the prerequisites of the successful use of modern missing data techniques, they can attempt to collect data systematically that can be used to justify the MAR assumption for important variables when a missing data problem occurs.

Therefore it can be assumed that modern missing data techniques mark a major advance in the methodology of quantitative social research and will have beneficial effects on crime research as well.

11. References

- Achenbach, Thomas M., Stephanie H. McConaughy, and Catherine T. Howell 1987: 'Child/Adolescent Behavioral and Emotional Problems: Implications of Cross-Informant Correlations for Situational Specificity,' *Psychological Bulletin* 101, pp. 213-232.
- Acock, Alan C. 2005. 'Working with Missing Values,' *Journal of Marriage and Family* 67, pp. 1012-1028.
- Agnew, Robert 1992: 'Foundation for a General Strain Theory of Crime and Delinquency,' *Criminology* 30, pp. 47-87.
- Allison, Paul D. 2002: *Missing Data*, Thousand Oaks, CA, London, and New Delhi: Sage.
- Azar, Beth 2002: 'Finding a Solution for Missing Data,' *APA Monitor on Psychology* 33, p. 70.
- Baier, Dirk, Susanne Rabold, Christian Pfeiffer, and Michael Windzio 2006: *Schülerbefragung 2005: Gewalterfahrungen, Schulabsentismus und Medienkonsum von Kindern und Jugendlichen in Thüringen. Abschlußbericht über eine repräsentative Befragung von Schülerinnen und Schülern der 4. und 9. Jahrgangsstufe* [Violence, Truancy, and Media Consumption of Children and Juveniles in Thuringia. Final Report on a Representative Survey of Fourth- and Ninth-Grade Students] Hanover: KFN.
- Becker, Rolf 2006: 'Selective Response to Questions on Delinquency,' *Quality and Quantity* 40, pp. 483-498.
- Bennett, Sarah, David P. Farrington, and L. Rowell Huesmann 2005: 'Explaining Gender Differences in Crime and Violence. The Importance of Social Cognitive Skills,' *Aggression and Violent Behavior* 10, pp. 263-288.
- Best, Joel 2001: *Damned Lies and Statistics. Untangling Numbers from the Media, Politicians, and Activists*, Berkeley, CA: University of California Press.
- Best, Joel 2004: *More Damned Lies and Statistics. How Numbers Confuse Public Issues*, Berkeley, CA: University of California Press.
- Brame, Robert, and Raymond Paternoster 2002: 'Missing Data Problems in Criminological Research,' *Journal of Quantitative Criminology* 19, pp. 55-78.
- Buck, S. F. 1960: 'A Method of Estimation of Missing Values in Multivariate Data Suitable for Use with an Electronic Computer,' *Journal of the Royal Statistical Society B* 22, pp. 302-306.
- Chen, Jiahua and Jun Shao 2000: 'Nearest Neighbor Imputation for Survey Data,' *Journal of Official Statistics* 16, pp. 113-131.
- Cohen, Jacob 1992: 'A Power Primer,' *Psychological Bulletin* 112, pp. 155-159.
- Cohen, Jacob, and Patricia Cohen 1983: *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 2nd ed., Hillsdale, NJ: Erlbaum.
- Coleman, Clive, and Jenny Moynihan 1996: *Understanding Crime Data. Haunted by the Dark Figure*, Buckingham and Bristol, PA: Open University Press.
- Collins, Linda M., Joseph L. Schafer and Chi-Ming Kam 2001: 'A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures,' *Psychological Methods* 6, pp. 330-351.

- Cronbach, Lee J. 1951: 'Coefficient Alpha and the Internal Structure of Tests,' *Psychometrika* 16, pp. 297-334
- Crowne, Douglas P., and David Marlowe 1960: 'A New Scale of Social Desirability Independent of Psychopathology,' *Journal of Consulting Psychology* 24, pp. 349-354.
- Davis, Mark H. 1980: 'A Multidimensional Approach to Individual Differences in Empathy,' *JSAS Catalog of Selected Documents in Psychology* 10, p. 85.
- Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin 1977: 'Maximum Likelihood from Incomplete Data via the EM Algorithm,' *Journal of the Royal Statistical Society B* 39, pp. 1-22.
- De Leeuw, Edith D., Joop Hox, and Mark Huismann 2003: 'Prevention and Treatment of Item Nonresponse,' *Journal of Official Statistics* 19, pp. 153-176.
- De Los Reyes, Andres, and Alan E. Kazdin 2004: 'Measuring Informant Discrepancies in Clinical Child Research,' *Psychological Assessment* 16, pp. 330-334.
- Eisner, Manuel, and Denis Ribeaud 2007: 'Conducting a Criminological Survey in a Culturally Diverse Context. Lessons from the Zürich Project on the Social Development of Children,' *European Journal of Criminology* 4, pp. 271-298.
- Elliott, Delbert S., David H. Huizinga, and Suzanne S. Ageton 1985: *Explaining Delinquency and Drug Use*, Beverly Hills, CA: Sage.
- Enders, Craig K. 2001: 'The Impact of Nonnormality on Full Information Maximum-Likelihood Estimation for Structural Equation Models with Missing Data,' *Psychological Methods* 6, pp. 352-370.
- Enzmann, Dirk 1995: *Gestresst, erschöpft oder ausgebrannt? Einflüsse von Arbeitssituation, Empathie und Coping auf den Burnoutprozeß* [Stressed, Exhausted, or Burned Out? Impacts of Working Conditions, Empathy, and Coping on the Burnout Process], Munich and Vienna: Profil.
- Enzmann, Dirk 2005: *Response Function Imputation and Cronbach's alpha*, mimeo.
- Enzmann, Dirk, and Olga Siegmunt 2005: 'Analyse der Jugendkriminalität in Deutschland und Russland: Ergebnisse einer Schülerbefragung in Wolgograd und Deutschland' [Analysis of Youth Crime in Germany and Russia. Results of a Survey among Students in Volgograd and Germany], *Standpunkt Sozial. Hamburger Forum für Soziale Arbeit* 16 (3), pp. 20-25.
- Enzmann, Dirk, Katrin Brettfeld, and Peter Wetzels 2003: 'Männlichkeitsnormen und die Kultur der Ehre. Empirische Prüfung eines theoretischen Modells zur Erklärung erhöhter Delinquenzraten jugendlicher Migranten' [Norms of Masculinity and Culture of Honor. Empirical Investigation of a Theoretical Model to Explain Elevated Rates of Delinquency among Juvenile Migrants], in: Dietrich Oberwittler and Susanne Karstedt (eds.): *Soziologie der Kriminalität* [Sociology of Crime], special issue 43 of the *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, Wiesbaden: VS, pp. 264-287.
- Farrington, David P. 2007: 'Childhood Risk-Factors and Risk-Focused Prevention,' in: Mike Maguire, Rod Morgan, and Robert Reiner (eds.): *The Oxford Handbook of Criminology*, 4th ed., Oxford: Oxford University Press, pp. 602-640.
- Farrington, David P., Bernard Gallagher, Lynda Morley, Raymond J. St. Ledger, and Donald J. West 1990: 'Minimizing Attrition in Longitudinal Research. Methods of Tracing and Securing Cooperation in a 24-Year Follow-up Study,' in: David Magnusson and

- Lars R. Bergman (eds.): *Data Quality in Longitudinal Research*, Cambridge: Cambridge University Press, pp. 122-147.
- Farrington, David P., Rolf Loeber, Magda Stouthamer-Loeber, Welmoet B. Van Kammen, and Laura Schmidt 1996: 'Self-Reported Delinquency and a Combined Delinquency Seriousness Scale Based on Boys, Mothers, and Teachers. Concurrent and Predictive Validity for African Americans and Caucasians,' *Criminology* 34, pp. 493-517.
- Fichman, Mark, and Jonathon N. Cummings 2003: 'Multiple Imputation for Missing Data. Making the Most of What You Know,' *Organizational Research Methods* 6, pp. 282-308.
- Frégier, Honoré Antoine 1840: *Des classes dangereuses et de la population dans les grandes villes et moyens de les rendre meilleures*, Paris: JB Ballière.
- Glynn, Robert J., Nan M. Laird and Donald B. Rubin 1993: 'Multiple Imputation in Mixture Models for Nonignorable Nonresponse with Follow-ups,' *Journal of the American Statistical Association* 88, pp. 984-993.
- Goyder, John 1987: *The Silent Minority. Nonrespondents on Sample Surveys*, Cambridge: Polity.
- Graham, John W. 1995: *Emcov Reference Manual*, <http://ftp.cac.psu.edu/pub/people/jwg4/dos/emcov.txt> (last accessed: Aug 20, 2007).
- Graham, John W. and Stewart I. Donaldson 1993: 'Evaluation Interventions with Differential Attrition. The Importance of Nonresponse Mechanisms and Use of Follow-Up Data,' *Journal of Applied Psychology* 78, pp. 119-128.
- Graham, John W., Patricio E. Cumsille, and Elvira Elek-Fisk 2003: 'Methods for Handling Missing Data,' in: John A. Schinka and Wayne F. Velicer (eds.), *Handbook of Psychology (Vol. 2: Research Methods in Psychology)*, Hoboken, NJ: John Wiley, pp. 87-114.
- Grasmick, Harold G., Charles R. Tittle, Robert J. Bursik, Jr., and Bruce J. Arneklev 1993: 'Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime,' *Journal of Research in Crime and Delinquency* 30, pp. 5-29.
- Hawkins, J. David, Todd Herrenkohl, David P. Farrington, Devon Brewer, Richard F. Catalano, and Tracy W. Harachi 1998: 'A Review of Predictors of Youth Violence,' in: Rolf Loeber and David P. Farrington (eds.): *Serious and Violent Juvenile Offenders. Risk Factors and Successful Interventions*, Thousand Oaks, CA, London, and New Delhi: Sage, pp. 106-146.
- Heidensohn, Frances, and Loraine Gelstorphe 2007: 'Gender and Crime,' in: Mike Maguire, Rod Morgan, and Robert Reiner (eds.): *The Oxford Handbook of Criminology*, 4th ed., Oxford: Oxford University Press, pp. 381-420.
- Holbrook, Allyson L., Jon A. Krosnick, Richard T. Carson, and Robert Cameron Mitchell 2000: 'Violating Conversational Conventions Disrupts Cognitive Processing of Attitude Questions,' *Journal of Experimental Social Psychology* 36, pp. 465-494.
- Honaker, James, Gary King, and Matthew Blackwell 2007: *Amelia II. A Program for Missing Data*, <http://gking.harvard.edu/amelia> (last accessed Oct 2, 2007).
- Horton, Nicholas J., and Ken P. Kleinman 2007: 'Much Ado About Nothing: A Comparison of Missing Data Methods and Software to Fit Incomplete Data Regression Models,' *The American Statistician* 61, pp. 79-90.

- Hosser, Daniela, and Dana Beckurts 2005: *Empathie und Delinquenz* [Empathy and Delinquency], KFN-Forschungsberichte No. 96, Hanover: KFN.
- Huisman, Mark 2000: 'Imputation of Missing Item Responses: Some Simple Techniques,' *Quality and Quantity* 34, pp. 331-351.
- Huisman, Mark, Boudien Krol, and Eric van Sonderen 1998: 'Handling Missing Data by Re-approaching Non-respondents,' *Quality and Quantity* 32, pp. 77-91.
- Huizinga, David, and Cynthia Jakob-Chien 1998: 'The Contemporaneous Co-Occurrence of Serious and Violent Juvenile Offending and Other Problem Behaviors,' in: Rolf Loeber and David P. Farrington (eds.): *Serious and Violent Juvenile Offenders. Risk Factors and Successful Interventions*, Thousand Oaks, CA, London, and New Delhi: Sage, pp. 47-67.
- Huizinga, David, and Delbert S. Elliott 1986: 'Reassessing the Reliability and Validity of Self-Report Delinquency Measures,' *Journal of Quantitative Criminology* 2, pp. 293-327.
- Jamshidian, Mortaza 2004: 'Strategies for Analysis of Incomplete Data,' in: Melissa Hardy and Alan Bryman (eds.): *Handbook of Data Analysis*, London, Thousand Oaks, CA, and New Delhi: Sage, pp. 113-130.
- Jasch, Michael 2004: 'Police and Prosecutions. Vanishing Differences between Practices in England and Germany,' *German Law Journal* 5, pp. 1208-1216.
- Jinn, Jann-Huei and Joe Sedransk 1989: 'Effect on Secondary Data Analysis of Common Imputation Methods,' *Sociological Methodology* 19, pp. 213-241.
- Jolliffe, Darrick, and David P. Farrington 2004: 'Empathy and Offending. A Systematic Review and Meta-Analysis,' *Aggression and Violent Behavior* 9, pp. 441-471.
- Jones, Edward E., and Harold Sigall 1971: 'The Bogus Pipeline. A New Paradigm for Measuring Affect and Attitude,' *Psychological Bulletin* 76, pp. 349-364.
- Junger-Tas, Josine, and Ineke Haen Marshall 1999: 'The Self-Report Methodology in Crime Research,' *Crime and Justice* 25, p. 291-367.
- Junger-Tas, Josine, Denise Ribeaud, and Maarten J. L. F. Cruyff 2004: 'Juvenile Delinquency and Gender,' *European Journal of Criminology* 1, pp. 333-375.
- Junger-Tas, Josine, Gert-Jan Terlouw, and Malcolm W. Klein 1994: *Delinquent Behavior among Young People in the Western World*, Amsterdam and New York: Kugler.
- Katz, Jack 1988: *Seductions of Crime. Moral and Sensual Attractions in Doing Evil*, New York: Basic.
- King, Gary, James Honaker, Anne Joseph, and Kenneth Scheve 2001: 'Analyzing Incomplete Political Science Data. An Alternative Algorithm for Multiple Imputation,' *American Political Science Review* 95, pp. 49-69.
- Kitsuse, John I., and Aaron V. Cicourel 1963: 'A Note on the Uses of Official Statistics,' *Social Problems* 11, p. 131-139.
- Kleck, Gary, Jongyeon Tark and Jon J. Bellows 2006: 'What Methods Are Most Frequently Used in Research in Criminology and Criminal Justice?' *Journal of Criminal Justice* 34, p. 147-152.
- Köllisch, Tilman, and Dietrich Oberwittler 2004: 'Wie ehrlich berichten männliche Jugendliche über ihr delinquentes Verhalten? Ergebnisse einer externen Validierung' [How Honestly do Male Adolescents Report their Delinquent Behaviour? Results of an Ex-

- ternal Validation of Self-reported Delinquency], *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 56, pp. 708-735.
- Kury, Helmut 2001: ‚Das Dunkelfeld der Kriminalität. Oder: Selektionsmechanismen und andere Verfälschungsstrukturen’ [*The Dark Figure of Criminality, Or, Selection Mechanisms and Other Tampering Structures], *Kriminalistik* 55, pp. 74-84.
- Levitt, Steven D. 1998: ‘The Relationship between Crime Reporting and Police. Implications for the Use of Uniform Crime Reports,’ *Journal of Quantitative Criminology* 14, pp. 61-81.
- Lipsey, Mark W., and James H. Derzon 1998: ‘Predictors of Violent or Serious Delinquency in Adolescence and Early Adulthood. A Synthesis of Longitudinal Research,’ in: Rolf Loeber and David P. Farrington (eds.): *Serious and Violent Juvenile Offenders. Risk Factors and Successful Interventions*, Thousand Oaks, CA, London, and New Delhi: Sage, pp. 86-105.
- Little, Roderick J. A. 1988: ‘A Test of Missing Completely at Random for Multivariate Data with Missing Values,’ *Journal of the American Statistical Association* 83, pp. 1198-1202.
- Little, Roderick J. A. 1992: ‘Regression with Missing X’s. A Review,’ *Journal of the American Statistical Association* 87, pp. 1227-1237.
- Little, Roderick J. A. and Donald B. Rubin 2002: *Statistical Analysis with Missing Data*, 2nd ed., Hoboken, NJ: John Wiley.
- Little, Roderick J. A. and Nathaniel Schenker 1995: ‘Missing Data,’ in: Gerhard Arminger, Clifford C. Clogg and Michael E. Sobel (eds.): *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, New York: Plenum, pp. 39-75.
- Loeber, Rolf, and David P. Farrington (eds.) 2001: *Child Delinquents. Development, Intervention, and Service Needs*, London, Thousand Oaks, CA, and New Delhi: Sage.
- Loeber, Rolf, and Marc LeBlanc 1990: ‘Toward a Developmental Criminology,’ *Crime and Justice* 12, pp. 375-437.
- Long, J. Scott 1997: *Regression Models for Categorical and Limited Dependent Variables*, Thousand Oak, CA, London, and New Delhi: Sage.
- Lösel, Friedrich 1975: *Handlungskontrolle und Jugenddelinquenz*, Stuttgart: Enke.
- Lüdtke, Oliver, Alexander Robitzsch, Ulrich Trautwein, and Olaf Köller 2007: ‘Umgang mit fehlenden Werten in der psychologischen Forschung. Probleme und Lösungen’ [Handling of Missing Data in Psychological Research. Problems and Solutions], *Psychologische Rundschau* 58, pp. 103-117.
- Maguire, Mike 2002: ‘Crime Statistics: The “Data Explosion” and Its Implications,’ in: Mike Maguire, Rod Morgan, and Robert Reiner (eds.): *The Oxford Handbook of Criminology*, 3rd ed., Oxford: Oxford University Press, pp. 322-375.
- Merton, Robert K., 1938: ‘Social Structure and Anomie,’ *American Sociological Review* 3, pp. 672-682.
- Messner, Steven F., and Richard Rosenfeld 2006: *Crime and the American Dream*, 4th ed., Belmont, CA: Wadsworth.
- Miller, Peter V., and Robert M. Groves 1985: ‘Matching Survey Responses to Official Records. An Exploration of Validity in Victimization Responses,’ *Public Opinion Quarterly* 49, pp. 366-380.

- Mislevy, Robert J., Albert E. Beaton, Bruce Kaplan, and Kathleen M. Sheehan 1992: 'Estimating Population Characteristics from Sparse Matrix Samples of Item Responses,' *Journal of Educational Measurement* 29, pp. 133-161.
- Mittag, Oskar, Thorsten Meyer, Ruth Deck, Thomas Kohlmann, Christine Matthis, Thorsten Schäfer, and Heiner Raspe 2005: 'Fehlende Werte in der Somatisierungsskala des SCL-90-R durch „Listenfragen-Effekt“. Häufigkeit in verschiedenen Surveys und Vergleich unterschiedlicher Imputationsmethoden' [Missing Data in the Somatisation Subscale of the SCL-90-R Due to a 'Checklist Effect'. Occurrence in Different Surveys and Results of Various Imputation Methods], *Gesundheitswesen* 67, pp. 155-158.
- Myers, Samuel L. 1980: 'Why Are Crimes Underreported? What is the Crime Rate? Does It Really Matter?,' *Social Science Quarterly* 61, pp. 23-43.
- Narayan, Sowmya, and Jon A. Krosnick 1996: 'Education Moderates Some Response Effects in Attitude Measurement,' *Public Opinion Quarterly* 60: pp. 58-88.
- Nye, F. Ivan, and James F. Short, Jr. 1957: 'Scaling Delinquent Behavior,' *American Sociological Review* 22, pp. 326-331.
- Nye, F. Ivan, James F. Short, Jr., and Virgil J. Olson 1958: 'Socioeconomic Status and Delinquent Behavior,' *American Journal of Sociology* 63, pp. 381-389.
- O'Brien, Robert M. 1996: 'Police Productivity and Crime Rates 1973-1992,' *Criminology* 34, pp. 183-207.
- Oberwittler, Dietrich, and Thomas Naplava 2002: 'Auswirkungen des Erhebungsverfahrens bei Jugendbefragungen zu „heiklen“ Themen. Schulbasierte schriftliche Befragung und haushaltsbasierte mündliche Befragung im Vergleich' [Impact of the data collection mode in youth surveys on sensitive topics. School-based pencil-and-paper surveys and household-based face-to-face surveys compared], *ZUMA-Nachrichten* 51, pp. 49-77.
- Paul, Bettina 2006: 'Ohne Geld und Bildung eher kriminell? Über die Schwierigkeiten des ursächlichen Zusammenhangs von Bildung, Armut und kriminellem Verhalten' [Without Money and Education More Likely to be Criminal? On the Problem of a Causal Relationship between Education, Poverty and Criminal Behavior], *Deutsche Polizei* 55, issue 2, pp. 6-11.
- Peugh, James L., and Craig K. Enders 2004: 'Missing Data in Educational Research. A Review of Reporting Practices and Suggestions for Improvement,' *Review of Educational Research* 74, pp. 525-556.
- Porterfield, Austin L. 1943: 'Delinquency and Its Outcome in Court and College,' *American Journal of Sociology* 49, pp. 199-208.
- Quetelet, Adolphe 1831: *Research on the Propensity for Crime at Different Ages*, published again in 1984, Cincinnati: Anderson.
- Raab, Gillian, Susan Purdon, and Kathy Buckner 2006: *Practical Exemplars on the Analysis of Surveys*, <http://www2.napier.ac.uk/depts/fhls/peas/index.htm> (last accessed Oct 05, 2007).
- Robins, Lee Nelken, and Kathryn Strother Ratcliffe 1980: 'The Long-Term Outcomes of Truancy,' in: Lionel Hersov and Ian Berg (eds.): *Out of School. Modern Perspectives in Truancy and School Refusal*, Chichester: John Wiley.
- Roth, Philip L. 1994: 'Missing Data. A Conceptual Review for Applied Psychologists,' *Personnel Psychology* 47, pp. 537-560.

- Roth, Philip L., Fred S. Switzer III, and Deborah M. Switzer 1999: 'Missing Data in Multiple Item Scales. A Monte Carlo Analysis of Missing Data Techniques,' *Organizational Research Methods* 2, pp. 211-232.
- Royston, Patrick 2004: 'Multiple Imputation of Missing Values,' *The Stata Journal* 4, pp. 227-241.
- Royston, Patrick 2005a: 'Multiple Imputation of Missing Values. Update,' *The Stata Journal* 5, pp. 188-201.
- Royston, Patrick 2005b: 'Multiple Imputation of Missing Values. Update of "ice,"' *The Stata Journal* 5, pp. 527-536.
- Rubin, Donald B. 1976: 'Inference and Missing Data,' *Biometrika* 63, pp. 581-592.
- Rubin, Donald B. 1987: *Multiple Regression for Nonresponse in Surveys*, New York: Wiley.
- Schaeffer, Nora Cate, and Stanley Presser 2003: 'The Science of Asking Questions,' *Annual Review of Sociology* 29, pp. 65-88.
- Schäfer, Herbert 1999: 'Das gesteuerte Dunkelfeld. Kritische Anmerkungen zu willkürlichen Beeinflussungen statistischer Dunkelfelder' [The Affected Dark Figure. Critical Remarks on Arbitrary Interferences with Statistical Dark Figures], *Kriminalistik* 53, pp. 805-810.
- Schafer, Joseph L. 1997: *Analysis of Incomplete Multivariate Data*, London etc.: Chapman & Hall.
- Schafer, Joseph L. 1999: *Norm 2.03*, available at <http://www.stat.psu.edu/~jls/misoftwa.html> (last accessed Aug 20, 2007).
- Schafer, Joseph L., and John W. Graham 2002: 'Missing Data. Our View of the State of the Art,' *Psychological Methods* 7, pp. 147-177.
- Schafer, Joseph L., and Maren K. Olsen 1998: 'Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective,' *Multivariate Behavioral Research* 33, pp. 545-571.
- Schafer, Joseph L., and Nathaniel Schenker 2000: 'Inference with Imputed Conditional Means,' *Journal of the American Statistical Association* 95, pp. 144-154.
- Schnell, Rainer 1997: *Nonresponse in Bevölkerungsumfragen. Ausmaß, Entwicklung und Ursachen* [Nonresponse in Social Surveys. Extent, Development, and Causes], Opladen: Leske und Budrich.
- Seidman, David, and Michael Couzens 1974: 'Getting the Crime Rate Down. Political Pressure and Crime Reporting,' *Law and Society Review* 8, pp. 457-494.
- Short, James F., and F. Ivan Nye 1957: 'Reported Behavior as a Criterion of Deviant Behavior,' *Social Problems* 5, pp. 207-213.
- Short, James F., and F. Ivan Nye 1958: 'Extent of Unrecorded Juvenile Delinquency. Tentative Findings,' *Journal of Criminal Law, Criminology, and Police Science* 49, pp. 296-302.
- Siegmunt, Olga 2005: *Jugendkriminalität in Rußland und der BRD. Ergebnisse einer Dunkelfeldstudie* [Youth Crime in Russia and Germany. Results of a Crime Survey], unpublished diploma thesis, University of Hamburg.

- Sijtsma, Klaas, and L. Andries Van der Ark 2003: 'Investigation and Treatment of Missing Item Scores in Test and Questionnaire Data,' *Multivariate Behavioral Research* 38, pp. 505-528.
- Simpson, Sally S., and Denise C. Herz 1999: 'Gender, Crime, and Criminal Justice', in: Janet Saltzman Chafetz (ed.): *Handbook of Sociology of Gender*, New York: Kluwer/Plenum, pp. 537-562.
- Skogan, Wesley G. 1976: 'Citizen Reporting of Crime. Some National Panel Data,' *Criminology* 13, pp. 535-549.
- Steffensmeier, Darrell, Hua Zhong, Jeff Ackerman, Jennifer Schwartz, and Suzanne Agha 2006: 'Gender Gap Trends for Violent Crimes, 1980 to 2003. A UCR-NCVS Comparison,' *Feminist Criminology* 1, pp. 72-98.
- Stoop, Ineke A. L. 2004: 'Surveying Nonrespondents,' *Field Methods* 16, pp. 23-54.
- Sutherland, Edwin H. 1940: 'White Collar Criminality,' *American Sociological Review* 5, pp. 1-12.
- Thornberry, Terence P., and Marvin D. Krohn 2000: 'The Self-Report Method for Measuring Delinquency and Crime,' in: Janet Reno, Daniel Marcus, Mary Lou Leary, and Julie E. Samuels (eds.): *Criminal Justice 2000* (Vol. 4: Measurement and Analysis of Crime and Justice), Washington, DC: National Institute of Justice, pp. 33-83.
- Thornberry, Terence P., and Marvin D. Krohn 2003: 'Comparison of Self-Report and Official Data for Measuring Crime,' in: John V. Pepper and Carol V. Petrie (eds.): *Measurement Problems in Criminal Justice Research*, Washington, DC: National Academies Press, pp. 43-94.
- Tittle, Charles R., Wayne J. Vilemez, and Douglas A. Smith 1978: 'The Myth of Social Class and Criminality. An Empirical Assessment of the Empirical Evidence,' *American Sociological Review* 45, pp. 643-656.
- Tourangeau, Roger 2004: 'Survey Research and Societal Change,' *Annual Review of Psychology* 55, pp. 775-801.
- Tourangeau, Roger, and Madeline E. McNeeley 2003: 'Measuring Crime and Crime Victimization. Methodological Issues,' in: John V. Pepper and Carol V. Petrie (eds.): *Measurement Problems in Criminal Justice Research*, Washington, DC: National Academies Press, pp. 10-42.
- Tourangeau, Roger, and Tom W. Smith 1996: 'Asking Sensitive Questions. The Impact of Data Collection Mode, Question Format, and Question Context,' *Public Opinion Quarterly* 60, pp. 275-304.
- Tourangeau, Roger, Lance J. Rips, and Kenneth Rasinski 2000: *The Psychology of Survey Response*, Cambridge: Cambridge University Press.
- Vach, Werner 1994: *Logistic Regression with Missing Values in the Covariates*, New York etc.: Springer.
- Van Buuren, Stef 2005: *Multiple Imputation Online*, www.multiple-imputation.com (last accessed: Aug 20, 2007).
- Van Buuren, Stef, Hendriek C. Boshuizen, and Dick L. Knook 1999: 'Multiple Imputation of Missing Blood Pressure Covariates in Survival Analysis,' *Statistics in Medicine* 18, pp. 681-694.

- Van Buuren, Stef, and Karin Oudshoorn 2000: *Multivariate Imputation by Chained Equations, MICE v1.0 User's manual*, TNO report PG/VGZ/00.038, Leiden: TNO.
- Van Kesteren, John, Pat Mayhew, and Paul Nieuwbeerta 2000: *Criminal Victimization in Seventeen Industrialized Countries. Key Findings from the 2000 International Crime Victims Survey*, The Hague: Ministry of Justice.
- Von Hippel, Paul T. 2004: 'Biases in SPSS 12.0 Missing Value Analysis,' *The American Statistician* 58, pp. 160-164.
- Von Hippel, Paul T. 2007: 'Regression with Missing Y's: An Improved Strategy for Analyzing Multiply Imputed Data,' *Sociological Methodology* 37, forthcoming.
- Watkins, Adam M., and Chris Melde 2007: 'The Effect of Self-Control on Unit and Item Nonresponse in an Adolescent Sample,' *Journal of Research in Crime and Delinquency* 44, pp. 267-294.
- West, Stephen G. 2001: 'New Approaches to Missing Data in Psychological Research,' *Psychological Methods* 6, pp. 315-316.
- Wetzels, Peter, Dirk Enzmann, Eberhard Mecklenburg and Christian Pfeiffer 2001: *Jugend und Gewalt. Eine repräsentative Dunkelfeldanalyse in München und acht anderen deutschen Städten* [Youth and Violence. A Representative Crime Survey in Munich and Eight Other German Cities], Baden-Baden: Nomos.
- Wilkinson, Leland, and the APA Task Force on Statistical Inference 1999: 'Statistical Methods in Psychology Journals. Guidelines and Explanations,' *American Psychologist* 54, pp. 594-604.
- Wilmers, Nicola, Dirk Enzmann, Dagmar Schaefer, Karin Herbers, Werner Greve, and Peter Wetzels 2002: *Jugendliche in Deutschland zur Jahrtausendwende. Gefährlich oder gefährdet? Ergebnisse wiederholter, repräsentativer Dunkelfelduntersuchungen zu Gewalt und Kriminalität im Leben junger Menschen 1998-2000* [Juveniles in Germany at the Turn of the Millenium. Dangerous or Endangered? Results of Repeated Representative Surveys on Violence and Crime in the Lives of Young People 1998-2000], Baden-Baden: Nomos.
- Wirtz, Markus 2004: 'Über das Problem fehlender Werte. Wie der Einfluß fehlender Informationen auf Analyseergebnisse entdeckt und reduziert werden kann' [On the Problem of Missing Data. How to Identify and Reduce the Impact of Missing Data on Findings of Data Analysis], *Rehabilitation* 43, pp. 109-115.
- Yuan, Ke-Hai, and Peter M. Bentler 2000: 'Three Likelihood-Based Methods for Mean and Covariance Structure Analysis with Nonnormal Missing Data,' *Sociological Methodology* 30, pp. 165-200.

Erklärung gemäß § 14 (8) Fakultätsmasterordnung

Ich versichere hiermit, daß ich die vorliegende Arbeit selbständig verfaßt und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Wörtlich oder dem Sinn nach aus anderen Werken entnommene Stellen habe ich unter Angabe der Quellen kenntlich gemacht. Die Arbeit habe ich vorher nicht in einem anderen Prüfungsverfahren eingereicht. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium.

Hamburg, den 15. Oktober 2007