



The development of delinquency during adolescence: a comparison of missing data techniques revisited

Kristian Kleinke¹ · Jost Reinecke² · Cornelia Weins³

Published online: 17 August 2020
© The Author(s) 2020, corrected publication 2021

Abstract

Conclusions about the individual development of delinquent behaviours during the life-course are often made by repeatedly interviewing the same respondents (i.e. panel data). Missing data, especially unit nonresponse and panel attrition are often a problem for the analysis of panel data, as they pose a threat to the validity of statistical inferences. Multiple imputation (MI) is a standard state-of-the-art technique to address these problems. However, until very recently, MI methods to impute highly skewed, zero-inflated and repeatedly measured count data such as the count of delinquent behaviours per year, were not available. Solutions that were often applied included data transformations and rounding, so that available MI methods, usually based on the multivariate normal model, could be applied. This approach has also been used by Reinecke and Weins (Qual Quant 47(6):3319–3334, 2013), who analysed delinquency data from an adolescents’ four-wave panel. Recent missing data research, however, suggests that these “normalizing” practices could be problematic and that imputation models with implausible distributional assumptions should generally be avoided when the empirical data depart too heavily from these assumptions. In the present paper, we re-analyse the data from Reinecke and Weins (2013) using MI models, where parametric assumptions of the imputation model are compatible to the subsequent analysis model (a growth curve model for zero-inflated count data). Results show that the chance of reporting zero punishable offences decreases over time, with a turning point at around the age of 15. Likewise, the versatility of delinquent behaviours increases early on in adolescence, and decreases later (reflecting the typical age-crime curve). Boys and students from the bottom-level branch of the German educational system exhibit a higher versatility in delinquent activity. A comparison of present results with the original ones from Reinecke and Weins (2013) corroborates recommendations e.g. by Yu et al. (Stat Methods Med Res 16(3):243–258, 2007) to opt for missing data methods with fitting distributional assumptions.

Keywords Missing data · Multiple imputation · Growth curve models · Development of delinquency · Age–crime relationship

✉ Kristian Kleinke
kristian.kleinke@uni-siegen.de

Extended author information available on the last page of the article

1 Introduction and overview

Several long-term criminological studies, such as the Cambridge Study (Farrington and West 1990), the Philadelphia Study (Tracy et al. 1990), the Rochester Study (Thornberry et al. 2003), or the CrimoC study [i.e. Crime in the modern City, e.g. Reinecke and Weins (2013); Boers et al. (2010); Seddig and Reinecke (2017)] examine the development of delinquent behaviours over time [see also Sampson and Laub (2005)]. One of the major problems affecting the analysis of longitudinal or panel data is *wave nonresponse* and *panel attrition*, meaning that some of the respondents could not or would not take part in one or more panel waves (i.e. wave nonresponse), or in all subsequent panel waves (i.e. panel attrition). In their analysis of four waves of the CrimoC data regarding the age-crime-relationship, Reinecke and Weins (2013) compared model results based on three missing data techniques that are often used to address the missing data problem: Listwise deletion (LD)—an ad hoc method, which simply excludes all cases with missing values from the analysis, and two more sophisticated methods: full information maximum likelihood estimation (FIML) and multiple imputation (MI) under the joint multivariate normal model (Schafer 1997; Schafer and Graham 2002). While LD makes the often highly implausible assumption that missingness in a certain variable is a completely random process, and does not depend on observed (or even unobserved) information in other variables, both MI and FIML explicitly allow that other variables can be related to the chance of *not* observing a certain variable. Both FIML and MI can make use of this available information, and the inter-relationships in the data (more precisely, the variables in the model), to estimate missing information and to estimate substantive model parameters.

Regarding their four-wave panel, Reinecke and Weins (2013) found that the chance of answering questions about delinquent behaviours depended on gender and school type, i.e. girls were more likely to answer these questions than boys, and also students from “Gymnasium”, the top-level branch of the German secondary educational system were more likely to complete the questionnaires in comparison to the other school types. Results based on LD can therefore be expected to be biased. For a detailed missing data analysis of the data used in Reinecke and Weins (2013), see also Kleinke et al. (2020), Chap. 5.

Furthermore, substantive results based on MI and FIML both showed the typical age-crime-curve [e.g. Moffitt (1993)], meaning that delinquency—on average—first increases over time, reaching a maximum at around the age of 15, then decreasing later on in adolescence. Additionally, the authors looked into effects of gender and school type on the age-crime relationship: Girls on average had a lower versatility of delinquent behaviours than boys, and students attending the intermediate branch (in German: Realschule) or top level branch (in German: Gymnasium) exhibited a lower versatility in criminal activity in comparison to students attending the lowest branch (in German: Hauptschule).¹ In their comparison of missing data methods, results of both FIML and MI furthermore showed a higher level of delinquent behaviours during adolescence, and a more pronounced curvilinear association between age and crime in comparison to LD. In addition, FIML and MI estimates showed a higher prevalence of delinquent behaviours of boys and of adolescents

¹ The German secondary educational system comprises three different branches, the top-level branch being *Gymnasium* with either 8 or 9 years of education (depending on the school and the federal state), which qualifies students for tertiary education at Universities or Universities of Applied Sciences. The intermediate branch is *Realschule* with 6 years of attendance, the bottom-level branch *Hauptschule* with 5 or 6 years of school attendance. The latter two branches prepare students for vocational training.

attending Hauptschule. Since it was also these groups who had a higher chance of *not* answering the questions about delinquent behaviours, estimated higher levels by MI and FIML can be deemed plausible. LD estimates can be assumed to be biased downwards.

Although general patterns of results were similar, when comparing FIML and MI estimates, parameter estimates nevertheless differed to some extent. The most likely explanation for these findings are differences in model assumptions: The substantive model was a growth curve model assuming a zero-inflated Poisson process. The imputation model, however, assumed a multivariate normal distribution—a highly implausible model assumption for a zero-inflated count variable. Since MI methods with fitting distributional assumptions were not available at that time, Reinecke and Weins (2013) adapted a procedure originally proposed by Schafer and Olsen (1998). During the imputation stage, they split the repeatedly measured zero-inflated count variables into two separate variables respectively, to deal with the high skewness in the variable, and to make the normality assumption (somewhat) more plausible. These were (a) a binary zero versus non-zero indicator, and (b) a zero-truncated count variable, indicating for the non-zero cases, which non-zero count this participant has. Then they imputed both variables jointly under the multivariate normal model. After the imputations were created, they re-integrated the two separate variables into one single variable: The indicator variables were rounded to 0 and 1, and 1's were replaced by the respective count (which was also rounded to the nearest integer value). For an in-depth discussion of this method and for further details, see Kleinke et al. (2020, Chap. 5). Note that data transformations and rounding could produce biased statistical inferences (Horton et al. 2003; Kleinke et al. 2020). Note furthermore, that the applied procedure could also not perfectly “normalize” the data. Scores were still skewed (but less extremely than before). As for example Yu et al. (2007) or Kleinke (2017) have shown, normal model MI could yield biased inferences, when empirical data depart from normality. We therefore can expect at least some bias in MI estimates reported in Reinecke and Weins (2013) due to violated distributional assumptions of the imputation model. Note again that at the time of writing, imputation techniques for zero-inflated count data were not yet generally available. Yu et al. (2007) recommend that using “MI methods with inappropriate distributional assumptions should be avoided when the data depart considerably from these assumptions” (p. 255).

Recently, Kleinke and Reinecke (2019) have proposed MI methods for zero-inflated clustered count data (based on zero-inflated Poisson or negative binomial models), and tested these methods in systematic Monte Carlo simulations [see also Kleinke et al. (2020), Chap. 6]. Additionally, also software for fitting substantive models nowadays provides greater modelling flexibility. Mariotti and Reinecke (2010) have shown that a zero-inflated negative binomial (NB) model usually has a better fit to the CrimoC data than the Poisson model, reported in Reinecke and Weins (2013). In the mean time personal computers have become fast enough to estimate rather complex (MI) models (e.g., zero-inflated negative binomial growth curve models with assumed interindividual variation of the growth factors).

Since both missing data methods with fitting distributional assumptions that are compatible to the subsequent analysis model, are now available, and the analysis of more complex substantive models is now computationally feasible, the aim of this paper is to re-analyse the data from Reinecke and Weins (2013), and discuss substantive model results in the light of new methods and models. In the present paper, we fit growth curve models to the CrimoC data assuming both a zero-inflated Poisson or NB process respectively and that allow starting levels of delinquency to vary across individuals. A congenial imputation model is adopted respectively.

The paper is structured as follows: Sect. 2 gives a brief introduction to missing data, MI and FIML estimation. The rationale of the imputation methods adopted in this article is introduced in Sect. 2.3. In Sect. 3 we provide background information regarding the CrimoC data, and about our substantive model. In Sect. 4 we discuss our updated model results, i.e. when imputation model and analysis model are congenial. We end with a discussion of results, limitations of the current analyses, and outline fruitful avenues for future research (Sect. 5).

2 Missing data, maximum likelihood estimation, and multiple imputation

Rubin (1976, 1987) distinguishes three missing data “mechanisms”, classifying values as being missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR). Values are MCAR, if the probability for the observed pattern of missing and not missing values in the dataset depends neither on variables whose values are observed nor on variables whose values are not observed. Missing data are called MAR, if this probability depends on observed information alone, but not in addition on unobserved information. MAR is sometimes also referred to as *conditionally random* missingness, which means that after controlling for all observed information in the data, the mechanism is in fact random. Finally, missing values are MNAR, if the probability for the observed pattern of missing and not missing values depends on unobserved information, even after conditioning on observed values.

While complete case analysis usually requires missing data to be MCAR, to produce unbiased statistical inferences, maximum likelihood methods and MI allow missing data to be MAR [e.g. Schafer and Graham (2002)]. If this assumption is not met, bias is to be expected. The magnitude of this bias depends on the amount and pattern of missing data. In this case special MNAR models (under strong and usually untestable assumptions) could be fitted [cf. Kleinke et al. (2020), Chap. 3].

2.1 Full information maximum likelihood (FIML)

One of the standard methods to address the missing data problem both in cross-sectional and longitudinal or panel data is maximum likelihood estimation (Schafer and Graham 2002), which is also referred to as full information (FIML) or direct maximum likelihood estimation. For an easily understandable introduction to maximum likelihood (ML) estimation both with and without missing data, see for example Enders (2010).

The basic estimation principle of ML is that the method repeatedly tries out different combinations of parameter values until it identifies a particular constellation of estimates that produces the best fit to the data (or in mathematical terms, the highest log-likelihood): The starting point for any ML analysis is to assume a parametric distribution for the population data (e.g. a normal distribution, or a Poisson distribution) and to specify a statistical model accordingly. Inserting the observed data of each individual and a set of parameter values into the corresponding density function returns a likelihood value for each case that quantifies the relative probability of drawing these data from the specified population. The sample log-likelihood is the sum of the individual log-likelihood values. It quantifies the probability of drawing the observed data from the specified population. Through an

iterative process the one particular constellation of parameters is then sought that maximizes the sample log-likelihood. If the data file contains missing values, individual cases contribute to their individual log-likelihood only to the extent to which they have observed data. The number of data points per case thus varies depending on the respective pattern of observed and unobserved values. Apart from that the basic estimation principle stays the same. FIML estimation can be enumerated among the model-based missing data methods. Estimates are usually unbiased, when the parametric model is correctly specified, when parametric assumptions are at least approximately met, and when the MAR assumption holds, i.e. missingness only depends on observed information in the model variables (Kleinke et al. 2020; Schafer and Graham 2002). Should missingness—however—depend on other variables in the data file (that are not part of the substantive model), bias is to be expected. Inclusion of extra variables only for the estimation of missing information, while not considering these variables in the substantive model, is straightforward in MI, but not in FIML estimation (Collins et al. 2001). This is possible, because MI is a two-stage process that clearly separates the imputation and the analysis stage, allowing different models in each step.

2.2 Multiple imputation (MI)

The basic idea of multiple imputation is to replace each missing value not only once, but multiple times based on some statistical model. The resulting m complete data files differ only in the formerly unobserved part. Each replacement of an incomplete value is equally plausible under the specified imputation model. In a second step, the m datasets are analysed by any standard technique or software for complete datasets. The m sets of obtained model results (parameter estimates and corresponding standard errors) are combined into an overall set of results, using simple combination rules (Rubin 1987; Barnard and Rubin 1999). The combined parameter estimate is simply the mean of the m estimates, while the standard error combines a within imputation and between imputation variance component, reflecting the extra estimation uncertainty due to missing data. There are two main theoretical frameworks to generate the imputations – often referred to as joint modelling (JM) and conditional modelling (CM). JM approaches require the specification of a joint distribution for *all* variables in the model. Schafer's (2016) `norm2` package for example, which is based on algorithms in Schafer (1997) assumes a multivariate normal distribution. In practice, it is often hard to find a joint distribution that reflects the individual distributions of the variables in the dataset well.

The second framework, CM [e.g. Raghunathan et al. (2001) or van Buuren (2018)] only hypothetically assumes that a common distribution exists. In CM, users specify an appropriate univariate model separately for each incompletely observed variable. The software then imputes each incomplete variable in turn (via an iterative process) according to the respective univariate model. This means, for an incomplete clustered count variable, a two-level Poisson or negative binomial model [cf. Hilbe (2011)] could be specified, while missing data in a binary variable such as gender could be filled in based on a logistic regression imputation model. It needs to be noted that while CM provides great modelling flexibility, from a theoretical perspective, the drawback is that it is not always clear, if the (implicitly) underlying joint probability distribution exists, and what the consequences are, if this is not the case. In practice, however, many simulations and applications have shown that the method works reasonably well (e.g., van Buuren et al. 2006). For a more

in-depth discussion of MI and advantages and disadvantages of the respective frameworks, see Kleinke et al. (2020).

2.3 Multiple imputation of incomplete count data

Both missing data theory (Rubin 1987) and missing data research suggest that the imputation model should be compatible to the subsequent analysis model. This means that relationships in the data that are of interest to the data analyst need to be reflected in the imputation model. But also distributional assumptions of the imputation model should be compatible to the subsequent analysis model, which ideally has a very good fit to the empirical data at hand. For example, Yu et al. (2007), Kleinke (2017) and Kleinke and Reinecke (2013) have demonstrated that adopting imputation methods and models with ill-fitting distributional assumptions can bias statistical inferences quite noticeably. This means that count data (such as the number of delinquent behaviours per year), which are often highly skewed, should usually *not* be imputed by methods relying on the normal model, or by proxy-methods, e.g. methods for categorical data such as polytomous regression [unless the number of categories is sufficiently small; see also the Monte Carlo simulation and discussions in Kleinke and Reinecke (2019), Chap. 3].

Kleinke and Reinecke (2019) recently have provided additional imputation functions for the conditional modelling MI software `mice` (van Buuren and Groothuis-Oudshoorn 2011), which are based on the most commonly used count data models [for a very comprehensive introduction to count data modelling, see Hilbe (2011)]. Package `countimp`² supports imputation of incomplete count data under different model assumptions: the classical Poisson model, the negative binomial model (NB), the zero-inflated Poisson model (ZIP), the zero-inflated negative binomial model (ZINB), and the Poisson or negative binomial hurdle model. In case of clustered or panel data, these models can also be fitted as two-level models. For classical Poisson regression, the probability of observing a certain count y_i given covariates x_i is

$$P(Y_i = y_i | x_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad y_i = 0, 1, 2, \dots \quad (1)$$

Poisson generalized linear models use a log link, i.e. they model the log of the expected count as a function of predictors x_i . The conditional mean is $E(y_i | x_i) = \lambda_i = e^{x_i \beta}$, the variance is assumed to be equal to the mean (assumption of equidispersion). Often, however, empirical data are overdispersed, which means that the variance is larger than expected by the Poisson model. In this case, the negative binomial model estimates an extra parameter that represents the amount of overdispersion in the data. There are several parametrizations of the NB model, a usual one being the notation as a gamma-Poisson mixture model, which estimates a shape parameter θ that represents the random variation in y_i that is not fully accounted for by x_i [for details, see Hilbe (2011)]. Conditionally on x_i , y_i is distributed as

² `countimp` is available from <https://www.github.com/kkleinke/countimp> and provides additional imputation functions for package `mice` in R.

$$P(Y_i = y_i | x_i) = \frac{\theta^\theta \mu_i^{y_i} \Gamma(\theta + y_i)}{\Gamma(y_i + 1) \Gamma(\theta) (\mu_i + \theta)^{\theta + y_i}}, \quad (2)$$

with conditional mean μ_i and conditional variance $\mu_i(1 + \alpha\mu_i)$, with dispersion parameter $\alpha = 1/\theta$. Note that the NB distribution and Poisson distribution are nested and the NB distribution converges to a Poisson distribution, when $\alpha \rightarrow 0$.

There are several ways to determine, which model has the better fit to the data [for an extensive discussion, see Hilbe (2011)]. In the present paper, we report test statistics for the null hypothesis, that the NB overdispersion parameter is zero. Additionally, Poisson and NB models can be compared by means of model fit statistics like the Akaike information criterion (AIC) or the Bayesian information criterion (BIC).

When empirical data contain more zero counts than would be expected by either the Poisson or the NB model, the data are called zero-inflated. In this case a zero-inflation model (Lambert 1992) can be fitted. Zero-inflation models are mixture models consisting of two model components—the zero model, usually a binomial model, which determines if an observational unit belongs to the count process or to the so-called “certain zero” process, and the count model, a Poisson or NB model, that determines which count (zero or non-zero) an observational unit has. Zero-inflation models allow two sources of zeros: They can stem from both the zero part or from the count part and can also have different sets of predictors. For example, we assume a group of students, who would never commit any punishable offences, perhaps due to their high moral standards. In addition, there might be another group of students, who *might* have committed an offence, but who nevertheless reported zero delinquent behaviours, because they were closely monitored by their parents after having committed an offence the year before. In contrast to zero-inflation models hurdle models assume only one source of zeros. They fit a zero-truncated Poisson or NB model to the non-zero observations. Since we do not assume offenders and non-offenders to be distinct categories that never overlap [see Seddig and Reinecke (2017)], we fit zero-inflation models rather than hurdle models to the empirical data.

In case of clustered (panel) data, the non-independence of residuals due to the clustered structure of the data needs to be taken into account. Linear mixed effects models [e.g. Bryk and Raudenbush (1992)], or equivalently, latent growth curve models [e.g. Bollen and Curran (2006)] are the method of choice when intraindividual as well as interindividual differences in the development across time are of interest. These models provide information about average starting levels and average levels of change across time (so-called “fixed effects”), but also allow predictors to test for differences in starting levels and patterns of change across time. Variances and covariances of individual starting levels and levels of change are also called “random effects”.

A detailed description regarding how the respective imputation functions generate the multiple imputations is given in Kleinke and Reinecke (2019). Basically, multiple imputation based on Rubin’s theory (1987) includes a stochastic component, that introduces an adequate amount of variability in the m imputations that reflects the estimation uncertainty due to missing data. Within the conditional modelling framework, the two solutions to introduce variability in the imputations are Bayesian regression [see for example Rubin (1987), pp. 169–170], and bootstrap regression: Bayesian regression first fits the specified model to the complete part of the data and obtains the posterior mean $\hat{\beta}$ and posterior variance $V(\hat{\beta})$ of the model parameters β . Secondly, new parameters are simulated from $\mathcal{N}(\hat{\beta}, V(\hat{\beta}))$. These new parameters are finally used to make predictions for the incomplete part of the data and to obtain the imputations. The bootstrap approach first draws

Table 1 Descriptive statistics of the delinquency index

Index	Mean	SD	Min	Max	Missing
2000	0.60	1.36	0	12	689
2001	1.06	1.99	0	15	397
2002	1.29	2.23	0	16	343
2003	1.00	1.76	0	16	579

Table 2 Frequency table of delinquent activity at wave 1 by gender and school type

Gender	Female			Male			Σ
	HA	RE	GY	HA	RE	GY	
School type							
Not delinquent	88	126	215	94	157	310	990
≥ 1 Delinquent behaviours	61	74	95	36	51	68	385
Missing	123	138	122	104	120	82	689

HA, RE, GY are the school type indicators, reflecting the bottom-level, intermediate level and top-level branch (Hauptschule, Realschule, Gymnasium)

a bootstrap sample from the observed cases and fits the specified model to this bootstrap sample. Secondly, the obtained model parameters are used to make predictions for the incomplete part of the data and to obtain the imputations. These steps are repeated $m - 1$ times. Imputation functions from package `countimp` are available as Bayesian regression and bootstrap regression variants (Kleinke and Reinecke 2019). Bootstrap regression is a suitable alternative whenever researchers are not happy to make the assumption that parameters can be simulated from $\mathcal{N}(\hat{\beta}, V(\hat{\beta}))$. In the present analyses, we use the Bayesian variants.

3 Data, imputation and substantive model

3.1 Data and measures

We reanalyse empirical panel data³ from four waves of the longitudinal research project *Crime in the modern City* (CrimoC, <https://www.crimoc.org>) The main focus of the study lies on the emergence and development of deviant and delinquent behaviours in adolescence and the social control surrounding it, i.e. both formal control—meaning the police and the judiciary, and informal control—referring to school and family (Boers et al. 2010; Seddig and Reinecke 2017). The target population of this analysis were students attending public schools in the town of Münster (about 300,000 inhabitants) in the western part of Germany. The survey of 7th graders (which this analysis is based on) included all 7th grade classes, and was first conducted in the year 2000. This cohort was interviewed annually until students reached the 10th grade in 2003. Data were collected by means of

³ The data set is included in package `countimp` and available from <https://github.com/kkleinke/countimp>.

self-administered classroom interviews and response rates ranged between 86 and 88% [see Table 1 in Reinecke and Weins (2013)]. Detailed information about missing data may be found in Reinecke and Weins (2013), Tables 2, 3 and 4. The sample for the present analysis consists of 1022 girls and 1042 boys attending all types of schools, who took part in at least two out of the four waves. 24.5% of students in the sample attended the low-level branch of the German secondary educational system (Hauptschule), 32.3% the intermediate branch (Realschule), and 43.2% the top-level branch (Gymnasium).

The dependent variable of interest in the present analyses is the count of delinquent behaviours. At each panel wave, students answered 16 questions about various offences like for example graffiti spraying, damaging property, robbery, purse snatching, assault with and without a weapon, drug abuse or drug trafficking [for details, see Table 8 in Reinecke and Weins (2013)]. Students indicated for each offence, if they committed this delinquent behaviour in the 12 months prior to the interview (1 = yes; 0 = no). The individual delinquency index was then computed as the sum of these 16 binary items, reflecting the versatility of delinquent activity. The index could range between 0 and 16. Higher index values indicate a more versatile criminal activity. Descriptive statistics of the delinquency index are given in Table 1 and confirm the curvilinear development of crime during the four years of data collection: Mean delinquency rates increase from 2000 to 2002, and decrease in 2003. Table 2 further differentiates delinquent activity as well as response patterns at the first panel wave by gender and school type, see also Reinecke and Weins (2013, Table 3) for a more in-depth discussion. There is a small relationship between gender and reporting delinquency at wave 1. Boys more often exhibit delinquent behaviours than girls, however, boys are somewhat also more likely to refuse to answer questions about delinquent activity in comparison to girls (Cramer's $V = .14$, $\chi^2(2) = 40.63$, $p < .001$). Furthermore, there is a small relationship between school type and reporting delinquency at the first panel wave—with students from the bottom level and from the intermediate branch being somewhat less likely to answer questions regarding delinquent activity in comparison to students from the top-level branch (Cramer's $V = .15$, $\chi^2(4) = 96.76$, $p < .001$).

3.2 Imputation and substantive model

Descriptive statistics in Tables 1 and 2 indicate a heavily skewed distribution of the delinquency indexes with a large number of zero counts at each panel wave. Most students reported that they did not commit any punishable offence during the previous year. When analysing count data, the skewed and discrete nature of the data (i.e. non-negative integer values) needs to be taken into account [cf. Erdman et al. (2008); Hilbe (2011)]. Mariotti and Reinecke (2010) fitted different growth curve and growth mixture models to the Crimoc data and found that Poisson or negative binomial (NB) models that cater for the large amount of zero counts have a good fit to the data.

In the present analysis, we re-examine the effects of gender and school type on the age-crime-relationship using newly available missing data methods. Based on results by Mariotti and Reinecke (2010) we assume that overdispersion in the data, as well as zero-inflation needs to be adequately addressed and that a ZINB model will have a better fit to the data than a ZIP model. We will present results of the age-crime-relationship based on assumed ZIP and ZINB processes.

The substantive model, which is also the model we use for data imputation, is illustrated in Fig. 1.

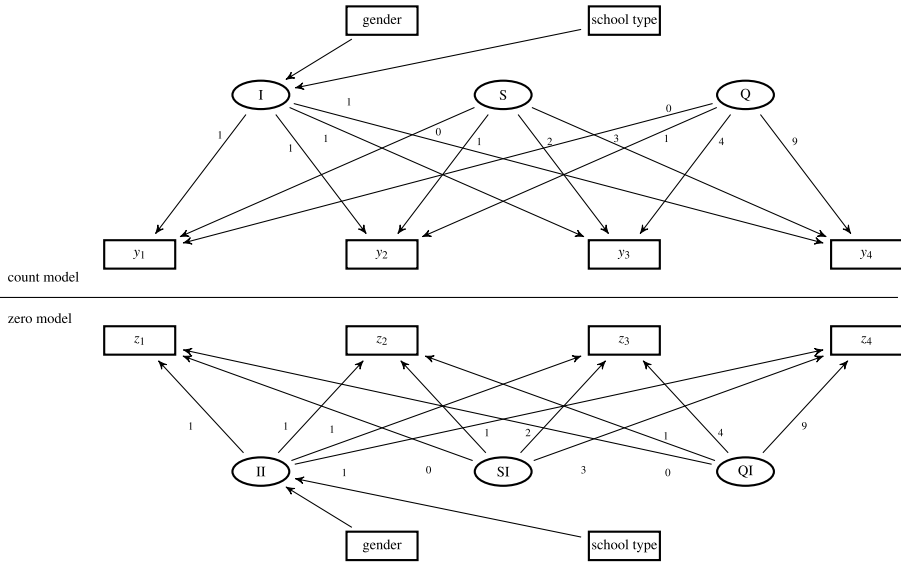


Fig. 1 Conditional zero-inflation growth curve model. *Note:* I, S, and Q are the growth factors representing the intercept (i.e. starting level) and linear and quadratic slopes respectively. II, SI, and QI are the counterparts in the inflation part of the model. Baseline levels of delinquency as well as the baseline chance of having a structural zero are predicted by gender and school type

Recall that zero-inflation models consist of two model components. The top-part of Fig. 1 displays the count model, the bottom part gives the zero-inflation component. Observed delinquency scores y_1 to y_4 (where subscripts 1 to 4 denote the respective panel wave) define latent growth factors I, S, and Q representing intercept, linear and quadratic slope, i.e. average starting levels of delinquency and the curvilinear development of delinquency across time respectively. Baseline levels of delinquency are regressed on gender and school type.

Analogously, in the inflation part, z_1 to z_4 are zero versus non-zero indicators, which define latent growth factors II, SI, and QI representing the chance of having a structural (or “certain”) zero across time. Again, the baseline chance of having a structural zero is regressed on gender and school type. Note that for identification purposes, the mean of II is fixed to zero.

While, generally speaking, growth curve models allow to estimate both average values (i.e. means or intercepts) of the growth factors, as well as their variances and covariances, with only four panel waves, models cannot be overly complex. Too complex models relative to the available sample size and the available panel waves would result in not or only weakly identified models and severe estimation problems.

To fit a quadratic growth curve model, a minimum of four timepoints is needed. In an unconditional model, 13 parameters need to be identified: the means of the intercept, the slope, and the quadratic slope (3 parameters), their variances (3 parameters) and covariances (3 parameters), and the residual variances of the outcome variables (4 parameters). An adequate sample size both at level-1 (the repeated measurements) and level-2 (the individuals) is needed to reliably estimate these growth model parameters. In a conditional model, there are additional parameters of the respective regressions to be estimated. While in our case the number of individuals is sufficiently large, the number of repeated

observations per individual is quite small. Since the estimated variances of the slope and quadratic slope are quite small, we fix these parameters to zero to stabilize the estimation process.

In the count part, we thus estimate intercepts (or means) of the growth factors, as well as the variance of individual baseline levels of delinquency, and predict these individual differences by gender and school type. In the zero model, we only estimate the average values of the growth factors. While our substantive model from Fig. 1 is a latent growth curve model, which will be estimated by the program *Mplus*,⁴ data imputation is done by package `countimp`, which fits generalized linear mixed effects models. However, the model from Fig. 1 can easily be rewritten as a linear mixed effects Poisson or NB model, which is fully compatible to the latent growth model:

The model equation in linear mixed effects modelling notation for the count part of the imputation model is

$$\begin{aligned} \ln(y_i) &= \beta_0 + \beta_1 T_i + \beta_2 T_i^2 + e_i(\text{level 1}) \\ \beta_{0j} &= \gamma_0 + \gamma_1 \text{FEMALE} + \gamma_2 \text{GY} + \gamma_3 \text{RE} + u_{0j}(\text{level 2}) \end{aligned} \quad (3)$$

where y_i is the respective delinquency score (in long format, which means that the repeated measurements are stacked upon another and stored in a single variable), β_0 is the intercept (labeled I in Fig. 1) – the average starting level of delinquency at wave 1 across all individuals, and β_1 (S), and β_2 (Q) being linear and quadratic slopes that reflect the average curvilinear development across time, with $T_i = (0, 1, 2, 3)$ being the time period. Here, 0 denotes the first panel wave. On the person-level (level 2), we furthermore predict individual differences in starting levels by gender (FEMALE) and school type (GY and RE being the school type indicators for the top-level and medium level branches, with the bottom-level branch being the reference category). e_i is the level-1 residual, u_{0j} is the level-2-residual, with i and j being level-1 and level-2 indices. The zero model equation—a binomial generalized linear model is

$$\text{logit}(z_i) = \beta_{1z} + \beta_{2z} T_i + \beta_{3z} T_i^2 + \beta_{4z} \text{FEMALE} + \beta_{5z} \text{GY} + \beta_{6z} \text{RE} + e_{iz}. \quad (4)$$

Here z_i represents a zero versus non-zero indicator variable (in long format). $\beta_{1z} - \beta_{3z}$ are the counterparts of the coefficients of the growth factors II, SI and QI in Fig. 1, representing the average chance for having a structural zero across time.

3.3 Data imputation

For data imputation, we select a multiple imputation model that is fully compatible to the subsequent substantive model of interest. This means that we choose an equivalent model class and estimate the same parameters as in the analysis model.

For the present analyses, we create two sets of $m = 100$ imputations by package `countimp` (Kleinke and Reinecke 2019). Imputation functions for the incomplete delinquency indexes are—depending on the subsequent analysis model—either `mice.impute.2l.zip` or `mice.impute.2l.zinb` assuming either a zero-inflated Poisson or NB

⁴ This has practical reasons. The stand-alone program *Mplus* (Muthén and Muthén 2017) is very user-friendly, as it supports automated repeated data analysis of the m imputed data sets and multiple imputation inference.

Table 3 Model fit statistics

	LL_M	LL_SD	AIC_M	AIC_SD	BIC_M	BIC_SD	aBIC_M	aBIC_SD
FIML ZINB	- 7457.84		14949.69		15045.44		14991.43	
FIML ZIP	- 7485.67		14997.34		15070.56		15029.26	
MI ZINB	- 9887.34	62.19	19808.67	124.39	19904.42	124.39	19850.41	124.39
MI ZIP	- 9928.34	55.57	19882.68	111.14	19955.90	111.14	19914.60	111.14

The lower part of the table displays means (M) and standard deviations (SD) of the respective criteria across the $m = 100$ imputations

LL, log-likelihood value; AIC, Akaike information criterion; BIC, Bayesian information criterion; aBIC, adjusted Bayesian information criterion

Table 4 Parameter estimates of the model based on MI

	ZINB				ZIP			
	EST	SE	<i>p</i>	FMI	EST	SE	<i>p</i>	FMI
<i>Growth factors</i>								
Mean of I	- 0.09	0.10	0.39	0.23	0.00	0.09	0.96	0.26
Mean of S	0.51	0.07	0.00	0.25	0.45	0.06	0.00	0.36
Mean of Q	- 0.14	0.02	0.00	0.25	- 0.13	0.02	0.00	0.32
Mean of SI	- 1.00	0.36	0.01	0.31	- 1.11	0.24	0.00	0.33
Mean of QI	0.16	0.13	0.22	0.35	0.19	0.08	0.02	0.32
Variance of I	1.60	0.10	0.00	0.16	1.61	0.10	0.00	0.15
<i>Regression coefficients</i>								
I ON FEMALE	- 0.91	0.08	0.00	0.14	- 0.90	0.08	0.00	0.13
I ON RE	- 0.20	0.10	0.04	0.13	- 0.19	0.10	0.05	0.16
I ON GY	- 0.80	0.10	0.00	0.12	- 0.80	0.10	0.00	0.16
II ON FEMALE	0.03	0.23	0.89	0.31	0.11	0.17	0.54	0.27
II ON RE	0.00	0.24	0.99	0.30	0.03	0.19	0.89	0.31
II ON GY	- 0.33	0.30	0.27	0.34	- 0.22	0.22	0.32	0.32
<i>Dispersion parameters</i>								
Delinquency (t_1)	0.32	0.13	0.01	0.27				
Delinquency (t_2)	0.16	0.06	0.01	0.26				
Delinquency (t_3)	0.04	0.04	0.34	0.23				
Delinquency (t_4)	0.11	0.06	0.04	0.28				

I is the intercept, S and Q the linear and quadratic slopes in the count part of the model. II, SI, and QI are the respective counterparts in the zero-inflation part of the model. The growth factors of both parts are regressed on FEMALE (the gender indicator), RE, and GY (the school type indicators)

process. Predictors gender and school type were completely observed. We choose the parametric Bayesian regression variants, since we assume that the parametric models fit the data sufficiently well and that parameters can be simulated from $\mathcal{N}(\hat{\beta}, V(\hat{\beta}))$. Convergence of the Gibbs sampler is monitored as outlined in van Buuren and Groothuis-Oudshoorn

Table 5 Parameter estimates of the model based on FIML

	ZINB				ZIP			
	EST	SE	EST/SE	<i>p</i>	EST	SE	EST/SE	<i>p</i>
<i>Growth factors</i>								
Mean of I	- 0.26	0.10	- 2.57	0.01	- 0.16	0.09	- 1.68	0.09
Mean of S	0.54	0.08	7.13	0.00	0.45	0.06	7.94	0.00
Mean of Q	- 0.14	0.02	- 6.54	0.00	- 0.12	0.02	- 7.34	0.00
Mean of SI	- 0.91	0.85	- 1.08	0.28	- 1.48	0.40	- 3.70	0.00
Mean of QI	0.12	0.30	0.39	0.70	0.25	0.15	1.68	0.09
Variance of I	1.80	0.10	17.76	0.00	1.81	0.10	17.98	0.00
<i>Regression coefficients</i>								
I ON FEMALE	- 0.92	0.08	- 12.07	0.00	- 0.91	0.08	- 11.93	0.00
I ON RE	- 0.21	0.10	- 2.14	0.03	- 0.20	0.10	- 2.05	0.04
I ON GY	- 0.80	0.09	- 8.50	0.00	- 0.80	0.09	- 8.49	0.00
II ON FEMALE	- 0.28	0.55	- 0.52	0.60	- 0.06	0.28	- 0.22	0.83
II ON RE	- 0.19	0.61	- 0.31	0.76	0.02	0.29	0.07	0.94
II ON GY	- 1.08	0.76	- 1.43	0.15	- 0.52	0.35	- 1.48	0.14
<i>Dispersion parameters</i>								
Delinquency (t_1)	0.60	0.17	3.47	0.00				
Delinquency (t_2)	0.11	0.06	1.90	0.06				
Delinquency (t_3)	0.00	0.04	0.00	1.00				
Delinquency (t_4)	0.11	0.05	2.15	0.03				

I is the intercept, S and Q the linear and quadratic slopes in the count part of the model. II, SI, and QI are the respective counterparts in the zero-inflation part of the model. The growth factors of both parts are regressed on FEMALE (the gender indicator), RE, and GY (the school type indicators)

(2011) by means of graphical inspection of the streams across the iterations. No convergence problems could be detected.

4 Results

The model in Fig. 1 is estimated by *Mplus* Version 8, which in addition to FIML estimation supports automated repeated data analysis of the m imputed data sets and provides combined point estimates, standard errors, and test statistics for the null hypothesis that the respective parameter is zero, as well as an estimate for the fraction of missing information – a statistic that gives an idea to what extent the respective combined parameter estimate is influenced by missing data. Additionally, *Mplus* also provides combined model fit statistics for the m sets of analyses, like AIC, BIC, or the adjusted BIC [for details about these statistics, see Kleinke et al. (2020), Appendix A]. Fit measures of the ZIP and ZINB models are displayed in Table 3, parameter estimates of the model are displayed in Tables 4 (MI) and 5 (FIML). All model fit statistics suggest that the ZINB model is slightly superior to the ZIP model, regardless of the missing data method. However, differences are not very pronounced. This is also reflected in the relatively small sizes of the overdispersion parameters for waves 2 to 4, ranging between .04 and

.16 (MI estimates, see Table 4), and ranging between 0 and .11 (FIML estimates, see Table 5). The estimate for the first wave is somewhat larger (.60 for FIML, and .32 for MI). The upper parts of Tables 4 and 5 display intercepts of the growth factors for the zero and count parts of the models respectively, the variance of the intercept in the count part, as well as coefficients of gender and school type predicting differences in starting levels of delinquency. Recall that zero-inflation models are mixture models: The logit part gives the odds of not being delinquent (so-called structural zeros), the Poisson part models the number of different delinquent behaviours exhibited in the previous year. Note that for model identification purposes, the intercept in the inflation part is fixed to zero. The negative slope (SI) suggests that the chance of having a structural zero decreases over time, while the small positive quadratic slope (QI) suggests that there is a turning point, meaning that the chance of having a structural zero starts to increase again later on in adolescence. In this regard, estimates of MI and FIML are highly similar. Furthermore, the chance of belonging to the certain-zero group does not depend on gender or school type. This is in contrast to findings by Reinecke and Weins (2013), which will be discussed later.

Likewise, the positive slope in the count model (S) suggests an increase in mean delinquency rates across time, while the negative quadratic slope (Q) means that there is a turning point, and mean delinquency rates start to decrease again later on in adolescence. The growth factor estimates of the ZINB models are somewhat smaller in comparison to the ZIP models, which is most likely a result of the mildly violated equidispersion assumption of the Poisson model.

Individual differences in baseline levels of delinquency can furthermore at least partially be explained by gender and school type. Girls on average exhibit fewer delinquent behaviours and students from the intermediate and top-level branch of the German educational system also commit fewer punishable offences in comparison to students from the bottom-level branch.

Note that overall, point estimates and standard errors of FIML and MI are quite similar. The only noticeable differences are the intercept estimates in the count model. Typically, MI and FIML are expected to yield comparable results, when (as in our case) compatible models are fitted. MI is a simulation based procedure and results can be expected to differ slightly with each run. With $m \rightarrow \infty$, results by MI and FIML could be expected to be equivalent, when imputation and analysis model are correctly specified and when both models are compatible [see for example Collins et al. (2001)]. Here, we created a set of $m = 100$ imputations. To see to what extent these differences in the intercept estimates affect model based predictions, we can compare predicted counts for certain groups that are of interest to us, e.g. boys attending the low-level branch of the German educational system, who can be expected to have the highest versatility in delinquent activity. For boys from Hauptschule, the count model predictions would be

$$\exp(-.16 + .45T_i + (-.12T_i^2)), T_i = (0, 1, 2, 3) \quad (5)$$

different delinquent activities based on the FIML estimates of the ZIP model in Table 5, resulting in an average estimate of 0.85 at wave 1, 1.19 at wave 2, 1.30 at wave 3, and 1.12 at wave 4. Based on Multiple Imputation (Table 4), we would expect

$$\exp(0 + .45T_i + (-.13T_i^2)), T_i = (0, 1, 2, 3), \quad (6)$$

resulting in an average count of 1.00, 1.38, 1.46, and 1.20 respectively. Application of MI (in this instance) yields a slightly higher versatility of delinquency in comparison to the FIML estimate. Note that this applies to this set of multiple imputations. It cannot be inferred that the MI point estimates will always be higher or that either the FIML point estimates or the MI estimates are more accurate, since the true population values are unknown.

Finally, all estimated fractions of missing information are sufficiently low (ranging between .12 and .35 for the ZINB model, and between .13 and .36 for the ZIP model). Schafer (1997) notes that estimates below .4 usually denote unproblematic inferences. With higher values, quality of statistical inferences depends to a far greater extent upon the correct specification of the imputation model and upon how well assumptions of the MI model (like the untestable MAR assumption or parametric assumptions of the imputation model) are met. In this scenario, even if missing data were MNAR, bias due to a violated MAR assumption will most likely be not very pronounced.

In summary, FIML and MI lead to the same substantive results concerning the development of delinquency. Point estimates and standard errors differ only slightly.

5 Discussion

Reinecke and Weins (2013) have stressed the need to use state-of-the-art missing data methods when analysing longitudinal criminological data: Usually, when analysing self-report data in a panel design, especially when questions are about sensitive topics such as delinquency, researchers have to deal with a non-negligible percentage of missing data. In their four-wave-panel regarding the versatility of delinquent behaviours, missingness in the delinquency scores ranged between 16.6 and 33.4%, and only 813 out of the 2064 cases were complete cases after four waves. Although complete case analysis is an attractive ad hoc missing data method, as it is quick and straightforward to apply, and implemented in any software for data analysis, inferences based on complete case analysis are prone to bias, when data are *not* missing completely at random—an assumption that is often violated like in the present analysis. In this case, inferences should be based on missing data methods that allow for violations of the MCAR assumption. Schafer and Graham (2002) enumerate ML estimation and MI among the state-of-the-art methods. From an applied researcher's point of view, the problem now is to find suitable missing data methods with fitting distributional assumptions for the problem at hand. Although FIML estimation is supported by most SEM software packages nowadays, FIML estimation of ZIP and ZINB models for clustered data still is not generally available. We used *Mplus* (Muthén and Muthén 2017) to fit the models. Here, FIML estimation is even the default setting. Hence, the application of FIML analysis has become quite as straightforward as selecting case deletion.

Application of MI on the other hand typically still involves a little more effort: imputing the data, repeated data analysis, and pooling of results. However, some steps could also be automatized (like for example using the automated repeated data analysis and pooling of results feature in *Mplus*). The more important problem is to find a suitable imputation function for the problem at hand. Especially MI models for zero-inflated and clustered count data have become available only very recently. To our best knowledge, such models are currently only available as an add-on to the open source multiple imputation package *mice* in R (Kleinke and Reinecke 2019).

The question now is, when to use ML, and when to use MI? Or does it even make any difference, which method is being used? In their comparison of MI and ML, Reinecke and

Weins (2013) already obtained similar, but not identical substantive model results. Differences in their parameter estimates could be explained by differences in model assumptions between their substantive model (a growth curve ZIP model) and the imputation model (a normal model). Since ZIP imputation models were not available then, Reinecke and Weins (2013) used typical data transformations and rounding to make the assumed multivariate normal distribution of the imputation model more plausible. Recent missing data research, however, suggests (e.g. Kleinke and Reinecke 2013; Kleinke 2017) that selecting an imputation model with ill-fitting distributional assumptions could do a lot of damage (i.e. produce biased statistical inferences). If imputation methods with fitting distributional assumptions are not yet available, it might therefore be better *not* to impute and to report results based on ML estimation instead. Regarding the results in Reinecke and Weins (2013), differences between ML and MI were only rather small and one could argue that the applied transformation and rounding procedure has worked sufficiently well. Since systematic Monte Carlo simulations are to our best knowledge not available, we would not *generally* recommend the workaround solution used in Reinecke and Weins (2013) for the analysis of incomplete clustered count data, since their method additionally did not cater for the clustered structure of the data (which was also not considered by their substantive model). Drechsler (2015) for example shows that ignoring the clustered structure of the data during data imputation can lead to biased estimates especially regarding the “random” part of the model. This was not problematic for the analysis reported in Reinecke and Weins (2013), since they were interested only in the fixed part of the model, but could have become a problem for the present analysis.

In the light of newly available methods for imputing incomplete clustered count data, and faster personal computers, we reanalysed the data from Reinecke and Weins (2013), and compared MI and ML models that were fully compatible, and that considered the clustered structure of the data. Again, we found highly similar substantial model results, which corroborates findings from earlier missing data research regarding the compatibility of MI and ML methods (e.g. Collins et al. 2001): when MI and FIML estimation is based on the same (or compatible) statistical models, using the same predictors, considering the same relationships within the data, and making the same distributional assumptions, highly similar results are to be expected—given that a sufficient number of imputations is created. In the present analysis, we created sets of $m = 100$ imputations for each analysis (ZINB and ZIP). Discussions regarding how many imputations should be generated may be for example found in Bodner (2008) or von Hippel (2003).

So why not simply use ML, which is more straightforward to apply in comparison to MI? ML estimation is a feasible method if missingness depends only on variables in the model—in our case for example on gender and school type. Had additional variables been identified that determine the chance of *not* answering the questions about delinquent behaviours, that are not of interest to the data analyst, then including these variables into the model only to predict missing information without changing the meaning of the coefficients of the other model variables would not have been possible. For such scenarios, MI is the method of choice since MI clearly separates the imputation stage and the analysis stage, allowing different models in each step.

Finally, there are some differences between the substantive models of the present analysis and the analyses reported in Reinecke and Weins (2013), who estimated only the fixed part of the model (i.e. the intercepts of the growth factors). Interindividual variation in starting levels was not considered. Coefficients reported in Reinecke and Weins (2013) therefore do not disentangle effects on the within-person level and between-person-level. In the present analysis we estimate interindividual variation in starting levels of

delinquency and predict them by gender and school type. Both studies found that as delinquency increases over time, the chance of having a structural zero decreases. Likewise, as the versatility of delinquent behaviours starts to fall again later on in adolescence, the chance of having a structural zero starts to increase again. Versatility of delinquent behaviours could be predicted by gender and school type. In the present analysis, however, the chance for having a structural zero at the first panel wave no longer depend on gender and school type—as it did in Reinecke and Weins (2013)—most likely because within-person and between-person effects have been disentangled in the present analysis and are already represented in the interindividual variation of intercepts. Additionally, we estimated overdispersion in the data. Fitting a ZINB model in comparison to a ZIP model improved model fit respectively. Apart from that, both results lead to the same conclusions.

Practical implications In this paper, we have fitted a conditional latent growth curve model. Growth models, panel regression models, as well as growth mixture models are among the most frequently used methods to describe the development of a variable of interest across time. Panel regression models (univariate or multivariate linear mixed effects models) are just a special case of latent growth models and can be rewritten as a growth model. MI and FIML methods discussed in this paper can therefore be applied to these models, as well without much ado. MI and FIML are expected to yield comparable results, when model assumptions are widely met. Growth mixture models, and other model classes that involve classification, were not considered in this paper. From a theoretical perspective, it is straightforward to get FIML estimates of such models, but it is *not* straightforward to apply multiple imputation in analyses that involve classification. The intention of MI is to yield widely unbiased statistical inferences on the population level, not to make sound predictions about class-membership on the individual level.

Limitations and future research With faster personal computers, it is nowadays feasible to fit both more complex MI models as well as more complex substantive models. Even for $m = 100$ models, parameters can be obtained by numerical integration within a reasonable amount of time. In the present paper, we therefore allowed starting levels of delinquency to vary across individuals. It would have been nice to also estimate variances (and covariances) of the other growth factors, both in the count and the zero-inflation part, and regress the other growth factors on gender and school type, as well. However, with only four panel waves, statistical models about the age-crime-relationship cannot be overly complex. Future research should corroborate present findings and extend the present analyses by considering data from more panel waves, which would allow to model the above mentioned additional quantities and relationships.

Funding Open Access funding enabled and organized by Projekt DEAL. The funding was provided by Deutsche Forschungsgemeinschaft (Grand Nos. Re832/4, Re832/8)

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barnard, J., Rubin, D.B.: Small-sample degrees of freedom with multiple imputation. *Biometrika* **86**(4), 948–955 (1999)
- Bodner, T.E.: What improves with increased missing data imputations? *Struct. Equ. Model.* **15**(4), 651–675 (2008)
- Boers, K., Reinecke, J., Seddig, D., Mariotti, L.: Explaining the development of adolescent violent delinquency. *Eur. J. Criminol.* **7**(6), 499–520 (2010)
- Bollen, K.A., Curran, P.J.: *Latent Curve Models: A Structural Equation Perspective*. Wiley, New York (2006)
- Bryk, A.S., Raudenbush, S.W.: *Hierarchical Linear Models*. Sage, Newbury Park (1992)
- Collins, L.M., Schafer, J.L., Kam, C.M.: A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychol. Methods* **6**(4), 330–351 (2001)
- Drechsler, J.: Multiple imputation of multilevel missing data—Rigor versus simplicity. *J. Educ. Behav. Stat.* **40**(1), 69–95 (2015)
- Enders, C.K.: *Applied Missing Data Analysis*. Guilford, New York (2010)
- Erdman, D., Jackson, L., & Sinko, A. (2008). Zero-inflated poisson and zero-inflated neg31 ative binomial models using the COUNTREG procedure. In: SAS Institute Inc. (ed.), *Proceedings of the SAS® Global Forum 2008 Conference*, Cary, NC, SAS Institute Inc (2008). Retrieved 18 October 2019 from <https://support.sas.com/resources/papers/proceedings/pdfs/sgf2008/322-2008.pdf>
- Farrington, D.P., West, D.J.: The Cambridge study in delinquent development: a longterm follow-up of 411 London males. In: Kerner, H.J., Kaiser, G. (eds.) *Kriminalität: Persönlichkeit, Lebensgeschichte und Verhalten*, pp. 115–138. Springer, Berlin (1990)
- Hilbe, J.M.: *Negative Binomial Regression*, 2nd edn. Cambridge University Press, Cambridge (2011)
- Horton, N.J., Lipsitz, S.R., Parzen, M.: A potential for bias when rounding in multiple imputation. *Am. Stat.* **57**(4), 229–232 (2003)
- Kleinke, K.: Multiple imputation under violated distributional assumptions: a systematic evaluation of the assumed robustness of predictive mean matching. *J. Educ. Behav. Stat.* **42**(4), 371–404 (2017)
- Kleinke, K., Reinecke, J.: Multiple imputation of incomplete zero-inflated count data. *Stat. Neerlandica* **67**(3), 311–336 (2013)
- Kleinke, K., Reinecke, J.: *Countimp version 2: a multiple imputation package for incomplete count data*. Technical Report, Siegen, Germany, University of Siegen, Department of Education Studies and Psychology (2019). Retrieved 22 October 2019 from <https://countimp.kkleinke.com>
- Kleinke, K., Reinecke, J., Salfrán, D., Spiess, M.: *Applied Multiple Imputation. Advantages, Pitfalls, New Developments and Applications*, Springer Nature, Cham (2020)
- Lambert, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics* **34**(1), 1–14 (1992)
- Mariotti, L., Reinecke, J. : Wachstums- und Mischverteilungsmodelle unter Berücksichtigung unbeobachteter Heterogenität: Empirische Analysen zum delinquenten Verhalten Jugendlicher in Duisburg [Growth models, mixture models, and unobserved heterogeneity: empirical analyses of juvenile delinquent behaviors in Duisburg]. Institut für sozialwissenschaftliche Forschung e.V, Münster (2010)
- Moffitt, T.E.: Adolescence-limited and life-course-persistent antisocial behavior: a developmental taxonomy. *Psychol. Rev.* **100**, 674–701 (1993)
- Muthén, L.K., Muthén, B.O.: *Mplus User's Guide*, 8th edn. Muthén & Muthén, Los Angeles (2017)
- Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., Solenberger, P.: A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* **27**(1), 85–95 (2001)
- Reinecke, J., Weins, C.: The development of delinquency during adolescence: a comparison of missing data techniques. *Q. Quant.* **47**(6), 3319–3334 (2013)
- Rubin, D.B.: Inference and missing data. *Biometrika* **63**(3), 581–590 (1976)
- Rubin, D.B.: *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York (1987)
- Sampson, R.J., Laub, J.H.: A life-course view of the development of crime. *Ann. Am. Acad. Polit. Soc. Sci.* **602**(1), 12–45 (2005)
- Schafer, J.L.: *Analysis of Incomplete Multivariate Data*. Chapman & Hall, London (1997)
- Schafer, J.L.: *norm2: analysis of incomplete multivariate data under a normal model (computer software manual)*. Retrieved 20 May 2016 from <https://CRAN.R-project.org/package=norm2> (R Package Version 2.0.1) (2016)
- Schafer, J.L., Graham, J.W.: Missing data: our view of the state of the art. *Psychol. Methods* **7**(2), 147–177 (2002)

- Schafer, J.L., Olsen, M.K.: Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivar. Behav. Res.* **33**(4), 545–571 (1998)
- Seddig, D., Reinecke, J.: Exploration and explanation of adolescent self-reported delinquency trajectories in the Crimoc study. In: Blokland, A., van der Geest, V. (eds.) *The Routledge International Handbook of Life-Course Criminology*, pp. 159–178. Routledge, New York (2017)
- Thornberry, T.P., Lizotte, A.J., Krohn, M.D., Smith, C.A., Porter, P.K.: Causes and consequences of delinquency. Findings from the Rochester Youth Development Study. In: Thornberry, T.P., Krohn, M.D. (eds.) *Taking Stock of Delinquency*, pp. 11–46. Kluwer, New York (2003)
- Tracy, P.E., Wolfgang, M.E., Figlio, R.M.: *Delinquency in Two Birth Cohorts*. Plenum, New York (1990)
- van Buuren, S.: *Flexible Imputation of Missing Data*, 2nd edn. CRC Press, Boca Raton (2018)
- van Buuren, S., Brand, J.P.L., Groothuis-Oudshoorn, C.G.M., Rubin, D.B.: Fully conditional specification in multivariate imputation. *J. Stat. Comput. Simul.* **76**(12), 1049–1064 (2006)
- van Buuren, S., Groothuis-Oudshoorn, K.: mice: multivariate imputation by chained equations in R. *J. Stat. Softw.* **45**(3), 1–67 (2011)
- von Hippel, P.T.: How many imputations are needed? A comment on Hershberger and Fisher (2003). *Struct. Equ. Model.* **12**(2), 334–335 (2003)
- Yu, L.M., Burton, A., Rivero-Arias, O.: Evaluation of software for multiple imputation of semi-continuous data. *Stat. Methods Med. Res.* **16**(3), 243–258 (2007)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Kristian Kleinke¹  · Jost Reinecke² · Cornelia Weins³

Jost Reinecke
jost.reinecke@uni-bielefeld.de

Cornelia Weins
cornelia.weins@rub.de

¹ Department of Education Studies and Psychology, University of Siegen, Adolf-Reichwein-Str. 2a, 57068 Siegen, Germany

² Faculty of Sociology, University of Bielefeld, 33501 Bielefeld, Germany

³ Faculty of Social Sciences, Ruhr-University Bochum, 44801 Bochum, Germany