Qual Quant (2013) 47:3319–3334
DOI 10.1007/s11135-012-9721-4

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

Jost Reinecke · Cornelia Weins

Published online: 19 May 2012

Abstract Conclusions on the development of delinquent behaviour during the life-course can only be made with longitudinal data, which is regularly gained by repeated interviews of the same respondents. Missing data are a problem for the analysis of delinquent behaviour during the life-course shown with data from an adolescents' four-wave panel. In this article two alternative techniques to cope with missing data are used: full information maximum likelihood estimation and multiple imputation. Both methods allow one to consider all available data (including adolescents with missing information on some variables) in order to estimate the development of delinquency. We demonstrate that self-reported delinquency is systematically underestimated with listwise deletion (LD) of missing data. Further, LD results in false conclusions on gender and school specific differences of the age–crime relationship. In the final discussion some hints are given for further methods to deal with bias in panel data affected by the missing process.

Keywords Full information maximum likelihood estimation · Multiple imputation · Growth curve models · Development of delinquency · Age–crime relationship

1 Introduction

Several long-term criminological studies, such as the Cambridge Study (Farrington and West 1990), the Philadelphia Study (Tracy et al. 1990) or the Rochester Study (Thornberry et al. 2003), are used to investigate the development of delinquent behaviour (e.g. the reader of Sampson and Laub 2005). One of the problems with the analysis of the development of delinquency is missing data: either the respondents could not participate or denied to cooperate in

J. Reinecke
Fakultät für Soziologie, Universität Bielefeld, 33501 Bielefeld, Germany
e-mail: jost.reinecke@uni-bielefeld.de

C. Weins (✉)
Fakultät für Sozialwissenschaft, Ruhr-Universität Bochum, 44801 Bochum, Germany
e-mail: cornelia.weins@rub.de

subsequent panel waves (panel attrition). There is a wide range of techniques to handle missing data. One of the most widespread methods is listwise deletion (LD), which excludes all cases with missing values on the variables under study. However, LD has several drawbacks. It discards a lot of information and results in biased parameter estimates if missingness is not completely at random. Full information maximum likelihood (FIML) and multiple imputation (MI) are the two state-of-the-art techniques to analyse incomplete data (Little and Rubin 2002). In contrast to LD they do not require missing data to be completely at random since they use all observed data to estimate parameters. FIML has been proposed as a method for estimating means and covariance matrices from incomplete data (Schafer and Graham 2002; Hofer and Hoffman 2007; Arbuckle 1996). As a second advanced strategy, MI of missing values is applied.

In this article LD, FIML and MI are used to analyse the development of delinquency during adolescence using a German panel study. Section 2 introduces the panel data and explores the missing data pattern and the missing data mechanism. Section 3 will briefly explain MI and FIML. To estimate the development of delinquency we use growth curve models. The variables measuring delinquency are count data and highly skewed to zero. Therefore, growth curve models are estimated under the assumption of a zero-inflated Poisson process (ZIP) (Sect. 4.1). Section 4.2 discusses results of different growth curve models assuming linear and quadratic growth of delinquency as well as the impact of sex and school type on the age–crime association. Section 5 presents the plausibility of assumptions, limitations and further research problems.

2 Data and missing data mechanism

2.1 Data

The empirical data for the analyses is taken from the longitudinal research project *Crime in the modern city*.[1] The main focus of the study is on the emergence and development of deviant and delinquent behaviour of juveniles and the social control surrounding it, that is, both formal, meaning the police and the judiciary, and informal, referring to school and family (Boers and Reinecke 2007; Wittenberg 2004).

The data set contains self-administered classroom interviews with pupils from the town of Münster, located north of Cologne in West Germany. The initial survey was conducted in the year 2000 on pupils in the 7th, 9th and 11th grade, considering all levels of schools in the community. The highest educational level in Germany is *Gymnasium* (‘high school’) with 9 years of school attendance in total. Two other school types exist beside high school: *Hauptschule* (‘basic school’) and *Realschule* (‘middle school’) with 6 years of school attendance. Those educational levels will be labeled hereafter as *high*, *medium* and *low*. The target population of this analysis includes pupils attending public schools in the Town of Münster.[2] Random samples of classes were chosen from the 9th and 11th grade of schools (cluster sampling). The survey among 7th graders includes all 7th grade classes. This cohort (7th grade) was interviewed annually until the pupils reached the 10th grade in the year 2003. We employ data from this cohort to analyse the development of delinquency. Response rates

[1] This interdisciplinary research project is conducted by the Universities of Münster (Institute of Criminology) and Bielefeld (Faculty of Sociology) and supported by the German National Science Foundation (DFG) under grant numbers Bo1234/6 and Re832/4 www.crimoc.org.

[2] Nearly all public schools participated over the whole time period.

Table 1 Size of population and realised interviews, cross-sections 2000–2003

Year	2000 (7th grade)	2001 (8th grade)	2002 (9th grade)	2003 (10th grade)
Population[a] size	2,105	2,181	2,251	2,077
Realised interviews	1,850	1,915	1,947	1,819

[a] Pupils attending public schools

Table 2 Patterns of missing values

Pattern	2000	2001	2002	2003	N
1.	t_1	t_2	t_3	t_4	813
2.	?	t_2	t_3	t_4	+262
3.	t_1	?	t_3	t_4	+38
4.	t_1	t_2	?	t_4	+10
5.	t_1	t_2	t_3	?	+184
6.	?	?	t_3	t_4	+293
7.	t_1	?	?	t_4	+49
8.	t_1	t_2	?	?	+264
9.	?	t_2	?	t_4	+20
10.	?	t_2	t_3	?	+114
11.	t_1	?	t_3	?	+17
(N)	(1,375)	(1,667)	(1,721)	(1,485)	$\sum = 2{,}064$

2,064 students with longitudinal information. Sample sizes at each wave are given in brackets

range between 86 and 88 %. Table 1 shows population sizes and realised interviews (Pöge 2005a, p. 14).

Due to strong confidentiality restrictions, it was not possible to collect pupils' home addresses. Instead an individual code was used to identify pupils in different waves. To generate the code, a specific code sheet was administered prior to the questionnaire in each panel wave. For example, the code sheet includes questions about specific letters corresponding to respondents' eye colour and hair colour and numbers of respondents' birthday (Pöge 2005b, 2008). The use of personal codes had two main drawbacks. First, identical codes belonging to different persons occurred. To deal with this problem additional information from the questionnaires was used to uniquely identify pupils. Furthermore not all pupils were able to exactly reproduce the code in subsequent waves. If the code was not useful to identify pupils over time (in the panel), individual questionnaires were checked manually. Since it was not possible to identify all students over time, we have cross-sectional data we cannot use in the panel. Note that cross-sectional data from wave 1 (in the year 2000) contains 1,850 cases (Table 1), but only 1,375 pupils could be identified across waves, that is we have longitudinal information for 1,375 pupils from wave 1 (see Table 2, last row).

Table 2 summarises the different missing data patterns of all respondents with valid information from at least two waves ($N = 2{,}064$). There are 813 pupils without missing values. These respondents took part in each survey, did not move to an area outside of the town and they were able to reproduce their encryption code in each wave. For 1,251 respondents, one or two panel waves are missing; for example there are 262 students with data in 2001–2003 but missing information in 2000. Several reasons can account for missing values: For some pupils it was not possible to assign interviews of these waves due to the above-mentioned problems with the encryption code, some pupils were not present in the school on the particular day of the survey, others refused to participate, and some relocated with their families (outside of the town Münster).

Delinquency was measured with several questions to get the prevalence (and incidence) of offenses[3] in the 12 months prior to the interview. The annual prevalence of the different offenses is provided in Table 8 in the Appendix.[4] Prevalences of the 16 different offenses are summed up to an index for each panel wave. The index has a range between zero and 16. Higher index values indicate more versatile criminal activity. On average, it is expected that the mean offense rate increases up to the third panel wave (average age of 15) and decreases thereafter. This curvilinear development (also called age–crime curve) is typical for adolescents aged between 14 and 18 years and is described as an *adolescent-limited* type of delinquent behaviour (e.g. Moffitt 1993).

With LD only cases without missing values (813 out of 2,064 cases) are analysed. 1,251 cases (and the available observations for these cases) would be discarded. LD inflates standard errors due to the loss of cases. Even worse LD produces biased parameters if the 813 cases with complete data are not a simple random sample of all 2,064 cases. To state the advantages of FIML and MI it is necessary to consider different *missing-data mechanisms*.

2.2 Missing data mechanisms

Missing data mechanisms relate the probability of missing values to the data. Rubin (1976) distinguished three different missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). Following the notation of Little and Rubin (2002) the vector of observed data is Y_{obs}; the vector of missing data is Y_{mis}: $Y = (Y_{obs}, Y_{mis})$. A missing data matrix $M(= M_{ij})$ provides the information if, for a person i, the value of a variable j is missing ($M_{ij} = 1$) or not. ψ denotes the parameter vector influencing the probability of missing data. θ contains the parameters of substantive interest, that are the parameters without missing data.

The missing data mechanism is MCAR when the probability of missing values on a variable does not depend either on the observed values Y_{obs} or the missing values Y_{mis}: $f(M|Y, \psi) = f(M|\psi)$. The assumption of MCAR is required with LD since observed values for cases with missing values on variables under study are discarded. Under MCAR complete cases are a simple random sample of all cases. The missing data mechanism is MAR when the probability of missingness depend on the observed values, but is unrelated to the missing values: $f(M|Y, \psi) = f(M|Y_{obs}, \psi)$.[5] MAR is a weaker assumption than MCAR. If the missing data mechanism is MAR (or MCAR) likelihood-based conclusions about the parameters of substantive interest (θ) are possible without information about the parameters ψ that govern the missing data process (Little and Rubin 2002, p. 119). Full information maximum likelihood and MI produce unbiased parameters under MAR. In contrast to LD all available data are employed. Finally, the process of missingness is NMAR when the process of missingness also depends on the missing values themselves: $f(M|Y, \psi) = f(M|Y_{mis}, Y_{obs}, \psi)$.

3 Vandalism: graffiti, damage to property. Violence: robbery, purse snatching, assault with and assault without a weapon. Property: burglary, fencing, theft of and out of cars, out of a vending machine, of bicycles, other theft, shoplifting. Drugs: Drug abuse, drug trafficking.

4 Item-nonresponse on offences is negligible. Therefore item-nonresponse is coded to 0 (no offense). This coding does not influence the results presented below.

5 Additionally it is required that θ and ψ are distinct parameters. The process of missingness is then referred to as nonignorable.

Table 3 Gender and school type

	Official statistics	All respondents	Respondents without missing data	Respondents with missing data
Male	52.1 %	51 %	43 %	55 %
Female	47.9 %	49 %	57 %	45 %
low	26.6 %	25 %	16 %	30 %
Middle	30.6 %	32 %	30 %	34 %
High	42.8 %	43 %	55 %	36 %
	$N = 2{,}105$	$N = 2{,}064$	$N = 813$	$N = 1{,}251$

Official statistics refer to the population in 2000
The three educational levels represent the German school types

The missing data mechanism cannot be ignored because the missing data contains information about the substantive parameters θ. If the process of missingness is NMAR it is necessary to model missingness with pattern mixture models or selection models (Enders 2011).

Which assumption about the missing process holds in our data? MCAR includes that, for instance, missingness on delinquency in the year 2000 in Pattern 2 of Table 2 is not related to the observed variables under study (delinquency in 2001–2003, school level, and sex) and not related to the missing values themselves (delinquency in 2000). MAR only requires, in this pattern, that missingness on delinquency in 2000 is not related to the missing values on delinquency in 2000. To establish whether missingness on delinquency is MCAR, adolescents with missing data and those without missing data are compared according to school type and sex (Table 3). Table 3 shows systematic differences in sex and educational level for both groups. Note that school type and sex are completely observed.[6] Of the variables under study only delinquency (at the four time-points) has missing values. Adolescents with missing data are overproportionally males. In addition, the percentage of students attending the lowest school type is nearly twice as high among adolescents with missing data than among adolescents without missing data. There is hardly any difference between official school statistics and sample statistics for all respondents with data from a minimum of two waves ($N = 2{,}064$).

Table 4 shows the means of the indices ($\bar{x}$). In addition, the percentage of adolescents who did not show delinquent behaviour in a particular year is included in the Table (% Zero). Of the 1,251 adolescents with one or two missing values, 562 persons have a valid value on delinquency in 2000, 854 in 2001 etc. It should be noted that the 813 adolescents with complete data have lower delinquency rates than the 1,251 adolescents with missing data. In addition, the percentage of non-delinquent adolescents is larger among students without missing data. Missing values on delinquency depend on the observed values of school-type, sex and delinquency. Obviously, the missing mechanism is not MCAR.

While it is simple to check the MCAR condition, this does not hold for MAR. MAR implies that missingness does not depend on the missing values. It is not possible to verify MAR because the missing values are not known. One exception is planned missing data, where the missing data mechanism is under the control of the researcher. Nevertheless, in our data it is plausible that the missing mechanism is MAR and not NMAR: The survey was

[6] Since we have at minimum two waves for each student and sex is a time-invariant characteristic, sex is known. Because changes between different school types are rare, we employ the first observed value to measure school type.

Table 4 Self-reported delinquency

	All respondents			Respondents without missing data			Respondents with missing data		
	$\bar{x}$	% Zero	N	$\bar{x}$	% Zero	N	$\bar{x}$	% Zero	N
2000	0.6	72	1,375	0.5	77	813	0.8	65	562
2001	1.1	61	1,667	0.7	68	813	1.4	53	854
2002	1.3	55	1,721	0.9	64	813	1.7	47	908
2003	1.0	57	1,485	0.8	63	813	1.2	49	672

% Zero Percentage of persons without delinquent behaviour in the previous year

conducted in each class in 2000–2003. If an adolescent did not participate in 2000, than he or she will be reinterviewed in the next year (*classroom survey*). Contrary to most panel studies, this design did not result in a monotone pattern of panel attrition.[7] At least two measurements of delinquency are in the data for all 2,064 adolescents. Since these measurements are strongly correlated, it is very plausible that missing information on self-reported delinquency, e.g. in t_1, can be explained by the observed delinquency at other time points, e.g. in t_2, t_3 or t_4 (Graham et al. 1997, 352 ff.).

3 MAR-methods

The impact of missing data on the development of delinquent behaviour will be evaluated by the comparison of model results based either on LD, FIML or MI. The primary advantage of MI and FIML is that both require less strict assumptions on the process of missingness (MAR) compared to traditional techniques like LD of missing data (MCAR). Even if the missing data mechanism is MCAR, MAR based methods should be favoured due to the loss of cases in LD, thereby making LD inefficient.

3.1 Multiple imputation

Instead of discarding observed values on incomplete cases (LD), MI uses all available information Y_{obs} (of complete and incomplete cases) to generate multiple plausible values for each missing value. A MI is a sequence of three steps: the imputation step, the analysis step, and the pooling step. First, multiple ($D > 1$) plausible values are generated for each missing value and then filled in. The imputation step results in $D > 1$ completed data sets without missing values. The D filled-in data sets only differ in the imputed values. Standard complete-data methods are then performed on each of the D imputed data sets separately. Finally, the D analyses are pooled to obtain a single set of results.

The most challenging step is to generate 'proper' imputations. First, proper imputations have to be stochastic to reproduce the variability of the data. For instance, regression imputation is not proper since all imputed values fall straight on the regression line, thereby overestimating the correlation. In contrast, stochastic regression imputation adds a random draw from the residual distribution to the predicted conditional mean to preserve the variability of the data. Second, parameters used to generate imputed values have to be randomly drawn

[7] If adolescents changed classes or schools within the city of Münster, they remained in the panel. Since school attendance is compulsory until the age of 18, it is very unlikely that an adolescent did not participate at all four panel waves, e.g. that we have no information at all.

from their Bayesian posterior distribution to reflect uncertainty about the true parameter values. The most widespread technique of realising imputations is data augmentation which assumes multivariate normality of the data (Schafer 1997). The algorithm iteratively cycles between the imputation (I) and the posterior (P) step. In the I-step, random draws for Y_{mis} are generated from the conditional distribution of the missing values, given the observed data and parameters: $Y_{mis}^{(t+1)} \sim p(Y_{mis}|Y_{obs}, \theta^{(t)})$. For the first I-step starting values of the parameters θ are required. Usually starting values are calculated with the EM-algorithm, alternatively listwise or pairwise deletion estimates may be used. The P-step randomly draws parameters from the complete-data posterior: $\theta^{(t+1)} \sim p(\theta|Y_{obs}, Y_{mis}^{(t+1)})$. To obtain the complete-data posterior, Y_{obs} is augmented with values of Y_{mis} from the preceding I-Step (Schafer 1997, p. 72). Repetitions of both steps result in a Markov chain $Y_{mis}^1; \theta^1; Y_{mis}^2; \theta^2; \ldots; Y_{mis}^t, \theta^t$ with t as the number of repetitions. After convergence (for large t), $Y_{mis}^{(t)}$ and $\theta^{(t)}$ are random draws of $P(\theta, Y_{mis}|Y_{obs})$ (Schafer 1997, 80). The algorithm converges if $\theta^{(t+k)}$ is independent of $\theta^{(t)}$, where k is the number of iterations. Convergence in a single chain can be assessed by time series or autocorrelation plots of parameters (Schafer 1997, Chap. 4.4). Because the imputed values in adjacent I-steps are correlated with one another, a sufficient number of I-steps have to be carried out between each imputation.[8] In many cases, a small number of three to five imputations yield standard errors that are not much larger than standard errors from an infinite number of imputations (Schafer 1997, pp. 106–107). However, such a small number of imputations may diminish statistical power for small effect sizes considerably (Graham et al. 2007).

Simulation studies show that data augmentation is robust to violations of the assumption of multivariate normality (Demirtas et al. 2008). Non-normal continuous variables may be normalised with appropriate transformations. It is also possible to impute dummy variables with data augmentation (Allison 2002). For mixed (continuous and categorical) data special algorithms are available (Schafer 1997; Raghunathan et al. 2001; Van Buuren 2007). Sequential regression imputation is implemented in SAS, Stata, and R. With sequential regression the imputation model depends on the distribution of a variable, e.g. continuous variables are imputed using a linear regression model, dichotomous variables using a logistic regression model and count data using a Poisson regression model. Since our variables measuring delinquency are count data with a large amount of zeros (see Table 4), we use a different approach (Schafer and Olsen 1999) explained below.

After imputing plausible values, each of the D completed data sets are analysed. Parameter estimates (e.g. coefficients from structural equation models) of these analyses are simply averaged to produce a single set of results: $\bar{\theta}_D = \frac{1}{D}\sum_{d=1}^{D} \hat{\theta}_d$ (Little and Rubin 2002, p. 86). To obtain correct standard errors, the average variance within the imputed data sets and the variance between estimates of these data sets are considered (Little and Rubin 2002, p. 86f.).[9]

The program NORM (Schafer 1997) implements data augmentation. We used NORM for Windows to impute missing values.[10] The imputation model includes delinquency rates at

[8] It is also possible to generate MIs with independent chains (parallel data augmentation).

[9] The within-imputation variance is the average of the D squared standard errors: $\bar{W}_D = \frac{1}{D}\sum_{d=1}^{D} se_d^2$. Within standard errors are calculated without missing values (too large n), thereby underestimating sampling error. The between-imputation variance quantifies the variability of the parameter estimates of different filled in data sets: $B_D = \frac{1}{D-1}\sum_{d=1}^{D} (\hat{\theta}_d - \bar{\theta}_D)^2$. The between-imputation variance reflects the uncertainty on the parameters due to missing data since the imputed data sets only differ in the imputed values. The total variance is: $T_D = \bar{W}_D + \frac{D+1}{D} B_D$. The MI standard error is the square root of the total variance: $se = \sqrt{T_D}$.

[10] The program NORM can be downloaded free of charge from the webpage http://www.stat.psu.edu/~jls. Data augmentation is also available in SAS, Stata (since version 11, see Stata 2011), and R.

t_1, t_2, t_3, t_4, sex and school type. Sex and school-type are completely observed. An imputation model may also contain variables which are highly correlated with missingness or with variables with missing values (Collins et al. 2001) even if these variables are not part of the substantive analysis. Such 'auxiliary' variables make MAR more realistic and thereby improve the estimation of plausible values. It is also possible (but more difficult) to incorporate auxiliary variables with FIML. To utilise delinquency measures of different time points for imputation, the data set is arranged in wide format (see Table 2), that is we have one record for each adolescent (Allison 2002, p. 74 ff.). Delinquency rates are zero-inflated: between 55 % (in the year 2002) and 72 % (in the year 2000) of the interviewees reported no offense at all. To take zero-inflation into account, we represent each delinquency rate with two variables in the imputation model (Schafer and Olsen 1999).[11] Sex is a dichotomous variable and school type (low, middle, high) is replaced by a set of two dummy variables (middle, high) in the imputation model.[12] For each missing value 10 values were imputed, i.e. we have 10 completed data sets, each with 2,064 students.[13]

3.2 Maximum likelihood estimation with missing data

ML estimation with incomplete data is well discussed (for example Dempster et al. 1977; Little and Rubin 2002). Allison (1987) and Muthén et al. (1987) have shown how the method applies to structural equation modelling by using multiple groups to separate missing data patterns. However, the multiple-group approach works only with a few distinct missing data patterns. Direct ML methods, also referred to as FIML, overcome those limitations. FIML is implemented in structural equation programs like AMOS, LISREL and M*plus*. The main difference between FIML and MI is that with FIML estimation of substantive parameters (e.g. coefficients in a growth curve model) and missing data handling are proceeded in one step. Like MI FIML uses all available data to estimate substantive parameters. With FIML, each case's contribution to the sample log likelihood is computed using all observed values for that case. Standard errors are obtained from the second derivatives of the FIML function via the so-called information matrix. With missing data the observed information matrix yields unbiased parameters under MAR and MCAR, whereas the expected information matrix can yield biased standard errors (Enders 2010, p. 102f.).

Several simulation studies confirm the feasibility of FIML estimation with incomplete data (Enders and Bandalos 2001; Enders 2001a,b). It has also been shown that FIML produces unbiased parameter estimates and standard errors when the missing process is MAR (Arbuckle 1996; Enders and Bandalos 2001; Enders 2001c). Using the same set of cases and variables, MI and FIML should produce similar parameter estimates (Schafer and Graham 2002; Schafer 1997). However, as mentioned above, the use of auxiliary variables is easier with MI. Using FIML therefore leads to a more restrictive use of auxiliary variables than using MI (Collins et al. 2001).

[11] Each delinquency rate Y_t was split into two variables, i.e. the imputation model included eight variables to capture delinquency. W_t measures if an adolescent reported no offence at time point t ($W_t = 0$) or at least one offence ($W_t = 1$). Z_t counts the number of offences greater zero and is missing for non-offenders. Integer values between 1 and 16 were imputed on Z_t. Imputations on W_t were rounded to zero or one. After the imputation W_t and Z_t were recoded to Y_t, where $Y_t = Z_t$ if $W_t = 1$ and $Y_t = 0$ if $W_t = 0$.

[12] Multivariate normality does not hold for these variables. Since both variables have no missing values this poses no problem (Schafer 1997, p. 203).

[13] We generated a single chain of 10,000 iterations, with 1,000 burn-in iterations and 1,000 between-imputation iterations.

4 Juvenile delinquency: results of LD, MI and FIML

The development of delinquent behaviour is analysed with latent growth curve models. Growth curve models are estimated first with full-information maximum likelihood, secondly using the 10 multiple imputed data sets (Sect. 3.1), and thirdly using solely complete cases ($N = 813$ adolescents, see Table 2).

4.1 Growth curve models

The growth curve model is specified for Poisson distributed variables because delinquent behaviour is measured as a count variable. Let $Y = 0, 1, 2 \ldots$ be a random variable for a given time interval and y be the number of observed occurrences.

The number of events in that interval is Poisson distributed with the following probability density function:

$$Pr(Y = y) = e^{-\mu} \left[\frac{\mu^y}{y!} \right] \tag{1}$$

The expected value of the Poisson distribution is $E(y) = \mu$ with $Var(y) = \mu$. The parameter μ is referred to as the mean rate of occurrences of events. Small values of μ yield high probability for zero occurrences of the random variable y. The higher the value of μ, the lower the skewness of the distribution. Here, a growth curve model will be used to explain the Poisson distributed count data[14]:

$$\ln(\mu_t) = \lambda_{1t}\eta_1 + \lambda_{2t}\eta_2 + \varepsilon_t \tag{2}$$

μ_t is the expected number of occurrences at time t. η_1 is the intercept, i.e. the mean rate of y at the starting point of a given time interval. η_2 is the linear slope describing the development of y, e.g. the linear increase or decrease of delinquent behaviour. ε_t is the error term of the measurements which can be held constant over time.

Empirical evidence in criminological research as well as our descriptive results for the cross sections with missing data (see Table 4) suggest that with increasing age delinquency during adolescence first rises and then declines. To capture a nonlinear relationship, Eq. (3) includes a quadratic term η_3:

$$\ln(\mu_t) = \lambda_{1t}\eta_1 + \lambda_{2t}\eta_2 + \lambda_{3t}\eta_3 + \varepsilon_t \tag{3}$$

If the number of zeros in the count variables is large, i.e. there is a high percentage of non-offenders, the ZIP originally proposed by Lambert (1992) is more appropriate (see Reinecke 2006). The ZIP model combines the Poisson model with a logit model to cover the zero inflation in the count variable y:

$$y_t \sim \begin{cases} 0 & \text{with probability } p \\ ln(\mu_t) & \text{with probability } 1 - p \end{cases} \tag{4}$$

Two parallel growth curve models are estimated simultaneously when zero inflation of the data is assumed. The first model in the upper panel of Fig. 1 contains the Poisson part of the variable y_t (variables y_1 to y_4). The means of the growth curve variables (η_1, η_2, η_3) are estimated. The second model in the lower panel of Fig. 1 refers to the Logit part of the variable y_t (Variables y_1^i to y_4^i). The mean of the intercept (η_1^i) is fixed to zero while the

[14] The growth curve model can also be formalised as a random effects model with the notation of multilevel equations (see Hox 2002).

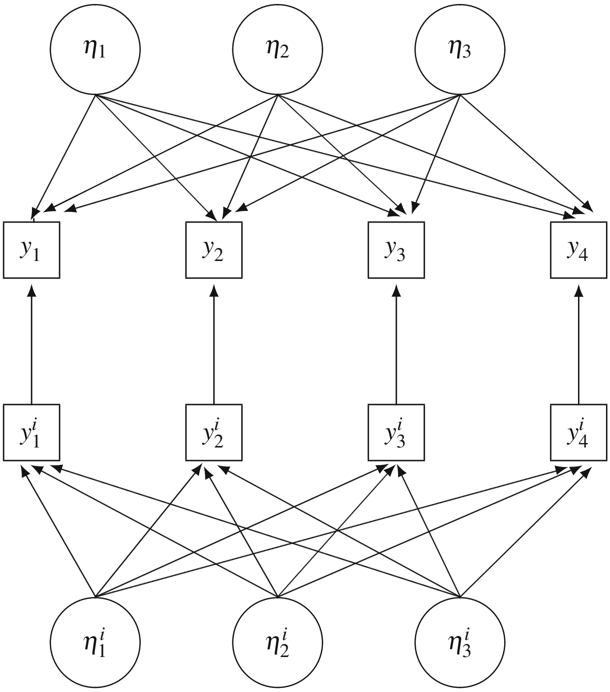

Fig. 1 Growth curve models for zero-inflated count data. Variables of the inflation part (logit model) are indicated by the suffix i

means of the linear and quadratic slope (η_2^i and η_3^i) are estimated. Logit and Poisson part of the growth curve model can be estimated simultaneously using M*plus* (Muthén and Muthén 2010).

Equations (2) and (3) will be applied in our analyses. Model restrictions are used in a conventional way: The factor loadings of the intercept η_1 are fixed to 1 over all four panel waves; factor loadings of the linear slope η_2 are fixed to the values 0 (t_1), 1 (t_2), 2 (t_3) and 3 (t_4); and factor loadings of the quadratic slope η_3 are fixed to the values 0 (t_1), 1 (t_2), 4 (t_3), and 9 (t_4). The same restrictions are used for the inflation part of the model. If the age–crime relationship is bell-shaped (*age–crime curve*), then the mean of the intercept and the mean of linear slope in the Poisson part of the model should be positive and the mean of the quadratic slope in the Poisson part should be negative.

4.2 Results

All Models are estimated with M*plus*. M*plus* allows a simultaneous analysis of several imputed data sets and provides combined point estimates and standard errors (see footnote 11). First, models without exogenous variables are estimated (*unconditional models*). Additionally, the impact of gender and school types (high, medium) on the development of delinquency is considered (*conditional models*). The fit measures of the different models are displayed in Table 5. With FIML and MI, the quadratic model specifications have superior fit measures, i.e. lower AIC and BIC values, than the linear ones. In contrast, with LD of

Table 5 Goodness of fit measures of the growth curve zero-inflated Poisson models

	Linear		Quadratic	
	Unconditional	Conditional	Unconditional	Conditional
FIML ($N = 2,064$)				
Log likelihood	−8997	−8637	−8934	−8578
N of parameter	4	16	6	24
AIC	18,002	17,305	17,881	17,204
BIC	18,025	17,395	17,914	17,339
Adj. BIC	18,012	17,345	17,895	17,263
MI ($N = 2,064$)				
Log likelihood	−12323	−11896	−12271	−11864
N of parameter	4	16	6	24
AIC	24,654	23,824	24,553	23,721
BIC	24,677	23,914	24,587	23,857
Adj. BIC	24,664	23,863	24,568	23,780
LD ($N = 813$)				
Log likelihood	−3793	−3670	−3787	−3660
N of parameter	4	16	6	24
AIC	7,594	7,372	7,585	7,368
BIC	7,613	7,448	7,613	7,480
Adj. BIC	7,601	7,397	7,594	7,404

missing data the fit measures for the linear and quadratic models differ only slightly. Including exogenous variables improves the model fit with FIML, MI, and LD.

Estimates of the unconditional quadratic growth curve models are presented in Table 6. As stated before, the ZIP model estimates two parts of the growth curve: The Logit part models the odds of not being delinquent (structural zeros), the Poisson part models the number of different offences (versatility). FIML, MI and LD results show the expected bell-shaped relationship between age and delinquency in adolescence: In the Poisson part (upper panel of Table 6) the mean of the intercept and the mean of the linear slope have positive signs and the mean of the quadratic slope has a negative sign. The effect sizes differ; with LD the mean rate of versatility is underestimated and LD estimates indicate a more flat age–crime curve in adolescence. Due to the loss of cases, standard errors are larger with LD.

Detailed results of the conditional quadratic growth curve models are presented in Table 7. Note that point estimates and standard errors of FIML and MI are quite similar. Firstly, we look at the influence of sex and school type on the prevalence of delinquency (Logit Part, lower panel of Table 7). The results of FIML and MI are more plausible, i.e. in accordance with criminological evidence, in comparison to the results of LD. According to FIML and MI, the estimated odds of not being delinquent at the beginning of the survey are about 50 % higher for girls than for boys.[15] In contrast, LD shows a non-significant effect of sex since delinquent boys are underrepresented with LD of missing data.[16] The odds of not being

[15] Using the coefficient estimated with the multiply imputed data sets: $(e^{.409} - 1) * 100$.

[16] Additional analyses using the multiple imputed data sets show that LD overestimates the percentage of non-delinquent boys, thereby obscuring sex differences.

Table 6 Parameter estimates of the unconditional growth curve ZIP-model

	FIML		MI		LD	
Poisson part						
Intercept (I)	.573**	(.064)	.681**	(.069)	.439**	(.101)
Slope (S)	.499**	(.075)	.374**	(.081)	.263*	(.112)
QSlope (Q)	−.151**	(.022)	−.113**	(.025)	−.066*	(.032)
Logit part						
Intercept (Ii)	–	–	–	–	–	–
Slope (Si)	−.482**	(.089)	−.397**	(.086)	−.430*	(.136)
QSlope (Qi)	.087**	(.027)	.064*	(.027)	.074	(.040)
N	2,064		2,064		813	

delinquent are also higher for students attending ‘high schools’ than for students attending ‘basic schools’. Attending a ‘high school’ has no significant impact on the intercept with LD of data.

Secondly, in the Poisson part of the model (upper panel of Table 7) gender and educational level influences the intercept (I). At the beginning of the observation period, girls have a lower versatility in comparison to boys, and students attending a high or medium level school type have a lower versatility than students attending ‘basic schools’, although the coefficient for ‘medium schools’ is not significant with MI. Gender and school type have no significant impact on the linear and quadratic slope (columns FIML and MI), i.e. on the curvature of the age–crime relationship.

To sum up, FIML and MI lead to the same results concerning the development of delinquency. Point estimates and standard errors of the conditional models differ only slightly. Although both methods—FIML and MI—are ceteris paribus asymptotically equivalent, one should note that the imputation model employed here was designed for continuous data with a large amount of zeros (Schafer and Olsen 1999) and not for zero-inflated count data. However, the violation of the distributional assumptions of the imputation model does not distort substantive results of the growth curve models.

5 Discussion

Criminological research has scrutinised the necessity of panel studies in order to analyse the development of delinquency (Sampson and Laub 2005). When self-reports of delinquent behaviour are analysed over time, the amount of missing data increases with the number of panel waves. The most widespread method used to cope with missing data is LD. With longitudinal data, LD results in a considerably high loss of cases. In our data, only 813 out of 2,064 adolescents have valid information at all four time points and therefore remain in the panel with LD. For substantive conclusions about the development of delinquency one has to assume that the remaining 813 respondents are a simple random subsample of the original sample. This assumption does not hold here. Descriptive results have shown that missing values on delinquency depend on sex, school type, and also (observed) delinquency at other time points. Females, adolescents attending a high school, and non-delinquent adolescents are overrepresented if one only analyses cases with complete longitudinal information.

Table 7 Parameter estimates of the conditional growth curve ZIP-model

	FIML		MI		LD	
Poisson part						
Intercept (I)	.957**	(.120)	.999**	(.110)	1.217**	(.192)
Slope (S)	.419**	(.142)	.339*	(.135)	−.203	(.232)
QSlope (Q)	−.119**	(.040)	−.096*	(.040)	.056	(.061)
Predictors						
Female → I	−.415**	(.146)	−.374**	(.134)	−.744**	(.209)
Medium → I	−.314*	(.152)	−.202	(.125)	−.608**	(.232)
High → I	−.356*	(.154)	−.373**	(.135)	−.676**	(.248)
Female → S	−.168	(.166)	−.167	(.165)	.316	(.234)
Medium → S	.189	(.182)	.111	(.153)	.437	(.282)
High → S	.090	(.182)	.118	(.171)	.479	(.274)
Female → Q	.061	(.046)	.053	(.046)	−.055	(.066)
Medium → Q	−.052	(.051)	−.034	(.047)	−.112	(.078)
High → Q	−.067	(.053)	−.058	(.052)	−.156*	(.076)
Logit part						
Intercept (Ii)	–	–	–	–	–	–
Slope (Si)	−.885**	(.199)	−.845**	(.182)	−1.405**	(.337)
QSlope (Qi)	.222**	(.060)	.210**	(.054)	.325**	(.102)
Predictors						
Female → Ii	.399*	(.161)	.409**	(.145)	−.120	(.256)
Medium → Ii	.056	(.195)	.018	(.168)	−.362	(.307)
High → Ii	.372*	(.182)	.322*	(.157)	.168	(.283)
Female → Si	.149	(.201)	.123	(.197)	.577	(.311)
Medium → Si	.391	(.247)	.421*	(.215)	.945*	(.404)
High → Si	.540*	(.239)	.578*	(.204)	.933*	(.377)
Female → Qi	−.016	(.060)	−.010	(.059)	−.094	(.091)
Medium → Qi	−.154*	(.073)	−.159*	(.062)	−.253*	(.119)
High → Qi	−.212**	(.074)	−.221**	(.064)	−.296**	(.114)
N	2,064		2,064		813	

The mean of the intercept (Ii) in the logit-part of the ZIP-model has to be fixed to zero
* $p < 0.05$, ** $p < 0.01$ (two-sided)

FIML and MI are the two state of the art techniques of handling missing data. Both methods use all observed information to estimate substantive parameters, thereby eliminating bias due to missing data which may be explained with observed data. Results of FIML and MI show a higher level of delinquent behaviour during adolescence, and a more pronounced curvilinear association between age and crime. In addition, FIML and MI estimates show a higher prevalence of delinquent behaviour of boys and of adolescents attending 'basic schools'. FIML and MI will generally produce similar parameter estimates given identical variables under study and consistency between the imputation and analysis model (Collins et al. 2001; Enders 2006). So far imputation routines for zero-inflated count data are not available in statistical software. Therefore FIML may be the method of choice for Poisson distributed variables with a high amount of zeros. However, with our data even an imputation routine not designed for count data (but accounting for a large amount of zeros) results in similar point estimates and standard errors as FIML. MI requires additional effort (imputation step) before substantive analyses, but it means that virtually any statistical analysis can be performed after the imputation.

FIML and MI are prone to bias when the MAR assumption is violated. We argued that MAR is plausible in the data under study due to the design (longitudinal data, no monotone pattern of dropout, see Sect. 2.2). MAR is violated if the dropout of respondents depends on the topic of the investigation. Under NMAR, selection models or pattern mixture models for growth curve models may be used (Enders 2011). However, these methods rely on untestable assumptions and are not robust against violations of these assumptions (Enders 2011, p. 15).

Appendix

Table 8 Annual prevalence of self-reported delinquency

	7th grade 13 years[a]		8th grade 14 years[a]		9th grade 15 years[a]		10th grade 16 years[a]	
	2000		2001		2002		2003	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Damage to property	126	(9.2)	193	(11.6)	216	(12.6)	130	(8.8)
Graffiti	57	(4.1)	187	(11.2)	213	(12.4)	96	(6.5)
Robbery	32	(2.3)	49	(2.9)	55	(3.2)	27	(1.8)
Purse snatching	1	(0.1)	14	(0.8)	15	(0.9)	5	(0.3)
Assault without weapon	100	(7.3)	179	(10.7)	195	(11.3)	138	(9.3)
Assault with weapon	8	(0.6)	40	(2.4)	33	(1.9)	15	(1.0)
Burglary	18	(1.3)	52	(3.1)	60	(3.5)	34	(2.3)
Fencing	70	(5.1)	111	(6.7)	135	(7.8)	93	(6.3)
Theft out of cars	6	(0.4)	27	(1.6)	30	(1.7)	17	(1.1)
Theft of cars	5	(0.4)	24	(1.4)	37	(2.1)	14	(0.9)
Theft out of a vending machine	32	(2.3)	49	(2.9)	55	(3.2)	21	(1.4)
Theft of bicycles	34	(2.5)	104	(6.2)	164	(9.5)	143	(9.6)
Shop-lifting	206	(15.0)	332	(19.9)	337	(19.6)	190	(12.8)
Other theft	28	(2.0)	55	(3.3)	68	(4.0)	38	(2.6)
Drug abuse	96	(7.0)	283	(17.0)	490	(28.5)	450	(30.3)
Drug trafficking	11	(0.8)	71	(4.3)	120	(7.0)	81	(5.5)
No offense	990	(72.0)	1011	(60.6)	946	(55.0)	846	(57.0)

[a] Mean age at the time of survey

References

Allison, P.D.: Estimation of linear models with incomplete data. Sociol. Methodol. **17**, 71–103 (1987)

Allison, P.D.: Missing Data. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-136. Sage, Thousand Oaks (2002)

Arbuckle, J: Full information estimation in the presence of incomplete data. In: Marcoulides, G.A., Schumacker, R.E. (eds.) Advanced Structural Equation Modeling: Issues and Techniques, pp. 243–277. Lawrence Erlbaum Associates, Mahwah (1996)

Boers, K., Reinecke, J. (eds.): Delinquenz im Jugendalter. Erkenntnisse einer Münsteraner Längsschnittstudie. Waxmann, Münster (2007)

Collins, L.M., Schafer, J.L., Kam, C.M.: A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol. Methods **6**, 330–351 (2001)
Demirtas, H., Freels, S.A., Yucel, R.M.: Plausibility of multivariate normality assumption when multiply imputing non-Gaussian continous outcomes: a simulation assessment. J. Stat. Comput. Simul. **78**, 69–84 (2008)
Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. B. **39**, 1–38 (1977)
Enders, C.K.: The impact of nonnormality on full information maximum-likelihood estimation for structural equation models with missing data. Psychol. Methods **6**, 352–370 (2001a)
Enders, C.K.: A primer on maximum likelihood algorithms available for use with missing data. Struct. Equ. Model. **8**, 128–141 (2001b)
Enders, C.K.: The performance of the full information maximum likelihood estimator in multiple regression models with missing data. Educ. Psychol. Meas. **61**, 713–740 (2001c)
Enders, C.K.: Analyzing structural equation models with missing data. In: Hancock, G.R., Mueller, R.O. (eds.) Structural Equation Modeling: A Second Course, pp. 269–314. IAP, Greenwich (2006)
Enders, C.K.: Applied Missing Data Analysis. The Guilford Press, New York (2010)
Enders, C.K.: Missing not at random models for latent growth curve analyses. Psychol. Methods **16**, 1–16 (2011)
Enders, C.K., Bandalos, D.L.: The relative performance of full information maximum likelihood estimation for missing data in structural equation models. Struct. Equ. Model. **8**, 430–457 (2001)
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)
Graham, J.W., Hofer, S.M., Donaldson, S.I., MacKinnon, D.P., Schafer, J.L.: Analysis with missing data in prevention research. In: Bryant, K.J., Windle, M., West, S.G. (eds.) The Science of Prevention. Methodological Advances From Alcohol and Substance Abuse Research, pp. 325–365. American Psychological Association, Washington (1997)
Graham, J.W., Olchowski, A.E., Gilreath, T.D.: How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev. Sci. **8**, 206–213 (2007)
Hofer, S.M., Hoffman, L.: Statistical analysis with incomplete data: a developmental perspective. In: Little, T.D., Bovaird, J.A., Card, N.A. (eds.) Modeling Ecological and Contextual Effects in Longitudinal Studies of Human Development, pp. 13–32. LEA, Mahwah (2007)
Hox, J.J.: Multilevel Analysis. Techniques and Applications. Lawrence Erlbaum, Mahwah (2002)
Lambert, D.: Zero-inflated Poisson regression with an application to defects in manufacturing. Technometrics **34**, 1–13 (1992)
Little, R.J., Rubin, D.B.: Statistical Analysis with Missing Data, 2nd edn. Wiley, Hoboken (2002)
Moffitt, T.E.: Adolescence-limited and life-course-persistent antisocial behaviour: a developmental taxonomy. Psychol. Rev. **100**, 674–701 (1993)
Muthén, L., Muthén, B.O.: Mplus. Statistical Analysis with Latent Variables. User's Guide (Version 6). Muthén & Muthén, Los Angeles (2010)
Muthén, B.O., Kaplan, D., Hollis, M.: On structural equation modeling with data that are not missing completely at random. Psychometrika **42**, 431–462 (1987)
Pöge, A.: Methodendokumentation der kriminologischen Schülerbefragung in Münster 2000–2003. Münster, Trier (2005a)
Pöge, A.: Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht. ZA-Information **8**, 50–69 (2005b)
Pöge, A.: Persönliche Codes "reloaded". Methoden - Daten - Analysen **2**, 59–70 (2008)
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**, 85–95 (2001)
Reinecke, J.: Longitudinal analysis of adolescents' deviant and delinquent behaviour. Applications of latent class growth curves and growth mixture models. Methodology **2**, 100–112 (2006)
Rubin, D.B.: Inference and missing data. Biometrika **63**, 581–592 (1976)
Sampson, R.J., Laub, J.H. : A life-course view of the development of crime. In: Sampson, R.J., Laub, J.H. (eds.) Developmental Criminology and Its Discontents. Annals of the American Academy of Political and Social Science, vol 602, pp. 12–45. Sage, Thousand Oaks (2005)
Schafer, J.L.: Analysis of Incomplete Multivariate Data. Chapman & Hall, Boca Raton (1997)
Schafer, J.L., Olsen, M.K.: Modeling and imputation of semicontinuous survey variables. Technical Report, 00-39, The Pennsylvania State University (1999)
Schafer, J.L., Graham, J.W.: Missing data: our view of the state of the art. Psychol. Methods **7**, 147–177 (2002)
Stata: Stata Multiple-Imputation Reference Manual. Stata Press, College Station (2011)

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/Plenum Publishers, New York (2003)

Tracy, P.E., Wolfgang, M.E., Figlio, R.M.: Delinquency in Two Birth Cohorts. Plenum, New York (1990)

Van Buuren, S.: Multiple imputation of discrete and continuous data by fully conditional specification. Stat. Methods Med. Res. **16**, 219–242 (2007)

Wittenberg, J.: Methodendokumentation der kriminologischen Schülerbefragung in Münster 2003 (Schriftenreihe "'Jugendkriminalität in der modernen Stadt – Methoden"' Nr. 7). Münster, Trier (2004)