An Analysis of Non-response and Attrition in the Zurich Project on Social Development from Childhood to Adulthood (z-proso)

Nora Eisner[1,2], Aja L. Murray[3], Manuel Eisner[3], Denis Ribeaud[4]

Submitted to **International Journal of Behavioral Development**

30 March 2018

Manuscript: 27 pages text, 5 tables, 1 figure

[1] Contacting author: nora.eisner@hotmail.com

[2] Department of Physics and Astronomy, University of Sheffield, Sheffield, UK

[3] Institute of Criminology, University of Cambridge, Cambridge, UK

[4] Jacobs Center for Productive Youth Development, University of Zurich, CH

ABSTRACT

Selective non-participation and attrition pose a ubiquitous threat to the validity of inferences drawn from observational longitudinal studies. We investigate various potential predictors for non-response and attrition of parents as well as young persons at different stages of a multi-informant study. Various phases of renewed consent from parents and young persons allowed for a unique comparison of factors that drive participation. The target sample consisted of 1675 children entering primary school at age seven in 2004. Seven waves of interviews, over the course of ten years, measured levels of problem behaviour as rated by children, parents and teachers. In the initial study recruitment, where participation was driven by parental consent, non-response was highest amongst certain socially disadvantaged immigrant minority groups. There were fewer significant group differences at wave 5, when young people could be directly recruited into the study. Similarly, attrition was higher for some immigrant background groups. Methodological implications for future analyses are discussed.

Keywords: Criminology, Attrition, Non-response, Study-Participation, Longitudinal Study

Longitudinal studies in normative samples are critical for making inferences about the developmental processes underpinning child and adolescent development. Non-random participation and attrition, whereby presence in a research sample at a given wave is directly or indirectly related to the characteristics under study, represent important challenges for such studies (e.g. Singer & Willet, 2003). Knowledge of the nature and severity of selective participation and drop-out is fundamental for applying appropriate corrections to data analyses and can inform strategies to ameliorate selective participation and drop-out in future studies and/or measurement waves. In this study, we therefore sought to identify key parent and child characteristics that predict non-participation at different stages in a 10-year multi-rater longitudinal study of youth development.

Child participation in longitudinal studies typically relies on parental consent until the child is of an age where they are legally able to provide informed consent. It is also common for parents to directly participate in the research as informants on their child's behaviour. As such, characteristics of both the parent and child can influence participation and drop-out in studies of child and adolescent development. These characteristics are often related to those under study (e.g. Audrain et al., 2002; Asendorpf et al., 2014; Noll et al., 1997; Ullebo et al., 2012), resulting in the effects due to non-random participation and non-random attrition undermining the representativeness of study samples. Moreover, non-random drop-out can introduce spurious developmental effects or mask genuine ones. Brame & Piquero (2003), for example, note that in longitudinal studies of delinquent behaviour, those with the highest levels of delinquency tend to be more likely to drop-out. In such cases, estimates of normative changes in delinquent behaviour over development will be negatively biased because of the selective loss of higher scoring participants.

To draw valid conclusions about developmental processes from longitudinal data, relations between participation and study outcomes need to be taken into account. Characterising non-random attrition has important implications for the interpretation of statistical results from an affected dataset, as well as for the application of appropriate mitigation strategies. Examples of such strategies include selection models, range restriction corrections, data weighting, maximum-likelihood, multiple imputation, pattern mixture models and random coefficient models (e.g. Asparouhov, 2005; Enders, 2011; Sackett & Yang, 2000; Schafer & Graham, 2002). It is important to note that each method comes with a different set of assumptions about the missing data mechanism. For example, maximum-likelihood estimation for missing data yields unbiased parameter estimates only if non-participation can be described as 'missing at random', that is, participation is related to the observed but not the unobserved data (Rubin, 1976). When participation is related to unobserved values over and above its relation to observed values this is known as 'missing not at random' (MNAR). Here, methods such as pattern mixture modelling or random coefficient modelling may be more appropriate; however, their utility depends on how closely correlated drop-out and the variables of interest are (e.g. Schafer & Graham, 2002). Even though in most cases there is insufficient information about those who are missing to identify participation mechanisms empirically, analysing patterns of drop-out can help researchers develop plausible hypotheses about these mechanisms. These kinds of analyses also have the potential to inform future study designs by providing a forecast of the profiles of individuals who may be most likely to drop-out. In future waves or studies, special strategies may need to be developed and/or additional resources channelled to their recruitment and retention (e.g. Eisner & Ribeaud, 2007). In this study, we therefore provide an analysis of participant factors predicting

non-participation and attrition in seven waves of a longitudinal study containing two intervention arms: the Zurich Project on Social Development from Childhood to Adulthood (z-proso).

Method

z-proso

Z-proso is an ongoing multi-rater longitudinal study of child and adolescent development with a particular focus on the development of crime and aggression. The study began in 2004 when the participants were entering their first year of school, aged 7. Parents provided data at four waves of interviews, when the children were aged 7, 8, 9, and 10 (labelled 'waves P1 to P4') and children provided self-report data when they were aged 7, 8, 9 ,10, 11, 13, 15 and 17 (labelled 'waves Y1 to Y7'). Additional information was provided by teachers when the children were aged 7, 8, 9, 10, 11, 13 and 15 (labelled 'waves T1 to T6').

Sample

At baseline, a stratified random sampling approach was used to define the target sample with schools as the randomisation units and stratification by school size and socioeconomic background. The target sample comprised all children entering first grade across 56 primary schools in the city of Zurich, Switzerland, corresponding to a total of 1,675 children. Lower socioeconomic neighbourhoods were slightly overrepresented. In 2004, when the study started, Zurich had a population of about 365,000, with a large proportion of immigrant-background residents. Zurich is an affluent city. The average GDP per capita was about USD 106,000 in 2004, and the unemployment rate was about 4%. Broadly representative of city demographics, the baseline target sample consisted of 39.3% German-speaking (mostly Swiss or German) primary caregivers. Over 60% of

primary caregiver were not native German speaking. The mean age of the target children at entry into primary school was 6.85 years, and the average number of siblings was 1.15.

Recruitment

Considerable efforts were employed in order to maximise participation of the baseline target sample, with a strong focus on recruiting caregivers with an immigrant background who may be less likely to agree to participate in research studies. Recruitment procedures are described in detail in Eisner & Ribeaud (2007). In brief, contact letters were written in the ten most commonly spoken languages with native speakers of these languages taking on the role of recruiting and interviewing participants. Monetary incentives, translated support letters from school authorities, and the inclusion of community stakeholders were also used to maximise participation. Bilingual information packs with study information and consent forms (available in German, Albanian, Portuguese, Serbian/Bosnian, Spanish, Tamil, Turkish, English, Croatian and Italian) were sent to all non-German-speaking primary caregivers. Prospective participants who did not respond to the initial information pack were contacted by phone. No upper limit on the number of trial calls to be made was imposed, and in some cases more than 20 attempts were necessary before contact was possible. Parents who could not be reached by phone were visited at home by a male and a female interviewer who explained the study in more detail. To further encourage participation, shopping vouchers worth 20 CHF were offered to parents for their participation. At the beginning of the interviews, parents were asked to sign an informed consent form for the participation of their child as well as the participation of the child's teacher. Non-respondents were re-contacted and asked to consent to their child's and his/her teacher's participation only. Recruitment at

waves 2 and 3 followed similar procedures with the informed consent obtained from parents at wave 1 covering the entire period from wave 1 to 3. Renewed consent, provided by the parents, was required at wave 4.

Several changes were made to the recruitment and assessment procedures in wave 5. First, parent interviews were no longer carried out. Second, at age 13, youth were able to actively consent to their own participation (Art. 16 of the Swiss Civil Code); however, parents were still provided the opportunity to opt their child out. Third, unlike in previous waves, the entire initial target sample defined at baseline could be re-contacted. Fourth, youth interviews could no longer be carried out during regular school hours and thus questionnaires were administered in classrooms outside of regular lesson times. In order to maximise participation, participants received monetary incentives worth 30 CHF and 50 CHF in waves 5 and 6 respectively. Wave 6 required renewed active youth consent. At this stage, the entire initial target sample could be re-contacted once again and the monetary incentive increased to 60 CHF. The same recruitment procedure as described above and preceding wave 5 was followed.

Given that the times of renewal of consent represented key attrition points, four key outcomes can be defined with respect to non-random participation and attrition: 1) baseline participation 2) participation in wave 5 3) drop-out in the wave 1 to 3 period 4) drop-out in the wave 5 to 7 period. We analysed parent and child participation separately given the previous evidence that patterns of participation may differ across these groups (Asendorpf et al., 2014).

Measures and Statistical Procedure

We evaluated a range of predictors of attrition reflecting the core theoretical themes of z-proso as well as additional possible risk factors for non-participation that could

159 inform recruitment and sampling design in future studies. For those who did not
160 participate at all, the only information available was on gender, primary caregiver
161 language, neighbourhood social class and neighbourhood familialism.

162 **Small (special needs) Class.** The school system in the city of Zurich differentiates
163 between regular and small classes. Small classes are intended to meet the specific needs
164 of children with difficulties such as developmental delays, behavioural problems,
165 language barriers, and/or learning difficulties. Typically, small classes have a size of ten
166 or less children, compared to around 20 in regular classes. The target sample comprised
167 9.1% of children attending a small class in year 1 of primary school.

168 **Mother Tongue of Primary Caregiver.** At the beginning of the study the City of
169 Zurich's School Department provided the z-proso team with a contact database of all the
170 study participants and their parents. The most reliable proxy for the cultural background
171 of the primary caregiver was the mother tongue as, unlike nationality, this remains
172 unaffected by naturalization. Nonetheless, the mother tongue does not differentiate
173 between, for example, caregivers of German or Swiss or of Portuguese or Brazilian
174 background, and can thus be ambiguous to some extent. In the present analyses we
175 distinguish nine groups, namely German or Swiss German (39.3%),
176 Serbian/Bosnian/Croatian (10%), Albanian (9%), Portuguese (7%), Tamil (5.3%), Italian
177 (5.3%), Spanish (5.1%), Turkish (4.5%), and 'other' languages (14.5%).

178 **Neighbourhood Social Class and Familialism.** Neighbourhood characteristics
179 were derived from census and other data that was systematically collected by the City of
180 Zurich's Office of Statistics. These data are aggregated at the level of the 212 statistical
181 zones. Neighbourhood social class was obtained as the factor score of three indicators,
182 namely the unemployment rate (2002), the (inverted) percentage of self-owned

households (2000), and the percentage of foreign nationals (2003). Similarly, mean household size (2000), the percentage of residents aged below 20 years (2002), and residential stability (measured as the percentage of households living in the same statistical zone as five years earlier) were used as the three indicators in order to obtain factor scores for neighbourhood familialism. The residential address of each study participant was then assigned to its corresponding statistical zone and its related factor scores.

Parent Education Level and Family Composition. Parental education was measured using a dichotomous variable indexing whether the primary caregiver possessed a university education or not. Single parent status was measured using a dichotomous item measuring whether the second parent (usually the father) was living in the same household as the primary caregiver. Based on this, 29.4% of households were defined as single parent households.

Child Behaviour. Child behaviour was measured using the Social Behaviour Questionnaire (SBQ; Tremblay et al., 1991) which captures prosocial behaviour, aggression, anxiety, depression, and ADHD. The z-proso study team developed a multi-informant version with matched items across a parent, a child and a teacher version. The parent questionnaire contained all questions (55 items) and was administered in computer assisted personal interviewing (CAPI) home interviews, offered in ten different languages. Responses were given on a 5-point Likert scale (‘never’ to ‘very often’).

Teachers were surveyed by mail. The teacher version used essentially the same items and answer format as the parent version, with some adaptations to the school context and some items from the full 55-item version omitted.

For children, a specially adapted self-administered multimedia version with 54 items was used in wave 1. Based on the concept of the "Dominique interactif" (Valla et al., 2000), each behaviour was depicted in a drawing representing either a boy or a girl, matched with the gender of the subject. A voice recorded on the laptop read out each item that was worded in an age-adequate language. The child could then answer the question by pushing a "yes" or a "no" button on the screen.

The SBQ distinguishes five major domains of children's social behaviour and scores were computed as averages of all items. Prosocial behaviour (e.g. 'shows sympathy to someone who has made a mistake') was measured with ten items in the parent (α = .77) version, seven items in the teacher version (α = .92) and ten items in the child version (α = .59). Symptoms of Anxiety/Depression (e.g. 'is too fearful or anxious', 'has trouble enjoying him/herself') were measured by nine items in the parent (α = .71) version, by seven items in the teacher version (α = .90) and by nine items in the child version (α = .62). Symptoms of attention deficit and hyperactivity measured by nine items in the parent (α = .79) version, by 8 items in the teacher version (α = .94) and by eight items in the child version (α = .58). The scale for non-aggressive conduct problems (e.g. 'steals at home', 'tells lies and cheats') comprised nine items in the parent (α = .68) version, six items in the teacher version (α = .81) and nine items in the child version (α = .60). Aggressive behaviour was measured by 12 items in the parent (α =.79) version, by 11 items in the teacher version (α = .93) and by 12 items in the child version (α =.72). In terms of item-level missingness, prior to the receipt of the dataset, within each set of SBQ variables (e.g. all parent-reported items within a wave; all self-reported items within a wave etc.), missing item scores had been singly imputed using an expectation-maximisation algorithm. The correlations between parent reports, youth self-reports, and

teacher reports are provided in Supplementary Table S1. These were modest at best ranging from .04 (teacher- and parent- reports of internalising) up to .33 (teacher- and parent- reports of ADHD).

Statistical Procedure

For descriptive purposes, simple logistic regressions with list-wise deletion were used to evaluate the relations between predictors and non-participation or drop-out without considering the effects of other predictors. Participation/retention was coded=0 and non-participation/drop-out was coded=1 such that odds ratios >1 reflect increased likelihood of non-participation or drop-out. Specifically, the odds ratios reflect the ratio of the odds of non-participation/drop-out at levels of the predictor separated by one unit. For example, an odds ratio of 2 would indicate that the odds of dropping out double for each unit increase in the predictor. Associated (unadjusted) *p*-values are reported for descriptive purposes. These analyses were conducted in R statistical software, using a logit link function in the glm function (R Core Team, 2016).

We then conducted a series of multiple regressions to evaluate the unique relations between each predictor and drop-out/attrition controlling for other predictors. These analyses were implemented in lavaan, again in R Statistical Software (Rosseel, 2012), this time using probit regression. Probit and logistic regression can both be used to model the prediction of dichotomous outcomes and generally result in the same conclusions. While logistic regression uses a logit function to model the probability that the outcome variable is equal to 1, probit regression uses an inverse standard normal cumulative distribution function. Probit regression coefficients can thus be interpreted as the difference in the cumulative normal probability of the outcome variable for a unit increase in the predictor. Here, probit regression was used for practical reasons, namely that it (but

not logistic regression) can currently be combined with full information maximum likelihood (FIML) estimation to account for missingness in lavaan. FIML provides unbiased parameter estimates provided data are missing at random (MAR). Predictors were entered using simultaneous entry.

To correct for multiple comparisons, we used the generalised Holm (1979) k-familywise error rate (FWER; Lehmano & Romano, 2005) method discussed by Keselman et al. (2011). This method was selected because it is less conservative than traditional FWER corrections which entail a substantial loss of statistical power when families of statistical tests are large. While it is important to control type 1 errors, the current study is somewhat exploratory in nature, and it was judged more problematic to fail to identify predictors that were associated with attrition than to falsely conclude that a predictor was related to attrition when it is not. This viewpoint, in turn, comes from the importance of recognising when and in what way attrition is non-random in order to guard against biases deriving from falsely assuming that attrition is random.

In the Holm k-FWER method, k-FWER is defined as the probability of rejecting at least k hypotheses H_i where i is a member of the set of true null hypotheses. Thus when k is one, this reduces to the traditional FWER correction i.e. that which controls the probability of rejecting at least one true null hypothesis. 2-FWER controls the probability of rejecting 2 or more true null hypotheses (implicitly tolerating one false positive), 3-FWER, the probability of rejecting 3 or more true null hypotheses and so on and so forth. Keselman et al. (2011) recommend selecting a value for k that is the nearest integer to $m\alpha$ where m is the number of tests conducted and α is the significance level. In total, 129 tests were conducted, giving a k value of 6. This was judged an acceptable level given the goals of the study.

The method involves considering the raw *p*-values for all the hypotheses, ordered from smallest to largest and making a sequential adjustment to the alpha value for each or, equivalently, to the *p*-value itself. We here report adjusted *p*-values rather than adjusted critical values on the assumption that the former will be more informative for most readers. Both unadjusted and adjusted *p*-values are reported in Table 2 for information; however, all inferences were made on the basis of adjusted *p*-values.

Results

Descriptive Statistics

Figure 1 shows the number of participating primary caregivers (subscript P) and youth (subscript Y) in each of the main data-collection waves. In wave 1, 1239 primary caregivers (74% of baseline target sample) participated and a further 121 parents provided consent for participation of the child only, meaning that 1360 young persons participated in wave 1. Between waves 1 and 2, attrition was low with 4.5% and 2.0% of primary caregivers and youth dropping out respectively.

The bold arrows in *Figure 1* show the pool of participants that could be re-contacted in each wave of the study, whilst the thin arrows highlight the number of subjects lost due to attrition between consecutive waves. In wave 4, for example, all participants that consented to participate in wave 1 could be re-contacted, resulting in a parent participation of N_P= 1075 (64% of baseline target sample) and young person participation of N_Y = 1147 (69%). The number of parents and youth lost due to attrition between waves 3 and 4 was N_P=116 (9.8% of the parents in Wave 3) and N_Y= 184 (14%). Unlike in the initial recruitment into wave 1, where 121 youth participated even though their parents did not, the number of 'youth-only' cases were small in wave 4. The total number of parents and youth who re-entered into wave 4 were N_P= 11 and N_Y = 10 respectively.

These participants missed one or more waves of data collection but subsequently continued to participate. This is depicted by dashed lines in figure 1. The lowest participation rate occurred following the request for renewed parental consent in wave 4. The highest number of participation followed in wave 6, where the whole initial target sample was re-contacted, with N_Y = 1446 (86% of baseline target sample). Finally, after wave 7, one participant requested to be withdrawn from the study and have their data removed from the database. This participant is represented in subsequent analyses as if they did not participate at baseline and remained a non-participant thereafter. Omitting this participant, N=1570, or 94% of the target sample participated in at least one of the seven waves.

Means/frequencies for each predictor of attrition, broken down by attrition status, are provided in Tables 1-3. These tables also provide simple logistic regression analyses to test the unadjusted effects of each predictor prior to controlling for other predictors. The sample sizes in the table also illustrate the level of variable-wise missingness at each wave.

Non-participation

Results of the multiple probit regression models assessing predictors of parent and youth non-participation in wave 1 (average age 7) and wave 5 (average age 13) are provided in Table 4. After correction for multiple comparisons, parent non-participation in wave 1 was significantly predicted by neighbourhood social class (b=-0.05) and several non-German first languages (ranging in effect from b=0.18 for Portuguese to b=0.30 for Tamil and Albanian). Youth non-participation in wave 1 was significantly predicted by the same non-German caregiver first language categories, except Tamil (ranging in effect from b=0.17 for Portuguese, Serbian-Croatian and Albanian up to b= 0.27 for Turkish).

The only predictor with a significant unique effect on youth participation at wave 5 was membership in a small class, which was associated with a greater probability of non-participation (b=0.12).

Results of the multiple probit regression models assessing predictors of dropping out are provided in Table 5. Although there were many significant effects on drop-out when considering bivariate models, there were few predictors that had significant unique effects on attrition in the multiple regression models after correcting for multiple comparisons. Being in the 'Other' language category significantly predicted both parent (b=0.13) and youth (b= 0.10) drop-out between waves 1 and 4, while being in the 'Serbian-Croatian' category predicted youth drop-out between waves 5 and 7 (b=0.10). As there were no strong a priori theoretical rationale for expecting higher-order effects, including interactions between predictors, these were not considered for any of the models.

Discussion

Longitudinal studies inevitably suffer from survey non-response and attrition. This not only results in a smaller dataset, thus reducing the power of the study, but also has the potential of introducing bias. We evaluated a number of different predictors of non-response and attrition, including characteristics of parents and children as rated by parents, children and teachers.

We found that child and parent non-participation and drop-out were more likely among children with primary caregivers who spoke languages other than the official regional language. This was in spite of specific and intensive efforts to encourage participation among non-German speaking, immigrant background parents (e.g. Eisner & Ribeaud, 2007). With the exception of Italian-speaking respondents, over 90% of non-German speaking households had at least one first generation immigrant parent. Thus,

non-native speaking primary caregivers is likely to be a proxy for belonging to an immigrant minority. In this respect, our results are consistent with those from other European survey-based studies that suggest that immigrant minorities are generally more difficult to contact and are more likely to decline to participate (Couper & Leeuw, 2003; van Goor et al., 2005; Kapteyn et al., 2006).

Even though we were unable to collect data on reasons for non-participation among these individuals, it is possible to speculate as to its cause. Some participants may have felt intimidated by the prospect of participating in an interview in a foreign country. Furthermore, sociocultural minorities are often in a more vulnerable position. Amongst other factors, their legal residence status may still be undecided, they may have had adverse experiences with immigration authorities or they may have faced persecution in their home countries, which could generalise to a lack of trust in research studies. This may be exacerbated by study topics that could be considered sensitive. Third, immigrant minority caregivers have, on average, a more limited educational background and lower socio-economic status (Eisner et al., 2007). Multiple studies have shown that educational level is an important predictor of participation in research (e.g. Stoop, 2005; Van Loon et al., 2003; Korkeila et al., 2001; Curtin, Presser & Singer, 2000). However, in this study whether or not caregivers had a university education was not a significant predictor of parent or youth drop-out over the early waves. However, we did not have parental education data for parents who did not participate in the study at all, meaning that we could only indirectly test whether this was associated with participation at wave 1 based on FIML estimates.

Finally, cultural factors may have played a role. Whilst in many countries it is common custom to participate in surveys and answer personal questions, in other cultures

giving personal information to a stranger is counter-normative (e.g., Johnson et al., 2002). Though we have no direct evidence that this was the case in z-proso, this could be one aspect to explore in future research. Overall, our results suggest that translation of invitation letters and additional efforts to contact immigrant background participants may be insufficient to mitigate the tendency to decline to participate and to drop out. Differing cultural attitudes and sources of apprehension must also be addressed as part of recruitment strategies. That said, there is some indirect evidence that the additional efforts invested in recruiting individuals from an immigrant minority status helped mitigate under-representation of these individuals in the sample. For example, at baseline, a higher proportion of those from an immigrant minority background were recruited via more 'active' methods (active telephone recruitment) than more 'passive' methods (participants answering an invitation by response slip; Eisner & Ribeaud, 2007). Specifically, 28% of the target sample from non-German speaking minorities were recruited via the return of a response slip, while 33.7% were recruited by telephone. In contrast, among the target sample of German-speakers 63.8% were recruited by reply slip and 24% via telephone contact. As home visit recruitments were extremely rare, they were collapsed with the telephone recruitment category. This suggests that investing additional efforts in recruiting immigrant minorities can improve response rates and that, in fact, these additional efforts may produce greater returns in immigrant minority groups than in non-minority groups.

Although they had no unique significant effect on non-participation, itt is also worth highlighting the bivariate association between child behaviour and non-participation as an area for potential future investigation. Based on bivariate analyses and according and teacher reports of child behaviour, child aggression, non-aggressive

conduct problems, ADHD, internalising and prosociality predicted parental and youth drop-out during the wave 1 to 4 phase. Specifically, both parents and youth were more likely to drop-out if the child showed higher levels of psychopathology, whereas parents reported a greater likelihood of drop-out where their child showed lower levels of psychopathology. These results are consistent with past research, which suggests that children exhibiting disruptive behaviour often show difficulties functioning within the school context (e.g. Barry et al., 2002). Moreover, in terms of participation in research studies, by definition children with high levels of ADHD symptomology may avoid and find it very difficult to engage in tasks that include sustained attention (APA, 2013). This will almost certainly include completing study measures. Further, it has been suggested that parents of children with conduct problems may have a tendency to 'disengage' from the child and their problem behaviour; an effect which would presumably carry over into participation in studies (e.g. Patterson, 1982). As such, studies seeking to retain the most disruptive children must take into account possible difficulties such as the child having a poor bond or negative associations with school context, difficulties completing measures and parents who are avoidant of or psychologically disengaged from their child's behaviour. In these cases, measures such as offering assessments outside of the school setting and breaking up the assessment into multiple sessions may help mitigate the loss of participants with ADHD and/or disruptive behaviours.

Notably, despite positive correlations between parent and teacher reports on all behaviours, parent reports of child behaviour suggested exactly the opposite pattern to that suggested by teachers. Parent-reports suggested that attrition was more likely when their child was well-adjusted. While it is common for different raters to disagree on child behaviour (e.g. De Los Reyes, 2011), that these two raters suggested completely

422 contradictory results is striking. One possibility is that children show context-specific
423 negative behaviour and that negative behaviour in the school environment is particularly
424 undermining of tendencies to engage with research projects conducted in the school
425 setting. Another possibility, however, is that socially desirable responding on the part of
426 parents with children displaying disruptive behaviours accounts for the counterintuitive
427 association between parent-reported child behaviour and study participation (e.g. Johnson
428 et al., 2012; Mundia, 2011; Eisner & Ribeaud, 2007). However, it is also important to
429 emphasise that in the multiple regressions, neither teacher nor parent-reported child
430 behaviour was uniquely associated was attrition. Thus, any effects of behaviour cannot
431 be easily disentangled from one another nor from the effects of other factors that may
432 affect non-participation, such as immigrant minority status discussed above.

433 Our result suggested that the dynamics of attrition and non-response changed once
434 children/youths could provide their own active consent, highlighting the need for different
435 retention methods based on the target sample. We found that being able to re-contact the
436 entire initial target sample going into waves 5 and 6 greatly improved our overall
437 participation, resulting in a highest young person participation rate occurring in wave 6.
438 We note that re-contacting entire target samples is not always a viable option, and
439 regulations may vary depending on location, nonetheless, we emphasise the positive
440 outcomes this can have on participation rates. Furthermore, it is important to identify
441 more vulnerable samples in the early stages of the study, so that long term retention plans
442 can be implemented in order to retain or regain participants.

443 Collectively, our results suggested that participation and attrition in z-proso are
444 related to characteristics of the children under study. As such, our data would not qualify
445 as missing completely at random (MCAR) rendering analysis methods involving listwise

or pairwise deletion inappropriate. Under the assumption of MAR, methods such as maximum likelihood estimation or multiple imputation could likely address parameter bias due to non-random attrition. The question of whether the data qualifies as MAR is more difficult to answer because this would require knowledge of unobserved data. However, given the comprehensiveness of the list of measures obtained in z-proso (see Ribeaud & Eisner, 2010) and the relatively low rates of attrition observed overall, we would argue that, provided a sufficient set of auxiliary variables is used, any additional attrition following a NMAR mechanism may have only a small effect. We, therefore, believe that users of z-proso dataset could be justified in employing methods such as multiple imputation or maximum likelihood estimation drawing on a range of auxiliary variables for testing developmental hypotheses. However, it is always best practice to use a range of methods, especially those that make different assumptions about mechanisms of missingness (e.g. pattern mixture modelling versus FIML estimation) to evaluate the sensitivity of results to the chosen method of dealing with missing data. In addition, our goal here was not to provide an exhaustive analysis of all the factors that may predict attrition for the purposes of aiming to achieve MAR but to focus on a smaller number of theoretically motivated predictors that may help researchers to understand why certain individuals may be more likely to drop-out than others. This in turn can help inform recruitment and retention in future studies and future waves of existing longitudinal studies. Thus, we would advise users of the dataset to consider an even broader range of predictors as auxiliary variables, as well as their non-linear effects and interactions.

Finally, it is important to consider the limitations of the current study. First, we did not ask participants to report on their reasons for non-response. This would have allowed us to understand proximal causes of non-participation. Second, we had only limited

information on those who declined to participate at baseline. Third, we included a limited number of predictors of attrition, focussing specifically on those that represented the core themes of z-proso and/or which were predicted to show a relation to attrition based on past research. Future research will be valuable in identifying other variables that may be related to selective non-participation.

Conclusion

Studying patterns of attrition has the potential to inform future study design, as well as provide information pertinent to the interpretation or correction of statistical effects in studies affected by non-random participation. In this study, we evaluated the characteristics associated with non-participation and attrition in a 10-year longitudinal study of child development. The design of the study allowed for a unique comparison of factors associated with non-participation of parents and youths as well as attrition due to various phases of renewed consent. This is because during the initial waves of the study the decisions on study participation were made by the primary caregiver of the target subject, whereas in later waves the youths themselves were able to provide informed consent. We found several predictors of non-participation and attrition, including characteristics of both the parent and child.

Our results showed that non-response and attrition was highest among certain immigrant background, non-native speaking parents, despite the significant efforts made to recruit these subjects. This result is in agreement with previous findings (e.g. Couper & Leeuw, 2003) and could be due to a number of reasons, reviewed above.

Child behaviour was also found to be a significant predictor of attrition in the bi-variate analyses. Between waves 1 and 4, aggression, non-aggressive conduct problems, ADHD, internalising and prosociality, as rated by teachers were all predictors of child

494 drop-out. This was in direct disagreement with the reporting of parents, which suggested
495 that children with lower levels of problem behaviours were less likely to participate. A
496 possible explanation for these contradictory results relates to social desirability bias,
497 where parents of children with disruptive behaviour are more likely to project a good
498 image of themselves by responding in the 'most desirable' way. A further explanation
499 could be context specific behaviour of the child, resulting in a disagreement between
500 parents and teachers.

501 **Acknowledgements**

502 We are grateful to the children, parents and teachers who provided data for the z-
503 proso study and the research assistants involved in its collection. Funding from the Jacobs
504 Foundation (Grant 2010-888) and the Swiss National Science Foundation (Grants
505 100013_116829 & 100014_132124) is also gratefully acknowledged.

References

American Psychiatric Association, *APA* (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Arlington, VA: APA.

Audrain, J., Tercyak, K.P., Goldman, P., & Bush, A. (2002). Recruiting adolescents into genetic studies of smoking behavior. *Cancer Epidemiology, Biomarkers & Prevention, 11*(3)*,* 249-252.

Asendorpf, J. B., Van De Schoot, R., Denissen, J. J., & Hutteman, R. (2014). Reducing bias due to systematic attrition in longitudinal studies: The benefits of multiple imputation. *International Journal of Behavioral Development*, *38*(5), 453-460.

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling, 12*(3), 411–434.

Barry, T. D., Lyman, R. D., & Klinger, L. G. (2002). Academic underachievement and attention-deficit/hyperactivity disorder: The negative impact of symptom severity on school performance. *Journal of School Psychology*, *40*(3), 259-283.

Brame, R., & Piquero, A. R. (2003). Selective attrition and the age-crime relationship. *Journal of Quantitative Criminology*, *19*(2), 107-127.

Couper, M. P., & Leeuw, E. D. d. (2003). Nonresponse in cross-cultural and cross national surveys. In J.A. Harkness, F. J. R. van de Vijver and P. P. Mohler (eds.) *Cross-Cultural Survey Methods*, 157–178. New York: Wiley.

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64*(4), 413-428.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *40*(1), 1–9.

Eisner, M., & and Denis, R. (2007). Conducting a criminological survey in a culturally diverse context: Lessons from the Zurich Project on the Social Development of Children. *European Journal of Criminology, 4*(3), 271-298.

Enders, C. K. (2011). Analyzing longitudinal data with missing values. *Rehabilitation Psychology*, *56*(4), 267-288.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65-70.

Johnson, T. P., Fenrich, M., & Mackesy-Amiti M. E. (2012). An evaluation of the validity of the Crowne–Marlowe need for approval scale. *Quality & Quantity, 46*(6), 1883-1986.

Johnson, T. P., O'Rourke, D., Burris, J., & Owens, L. (2002). Culture and survey nonresponse. In R. M. Groves, D. A. Dillman, J. L. Eltinge and R. J. A. Little (eds) *Survey Nonresponse*, 55–69. New York: Wiley.

Kapteyn, A., Michaud, P-C., Smith, J., & van Soest, A. (2006). *Effects of attrition and non-response in the Health and Retirement Study*. Discussion paper from the Institute for the Study of Labor (IZA), Bonn, Germany.

Keselman, H. J., Miller, C. W., & Holland, B. (2011). Many tests of significance: New methods for controlling type I errors. *Psychological Methods*, *16*(4), 420-431.

Korkeila, K., Suominen, S., Ahvenainen, J., Ojanlatva, A., Rautava, P., Helenius, H., & Koskenvuo, M. (2001). Non-response and related factors in a nation-wide health survey. *European Journal of Epidemiology*, *17*(11), 991-999.

Lehmann, E. L., & Romano, J. P. (2012). Generalizations of the familywise error rate. In *Selected Works of EL Lehmann*(pp. 719-735). Springer, Boston, MA.

Mundia, L. (2011). Social desirability, non-response bias and reliability in a long self-report measure: illustrations from the MMPI-2 administered to Brunei student teachers. *Educational Psychology*, *31*(2), 207-224.

Noll, B., Zeller, M.H., Vannatta, K., Bukowski, W.M., & Davies, W.H. (1997). Potential Bias in classroom research: Comparison of children with permission and those who do not receive permission to participate. *Journal of Clinical Child Psychology, 26*(1), 36-42.

Patterson, G. R. (1982). *Coercive family process: A social learning approach* (Vol. 3). Eugene, OR: Castalia.

R Core Team (2016). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Ribeaud, D., & Eisner, M. (2010). Risk factors for aggression in pre-adolescence: Risk domains, cumulative risk and gender differences-Results from a prospective longitudinal study in a multi-ethnic urban sample. *European Journal of Criminology*, *7*(6), 460-498.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. URL http://www.jstatsoft.org/v48/i02/.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, *63*(3), 581–592.

Sackett, P. R., & Yang, H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology*, *85*(1), 112–118.

Schafer, J.L. & Graham, J.W (2002). Missing data: our view of the state of the art. *Psychological Methods, 7*(2), 147–177.

Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.

Stoop, I. A. L. (2005). *The hunt for the last respondent. Nonresponse in sample surveys*. The Hague: Social and Cultural Planning Office of the Netherlands.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*(3), 285-300.

Ullebø, A. K., Posserud, M. B., Heiervang, E., Obel, C., & Gillberg, C. (2012). Prevalence of the ADHD phenotype in 7-to 9-year-old children: effects of informant, gender and non-participation. *Social Psychiatry and Psychiatric Epidemiology, 47*(5), 763–769.

Valla, J.P., L. Bergeron, M. St-Georges et C. Berthiaume (2000). Le Dominique Interactif: présentation, cadre conceptuel, propriétés psychométriques, limites, et utilisations. *Revue Canadienne de Psycho-éducation*, *29*(2), 327-347.

Van Goor, H., Jamsma, F. and Veenstra, R. (2005). Differences in undercoverage and nonresponse between city neighbourhoods in a telephone survey. *Psychological Reports, 96*(3), 867–878.

Van Loon, A.J.M., Tijhuis, M., Picavet., H.S.J., Surtees P.G. and Ormel, J. (2003) Survey non-response in the Netherlands: Effects on prevalence estimates and associations. *Annals of Epidemiology*, *13*(2), 105–110.

Figures

Figure 1: Flow of subjects across seven waves of interviews. Highlighting the rate of attrition between consecutive waves as well as the number of subjects re-entering the study from previous waves.

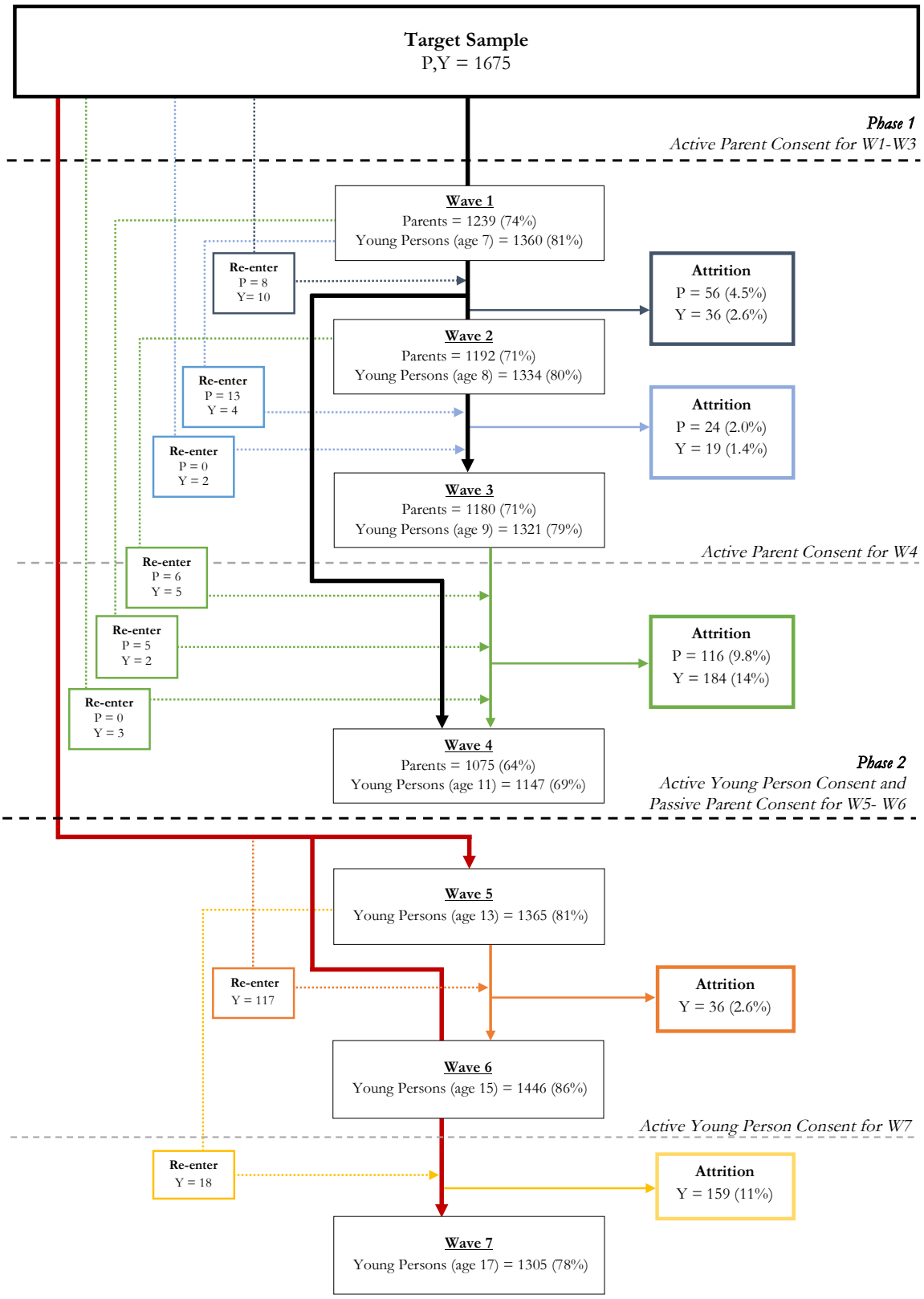

Table 1: Means/frequencies of participant background predictors and bivariate associations (ORs) with baseline response status

Predictor	Non-Response P1				Non-Response Y1				Non-Response Y5			
	Participants	Non-participan	OR	*p*	Participants	Non-participant	OR	*p*	Participants	Non-participant	OR	*p*
Class size												
Normal	1139	378	1.70	<.01*	1244	273	1.60	.02*	1254	263	1.96	<.001
Small	101	57			117	41			112	46		
Gender												
Male	643	226	1.01	.92	696	174	0.84	.17	703	167	0.90	.41
Female	596	209			665	140			663	142		
Caregiver higher education	-	-	-	-	-	-	-	-	-	-	-	-
Neighbourhood Social Class	-0.34	-0.67	0.65	<.01*	-0.37	-0.67	0.67	<.001*	-0.42	-0.47	0.93	.33
Neighbourhood Familialism	0.54	0.51	0.96	.51	0.53	0.56	1.03	.61	0.52	0.60	1.08	.20
Language												
Albanian	80	70	5.07	<.001	105	45	3.55	<.001	120	30	1.32	.23
Other	168	39	1.34	.16	184	23	1.04	.89	158	49	1.64	.01*
German[a]	562	97	-	-	588	71	-	-	554	105	-	-
English	23	3	0.76	.65	23	3	1.08	.90	22	4	0.96	.94
Italian	66	22	1.93	.01*	73	15	1.70	.09	69	19	1.45	.18
Portuguese	78	40	2.97	<.001	84	34	3.35	<.001	100	18	0.95	.85
Serbian-Croatian	105	62	3.42	<.001	117	50	3.54	<.001	138	29	1.11	.65
Spanish	63	22	2.02	.01*	70	15	1.77	.07	59	26	2.33	<.001
Turkish	44	32	4.21	<.001	46	30	5.40	<.001	61	15	1.30	.40
Tamil	48	41	4.95	<.001	67	22	2.72	<.001	77	12	0.82	.55

Note. [a]Reference category. *Significant at alpha<.05. N = 1665 – 1675.

Table 2: Means/frequencies of participant background predictors and bivariate associations (ORs) with attrition status

Predictor	Attrition P1 – P4 (N = 1240)				Attrition Y1 – Y4 (N = 1217 - 1361)				Attrition Y5 – Y7 (N = 1054 - 1366)			
	Retained	Dropped	OR	*p*	Retained	Dropped	OR	*p*	Retained	Dropped	OR	*p*
Class Size												
Normal	982	157	1.09	.77	1046	198	1.66	.03*	1119	135	1.38	.26
Small	86	15			89	28			96	16		
Gender												
Male	556	89	1.01	.94	578	118	0.95	.73	610	93	0.63	.01*
Female	512	83			557	108			605	58		
Caregiver Higher Education												
Higher Education	780	136	0.72	.10	761	136	0.75	.17	689	90	0.31	<.001*
No Higher Education	288	36			282	38			264	11		
Neighbourhood Social Class	-0.32	-0.49	0.82	.03*	-0.35	-0.52	0.81	.01*	-0.38	-0.62	0.73	.01*
Neighbourhood Familialism	0.54	0.57	1.03	.74	0.52	0.58	1.06	.41	0.54	0.54	1.01	.87
Language												
Albanian	60	20	3.75	<.001	79	26	2.56	<.001	96	24	3.24	<.001
Other	131	40	3.40	<.001	140	48	2.52	<.001	151	15	1.23	.52
German[a]	516	46	-	-	521	67	-	-	515	39	-	-
English	20	3	1.69	.41	22	1	0.35	.32	19	3	2.13	.24
Italian	60	6	1.12	.80	65	8	0.96	.92	61	8	1.77	.16
Portuguese	63	15	2.63	<.001	71	13	1.42	.27	82	18	3.04	<.001
Serbian-Croatian	84	21	2.81	<.001	86	31	2.80	<.001	110	28	3.45	<.001
Spanish	51	12	2.59	.01*	54	16	2.30	.01*	54	5	1.11	.84
Turkish	37	7	2.13	.09	38	8	1.64	.22	54	7	1.74	.20
Tamil	46	2	0.45	.28	59	8	1.04	.92	73	4	0.75	.59

Note. HE= higher education. [a]Reference category. *Significant at alpha<0.05

Table 3: Means of participant behavioural predictors and bivariate associations (ORs) with attrition status

Predictor	Attrition P1 – P4				Attrition Y1 – Y4				Attrition Y5 – Y7			
	Retained	Dropped out	OR	*p*	Retained	Dropped out	OR	*p*	Retained	Dropped out	OR	*P*
Parent reports (N_{P1-P4} = 1230, N_{Y1-Y4} = 1207, N_{Y5-Y7} = 1044)												
Prosocial Behaviour	2.56	2.66	1.43	.02*	2.57	2.58	1.04	.79	2.57	2.59	1.07	.73
Internalising	0.72	0.61	0.59	.01*	0.71	0.67	0.80	.22	0.71	0.64	0.68	.11
ADHD	1.23	1.13	0.78	.07	1.21	1.22	1.01	.93	1.22	1.16	0.86	.37
Non-aggressive Conduct Disorder	0.61	0.48	0.38	<.001*	0.60	0.55	0.68	.08	0.60	0.53	0.59	.07
Aggression	0.62	0.51	0.53	<.001*	0.61	0.57	0.78	.21	0.61	0.55	0.70	.17
Self-reports (N_{P1-P4} = 1215, N_{Y1-Y4} = 1359, N_{Y5-Y7} = 1156)												
Prosocial Behaviour	0.82	0.82	0.91	.84	0.82	0.80	0.53	.11	0.82	0.82	1.06	.92
Internalising	0.41	0.43	1.32	.44	0.41	0.42	1.29	.41	0.41	0.41	0.89	.78
ADHD	0.17	0.17	0.95	.91	0.17	0.19	1.84	.10	0.17	0.17	0.79	.66
Non-aggressive Conduct Disorder	0.21	0.22	1.14	.77	0.21	0.23	1.79	.13	0.22	0.22	1.29	.63
Aggression	0.18	0.16	0.63	.37	0.17	0.18	1.10	.82	0.17	0.17	0.77	.66
Teacher reports (N_{P1-P4} = 1214, N_{Y1-Y4} = 1338, N_{Y5-Y7} = 1144)												
Prosocial Behaviour	2.20	2.13	0.90	0.3	2.19	2.05	0.81	.02*	2.20	2.06	0.81	.08
Internalising	0.85	0.90	1.09	.44	0.86	0.97	1.21	.04*	0.85	0.90	1.09	.50
ADHD	1.20	1.48	1.30	<.001*	1.20	1.45	1.27	<.001*	1.22	1.30	1.09	.39
Non-aggressive Conduct Disorder	0.30	0.43	1.58	<.001*	0.30	0.45	1.74	<.001*	0.31	0.36	1.25	.25
Aggression	0.55	2.13	1.39	<.001*	0.56	0.75	1.45	<.001*	0.57	0.62	1.12	.41

Note. *Significant at alpha<.05.

Table 4: Multiple probit regression results predicting non-response by participant background

Predictor	Outcome= Non-Response P1				Outcome= Non-Response Y1				Outcome= Non-Response Y5			
	b	**SE**	***p***	**6-FWER Holm corrected *p***	**b**	**SE**	***p***	**6-FWER Holm corrected *p***	**b**	**SE**	***p***	**6-FWER Holm corrected *p***
Small Class ***(reference category = small class)***	0.04	0.04	.31	1.00	0.03	0.03	.36	1.00	0.12	0.03	<.001*	.01*
Gender ***(reference category = female)***	0.01	0.02	.78	1.00	-0.02	0.02	.20	1.00	-0.01	0.02	.55	1.00
Caregiver Higher Education	0.10	0.05	.03	.60	-0.02	0.04	.60	1.00	0.01	0.03	.70	1.00
Neighbourhood Social Class	-0.05	0.01	<.001*	.02*	-0.02	0.01	.20	1.00	0.00	0.01	.96	1.00
Neighbourhood Familialism	0.00	0.01	.67	1.00	0.00	0.01	.67	1.00	0.01	0.01	.24	1.00
Language ***(reference category = German)***												
Albanian	0.30	0.04	<.001*	<.001*	0.17	0.04	<.001*	<.001*	0.03	0.04	.34	1.00
Other	0.03	0.03	.40	1.00	0.00	0.03	.91	1.00	0.07	0.03	.02*	.36
English	-0.04	0.08	.60	1.00	0.01	0.08	.87	1.00	-0.01	0.08	.88	1.00
Italian	0.09	0.05	.06	1.00	0.05	0.04	.23	1.00	0.06	0.04	.21	1.00
Portuguese	0.18	0.04	<.001*	<.001*	0.17	0.04	<.001*	<.001*	-0.01	0.04	.82	1.00
Serbian-Croatian	0.21	0.04	<.001*	<.001*	0.17	0.03	<.001*	<.001*	0.02	0.03	.65	1.00
Spanish	0.09	0.05	.06	1.00	0.05	0.04	.22	1.00	0.13	0.05	<.001*	.06
Turkish	0.25	0.05	<.001*	<.001*	0.27	0.05	<.001*	<.001*	0.03	0.05	.55	1.00
Tamil	0.30	0.05	<.001*	<.001*	0.12	0.05	.01*	0.22	-0.04	0.05	.35	1.00

Note. HE = higher education. *Significant at $p<.05$;

Table 5: Multiple probit regression results predicting attrition by participant background and behaviour

Predictor	Outcome= P1-P4 attrition				Outcome= Y1-Y4 attrition				Outcome= Y5-Y7 attrition			
	b	**SE**	***p***	**6-FWER Holm corrected *p***	**b**	**SE**	***p***	**6-FWER Holm corrected *p***	**b**	**SE**	***p***	**6-FWER Holm corrected *p***
Small Class *(reference category= small class)*	-0.02	0.04	.67	1.00	0.05	0.04	.23	1.00	0.02	0.03	.51	1.00
Gender *(reference category= female)*	0.00	0.02	.95	1.00	0.02	0.02	.51	1.00	-0.04	0.02	.04*	.65
Parental Higher Education	-0.01	0.02	.81	1.00	0.00	0.03	.96	1.00	-0.07	0.02	<.001*	.09
Neighbourhood Social Class	0.00	0.01	.70	1.00	0.00	0.01	.81	1.00	-0.01	0.01	.57	1.00
Neighbourhood Familialism	0.00	0.01	.66	1.00	0.01	0.01	.48	1.00	0.00	0.01	.73	1.00
Language *(reference category= German)*												
Albanian	0.12	0.04	.01*	.19	0.10	0.04	.02*	0.44	0.09	0.03	.01*	.19
Other	0.13	0.03	<.001*	<.001*	0.10	0.03	<.001*	.03*	0.00	0.03	.95	1.00
English	0.02	0.07	.75	1.00	-0.10	0.08	.22	1.00	0.06	0.06	.32	1.00
Italian	0.00	0.04	.98	.98	-0.02	0.05	.70	1.00	0.04	0.04	.29	1.00
Portuguese	0.07	0.04	.10	1.00	0.01	0.05	.89	1.00	0.09	0.04	.02*	0.29
Serbian-Croatian	0.07	0.04	.07	1.00	0.11	0.04	<.001*	.09	0.10	0.03	<.001*	.04*
Spanish	0.08	0.05	.09	1.00	0.08	0.05	.09	1.00	-0.02	0.04	.70	1.00
Turkish	0.05	0.06	.33	1.00	0.04	0.06	.53	1.00	0.02	0.04	.71	1.00

Tamil	-0.08	0.05	.14	1.00	-0.02	0.05	.72	1.00	-0.06	0.04	.17	1.00
Parent reports												
Prosocial Behaviour	0.02	0.02	.33	1.00	0.00	0.02	0.88	1.00	0.00	0.02	0.84	1.00
Internalising	-0.03	0.02	.20	1.00	-0.03	0.03	.33	1.00	-0.03	0.02	.22	1.00
ADHD	-0.01	0.02	.80	1.00	0.00	0.02	.87	1.00	0.00	0.02	.90	1.00
Non-aggressive Conduct Disorder	-0.07	0.04	.05*	.91	-0.03	0.04	.46	1.00	-0.02	0.04	.51	1.00
Aggression	0.00	0.03	.93	1.00	0.00	0.03	.95	1.00	0.00	0.03	.95	1.00
Self-reports												
Prosocial Behaviour	-0.02	0.06	.74	1.00	-0.06	0.06	.35	1.00	0.02	0.06	.73	1.00
Internalising	0.05	0.05	.32	1.00	0.05	0.05	.36	1.00	0.02	0.04	.68	1.00
ADHD	-0.04	0.07	.61	1.00	0.03	0.07	.68	1.00	-0.03	0.06	.60	1.00
Non-aggressive Conduct Disorder	0.05	0.07	.46	1.00	0.05	0.07	.45	1.00	0.00	0.07	.96	1.00
Aggression	-0.05	0.08	.48	1.00	-0.05	0.08	.49	1.00	0.01	0.07	.90	1.00
Teacher reports												
Prosocial Behaviour	-0.01	0.01	.66	1.00	-0.01	0.01	.35	1.00	-0.02	0.01	.09	1.00
Internalising	-0.02	0.01	.24	1.00	0.00	0.01	.74	1.00	0.00	0.01	.95	1.00
ADHD	0.02	0.01	.10	1.00	0.01	0.01	.71	1.00	0.00	0.01	.72	1.00
Non-aggressive Conduct Disorder	0.03	0.03	.38	1.00	0.04	0.03	.22	1.00	0.01	0.03	.68	1.00
Aggression	0.02	0.02	.41	1.00	0.03	0.02	.24	1.00	-0.01	0.02	.53	1.00

Note. *Significant at $p<.05$.

Supplementary Table 1: Bivariate cross-correlation between behaviour problem measures

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
Parent reports															
1.Prosocial Behaviour	-														
2.Internalising	-	-													
3.ADHD	-0.2	0.42	-												
4.Non-aggressive Conduct Disorder	- 0.27	0.35	0.51	-											
5.Aggression	-	0.33	0.42	0.55	-										
Self-reports															
6. Prosocial Behaviour	0.17	-	-	-0.06	-0.06	-									
7.Internalising	0.09	0.08	0.02	-0.01	-0.02	0.22	-								
8.ADHD	-	0.09	0.1	0.1	0.1	-0.05	0.35	-							
9. Non-aggressive Conduct Disorder	- 0.06	0.07	0.14	0.14	0.12	-0.15	0.15	0.52	-						
10.Aggression	-	0.05	0.11	0.15	0.17	-0.08	0.36	0.57	0.53	-					
Teacher reports															
11. Prosocial	0.18	-	-0.1	-0.13	-0.06	0.15	0.06	-0.09	-0.15	-0.09	-				
12.Internalising	-	0.14	0.11	0.03	0	-0.06	0.04	0.08	0.09	0.07	-0.23	-			
13. ADHD	-	0.04	0.33	0.16	0.1	-0.11	0.05	0.18	0.2	0.15	-0.25	0.36	-		
14. Non-aggressive Conduct Disorder	- 0.02	0.04	0.24	0.18	0.12	-0.11	0.01	0.15	0.18	0.11	-0.32	0.38	0.63	-	
15. Aggression	-	0.03	0.21	0.13	0.17	-0.06	0.01	0.16	0.15	0.12	-0.3	0.34	0.56	0.69	-

Note. Cross-informant correlations on same phenotype.