1 August 2019

Deliverable Report for the Economic and Social Research Council (ESRC) grant ES/T001224/1

Theoretical Sampling Design Options for a New Birth Cohort: An Accelerated Longitudinal Design Perspective

Prepared by:

Natalie Shlomo[1], Chris Skinner[2], Myong Sook Kim[1]

[1] Social Statistics Department, School of Social Sciences, University of Manchester
Contact Author Email: Natalie.Shlomo@manchester.ac.uk

[2] Department of Statistics, The London School of Economics and Political Science

Executive Summary:

- The key difference between an accelerated longitudinal design and the single cohort study is that an accelerated longitudinal design incorporates multiple cohort groups at the onset of the study, enabling the duration of the study to be shortened. This investigation focuses on aspects of an accelerated longitudinal design compared to a single cohort design, particularly for following babies, children and youth into adulthood. We provide a comprehensive literature review and investigate aspects of the designs through a theoretical assessment and simulation study.
- The theoretical assessment of the precision of basic longitudinal analyses under single, 2-cohort and 3-cohort designs shows that sample sizes can be compromised especially at the extreme ages of the cohorts. This problem can be addressed by refreshment samples that not only ensure representativity but also compensate for left- censored data.
- The simulation study shows that the accelerated longitudinal designs permit analysis across a wider age span for a given duration of the study. The precision of parameter estimation for a multilevel growth curve model is similar for the alternative designs. If cohort effects are present in the study, it is important to include cohort main effects and their interactions in the model. For small sub-groups, the level of precision can be compromised due to small sample sizes.
- The overall conclusion of this investigation regarding the suitability of an accelerated longitudinal design for the future of longitudinal study in the UK is positive. We recognize, nevertheless, that, in coming to a decision about the longitudinal design, careful consideration is needed of the different types of analysis undertaken by substantive researchers and their additional complexities under an accelerated longitudinal design. In addition, it may be important to consider other aspects of survey methodology impacted by an accelerated longitudinal design when coming to a decision, but these are outside the scope of this report.
- From the review of international surveys of children and young people, it is clear that multiple cohort designs have been successfully implemented in different countries and the UK can learn from such experiences whilst considering specific needs, for example, boosting the 2016 'missed' cohort from the cancelled Life Style Study.

Report for the ESRC grant: ES/T001224/1

Theoretical Sampling Design Options for a New Birth Cohort: An Accelerated Longitudinal Design Perspective

Prepared by: Natalie Shlomo[1], Chris Skinner[2], Myong Sook Kim[1]

[1] Social Statistics Department, School of Social Sciences, University of Manchester
Contact Author Email: Natalie.Shlomo@manchester.ac.uk

[2] Department of Statistics, The London School of Economics and Political Science

1. Introduction

The 2017 Longitudinal Studies Strategic Review by Davis-Kean, Chambers, Davidson, Kleinert, Ren and Tang (2018) made a series of recommendations on innovative ways to enhance and to invest in opportunities for longitudinal research in the UK. Since the discontinuity of Life Study in 2016 for various reasons, which was originally intended to collect data from over 80,000 babies born during 2014-2018, nationally-representative data on children born after 2000/2001 are currently unavailable. Whilst there is longitudinal data available from current and ongoing cohort studies and other longitudinal data collected, for example Understanding Society (including youth data for 10-15-year-olds) and the English Longitudinal Study of Aging, the lack of UK-level longitudinal data for youngest generations' childhood and youth years is concerning and needs urgent action. For this reason we focus our attention in this report to the early years.

The challenges in designing a new large scale longitudinal birth cohort study are obtaining an optimal balance between analytic benefits and deciding the most cost-efficient way of conducting such a study, in particular, controlling the duration of the study. One of the key recommendations to the ESRC was to commission a new birth cohort with an alternative approach based on an accelerated longitudinal design (Davis-Kean, et al. 2018, page 7). The key difference between an accelerated longitudinal design and the usual single cohort study is that an accelerated design incorporates multiple cohort groups at the outset of the study, enabling the duration of the study to be shortened.

In this report, we provide a literature review and investigate theoretical and quantitative aspects of an accelerated longitudinal design versus a traditional single cohort. We compare designs with respect to the precision of different kinds of statistical analyses, sample sizes, comparability of measures across cohorts and across sweeps, missing data and attrition.

Section 2 provides a literature review introducing the alternative designs and their properties and includes rationale for implementing an accelerated longitudinal design. The literature review is also extended in Appendix A with an overview of the most relevant international studies using accelerated longitudinal designs, case studies and applications. Section 3 presents an assessment of alternative design options and includes both a theoretical assessment for basic longitudinal analyses with more results found in Appendix B and a simulation study for growth curve estimation where the results and findings are presented in Appendix C. We conclude in Section 4 with a discussion and also mention other relevant survey methodology considerations in an accelerated longitudinal design but are currently out-of-scope of this report.

2. Alternative longitudinal designs and literature review

This section discusses alternative longitudinal designs with a focus on the early years from birth to adulthood.

2.1 Single cohort design approach

The most recent birth cohort study covering the whole of the UK is the Millennium Cohort Study (MCS). The MCS follows approximately 19,000 children born in the UK during 2000-2001 (Joshi and Fitzsimons, 2016). The oversampling allows the study of children from disadvantaged and ethnic minority families. The MCS is regarded as a *prospective* cohort study as participants are followed up "longitudinally" over a period of time. To date, data have been collected over seven assessment points when respondents were 9 months, 3, 5, 7, 11 and 14 years old (Table 1). The simplicity in survey instruments in a single cohort study is of great advantage in a practical sense, as it does not have a different set of questionnaires and unique measurement schedules that are varying by respondents' age.

Along with the longitudinal observation of individuals, the large sample size, the geographical coverage of the whole UK[1], and oversampling of vulnerable sub-groups make MCS a valuable resource for scientific enquiries concerning many aspects of child development over the life course (Connelly and Platt, 2014; Joshi and Fitzsimons, 2016). MCS has been used in a wide range of research domains[2] including social disadvantage (e.g. linking parental education and family income to children's cognitive ability development gaps by Brown and Sullivan, 2014); gender-specific trajectories of behavioural disorders (Gutman et al., 2018), child health (e.g. mother's health-related behaviours by Ward et al., 2007), developmental psychology (e.g. dynamics of family structure and socio-emotional well-being in the early years by Pearce et al., 2014). MCS data also provides evidence of intergenerational transmission of worklessness (Schoon et al., 2012).

Table 1. MCS sweeps (prospective single cohort, data collection during 2001-2018)

Wave	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7
Year	2001	2004	2006	2008	2012	2015	2018 (ongoing)
Age	9 months	3	5	7	11	14	17

A longitudinal study consisting of a single cohort, such as MCS, entails a relatively slow process in terms of data availability. While waiting for children to reach certain ages of interest, a single cohort approach suffers from major drawbacks in three aspects: attrition, administrative costs of tracking individuals over time, extended assessment time points and associated burden borne by survey participants. As of wave 6, 61% of the initial sample remained in the MCS due to dropouts (Joshi and Fitzsimons, 2016). The dependence on cohort members' maturation in a single cohort population also poses challenges for analysts. If one is interested in the adolescent development psychology for age 7 through 17 among those born in 2000 across the UK, for instance, one needs to allow 10 years for data availability under the current data collection schedules (i.e., 2008-2018). This time lag can be of concern due to the changing nature of social environments. Analyses results may be less timely and relevant to younger cohorts if there are fast-paced changes in childhood environments (Nicholson, Sanson, Rempel, Smart, and Patton, 2002). In addition, Sanson (2002) pointed out the nature of measurement techniques, which can be time-sensitive; measured implemented at the

[1] Another prospective single cohort study, Next Steps (https://cls.ucl.ac.uk/cls-studies/next-steps/), or the Longitudinal Study of Young People in England (LSYPE) previously, covers England only. Starting at age 14 in 2004, the original survey participants of 15,770 were followed over seven waves (annually between 2005-2010, then 2015). The Next Steps data's is linked to National Pupil Database (NPD), which contain cohort members' individual scores at Key Stage 2, 3, and 4. The incorporation of the UK geographical coverage would enhance the usability of Next Steps.

[2] Full publications using MCS can be found in www.bibliography.cls.ucl.ac.uk/bibliography.

onset of the study may become obsolete over time, and may no longer be cutting edge with the progression of the study. Nevertheless, instruments tend to remain constant to ensure consistency of the collected longitudinal data in a single cohort study.

2.2 Accelerated longitudinal design: Multiple cohort approach

In this section, we introduce the accelerated longitudinal design and design considerations when planning such designs for a birth cohort study. We also provide illustrative examples of influential overseas studies, a brief discussion of empirical studies using nationally-representative surveys of children and youth and international practices for researching vulnerable sub-groups.

2.2.1 Introducing the accelerated longitudinal design

In an accelerated longitudinal design, multiple samples of individuals in different age groups are studied at the outset and followed forward repeatedly over a period of time. The idea of combining the strength of longitudinal and cross-sectional data was introduced by Bell (1953). In child development trajectories, Bell (1953, 1954) demonstrated the "convergence" approach as a means of meeting research needs not satisfied by either cross-sectional or longitudinal design. The similar development of the accelerated longitudinal design was also seen in psychology (Schaie, 1965, Nesselroade and Baltes, 1979, Meredith and Tisak, 1990). More recently, Tonry, Ohlin and Farrington (1991) were among the first to provide a comprehensive review of accelerated longitudinal designs.

Figure 1 presents an illustrative example comparing a single cohort design to an accelerated longitudinal design with multiple cohorts.

Figure 1. Longitudinal data structured by year and age (four annual measurements)

			Year (wave, T=4)			
Design	*Cohort*	*Birthyear*	2020	2021	2022	2023
Single-cohort (SC)	C4	2020	0	1	2	3
Multiple cohort (A)	C4	2020	0	1	2	3
	C3	2019	1	2	3	4
	C2	2018	2	3	4	5
	C1	2017	3	4	5	6

Data Re-Structured by Age

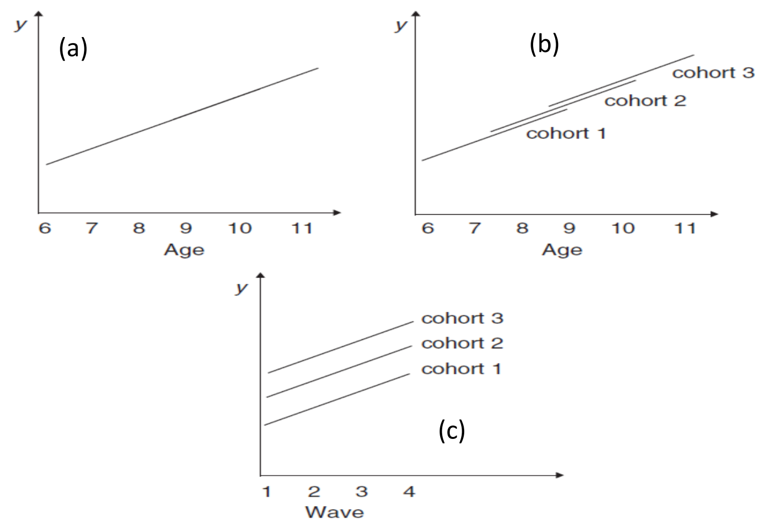
						Age (T=6)				
	Cohort	*Birthyear*	Age 0	Age 1	Age 2	Age 3	Age 4	Age 5	Age 6	Age 7
Multiple cohort (B)	C4	2020	0	1	2	3	.	.	.	.
	C3	2019	.	1	2	3	4	.	.	.
	C2	2018	.	.	2	3	4	5	.	.
	C1	2017	.	.	.	3	4	5	6	.

Note: Illustrative example using four waves of data in prospective cohort studies. Respondents aged 0-3 at the initial survey (in year 2020) in the multiple cohort design. Age is calculated by Year-Birth year.

As shown in Figure 1, a single cohort design (SC) only samples one group of children all born in the same year 2020. Figure 1 also shows an accelerated longitudinal design comprised of four cohort groups of differing ages. In the example, individuals at multiple ages from 0 to 3, relating to four birth cohorts, are recruited in the first data collection point (year 2020) and followed up over 4 waves.

A similar study with three cohorts, starting at ages 6, 7 and 8 at the first data collection point and 4 waves is illustrated in Figure 2. The effects of cohort, age and period (wave) on the development of an outcome variable Y are shown in different ways.

Figure 2. Accelerated 3-cohort design starting at ages 6,7 and 8 (Bollen and Curran 2006 p. 78)



In Figure 2, panel (b), adjacent cohort groups are overlapped according to age[3]. Groups may be linked at their overlapping time points to approximate an overall longitudinal trajectory. The maximum number of overlapping groups is associated with the extent of observation periods. The example in Figure 2, panel (b) demonstrates that four assessments of individuals over time allow up to three overlaps at ages 8, 9 and 10.

A limitation of accelerated designs displayed in panel (b) in Figure 2 is the fluctuating number of overlapping groups by age. At both extremes of age 6 and 11, it is solely a single birth cohort group contributing to the analysis outcome with a much reduced sample size compared to a single cohort. Therefore, it is essential to consider the breakdown of each age group at the design stage carefully in the calculation of the sample size, with particular attention to the extremes where overlapping is not possible by design. If research interest surrounds vulnerable groups of children, for instance, this aspect should be fully taken into account in sampling decisions to allow adequate sample size for meaningful group-based analyses. More details of sample size implications are described in Section 3.1 and in Appendix B.

Although the illustrated example in Figure 1 does not recruit a new younger cohort group (birth year 2021 onwards) in the subsequent measurement points, it can be a possible sampling strategy (see Sanson, 2002) to provide more data to overlap, as part of an attempt to address attrition or shrinking sample size. Other strategies that can increase flexibility in design may be exploring options in the variation in assessment intervals, the number of cohort groups, and the total duration of the study.

[3] The data rearrangement is often a useful strategy in development contexts because it is *age* which has more substantial meaning in measuring developmental trajectories over time, rather than calendar year that represents assessment time (Bollen & Curran, 2006). It is a researchers' role to arrange the data in an appropriate structure to meet the study goal, as reference point (for instance, year 2020 in Figure 1) may vary by studies.

The key feature of the multiple cohort design is the ability to shorten the overall length of the study, whilst covering a given age range. Alternatively, it enables a wider age range to be covered by a study of given length. In Figure 1, the age range available from a single cohort study is from 0 to 3 (T=4), as data covering age 4, 5, and 6 are unavailable in 2023. In an accelerated longitudinal design with multiple cohorts, the age range "accelerates" up to 6 with the same data collected during the 2020-2023 period. The potential benefit to researchers is that it may be to assess developmental trajectories of children for an extended age span (0-6, rather than 0-3) and observe interesting findings at an earlier stage, without having to wait for a single cohort to mature (Sanson, 2002: p. 42).

As highlighted in Figure 2, panel (b), the accelerated longitudinal design 'pieces together' trajectories from cohorts 1 through 3. If there is no evidence of statistically significant cohort effects, one can report a common trajectory across cohorts (Duncan, Duncan, and Hops, 1996). On the other hand, if cohort effects do exist, a common trajectory may not be a useful representation across the whole age range and, instead, cohort-specific trajectories may be reported. In any case, it is clear that the accelerated longitudinal design may add complexities for substantive researchers who will need to consider cohort effects and other design issues into their analyses.

The aspect of utilising shorter study spans can also be one of the main shortcomings of the accelerated design compared to the single-cohort longitudinal design. The concerns include limited capacity to observe within-individual developmental sequence and continuity (Duncan et al. 2011; Raudenbush and Chan, 1992), and limited ability to capture long-term causal effects (Baltes, 1968; Farrington, 1991). For example, if the social environment is changing, then for a cohort born in 2022 and followed for 8 years, it may not be appropriate to expect that one can learn about how this cohort will be in their teens by 'borrowing' from an older cohort born in say, 2012 who have been followed between ages 10 and 18. A possible solution may be using retrospective data, which can then be cross-checked with later prospective data. For instance, individuals may be asked whether they had *ever* engaged in specified activities, and if so, at what age they had engaged in those activities. The collection of retrospective data can offer the basis for an investigation of intergenerational effects by linking parental factors to children's developmental trajectories (Farrington, 1991).

On the other hand, some analysts advocate that the capacity to anticipate long-term trends at an earlier stage is important in the sense that it enables timely interventions, if necessary. According to Bell (1954), the rapid approximations of longitudinal results are important as social changes over an extended period of time may alter the value and implications of studies. Sanson (2002) argues that due to the shorter time required for data availability in the accelerated longitudinal design, there is less concern about theories, instruments and policy issues being outdated. Whilst in a single-cohort design, instruments tend to remain constant to maintain consistency of the collected data, multiple cohort designs will typically have different instruments for each cohort while still needing to ensure comparability of collected data between early and later cohorts.

One advantage of the accelerated longitudinal design comprising of overlapping cohorts (panel (b), Figure 2) over the single cohort design is that it offers the opportunity of studying more than one time metric simultaneously and in addition, allows the possibility to untangle age and cohort, in order to study age, which is the optimal measure of temporal change, independent of potential cohort effects. It enables cohort effects to be analysed separately from age effects and also allows generalization to other cohorts. If we take "cohort" as birth year and "period" as calendar year (Rabe-Hesketh and Skrondal, 2012, p. 239), age, period and cohort are related through the identity

$$Age = Period - Cohort$$

The confounding of age and period and the inability to generalise to earlier or later birth cohorts are basic problems for single cohort studies (Farrington, 1991). In a single cohort design where birth year is fixed, age and period are intrinsically confounded so that age effects are effectively identical to period effects. Similarly, with cross-sectional data, period is fixed and age and cohort are intrinsically confounded (Glenn, 1976). The pooling of data from several birth cohort studies can help overcome these problems although this may require larger samples sizes and there needs to be sufficient overlap to be able to test for age and cohort effects, for example by testing differences in linear or quadrative slopes in a growth curve model between adjacent cohort groups. One application to overcome the drawbacks of a single cohort design in the UK is in McMunn et al. (2015) where cohort-specific work-family life courses of three cohorts of British women and men between the ages of 16 and 42, were investigated using the National Survey of Health and Development 1946 birth cohort, the National Child Development Study 1958 birth cohort and the British Cohort Study 1970 birth cohort. However, the history of birth cohort studies in the UK since 1970 has been uneven, as noted earlier, especially in terms of varying time lags.

As mentioned, the accelerated longitudinal design can create complexities that may be difficult for the substantive focused researcher. Cohort effects often emerge in response to dramatic system changes (for example, the introduction of pre-school care), and this may impact on one part of the cohort but not another. This complexity in pooling data from different cohorts may not be as straight-forward as simply accounting for additive cohort effects in statistical models to capture their differences. If cohort effects are identified in a multiple cohort design, sample sizes may be too small and may reduce the power of statistical analysis. For these reasons, it is important to ensure large sample sizes for each cohort to allow for testing and compensating for cohort effects. As seen in McMunn et al. (2015), there is evidence that combining different single year birth cohorts work. The accelerated longitudinal design may be seen as an extension of this idea of pooling existing single cohort studies in a more planned and structured way.

2.2.2 Examples of studies with accelerated longitudinal designs

We review the literature on multiple-aged nationally-representative surveys that employ accelerated longitudinal designs in Appendix A. The literature shows a mixture in surveys with varying degree of representativeness. While some empirical research used either primary or non-representative data, often limited to a certain population or target geographical locations, other studies capitalised on nationally-representative samples. We examine the most relevant overseas nationally-representative longitudinal surveys employing multiple cohort samples in Section A.1 of Appendix A. We place emphasis on a focused review of these longitudinal studies from Australia, US, Canada and Germany, and the list of studies is by no means exhaustive. In addition, Section A.2 of Appendix A includes applications and case studies based on accelerated longitudinal designs for child development and Section A.3 includes a review of dedicated studies for vulnerable sub-groups.

2.2.3 Design considerations for an accelerated longitudinal study

The design of an accelerated longitudinal study requires careful consideration with respect to:

- number of cohorts, including the ability to test cohort effects
- number of participants per age- cohort group
- extent of overlap between cohorts and the frequency and timing of measurements
- constraints on the duration of the study
- measuring instruments and their comparability between cohorts

- precision of estimates at national and regional levels
- power in statistical modelling for addressing key research questions particularly for vulnerable sub-groups and cohorts
- costs.

In general, the literature shows no consensus on the above. According to Farrington (1991), it is desirable that a project be limited to a total length of 10 years. When we allow for piloting and analyses stages, the follow-up period can take a maximum of 7 or 8 years. If there's a need for analyses based on a single cohort group, for various reasons, each cohort should meet the minimum required number of respondents, although the cut-off point is unclear. Galbraith, Bowden and Mander (2017) present an analysis of accelerated longitudinal designs with respect to cohort effects and the impact of costs and dropouts on the power of testing and precision of estimation parameters. They found that as duration-related costs increase relative to recruitment costs, the best designs shift towards shorter duration with the cross-sectional design being the optimal design. In addition, for designs with the same duration but differing interval between measurements, they found that there was a cut-off point for measurement costs relative to recruitment costs relating to the frequency of measurements.

Advice on the desirable number of overlaps among adjacent birth cohort groups is also ambiguous. While some point out the inadequacy of one or two overlaps between immediately adjacent birth cohort groups (Anderson, 1993, Rogosa, 1988), others advise no fewer than five or even six, a much higher threshold (Raudenbush and Chan, 1993; Tonry, Ohlin, and Farrington, 1991). Using a small sample of Australian adolescent school children between age 12 through 14, Watt (2008) investigated common developmental trajectories of Math and English scores among three cohort groups. The data covers calendar years 1995-1998. The sample size for each cohort group was less than 500 individuals (Total N=1,323), and the sample was selected from three upper-middle class schools. The challenges in Watt's study were that it was small with 2-3 overlaps per age group, model complexity arising from using quadratic as well as cubic growth curve parameters, and the examination of gender-specific growth curve models, using less than 500 subjects in each cohort group.

A similar issue associated with small overlapping time points was discussed by Raudenbush and Chan (1993). Using the nationally-representative National Youth Survey (NYS), Raudenbush and Chan (1993) demonstrated trajectories of delinquency among two groups (age 11-15, N=239 and age 14-18, N=245 respectively). Raudenbush and Chan (1993) find that their survey design permitting only two overlaps at age 14 and 15 at each time point, is ill-equipped to estimate cohort heterogeneity in quadratic and cubic rates. Raudenbush and Chan (1993) further advocated higher cut-off points suggested by Tonry et al. (1991).

In power analysis, using 1,000 subjects per cohort, Tonry et al. (1991) concluded that less than five or six overlapping time points is unacceptable for verifying cohort effects, due to multicollinearity. The cut-off time point suggested by Tonry et al. is substantially higher than that of Anderson (1993) and Rogosa (1988). According to Farrington (1991), longer periods of overlap between cohorts is optimal. Miyazaki and Raudenbush (2000) are also in agreement and show that the power of testing cohort effects increases with larger sample sizes which occur at the overlapping time points.

Much of the studies mentioned in the literature review relate to growth curve analysis (see Sections A2 and A3 in Appendix A). With recent advances in methodology and computational capacity, the use of growth curve modelling has grown dramatically. The growth curve model can handle an accelerated cohort design sample with ease, facilitating enquiries regarding developmental trajectories. Recently, growth curve models have been extended to latent class trajectory analyses thus enhancing research scope. The examination of cohort effects can vary depending on whether the multilevel (MLM) or structural equation modelling (SEM) approach is used. In MLM, Miyazaki and Raudenbush (2000) demonstrate age and cohort (age x cohort) interaction effects empirically to test for cohort-specific heterogeneity, which can result from demographic and historical differences between cohorts. In the SEM framework, multi-group analysis is conducted by fitting cohort-specific group models simultaneously (Duncan et al., 2011).

The contexts of using the growth curve method vary from children or adolescents' development contexts (Duncan, Duncan, and Hops, 1996; Duncan, Duncan, and Strycker, 2001; Jacobs, Lanza, Osgood, Eccles, and Wigfield, 2002; Miyazaki and Raudenbush, 2000; Prinzie, Onghena, and Hellinckx, 2006; Raudenbush and Chan, 1993; Watt, 2008), to life span developmental psychology covering adults (Muthén and Muthén, 2000) and older population (Finkel, Reynolds, McArdle, and Pedersen, 2007; Gerstorf et al. 2011, McArdle, Ferrer-Caja, Hamagami, and Woodcock, 2002; Orth, Trzesniewski, and Robins, 2010). We note that growth curve modelling is often used in traditional longitudinal studies as well, such as the English Longitudinal Study of Aging (ELSA), where they can identify cohort/period effects at a given age (see examples in: Weber, 2016; Tampubolon and Maharani, 2018; Zaninotto, et al., 2018). We examine a multilevel growth curve model and the impact of additive cohort effects under an accelerated longitudinal design for a small sub-group in the simulation study in Section 3.2 and Appendix C.

Although an accelerated longitudinal design sample is often used in growth curve models, other traditional research methods to address important research questions can be used using these samples, particularly for the case where the variations in the age-specific development trajectories are too small. For instance, Canada's rich data, National Longitudinal Survey of Children and Youth (NLSCY) is used under many conventional research methods (e.g. Baker and Milligan, 2016; Strohschein and Gauthier, 2018). In general, as the sample is composed of individuals in multiple age-cohorts, the accelerated longitudinal design may offer a greater flexibility in analytic potential for addressing societal research questions within a shorter time frame compared to a single cohort design.

Regarding missing data and attrition, an important feature in the accelerated longitudinal design is "planned missingness". As illustrated in Figure 1, areas marked with a dot (.) indicate *intentional* missingness by design, due to unplanned data collection. Therefore, for missing data in the age range which is out of the scope of the study, we can assume missing completely at random (MCAR) (Bollen and Curran, 2006; Enders, 2010). Under the MCAR mechanism, the probability of missing data on a variable Y is *un*related to the value of Y itself and *un*related to the values of any other observed variables (Allison, 2001, p. 3). Although MCAR is a strong assumption, it is reasonable to assume when planned missing data is part of the research design. However, it should be noted that a missing outcome variable that arises in the study age span to be covered (shaded areas) is typically assumed under the missing at random (MAR) assumption, as in growth curve modelling (Bollen and Curran, 2006). The growth curve models often rely on accelerated multiple cohort samples, and have an ability to yield valid estimations under both MCAR and MAR conditions. In some circumstances, one needs to address the remaining selection bias, under missing not at random MNAR scenarios. For detailed discussions on the assumptions of growth curve modelling, see elsewhere (Bollen and Curran, 2006; Enders 2010; Stoel, Van Den Wittenboer, and Hox, 2003; Cheung, 2013).

In the literature review in Section A.1 of Appendix A, missing data and attrition rates are reported for the major studies of those employing longitudinal accelerated designs. In a single cohort design, missing data typically occurs in wave 1 of the study, and for those responding in wave 1, attrition rates increase as the cohort is investigated over time. This problem is often mitigated by periodic refreshment samples throughout the course of the survey. For an accelerated longitudinal design, there is an implication that there may be a potential reduction in sample attrition over time owing to the shorter study span, which is linked to effective management in administrative costs and burden borne by participants (Sanson, 2002).

3. Quantitative assessment of alternative design options

In this section we provide a theoretical assessment of precision in an accelerated longitudinal design (also in Appendix B) and a simulation study to assess the ability to model a growth curve under these designs with results in Appendix C.

3.1 Theoretical precision of basic longitudinal analyses

A fairly wide class of methods of longitudinal analysis are based on data across a fixed interval of ages, denoted here age s up to age t ($s<t$). One example is regression analysis of a variable measured at age t with covariates measured at age s or at ages between s and t. Another is a comparison of the distribution of a variable measured at age t across sub-groups defined according to a variable measured at age s. We assume here no missing data and that the standard error of any estimate is proportional to the reciprocal of the square root of the size of the sample of cases which are observed at all ages s up to t. The precision of a longitudinal analysis is a function of this sample size and we present the sample sizes below for the case of 8 measurements per cohort. In Appendix B we present the case of 5 measurements per cohort.

Designs Considered for Comparison:

Single cohort design – sample of n babies observed at age 0 at base year and followed up at 2 year intervals until age 14.

2- cohort design - 2 cohorts starting in same base year: cohort 1 is sample of $n/2$ babies aged 0, cohort 2 is sample of $n/2$ children aged 6. We take 7 further measurements on each cohort at 2 year intervals. Hence measurements at following ages:

Cohort 1 – 0, 2, 4, 6, 8, 10, 12, 14. Cohort 2 – 6, 8, 10, 12, 14, 16, 18, 20.

3- cohort design -3 cohorts starting in same base year: cohort 1 is sample of $n/3$ babies aged 0, cohort 2 is sample of $n/3$ children aged 6, cohort 3 is sample of $n/3$ children aged 12. We take 7 further measurements on each cohort at 2 year intervals. So measurements at following ages:

Cohort 1 – 0, 2, 4, 6, 8, 10, 12, 14. Cohort 2 – 6, 8, 10, 12, 14, 16, 18, 20.

Cohort 3 – 12, 14, 16, 18, 20, 22, 24, 26.

We consider the precision of statistical analyses after eight waves of observation, that is after all the above measurements are taken.

Comparison of Designs after 8 Waves of Observation:

Tables 2, 3 and 4 show the sample sizes for longitudinal analysis between variables measured at ages s and t ($s<t$) for the single cohort, 2- cohort and 3- cohort designs, respectively.

Table 2. Sample sizes for longitudinal analysis between variables measured at ages s and t ($s<t$) under 1 cohort design after 8 waves of observation

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	n	n	n	n	n	n	n				
2		n	n	n	n	n	n				
4			n	n	n	n	n				
6				n	n	n	n				
8					n	n	n				
10						n	n				
12							n				
14											

Table 3. Sample sizes for longitudinal analysis between variables measured at ages s and t ($s<t$) under 2 cohort design after 8 waves of observation. Higher sample sizes are shaded darker.

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$				
2		$n/2$	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$				
4			$n/2$	$n/2$	$n/2$	$n/2$	$n/2$				
6				n	n	n	n	$n/2$	$n/2$	$n/2$	
8					n	n	n	$n/2$	$n/2$	$n/2$	
10						n	n	$n/2$	$n/2$	$n/2$	
12							n	$n/2$	$n/2$	$n/2$	
14								$n/2$	$n/2$	$n/2$	
16									$n/2$	$n/2$	
18										$n/2$	
20											

Table 4. Sample sizes for longitudinal analysis between variables measured at ages t and $s<t$ under 3 cohort design after 8 waves of observation. Higher sample sizes are shaded darker.

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	$n/3$	$n/3$	$n/3$	$n/3$	$n/3$	$n/3$	$n/3$				
2		$n/3$	$n/3$	$n/3$	$n/3$	$n/3$	$n/3$				
4			$n/3$	$n/3$	$n/3$	$n/3$	$n/3$				
6				$2n/3$	$2n/3$	$2n/3$	$2n/3$	$n/3$	$n/3$	$n/3$	
8					$2n/3$	$2n/3$	$2n/3$	$n/3$	$n/3$	$n/3$	
10						$2n/3$	$2n/3$	$n/3$	$n/3$	$n/3$	
12							n	$2n/3$	$2n/3$	$2n/3$	$n/3$
14								$2n/3$	$2n/3$	$2n/3$	$n/3$
16									$2n/3$	$2n/3$	$n/3$
18										$2n/3$	$n/3$
20											$n/3$

Advantages of multiple cohort designs:

(i) For 2- and 3- cohort designs, time is ‘accelerated’, enabling longitudinal analyses with $t=16,18,20$ for any value of s ($s<t$) of 6 or higher under the 2- cohort design, and

$t = 22, 24, 26$ for any value of s ($s < t$) of 12 or higher under the 3- cohort design. The single cohort design will not extend to these age ranges.

(ii) Cohort effects can be estimated from the overlapping age range 8 to14 for the 2- cohort designs. Cohort effects can also be estimated for the same age range and the additional age range 16 to 20 for pairs of cohorts in the 3- cohort design albeit with less power. In addition, cohort effects can be estimated across all 3 cohorts at the age of 14.

(iii) For longitudinal analyses with $t = 16, 18, 20$ for any value of s ($s < t$) of 12 or higher the sample size is highest for the 3- cohort design;

Disadvantages of multiple cohort designs:

(i) The sample size is halved for longitudinal analyses with t up to age 14, for values of s ($s < t$) of 0, 2 or 4 under the 2- cohort design, and the sample size is reduced to a third under the 3- cohort design. This is the problem of 'extremes' mentioned in section 2.2.1.

(ii) The 3- cohort design has a clear disadvantage over both 1 cohort and 2 cohort designs since for longitudinal analyses with t up to 14 and s ($s < t$) of 0-10, the sample size is least under the 3- cohort design.

3.2 Simulation study of growth curve analyses

The assumption in Section 3.1 that precision is a function of the size of the sample across a given age interval is reasonable for basic longitudinal analyses but not for more sophisticated ones. Here we consider a multilevel modelling approach to growth curve analysis and use a simulation study to examine the precision of parameter estimates as well as mean predicted values under alternative designs. In the first simulation, Simulation A, we assume no cohort effects in the population and in the second, Simulation B, we assume there is a cohort effect in a population for a 2- cohort design. We assume that in each cohort, there are 8 waves of observations similar to the setting in Section 3.1 although here we assume annual measurements for ease of interpretation.

Since we are interested in vulnerable sub-groups, the population and sample size for the simulation study are small: N=100,000, n=1,000, respectively assuming a 1/100 sample fraction. Within this group, we also study smaller sub-groups where we generate a random variable at 50% (denoted 'sex') and at 10% (denoted 'ethnic' for ethnic minority). Given the small sample sizes we do not assume missing data at this point.

The results of the simulation are in Appendix C and we include the summary here:

The simulation study shows that the precision with which the parameters of the multilevel growth curve model are estimated is fairly similar for the alternative designs. This is the case for all the parameters, including both the coefficients (fixed effects) and the parameters in the variance matrix of the random effects. When cohort effects are absent (Simulation A) there were two parameters (the intercept and age slope) where there was a slight loss of precision with the 2 and 3 cohort designs, but when cohort effects do exist (Simulation B) there were some coefficients where some estimates for the 2- cohort design were more precise. In particular, the simulation study showed that it is important to include cohort effects in the model for growth curve analysis. The mean predicted values of the growth curve for the overall total showed good precision compared to the true values in both populations of Simulation A and Simulation B, but the growth curves for the smaller sub-groups and in particular for the ethnic minority showed larger deviations. Overall, these differences in precision seem rather less important than the qualitative differences between the designs. Thus, the 2- and 3- cohort designs enable cohort effects to be estimated, although we found in Table C.6 that the power with which such cohort effects can be detected may not be large. More importantly, if we are choosing between the kinds of designs considered here, the fact that the single cohort design covers a shorter age range may be viewed as an important limitation for some kinds of analysis. However, we have found that fairly good estimation of the parameters of the multilevel growth curve model can be achieved even for the single cohort design with a limited age up to 7.

4. Conclusions

In this report, we have focused on the advantages and disadvantages of an accelerated longitudinal design for following babies and children and their transition into youth and adulthood compared to a single cohort design. From the review of international surveys of children and young people in Appendix A, it is clear that such multiple cohort designs have been successfully implemented in different countries. An important advantage of accelerated longitudinal designs is that they permit analyses across a wider age span for a given duration of study, although a corresponding disadvantage is that the sample size may be reduced for some parts of the age span, especially at the extreme ages. This problem has been addressed by refreshment samples in some of the studies described in Appendix A. We note that the use of refreshment samples are often used in longitudinal and birth cohort studies to ensure representativeness and left-censored data. We have also seen in the simulation study that the parameters of growth curve models, defined across the whole age span, can still be estimated with similar precision to single cohort designs from an accelerated design for the case where cohort effects are additive. Another advantage of accelerated longitudinal designs is that they enable the study of different birth cohorts in a single study. This use of multiple cohorts also enables cohort effects to be estimated, including age-cohort interactions, unlike in a single cohort design. On the other hand, it does imply that researchers need to consider more complex forms of analysis, which take account appropriately of such effects when analysing data from an accelerated longitudinal study.

Although the theoretical scope of this report has not included practical survey methodology considerations, we recognize their importance for design choices and comment briefly on them here:

- **Sample frames:** Unlike many other countries, the UK does not have a population register. A population spine constructed from census/administrative data sources was a recommendation in the 2017 Longitudinal Studies Strategic Review (Davis-Kean, et al., 2018), but in the absence of such a spine and considering our focus in this report on ages from birth into adulthood, we need to consider sampling frames from administrative data sources. Similar to other UK birth cohort studies, babies can be sampled from birth registers. Other cohorts starting at later ages could be sampled from the UK school census. However, there are limitations to this census since it does not include children in private schools or those home-schooled. For these populations, appropriate supplemental samples will need to be drawn, such as through an area- based cluster sampling design where pupils may be sampled from schools within sampled areas that are outside of the school census domain. In addition, there may be particular vulnerable sub-groups which will need to be identified and an appropriate sampling frame developed. Note that sample boosting and surveys drawn from multiple sample frames need careful methodological considerations in the calculation of the design weights and final survey weights. In Appendix A, there are examples of accelerated longitudinal designs in international studies, including supplemental surveys for specialized sub-groups, with examples of sampling frames for those countries without a population register or spine.

- **Recruitment and survey instruments:** Compared to a single cohort study, the simultaneous recruitment of multiple cohorts, the need for different instruments/ questionnaires for each cohort and the management of future follow-up in an accelerated longitudinal design introduces complexities and challenges. Multiple questionnaires/instruments for the different cohorts need to be developed and tested and more time and resources devoted to interviewer training so that interviewers can deliver multiple instruments to different cohorts. The associated budget and resources required may exceed that of the single cohort design with the same overall sample size.
- **Sample sizes:** The Millennium Cohort Study (MCS) sampled up to 20,000 babies and this reflected the given budget and desired precision at that time. Compared to such a single cohort study, there are other considerations of fixed and rolling costs that need to be accounted for in an accelerated design with 2 or more cohorts. Fixed costs in the planning stages of an accelerated design need to account for developing multiple sample frames, sampling and recruiting multiple cohorts, development of different questionnaires/instruments and their data capture and interviewer training with a focus on multiple instruments. Rolling costs at each wave of the survey need to account for the tracking of individuals in different cohorts between waves and the data collection with multiple instruments which may be more challenging than for a single cohort design. However, given the fixed number of measurements in an accelerated longitudinal design, rolling costs needed for tracking individuals may be reduced compared to a single cohort design. With a carefully planned budget framework to undertake an accelerated longitudinal design and the desired precision of estimates (particularly for vulnerable sub-groups), sample size calculations need to consider the number of cohorts, the number of overlapping ages, assumptions on the initial wave 1 missing data and subsequent attrition rates depending on the number of measurements. In addition, sample boosting may be required to target vulnerable sub-groups similar to the MCS.

The overall conclusion of this investigation regarding the suitability of an accelerated longitudinal design for the future of longitudinal study in the UK is positive. We recognize, nevertheless, that, in coming to a decision about the longitudinal design, careful consideration is needed of the different types of analysis undertaken by substantive researchers and other aspects of survey methodology, as mentioned above. The NLSCY of Canada appears to offer an ideal study design for child development. An alternative choice would be to adapt the German KiGGS survey to the UK's research needs, permitting cross-sectional as well as longitudinal investigations. For example, five birth cohort groups of children aged 0, 3, 5, 7, 9 at the first assessment with relatively short 2-3 year intervals over 5 occasions of data collection results in a total study period of 10 years. As shown from Canada's NLSCY and KiGGS samples, a top-up of newborns in each data collection will be a useful strategy with regards to maintaining the level of sample representativeness, as well as testing for cohort effects. With 5 multiple cohorts in this example, sample sizes of at least 3000 to 4000 per cohort would be appropriate and would allow sample boosting for vulnerable sub-groups and would be large enough to compensate for wave 1 missing data and attrition levels over subsequent waves. In terms of the wider strategy for the future of longitudinal studies in the UK, including previous birth cohort studies in 1946, 1958, 1970 and 2000, we could consider boosting the sample of children born around 2016 to compensate for the cancellation of the Life Study.

Acknowledgments: We thank Professor Tarani Chandola and Professor James Nazroo from the University of Manchester for their helpful comments to this report.

APPENDIX A: Literature Review

A.1. International surveys using nationally-representative samples

The utility of multiple cohort samples in accelerated longitudinal designs can be seen in many international surveys, and the survey instruments vary substantially: The number of cohorts, ranging from 2 (e.g. LSAC) to 11 or larger (e.g. NLSCY); the sample size per each cohort; decisions to recruit later cohorts; and the total duration of the study.

Growing Up in Australia: Longitudinal Study of Australian Children (LSAC)

Australia embarked on *Growing Up in Australia*: Longitudinal Study of Australian Children (LSAC[4]) in 2004 as a 2-cohort design following extensive research with a primary focus on new research directions for Children's health and development (Sanson, 2002). LSAC consists of 2 cohorts: B (baby, age 0-1) cohort and K (kindergarten, age 4-5) cohort. The data collection occurred in 2-year intervals during the 2004-2016 period, and 7 waves of data are currently available. The LSAC obtained information for over 10,000 young children (5,107 infant and 4,983 child cohort) at the start of the survey (Australian Institute of Family Studies, 2015, Table 9). The retention rate from wave 1 is 74% for B cohort and 71%, for K cohort, respectively.

Figure A.1. Longitudinal study of Australian children (LSAC), 2004-2016

						Age				
Birthyear	Cohort	0-1	2-3	4-5	6-7	8-9	10-11	12-13	14-15	16-17
2003-2004	cohort B	0-1	2-3	4-5	6-7	8-9	10-11	12-13	.	.
		(2004)	(2006)	(2008)	(2010)	(2012)	(2010)	(2016)		
1999-2000	cohort K	.	.	4-5	6-7	8-9	10-11	12-13	14-15	16-17
				(2004)	(2006)	(2008)	(2010)	(2012)	(2010)	(2016)

Note: Data collection years shown in brackets. Data linkage: school (selective Years ranging 3-12), vocational education and training, higher education). Oversampling: None (rejected due to inefficient use, preferring separate intensive studies).

Figure A.1 demonstrates a restructured data format with same age groups stacked up. This set up allows an investigation of 17-year long child development trajectories, and permits testing of cohort effects with the evident 8-year overlaps between age 4-5 and 12-13 between B and K cohort. The study, however, still fails to allow comparisons among dual cohorts concerning age before 4. This may depend on funding plans, but studies in Canada have allowed top-up samples for newly-born children in the subsequent waves. LSAC does not employ oversampling. The Australian funding body and expert groups preferred utility of separate intensive studies as an efficient way of allocating funding. Galbraith et al. (2017) have used LSAC data in their analyses on the impact of dropout on the statistical power in the accelerated multiple cohort design.

Longitudinal Surveys of Australian Youth (LSAY)

Longitudinal Surveys of Australian Youth (LSAY[5]) started collecting data in 1995. LSAY incorporates two existing multiple-aged large scale national surveys, Youth in Transition Cohort (YITS) 1978 and Australian Youth Survey (AYS) 1991. The aim of the national panel survey, LSAY is to gather information about health, education, work, and social activities among adolescents, who are enrolled in school, from age 15 (Year 9). The new cycle of data collection for later cohorts occurred

[4] https://growingupinaustralia.gov.au/
[5] https://www.lsay.edu.au/aboutlsay

within 2-6 year intervals: 1995, 1998, 2003, 2006, 2009, 2015 and 2017, adding over 10,000 new participants in Year 9 in the subsequent waves. Once a new cohort is surveyed initially, the participants were contacted annually until they mature to the age of 25, over a 10-assessment period (see Figure A.2).

Figure A.2. Longitudinal Surveys of Australian Youth (LSAY), 1995-2015

Cycles:		1st	2nd	3rd	4th	5th	6th
Birthyear	Cohort	1995	1998	2003	2006	2009	2015
2000	Y15						15
1994	Y09					15	
1991	Y06				15		
1988	Y03			15		Annual Follow-up Until Age	
1983	Y98		15				
1980	Y95	15					

LSAY yields rich longitudinal data as can be seen in Figure A.3. Here, the discrete data collection schedule for each birth cohort group is visible. The 10-year follow up period for each cohort allows longer than suggested overlaps of six time points between adjacent cohorts, in testing potential cohort effects (Raudenbush and Chan, 1993, Tonry et al. 1991). The consideration to be needed here may be the treatment of period effects, given the data collection spanning over 20 years.

Figure A.3. Longitudinal Surveys of Australian Youth (LSAY), 1995-2015

Birthyear	Cohort	Age										
2000	Y15	15 (2015)	16 (2016)	17 (2017)	18 (2018)	19 (2019)	20 (2020)	21 (2021)	22 (2022)	23 (2023)	24 (2024)	25 (2025)
1994	Y09	15 (2009)	16 (2010)	17 (2011)	18 (2012)	19 (2013)	20 (2014)	21 (2015)	22 (2016)	23 (2017)	24 (2018)	25 (2019)
1991	Y06	15 (2006)	16 (2007)	17 (2008)	18 (2009)	19 (2010)	20 (2011)	21 (2012)	22 (2013)	23 (2014)	24 (2015)	25 (2016)
1988	Y03	15 (2003)	16 (2004)	17 (2005)	18 (2006)	19 (2007)	20 (2008)	21 (2009)	22 (2010)	23 (2011)	24 (2012)	25 (2013)
1983	Y98	15 (1998)	16 (1999)	17 (2000)	18 (2001)	19 (2002)	20 (2003)	21 (2004)	22 (2005)	23 (2006)	24 (2007)	25 (2008)	26 (2009)
1980	Y95	15 (1995)	16 (1996)	17 (1997)	18 (1998)	19 (1999)	20 (2000)	21 (2001)	22 (2002)	23 (2003)	24 (2004)	25 (2005)	26 (2006)

Note: Shaded areas indicate future data collection schedules. Some pupils were still 14 years of age during 1995 and 1998 surveys, so the follow-up extended to ensure all respondents had reached 25 years of age.

Another important aspect of LSAY concerns attrition rates. Among the latest cohort of 14,251 participants, Y09, by the fourth annual measurement in 2012, less than half remained in the survey (6,541) before further shrinking to 3,518 when the respondents reach 23 years of age (NCVER, 2018). The trend was broadly similar to Y98 cohort; from 14,117 in 1998, then just over half at 7,762 in 2002 and the final assessment period at 3,596 participants (NCVER, 2013b).

The study population of LSAY was randomly selected from two Year 9 classes from a national sample consisting of 300 schools. Sampling size was adjusted by states; Students from small states were over-sampled, and those from larger states were under-sampled (NCVER, 2013a).

Since 2003, LSAY has been integrated with the OECD's PISA, the Programme for International Student Assessment, a comparative study of academic achievement among nationally representative sample of 15-year-old students. Geographical location, gender, socioeconomic background were used as strata. Due to the PISA's oversampling, students from smaller jurisdictions and indigenous populations are included in LSAY. LSAY top-up sample includes 746 respondents in 2017.

National Longitudinal Survey of Children and Youth (NLSCY)

The National Longitudinal Survey of Children and Youth (NLSCY[6]) was Canada's long-run study that ran from early 1990s to late 2000s. NLSCY contacts children from birth through early adulthood. The initial sample is comprised of 22,831 children, who were aged 0 to 11 at the time of recruitment which is 12 cohorts at the onset of the study in one year age bands. The participating children can remain in the study until they reach the age of 25. Commenced in 1994, the study is conducted every two years and discontinued in 2008, completing 8 waves of data. At the final assessment period, the respondents were aged between 14 and 25, maintaining 57%-80% of the original sample (Statistics Canada, 2019a).

The notable feature of the NLSCY is the top-up sample called, Early Child Development (ECD), later born age 0 and 1 groups, being recruited in the subsequent survey. The ECD for age 0-1 sample varied between 4,000 and 5,500 except for the survey year 1998, when nearly 10,000 youngest children (age 0-1) were recruited. The consistency in 2-year intervals of assessments and continuity maintained in the NLSCY data collection during a 14-year time period (1994 through 2008, see Figure A.4), providing a wealth of information from birth through early adulthood (age 0 through age 25) and sets an excellent example of a multiple-aged nationally representative sample. It needs to be noted that the complexity of instruments adds difficulties in understanding the operationalisation. Samples were selected from the Labour Force Survey's (LFS) sample of respondent households (Statistics Canada, 2019a) rather than from schools, and Birth Registry data for some 0-5-year-olds. No oversampling was implemented.

Figure A.4. Overview of National Longitudinal Survey of Children and Youth, NLSCY

			Data Collection Schedule (1994-2008, 8 waves)							
			Wave 1	Wave 2	Wave 3	Wave 4	Wave 5	Wave 6	Wave 7	Wave 8
	Birthyear	Age in 1994	1994	1996	1998	2000	2002	2004	2006	2008
Early Child Development (ECD) Cohorts	2008	*-14*								**0**
	2007	*-13*								**1**
	2006	*-12*							**0**	2
	2005	*-11*							**1**	3
	2004	*-10*						**0**	2	4
	2003	*-9*						**1**	3	5
	2002	*-8*					**0**	2	4	6
	2001	*-7*					**1**	3	5	7
	2000	*-6*				**0**	2	4	6	
	1999	*-5*				**1**	3	5	7	
	1998	*-4*			**0**	2	4		8	
	1997	*-3*			**1**	3	5		9	
	1996	*-2*		**0**	2	4				
	1995	*-1*		**1**	3	5				
Original Sample	1994	*0*	0	2	4	6	8	10	12	14
	1993	*1*	1	3	5	7	9	11	13	15
	1992	*2*	2	4	6	8	10	12	14	16
	1991	*3*	3	5	7	9	11	13	15	17
	1990	*4*	4	6	8	10	12	14	16	18
	1989	*5*	5	7	9	11	13	15	17	19
	1988	*6*	6	8	10	12	14	16	18	20
	1987	*7*	7	9	11	13	15	17	19	21
	1986	*8*	8	10	12	14	16	18	20	22
	1985	*9*	9	11	13	15	17	19	21	23
	1984	*10*	10	12	14	16	18	20	22	24
	1983	*11*	11	13	15	17	19	21	23	25

Note: Reconstructed by author based on NLSCY microdata user guide, cycle 8 p. 18 Figure 1 and p. 27 Table 2 (Statistics Canada, 2019a). Early Child Development (ECD) Cohorts, 0-1 year of age in recruitment year, joined

[6] http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=56797

since 1996 (shown in box). Top-up sample of new entrants aged 2-5 join waves 6, 7 and 8, and the returning respondents aged 2-5 were re-interviewed. Blank areas indicate planned missing.

Figure A.5 shows the data organisation scheme arranged in the accelerated cohort design. The first data collection period is shown as shaded area diagonally. By inspecting the diagonal dimension of Figure A.5, one can apply this figure in deciding the time frame of any desirable study design. Figure A.5 also illustrates that both cross-sectional and longitudinal data, for 0-5 age groups in particular and for other target age groups are available, and the possibility of data augmentation with additional data in the subsequent waves. We can see that there are 8 overlapping groups of children aged 0-5, which is ideal for testing cohort effects.

Figure A.5. NLSCY structured by age

Birth year	Age																									
2008	**0**																									
2007		**1**																								
2006	**0**		2																							
2005		**1**		3																						
2004	**0**		2		4																					
2003		**1**		3		5																				
2002	**0**		2		4		6																			
2001		**1**		3		5		7																		
2000	**0**		2		4		6																			
1999		**1**		3		5		7																		
1998	**0**		2		4				8																	
1997		**1**		3		5				9																
1996	**0**		2		4																					
1995		**1**		3		5																				
1994	0		2		4		6		8		10		12		14											
1993		1		3		5		7		9		11		13		15										
1992			2		4		6		8		10		12		14		16									
1991				3		5		7		9		11		13		15		17								
1990					4		6		8		10		12		14		16		18							
1989						5		7		9		11		13		15		17		19						
1988							6		8		10		12		14		16		18		20					
1987								7		9		11		13		15		17		19		21				
1986									8		10		12		14		16		18		20		22			
1985										9		11		13		15		17		19		21		23		
1984											10		12		14		16		18		20		22		24	
1983												11		13		15		17		19		21		23		25

Early Child Development (ECD) Cohorts

Wave 1 Wave 3 Wave 5 Wave 8

The accelerated multiple cohort design sample was analysed by Côté et al. (2006) who explored the development of physical aggression from toddlerhood to pre-adolescence. Another study using accelerated multiple cohort sample concerns protective factors of alcohol use trajectories among Canadian Aboriginal adolescents with a age ranging from 12 to 23 (Rawana and Ames, 2012).

Following the success of NLSCY[7], Canada undertook a cross-sectional survey, Survey of Young Canadians (SYC) in 2010. The nationally representative SYC consists of child development data for children between age 0 and 9. The restriction of one child participant per household resulted in 17,000 participants. One of the objectives was to produce early child development indicators at the regional level, by age for younger children between 1 and 5 years of age. The need for national-level data for 6-9-year old children led to the extension of age range in SYC (Statistics Canada, 2019b).

[7] See https://crdcn.org/datasets/nlscy-national-longitudinal-survey-children-and-youth for publications using NLSCY data.

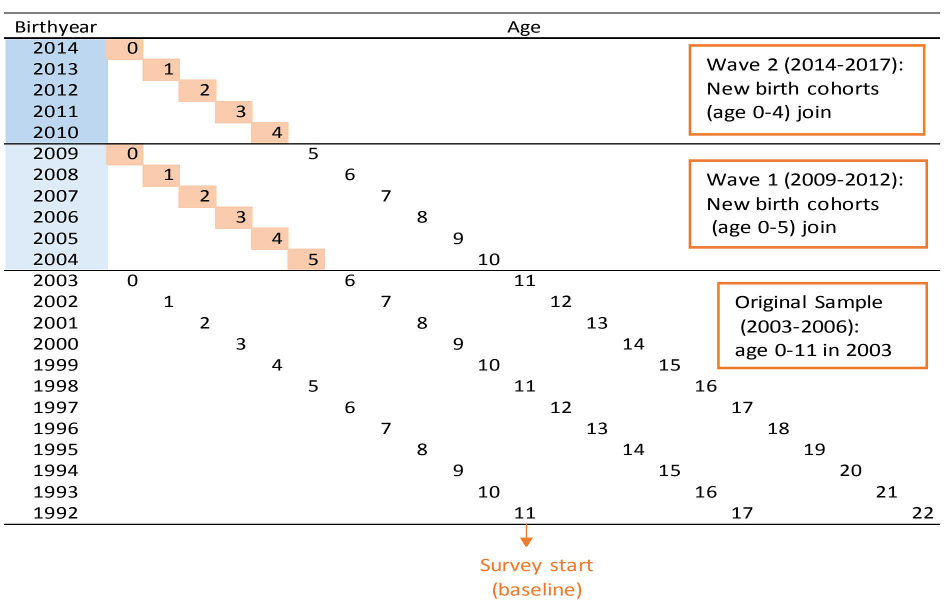
German Health Interview and Examination Survey for Children and Adolescents (KiGGS)

The German Health Interview and Examination Survey for Children and Adolescents (KiGGS[8]) has collected data since 2003. KIGGS gathered data for 17,641 children age 0-11 (12 cohorts), encompassing a large age range at the beginning of the survey (Kamtsiuris, Lange, and Schaffrath, 2007). Following the baseline survey during inception, KIGGS implemented two follow-up surveys in 5-6 year intervals; 2009-2012 (labelled as wave 1 according to KiGGS), then 2014-2017 (wave 2) (Klipker, Baumgarten, Göbel, Lampert, and Hölling, 2018).

The population-based survey, KiGGS is primarily intended to produce cross-sectional data covering from age 0 through 17, and the decision was the result of lengthy discussions among expert groups with the cross-sectional aspect chosen as the optimal design. KiGGS also releases longitudinal components for public use. It is notable that KiGGS further recruits later younger cohorts in subsequent waves (6 cohorts at ages 0 to 5 in wave 1 and 5 cohorts at ages 0 to 4 in wave 2), as seen in Canada's NLSCY sampling strategies. What makes KIGGS unique is that along with new younger cohorts, it invites *older* individuals in waves 1 and 2, extending the age coverage up to 24 and 31, respectively. KiGGS data has been used in accelerated multiple cohort design samples in studying long-term ADHD symptoms among children and adolescents aged 7-19 years by Döpfner et al. (2015).

For the purpose of this report, we restructured the data, ignoring the age range over 22 (see Figure A.6). Given the data contains 3 assessments during the period between 2003 and 2014, this data allows three overlaps in each age, and data for the total age span of 22 years (age 0-22) can be obtained within 11 years from the onset of the survey. This approach increases feasibility of the study as implementation costs will be significantly lower than Canada's survey, NLSCY.

Figure A.6. Child and Youth data, adapted from KiGGS study



[8] https://www.kiggs-studie.de/english/survey/kiggs-overview.html

In the US, accelerated multiple cohort designs, studying adolescents' behaviours in particular, has been extensively used using national samples, notably the National Longitudinal Survey of Youth (NLSY[9]) and a rather smaller scale survey the National Youth Survey (NYS).

National Longitudinal Survey of Youth (NLSY) 79

NLSY[10] is one of the well-established long-running longitudinal studies for adolescents in the US. Initiated in 1979, NLSY offers nationally representative sample of 12,686 men and women born between 1957 and 1964 (age 14-22 when first surveyed). The eight birth cohort groups were followed annually through 1994, and currently re-invited every two years. The successor NLSY97 began in 1997 and contains information for individuals aged 12-17 in the beginning of the survey. Currently, "round" 1 (1997-1998) through 17 (2015-2016) data is available.

The children who were born to NLSY79 female respondents were followed in a separate survey. Initiated in 1986, NLSY79 *Children and Young Adults* contains both child-specific information as well as their mothers. As of 2014, 11,521 children have been identified as having been born to the original 6,283 NLSY female participants (Bureau of Labor Statistics, 2019). Once NLSY79 children reach 15 years of age, they become part of the NLSY Young Adult sample.

For applications of NLSY, see Muthén and Muthén, who applied accelerated cohort designs in NLSY data estimating trajectories of drinking and alcohol related problems for individuals aged 18 through 37 (Muthén and Muthén, 2000).

National Youth Survey (NYS)

NYS[11] is sponsored by the National Institute of Mental Health in the US. It contains data on young individuals from age 11 to 17 at the beginning of the study. Since the first wave of the survey was conducted in 1976, NLS follows up the original 1,725 participants annually until 1980, then 1983 and 1987, providing 7 waves of data for public use. Given the 7 cohort groups, sample size is rather small (less than 300 per age). The total 11-year study allows to study developmental trajectories from early adolescence to early adulthood. The fact that NYS recruits 7 cohort groups and annual assessment for the first years leads to accumulation of rich data in a short space of time. The survey does not offer refreshment samples of younger cohort groups. Despite the small sample, NYS is frequently used. An example is understanding adolescents' delinquent behaviours[12] (e.g. Gunnison, 2015; Li, Barrera, Hops, and Fisher, 2002; Raudenbush and Chan, 1993).

A.2. Applications and Case Studies based on Accelerated Longitudinal Designs for Child Development

Table A.1 demonstrates varying operationalisations of accelerated longitudinal designs through growth curve models, primarily for children and young adults. As shown, relatively small sample sizes (100-300) from national representative studies such as NYS (from US) were used in the applications, with the exception of Côté et al. (2006) and Muthén and Muthén (2000). The noticeable variation is in the number of cohort groups, ranging from 2 (NYS from US, LSAC from

[9] NLSY was introduced in the earlier strategic review report by Martin et al. (2006, p. 23) as an exemplar study of age cohorts studies.

[10] https://www.nlsinfo.org/content/cohorts/nlsy79

[11] https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8375/datadocumentation#

[12] More publications using NYS, see https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/8375/publications

Australia) to 13 (KiGGS from Germany). The features of the study design are likely to be guided by research questions and characteristics of the sample.

For older adults' development studies (not shown in Table A.1), see life time development in intellectual abilities for individuals aged from 2 to 95 years (McArdle, Ferrer-Caja, Hamagami, and Woodcock, 2002), self-esteem development of individuals from 25 to 104 years of age (Orth, Trzesniewski, and Robins, 2010), and cognitive aging among elderly (Finkel, Reynolds, McArdle, and Pedersen, 2007; Gerstorf et al. 2011[13]). Growth curve modelling is often used in traditional longitudinal studies as well, particularly using the English Longitudinal Study of Aging (ELSA).

Table A.1. Literature of use of growth curve modelling in national samples having accelerated longitudinal designs

Empirical Study	NR	ACS (domain)	Birth Cohort Groups	Initial Age	Waves of Data	Per Age (N)
Watt (2008)	No	Yes (Education)	3	Age 12-14	4	<500
Raudenbush and Chan (1993)	NYS	Yes (Behaviour)	2	Age 11 and 14	4	<300
Duncan, Duncan and Strycker (2001)	NYS	Yes (Alcohol use)	4	Age 11-14	5	App. <200
Prinzie, Onghena and Hellinckx (2006)	No	Yes (Aggression)	4	Age 4-7	3	<200
Duncan, Duncan and Hops (2001)	No	Yes (Alcohol use)	4	Age 12-15	3	Min. 92 Max. 146
Galbraith, Bowden and Mander (2017)	LSAC	Yes (Child development)	2	Age 0-1 and Age 4-5	7	5,107 (infant) 4,983 (child)
Muthén and Muthén (2000)	NLSY97	Yes (Alcohol use)	8	Age 18-25	7	Min. 833 Max. 4624
Côté, Vaillancourt, LeBlanc, Nagin and Tremblay (2006)	NLSCY	Yes, Group-based trajectories (Aggression)	10	Age 2-5	6	App. 1,000
Rawana and Ames (2012)	NLSCY	Yes (Alcohol use)	11	Age 12-13	6	>300
Döpfner, Hautmann, Görtz-Dorten, Klasen, Ravens-Sieberer, The BELLA study group (2015)	KiGGS	Yes (ADHD symptoms)	13	Age 7-16	3	Min. 101 Max. 640

Note: NR=whether the study is Nationally Representative sample. Per Age=non-overlapped, single birth cohort group. ACS=accelerated longitudinal design. NYS=National Youth Survey (location: US), LSAC=Growing Up in Australia: Longitudinal Study of Australian Children (Australia). NLSY97= National Longitudinal Survey of Youth 97 (US). NLSCY=National Longitudinal Survey of Children and Youth (Canada). KiGGS=German Health Interview and Examination Survey for Children and Adolescents (Germany)

[13] Gerstorf et al. (2011) investigates cohort differences (those born during 1886-1913 vs 1914-1948) among 50- and 80- year old individuals' cognitive decline.

A.3. Dedicated studies for vulnerable groups

The international longitudinal surveys using multiple cohort samples have adopted strategies that ensures population-representative data. Oversampling for particular vulnerable groups can be one of the considerations in the survey design for more focused research. In addition, among the international surveys, there are those that uniquely assess specific target populations and can provide focused data.

In Canada and Australia, separate surveys aiming to study specific marginalised populations exist. For instance, Canada's history on disability-specific surveys goes back to 1980s (e.g. Canadian Health and Disability Survey) (Arim, Findlay, and Kohen, 2016). For children, Aboriginal Children's Survey (ACS) began data collection for Aboriginal children under the age of six years in 2006. Similarly, Australia has launched a mixed-methods cohort study: 'Next Generation: Youth Well-being Study' with plans to recruit 2,250 Aboriginal adolescents aged 10-24 between April 2018 to June 2020, from rural, remote and urban communities in Central Australia, Western Australia and New South Wales, in order to assess their overall health and well-being (Gubhaju et al., 2019).

Other recommendations include a consideration of surveys targeting specific age groups, such as pre-schoolers, to facilitate in-depth assessments of vulnerable individuals. The Early Development Instrument (EDI) Data from Canada is one example: https://edi.offordcentre.com/.The EDI is a response to recommendations on the use of new knowledge about brain development of children from "Early Years Study" in 1999. The EDI is built on instruments used in the country's existing population-level survey NLSCY. By reflecting newer developments in school readiness and developmental health, EDI has been in operation since 2003. EDI studies 5-year-old kindergarten children in Canada, by using a nationally-representative questionnaire, which is completed by kindergarten teachers. Kindergarten is chosen as 90% of the eligible children attend. Presently, data from 2004 to 2014 is available. Oversampling is not used. EDI is linked to neighbourhood and socioeconomic data, creating the Canadian Neighbourhoods and Early Child Development (CanNECD) database. EDI's ambition is to identify "developmentally" vulnerable children at an earlier stage prior to formal education. EDI is increasingly adapted and piloted worldwide, including Australia, New Zealand, and US. Currently, EDI is also piloted in England and Scotland (https://edi.offordcentre.com/about/what-is-the-edi/). The EDI data are collated and managed by the Offord Centre for Child Studies, at McMaster University in Ontario, Canada.

Canada's population-based EDI was adopted in Australia in 2009 with the aim of guiding national and state policy and informing program development. Australia adopted Canada's EDI and Australian Early Development Census (AEDC) was created in 2009, and further data collection occurred in 2012 and 2015 (Boller and Harman-Smith, 2019).

APPENDIX B: Theoretical precision of basic longitudinal analyses after 5 Waves of Observation

In Section 3.1 we compared the 1- cohort, 2- cohort and 3- cohort designs after 8 waves of observations. Here we present the same analysis after 5 waves of observation where we assume that we have measurements at the following ages:

1- cohort design: ages 0, 2, 4, 6, 8;

2- cohort design: ages 0, 2, 4, 6, 8 in cohort 1 and ages 6, 8, 10, 12, 14 in cohort 2;

3- cohort design: ages 0, 2, 4, 6, 8 in cohort 1, ages 6, 8, 10, 12, 14 in cohort 2 and ages 12, 14, 16, 18, 20 in cohort 3.

Comparison of 2- cohort design with single cohort design

Tables B.1 and B.2 show the sample sizes for longitudinal analysis between variables measured at ages s and t ($s < t$) for the single cohort and 2- cohort designs, respectively. We find the following advantages and disadvantages of the 2- cohort design.

Advantages of 2- cohort design over single cohort design:

(i) time is 'accelerated', enabling longitudinal analyses with $t = 10, 12, 14$ for any value of s ($s < t$) of 6 or higher. For such cases longitudinal analysis, it is not possible for the single cohort design.

(ii) cohort effects can be estimated at the overlapping ages 6 and 8.

Disadvantage of 2- cohort design over single cohort design:

(iii) The sample size is halved for longitudinal analyses with t up to age 8, for values of s ($s < t$) of 0, 2 or 4.

Table B.1 Sample sizes for longitudinal analysis between variables measured at ages s and t ($s < t$) under 1- cohort design after 5 waves of observation

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	n	n	n	n							
2		n	n	n							
4			n	n							
6				n							
8											
10											
12											
14											

Table B.2 Sample sizes for longitudinal analysis between variables measured at ages s and t ($s < t$) under 2- cohort design after 5 waves of observation. Higher sample sizes are shaded darker.

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	$n/2$	$n/2$	$n/2$	$n/2$							
2		$n/2$	$n/2$	$n/2$							
4			$n/2$	$n/2$							
6				n	$n/2$	$n/2$	$n/2$				
8					$n/2$	$n/2$	$n/2$				
10						$n/2$	$n/2$				
12							$n/2$				
14											

Comparison of 3- cohort design with 2- cohort and single cohort designs

Table B.3 provides properties of the 3- cohort design analogous to Tables B.1 and B.2. The pros and cons of the 3- cohort design are as follows.

Table B.3 Sample sizes for longitudinal analysis between variables measured at ages t and $s < t$ under 3- cohort design after 5 waves of observation. Higher sample sizes are shaded darker.

	t										
s	2	4	6	8	10	12	14	16	18	20	22
0	$n/3$	$n/3$	$n/3$	$n/3$							
2		$n/3$	$n/3$	$n/3$							
4			$n/3$	$n/3$							
6				$2n/3$	$n/3$	$n/3$	$n/3$				
8					$n/3$	$n/3$	$n/3$				
10						$n/3$	$n/3$				
12							$2n/3$	$n/3$	$n/3$	$n/3$	
14								$n/3$	$n/3$	$n/3$	
16									$n/3$	$n/3$	
18										$n/3$	
20											

Advantages of 3- cohort design over both 1- cohort and 2- cohort designs:

(iv) time is accelerated further, enabling longitudinal analyses with $t = 16, 18, 20$ for any value of s ($s < t$) of 12 or higher. For such cases longitudinal analysis is not possible for the single cohort or 2- cohort designs;

(v) for longitudinal analyses with $s = 12$ and $t = 14$ the sample size is highest for the 3- cohort design;

(vi) cross-sectional cohort effects, across pairs of cohorts, can be estimated for ages 8 and 14.

Advantage of 3- cohort design over single cohort design, but disadvantage relative to 2- cohort design:

(vii) Longitudinal analyses is feasible with $t = 10, 12, 14$ for any value of s ($s < t$) of 6 or higher, unlike single cohort design. However, sample size is less than for 2- cohort design.

Disadvantage of 3- cohort design over both 1- cohort and 2- cohort designs:

(viii) For longitudinal analyse with t up to 8 and s ($s < t$) of 0-6, the sample size is least for the 3-cohort design.

Note that there are no situations where it is possible to estimate cohort effects from the 2- cohort design but not the 3- cohort design.

APPENDIX C: Results of Simulation Study

The purpose of the simulation study introduced in Section 3.2 is to assess whether we can estimate a growth curve using a multilevel model based on an accelerated longitudinal design with multiple cohorts, and examine the precision of parameter estimates of the models and their mean predicted values.

In the first simulation, Simulation A, we assume no cohort effects in the population and in the second simulation, Simulation B, we assume that there is a cohort effect in a population under a 2-cohort design. We assume that in each cohort, there are 8 waves of observations similar to the setting in Section 3.1 although here we assume annual measurements for ease of interpretation.

Since we are interested in vulnerable sub-groups, the population and sample size for the simulation study are small: N=100,000, n=1,000 assuming a 1/100 sample fraction. Within this group, we also study smaller sub-groups where we generate a random variable at 50% (denoted 'sex') and at 10% (denoted 'ethnic' for ethnic minority). We assess the power of statistical testing for cohort effects under these conditions.

Simulation A: No cohort effects in the population

We generate annual values of a dependent outcome variable for ages 0 to 19 for a single population according to a quadratic multilevel growth curve model. The model includes fixed sex and ethnic minority effects as well as random effects for both the intercept and slope. Age is centred at 9.5. The parameter values for generating the population are listed in the second column of Table C.1. Note that the variable 'cohort' under the 2-cohort design is a dummy variable with a value of 1 defined for the first cohort and 0 for the second cohort, and 'cohort1' and 'cohort2' under the 3-cohort design are dummy variables with a value of 1 defined for the first cohort and 0 otherwise, and a value of 1 defined for the second cohort and 0 otherwise, respectively.

We then draw 500 simple random samples of size 1,000 and repeat the steps described below three times for each of the different cohort designs: a single cohort design, a 2- cohort design and a 3- cohort design. We assume 8 years of observations so that the single cohort includes observations 0 to 7, the 2-cohort design includes observations 0 to 7 in the first cohort and 6 to 13 in the second cohort, the 3-cohort design includes observations 0 to 7 in the first cohort, 6 to 13 in the second cohort and 12 to 19 in the third cohort. For each design, with a sample size of 1,000 and 8 annual observations, the long format dataset for fitting the multilevel growth curve model contains a total of 1000×8=8000 observations. Note that under a single cohort, we are able to model the growth curve from 0 to 7, under the 2-cohort design, we model the growth curve from 0 to 13 and under the 3-cohort design, we model the growth curve from 0 to 19. Given the assumption of a small sample size and smaller sub-groups, we do not assume missing data at this stage.

The steps are:
1. Estimate model parameters for each of the 500 samples under each design using the same covariates as in the true model (assuming no model misspecification). For the 2- and 3- cohort designs, the model was fitted with and without cohort effects.
2. Assess the precision of the parameter estimates.
3. Calculate the mean predicted values of the growth curve averaged across the 500 samples under each design.
4. Assess the precision of the mean predicted values
5. Conduct a statistical test to compare the mean of the outcome variable at the overlapping ages for the 2-cohort and 3-cohort designs.

Table C.1 presents the averages of the parameter estimates over the 500 samples and their standard errors. Standard errors are calculated from the simulation standard deviation of the parameter estimates across the 500 samples divided by the square root of 500. We present the results in Table C.1 under the 2-cohort and 3-cohort designs with the cohort effects included in the multilevel growth curve model as there was little difference between the estimates for other parameters compared to the model without cohort effects.

Table C.1. Average parameter estimates of growth curve with standard errors (Simulation A, no cohort effects, 500 samples)

	True Parameters	Single cohort		2-cohort		3- cohort	
		Parm	SE	Parm	SE	Parm	SE
Fixed Effects							
Intercept	0.85	0.8524*	0.00069	0.8527*	0.00080	0.8535*	0.00088
Cohort	none	-	-	-0.0016	0.00092	-	-
Cohort1	none	-	-	-	-	-0.0012	0.00104
Cohort2	none	-	-	-	-	-0.0029*	0.00108
Ethnic	-0.008	-0.0043*	0.00151	-0.0046*	0.00147	-0.0052*	0.00149
Sex	-0.08	-0.0813*	0.00092	-0.0813*	0.00089	-0.0811*	0.00090
Age	0.45	0.4504*	0.00031	0.4506*	0.00037	0.4499*	0.00041
Ethnic*Age	-0.00002	-0.0010	0.00074	-0.0011	0.00074	-0.0013	0.00075
Sex*Age	-0.0002	-0.0002	0.00043	-0.0001	0.00043	-0.0001	0.00043
Cohort*Age	none	-	-	-0.0007	0.00042	-	-
Cohort1*Age	none	-	-	-	-	0.0005	0.00057
Cohort2*Age	none	-	-	-	-	0.0002	0.00051
Age_squared	-0.01	-0.0100*	0.00001	-0.0100*	0.00001	-0.0100*	0.00001
Random Effects (Unstructured Variance Matrix)							
UN(1,1)	0.08	0.0797*	0.00017	0.0800*	0.00016	0.0800*	0.00017
UN(2,1)	0.004	0.0041*	0.00006	0.0041*	0.00006	0.0041*	0.00006
UN(2,2)	0.02	0.0201*	0.00004	0.0200*	0.00004	0.0200*	0.00004
Residual	0.01	0.0100*	0.00001	0.0100*	0.00001	0.0100*	0.00001

(*) denotes significance at 5% significance level

In Table C.1 we see good parameter estimation under all designs with little differences in the precision of the estimates. Only the intercept and slope (age) showed increasing standard errors across the different designs. The main effect of ethnic minority showed deviations from the true parameters, likely due to the small sample size. The interactions of ethnic minority and sex with age were not significant for any of the designs. In addition, cohort effects and their interactions with age were largely non-significant but this was expected as no cohort effects were introduced into the population through the generating model. However, one exception was the main effect of cohort2 with a confidence interval of (-0.00498, -0.00074).

Table C.2 shows the true growth curve mean values at each age for the whole population and the sex and ethnic minority subgroups. For each sample and age, the mean of the predictions of the growth curve was calculated and then averaged over the 500 samples. We present in Table C.2 the percent relative absolute deviation of these averaged mean growth curve predictions from the true mean values based on the single, 2- cohort and 3 -cohort designs. We note that all true mean growth curve values are included in the confidence intervals of the averaged mean growth curve predictions according to the simulation standard errors.

Table C.2 True mean values of the growth curve for total and subgroups (Sex 50% and Ethnic Minority 10%) with percent relative absolute deviations of averaged mean growth curve predictions under each cohort design (Simulation A, no cohort effects, 500 samples)

Age	True Mean Values			Percent Relative Absolute Deviation of Estimated Mean Predicted Values								
				Single Cohort			2-cohort			3- cohort		
	All	Sex	Ethnic	All	Sex	Ethnic	All	Sex	Ethnic	All	Sex	Ethnic
0	-4.37	-4.41	-4.40	0.01	0.07	0.05	0.01	0.00	0.06	0.03	0.20	0.11
1	-3.74	-3.78	-3.77	0.02	0.08	0.01	0.00	0.01	0.03	0.04	0.21	0.08
2	-3.13	-3.17	-3.16	0.02	0.07	0.02	0.00	0.01	0.04	0.04	0.21	0.10
3	-2.54	-2.58	-2.58	0.01	0.09	0.03	0.01	0.01	0.00	0.04	0.23	0.05
4	-1.97	-2.01	-2.01	0.01	0.09	0.02	0.01	0.00	0.02	0.04	0.24	0.08
5	-1.42	-1.46	-1.46	0.01	0.12	0.07	0.02	0.01	0.01	0.04	0.28	0.06
6	-0.89	-0.93	-0.93	0.05	0.10	0.06	0.03	0.11	0.03	0.05	0.21	0.01
7	-0.38	-0.42	-0.42	0.00	0.27	0.09	0.01	0.23	0.08	0.06	0.36	0.08
8	0.11	0.07	0.07				0.09	1.75	5.73	0.68	0.61	3.67
9	0.58	0.54	0.54				0.04	0.08	0.48	0.14	0.06	0.10
10	1.03	0.99	0.98				0.08	0.00	0.35	0.12	0.08	0.08
11	1.46	1.42	1.41				0.05	0.03	0.26	0.08	0.09	0.02
12	1.87	1.83	1.82				0.03	0.05	0.17	0.02	0.07	0.10
13	2.26	2.22	2.21				0.06	0.09	0.12	0.02	0.02	0.05
14	2.63	2.59	2.58							0.04	0.19	0.29
15	2.98	2.94	2.93							0.06	0.20	0.30
16	3.31	3.27	3.26							0.03	0.17	0.26
17	3.62	3.58	3.56							0.05	0.19	0.33
18	3.91	3.87	3.85							0.05	0.19	0.31
19	4.18	4.14	4.12							0.04	0.20	0.32
Total Absolute Percent Relative Deviation				0.13	0.89	0.36	0.46	2.39	7.39	1.66	4.02	6.43
Average Absolute Percent Relative Deviation				0.02	0.11	0.04	0.03	0.17	0.53	0.08	0.20	0.32

In Table C.2, we see that whilst the single cohort design only allows estimation up to the age of 7, the 2-cohort design allows estimation up to the age of 13 and the 3-cohort design allows estimation up to the age of 19 under the same time frame. Up to the age of 7, there is little difference between the designs with respect to their percent relative absolute deviations from the true values. The largest percent relative absolute deviation occurs at the smallest value of the growth curve at age 8 since in this calculation, the absolute relative deviation is divided by the true value in the population which is very small (0.11 for the total and 0.07 for the sex and ethnic minority subgroups). In particular, the ethnic minority (10% of the population) growth curve has a substantial deviation from the true value at this age and the older ages also show larger deviations for both the 2-cohort and 3-cohort designs. Note that for the 2-cohort design, ages 6 and 7 have a larger sample size compared to the other ages due to the overlap and for the 3-cohort design, ages 6 and 7 and ages 12 and 13 have larger sample sizes compared to the other ages. We therefore expect that at these ages, the growth curve estimation should be more precise compared to other ages. However, we see a larger deviation for the sex growth curve for ages 6 and 7. In general, the 3-cohort design shows slightly less precision than the single and 2-cohort designs but this is outweighed by the ability to estimate a growth curve of 20 years within the same time span of 8 years of measurement.

In Table C.3, we test whether there is a significant difference between cohorts in the mean of the outcome variable that was used to estimate the growth curve models shown in Tables C.1 and C.2 using a t-test where the null hypothesis is no difference in means. The aim is to test whether there is a cohort effect at the overlapping ages of the cohorts. For the 2-cohort design, the overlapping ages are 6 and 7, and for the 3 cohort design, the overlapping ages are 6 and 7 and ages 12 and 13. Table C.3 shows the number of times the null hypothesis was rejected in the 500 samples (significant differences in the means), the average mean difference across the 500 samples and simulation standard errors. Note that none of the average mean differences in Table C.3 was significantly different than the value of 0.

Table C.3. Results of t-test for testing differences in mean outcome variable across cohorts in 2-cohort and 3-cohort designs (Simulation A, no cohort effects, 500 samples)

Age	Percent rejections of null hypothesis of equal means	Average Mean Difference (SE)
2- cohort design		
Age 6	6.2	-0.0005 (0.0016)
Age 7	4.6	-0.0006 (0.0013)
3- cohort design		
Age 6	7.2	0.0007 (0.0021)
Age 7	7.2	0.0016 (0.0017)
Age 12	5.6	-0.0024 (0.0017)
Age 13	4.8	-0.0023 (0.0021)

In Table C.3, we see that the average mean difference is not significantly different from zero for each of the overlapping ages according to the simulation standard errors and that the expected number of rejections of the null hypothesis of equal means at 5% is largely respected.

Simulation B: Cohort effects in the population

We repeat the simulation for the single and 2-cohort designs where the true model for generating the population now has a cohort effect and a cohort × age interaction. We continue with a small population of size N=100,000, for which values of the dependent outcome variable are generated across 2 cohorts as well as across ages 0 to 19. The true parameters for generating the population are listed in the second column of Table C.4. Note that the variable 'cohort' under the 2-cohort design is a dummy variable with a value of 1 defined for the first cohort and 0 for the second cohort. This simulation also includes the sex sub-group at 50% and the ethnic minority sub-group at 10% similar to Simulation A.

We drew 500 samples of size 1,000. For the single cohort design we drew the whole sample from the first cohort and for the 2-cohort design we drew half of the sample from the first cohort and half of the sample from the second cohort. Again we assume 8 years of observations so that the single cohort includes observations 0 to 7 and the 2- cohort design includes observations 0 to 7 in the first cohort and 6 to 13 in the second cohort. Given the assumption of small sample sizes representing vulnerable sub-groups, we do not assume missing data at this stage.

To illustrate the difference between Simulation A and Simulation B, we present in Figure C.1 the mean growth curve for one of the samples for the 2-cohort design from Simulation A without cohort effects (panel (a)) and Simulation B with cohort effects where the cohort effects and interaction with age are included in the growth curve model (panel (b)). In panel (c) and panel (d) of Figure C.1, we zoom in on the overlapping age groups at ages 6 and 7 for Simulation A and Simulation B respectively. Panel (d) clearly shows the cohort effects at the overlapping ages and the need to account for the cohort effects in the multilevel growth curve models.

Figure C.1 Mean growth curve for 2-cohort designs from Simulation A (no cohort effects) (panel (a)) and Simulation B (cohort effects) (panel (b)) for one sample. Overlapping ages of Simulation A are in panel (c) and overlapping ages of Simulation B are in panel (d)

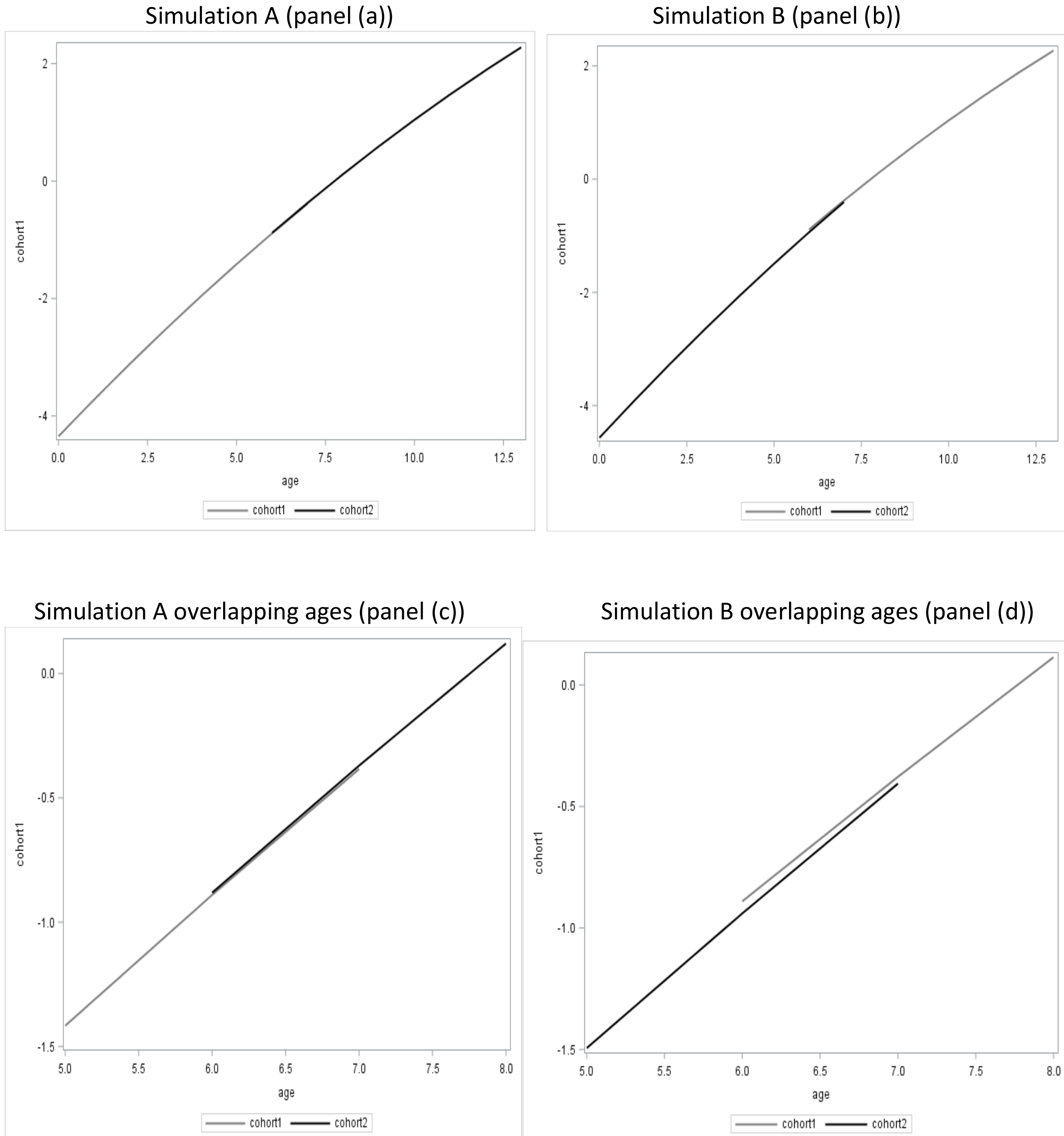


Table C.4 presents the average parameter estimates and their standard errors for the population where cohort effects have been added. Standard errors are calculated from the simulation standard deviation of the parameter estimates across the 500 samples divided by the square root of 500. We present the estimates in Table C.4 for the 2-cohort design with and without the cohort effects included in the model.

Table C.4. Average parameter estimates of growth curve with standard errors (Simulation B, with cohort effects, 500 samples)

	True para-meters	Single		2-cohort – without cohort effects		2-cohort – with cohort effects	
		Parm	SE	Parm	SE	Parm	SE
Fixed Effects							
Intercept	0.85	0.8931*	0.00067	0.8676*	0.00056	0.8535*	0.00066
Cohort	0.04	-	-	-	-	0.0369*	0.00089
Ethnic	-0.008	-0.0031	0.00161	-0.0077*	0.00149	-0.0077*	0.00149
Sex	-0.08	-0.0876*	0.00090	-0.0826*	0.00084	-0.0825*	0.00084
Age	0.45	0.4708*	0.00030	0.4593*	0.00027	0.4509*	0.00035
Ethnic*Age	-0.00002	0.0007	0.00072	0.0003	0.00072	0.0003	0.00072
Sex*Age	-0.0002	0.0003	0.00043	-0.0012*	0.00042	-0.0012*	0.00042
Cohort*Age	0. 02	-	0.00000	-	-	0.0204*	0.00044
Age_squared	-0.01	-0.0100*	0.00001	-0.0103*	0.00001	-0.0100*	0.00001
Random Effects (Unstructured Variance Matrix)							
UN(1,1)	0.08	0.0794*	0.00016	0.0802*	0.00017	0.0799*	0.00017
UN(2,1)	0.004	0.0039*	0.00006	0.0041*	0.00006	0.0039*	0.00006
UN(2,2)	0.02	0.0200*	0.00004	0.0200*	0.00004	0.0199*	0.00004
Residual	0.01	0.0100*	0.00001	0.0100*	0.00001	0.0100*	0.00001

(*) denotes significance at 5% significance level

In Table C.4, under a population generated with cohort effects according to the parameters in the second column of the table, the single cohort design which was drawn from the first cohort only (and had a dummy variable equal to 1) shows good parameter estimation for the main effects and age squared when combining the true values of the cohort effect of 0.04 with the intercept of 0.85 and the cohort*age interaction of 0.02 with the age effect of 0.45. Other parameter estimates are not well-estimated under the single cohort design, particularly the interactions of the ethnic minority and sex with age where there is a change in sign. However, these interaction terms were not significant.

In Table C.4, comparing the 2-cohort design with and without cohort effects included in the multilevel growth curve model, we see slight bias in the intercept and slope (age) when ignoring the cohort effects. These parameter estimates are more precise when cohort effects are added into the model. The cohort effect and the interaction of cohort and age are indeed significant as expected in this simulation study. The ethnic minority interaction with age is not well-estimated similar to the result under the single cohort, but the parameter estimate is not significant. The interaction of sex with age is better estimated and is indeed significant under the 2-cohort design.

Table C.5 shows the true growth curve mean values at each age for the whole population and for the sex and ethnic minority sub-groups. For each sample and at each age, the mean of the predictions of the growth curve was calculated and then averaged over the 500 samples. We present in Table C.5 the percent relative absolute deviation of these averaged mean growth curve predictions from the true mean values based on the single and 2- cohort design (with cohort effects added to the model). We note that all true mean growth curve values are included in the confidence intervals of the averaged mean growth curve predictions according to the simulation standard errors.

Table C.5. True mean values of the growth curve for total and subgroups (Sex 50% and Ethnic Minority 10%) with percent relative absolute deviations of averaged mean growth curve predictions under single and 2-cohort design (Simulation B, with cohort effects, 500 samples)

Age	Single Cohort						2- cohort design					
	True Mean Values			Percent Relative Absolute Deviation from Mean Predicted Values			True Mean Values			Percent Relative Absolute Deviation from Mean Predicted Values		
	All	Sex	Eth-nic	All	Sex	Eth-nic	All	Sex	Eth-nic	All	Sex	Eth-nic
0	-4.53	-4.57	-4.58	0.00	0.01	0.05	-4.53	-4.57	-4.58	0.05	0.04	0.15
1	-3.88	-3.92	-3.93	0.02	0.02	0.10	-3.88	-3.92	-3.93	0.02	0.02	0.21
2	-3.25	-3.29	-3.30	0.01	0.02	0.06	-3.25	-3.29	-3.30	0.04	0.02	0.19
3	-2.63	-2.68	-2.68	0.02	0.04	0.04	-2.63	-2.68	-2.68	0.02	0.00	0.19
4	-2.04	-2.09	-2.10	0.01	0.00	0.08	-2.04	-2.09	-2.10	0.03	0.04	0.10
5	-1.47	-1.52	-1.52	0.04	0.01	0.11	-1.47	-1.52	-1.52	0.07	0.05	0.34
6	-0.92	-0.97	-0.97	0.15	0.11	0.14	-0.91	-0.95	-0.95	0.09	0.20	0.11
7	-0.39	-0.44	-0.44	0.02	0.11	0.31	-0.38	-0.43	-0.43	0.07	0.31	0.03
8							0.11	0.07	0.06	1.22	3.02	7.22
9							0.58	0.54	0.53	0.25	0.43	0.62
10							1.03	0.99	0.98	0.11	0.15	0.04
11							1.46	1.42	1.42	0.14	0.07	0.20
12							1.87	1.83	1.82	0.14	0.08	0.05
13							2.26	2.22	2.21	0.11	0.02	0.16
Total Absolute Percent Relative Deviation				0.26	0.32	0.89	Total Absolute Percent Relative Deviation			2.36	4.43	9.61
Average Absolute Percent Relative Deviation				0.03	0.04	0.11	Average Absolute Percent Relative Deviation			0.17	0.32	0.69

In Table C.5, we see that whilst the single cohort design only allows estimation up to the age of 7, the 2-cohort design allows estimation up to the age of 13. Under the single cohort design, we see low percent relative absolute deviations from the true values up to the age of 7 with the ethnic minority having larger deviations compared to the total and the sex group. Similar to Table C.2 under the 2-cohort design, the largest deviation occurs at the smallest value of the growth curve at age 8 since the calculation involves dividing by the true value which is 0.11 for the total, 0.07 for the sex group and 0.06 for the ethnic minority group. The ethnic minority group have the largest deviations under the 2-cohort design as well. Also, compared to Table C.2 we see larger deviations under the 2-cohort design where the population was generated with a cohort effect compared to the case where no cohort effect was introduced into the population in Simulation A. Under the 2-cohort design the sample size is larger at ages 6 and 7 compared to the other ages, but we do not see a particular improvement in the mean prediction of the growth curve at these ages.

Table C.6 contains the test for significant differences between cohorts in the mean of the dependent outcome variable that was used to estimate the growth curve models shown in Tables C.4 and C.5 using a t-test where the null hypothesis is no differences in means. The aim is to test whether there is a cohort effect at the overlapping ages of the cohorts. Table C.6 shows the number of times the null hypothesis was rejected in the 500 samples (significant differences in the means), the average mean difference across the 500 samples and simulation standard errors.

Table C.6. Results of t-test for testing differences in mean outcome variable across cohorts in 2-cohort design (Simulation B, with cohort effects, 500 samples)

Age	Percent rejections of null hypothesis of equal means	Average Mean Difference (SE)
2 cohort design		
Age 6	15.4%	-0.0354* (0.0015)
Age 7	7.6%	0.0134* (0.0012)

(*) denotes significance at 5% significance level

In Table C.6, we see that the average mean difference is significantly different from zero for each of the overlapping ages according to the simulation standard errors and that the expected number of rejections of the null hypothesis of equal means is greater than 5%. Further work will be to assess the power of statistical testing on larger sample sizes.

We summarize the findings of the simulation study in Section 3.2.

Bibliography

Allison, P. D. (2001). *Missing data*. London: Sage publications.

Anderson, E. R. (1993). Analyzing change in short-term longitudinal research using cohort-sequential designs. *Journal of Consulting and Clinical Psychology*, *61*(6), 929–940.

Arim, R. G., Findlay, L. C., and Kohen, D. E. (2016). What statistics Canada survey data sources are available to study neurodevelopmental conditions and disabilities in children and youth? *University of Calgary. The School of Public Policy*, *9*(29), 36.

Australian Institute of Family Studies. (2015, November). *Longitudinal Study of Australian Children Data User Guide*. Retrieved from https://growingupinaustralia.gov.au/sites/default/files/userguide.pdf

Baker, M., and Milligan, K. (2016). Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries. *Journal of Human Capital*, 43.

Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, *11*, 145–171.

Bell, R. Q. (1953). Convergence: An Accelerated Longitudinal Approach. *Child Development*, *24*(2), 145–152.

Bell, R. Q. (1954). An Experimental Test of the Accelerated Longitudinal Approach. *Child Development*, *25*(4), 281–286.

Bollen, K. A., and Curran, P. J. (2006). *Latent Curve Models: A Structural Equation Perspective*. https://doi.org/10.1002/0471746096

Boller, K., and Harman-Smith, Y. (2019). *Measuring Up: Learning About Improving Equity from Australia's Early Development Census*. Princeton, NL: Mathematica Policy Research.

Brown, M. and Sullivan, A. (2014). Cognitive development. In L. Platt (Ed.), *Millennium Cohort Study: Initial findings from the age 11 survey*. Retrieved from http://www.cls.ioe.ac.uk/page.aspx? and sitesectionid=1330 and sitesectiontitle=MCS+age+11+initial+findings

Bureau of Labor Statistics. (2019). NLSY79 Children and Young Adults. Retrieved 17 May 2019, from https://www.bls.gov/nls/nlsy79ch.htm

Cheung, Y. B. (2013). *Statistical analysis of human growth and development*. New York: Chapman and Hall.

Connelly, R., and Platt, L. (2014). Cohort Profile: UK Millennium Cohort Study (MCS). *International Journal of Epidemiology*, *43*(6), 1719–1725. https://doi.org/10.1093/ije/dyu001

Côté, S., Vaillancourt, T., LeBlanc, J. C., Nagin, D. S., and Tremblay, R. E. (2006). *The Development of Physical Aggression from Toddlerhood to Pre-Adolescence: A Nation Wide Longitudinal Study of Canadian Children*. *34*(1), 68–82.

Davis-Kean, P., Chambers, R. L., Davidson, L. L., Kleinert, C., Ren, Q., and Tang, S. (2018). *Longitudinal Studies Strategic Review: 2017 Report to the Economic and Social Research Council*. Economic and Social Research Council.

Döpfner, M., Hautmann, C., Görtz-Dorten, A., Klasen, F., Ravens-Sieberer, U., and The BELLA study group. (2015). Long-term course of ADHD symptoms from childhood to early adulthood in a community sample. *European Child and Adolescent Psychiatry*, *24*(6), 665–673. https://doi.org/10.1007/s00787-014-0634-8

Duncan, S. C., Duncan, T. E., and Hops, H. (1996). Analysis of longitudinal data within accelerated longitudinal designs. *Psychological Methods*, *1*(3), 236.

Duncan, S. C., Duncan, T. E., and Strycker, L. A. (2001). Qualitative and Quantitative Shifts in Adolescent Problem Behavior Development: A Cohort-Sequential Multivariate Latent Growth Modeling Approach. *SPH P*, 8.

Duncan, T. E., Duncan, S. C., and Strycker, L. A. (2011). *An introduction to latent variable growth curve modeling: Concepts, issues, and application* (2nd ed.). Psychology Press.

Enders, C. K. (2010). *Applied missing data analysis*. New York: The Guilford Press.

Farrington, D. P. (1991). Longitudinal research strategies: Advantages, problems, and prospects. *Journal of the American Academy of Child and Adolescent Psychiatry*, *30*(3), 369–374.

Finkel, D., Reynolds, C. A., McArdle, J. J., and Pedersen, N. L. (2007). Cohort Differences in Trajectories of Cognitive Aging. *The Journals of Gerontology: Series B*, *62B*(5), P286–P294. https://doi.org/10.1093/geronb/62.5.P286

Fitzmaurice, G. M., Laird, N. M., and Ware, J. H. (2004). *Applied longitudinal analysis*. Boston, MA: Wiley-Interscience.

Galbraith, S., Bowden, J., and Mander, A. (2017). Accelerated longitudinal designs: An overview of modelling, power, costs and handling missing data. *Statistical Methods in Medical Research*, *26*(1), 374–398. https://doi.org/10.1177/0962280214547150

Gerstorf, D., Ram, N., Hoppmann, C., Willis, S. L., and Schaie, K. W. (2011). Cohort differences in cognitive aging and terminal decline in the Seattle Longitudinal Study. *Developmental Psychology*, *47*(4), 1026–1041. https://doi.org/10.1037/a0023426

Glenn, N. D. (1976). Cohort Analysts' Futile Quest: Statistical Attempts to Separate Age, Period and Cohort Effects. *American Sociological Review*, *41*(5), 900–904. https://doi.org/10.2307/2094738

Gubhaju, L., Banks, E., Ward, J., D'Este, C., Ivers, R., Roseby, R., … Eades, S. J. (2019). 'Next Generation Youth Well-being Study:' understanding the health and social well-being trajectories of Australian Aboriginal adolescents aged 10–24 years: study protocol. *BMJ Open*, *9*(3), 1–8. https://doi.org/10.1136/bmjopen-2018-028734

Gunnison, E. (2015). Investigating Life Course Offender Subgroup Heterogeneity: An Exploratory Latent Class Analysis Approach. *Women and Criminal Justice*, *25*(4), 223–240. https://doi.org/10.1080/08974454.2014.989300

Gutman, L. M., Joshi, H., Parsonage, M., and Schoon, I. (2018). Gender-specific trajectories of conduct problems from ages 3 to 11. *Journal of Abnormal Child Psychology*, *46*(7), 1467–1480.

Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., and Wigfield, A. (2002). Changes in Children's Self-Competence and Values: Gender and Domain Differences across Grades One through Twelve. *Child Development*, *73*(2), 509–527. Retrieved from JSTOR.

Joshi, H., and Fitzsimons, E. (2016). The Millennium Cohort Study: the making of a multi-purpose resource for social science and policy. *Longitudinal and Life Course Studies*, *7*(4), 409–430.

Kamtsiuris, P., Lange, M., and Schaffrath, A. R. (2007). The German Health Interview and Examination Survey for Children and Adolescents (KiGGS): sample design, response and nonresponse analysis (in German). *Bundesgesundheitsbl*, *50*(5/6), 547–556.

Klipker, K., Baumgarten, F., Göbel, K., Lampert, T., and Hölling, H. (2018). Mental health problems in children and adolescents in Germany. Results of the cross-sectional KiGGS Wave 2 study and trends. *Journal of Health Monitoring*, *3*(3), 34–41. https://doi.org/10.17886/rki-gbe-2018-084

Li, F., Barrera, M., Hops, H., and Fisher, K. J. (2002). The Longitudinal Influence of Peers on the Development of Alcohol Use in Late Adolescence: A Growth Mixture Analysis. *Journal of Behavioral Medicine*, *25*(3), 293–315.

Martin, J., Bynner, J., Kalton, G., Boyle, P., Goldstein, H., Gayle, V., … Piesse, A. (2006). *Strategic Review of Panel and Cohort Studies* (pp. 1–85). London: Economic and Social Research Council.

McArdle, J. J., Ferrer-Caja, E., Hamagami, F., and Woodcock, R. W. (2002). Comparative longitudinal structural analyses of the growth and decline of multiple intellectual abilities over the life span. *Developmental Psychology*, *38*(1), 115–142.

McMunn, A., Lacey, R., Worts, D., McDonough, P., Stafford, M., Booker, C., … Sacker, A. (2015). De-standardization and gender convergence in work–family life courses in Great Britain: A multi-channel sequence analysis. *Advances in Life Course Research*, *26*, 60–75.

Meredith, W., and Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122.

Miyazaki, Y., and Raudenbush, S. W. (2000). Tests for linkage of multiple cohorts in an accelerated longitudinal design. *Psychological Methods*, *5*(1), 44–63. https://doi.org/10.1037//1082-989X.5.1.44

Muthén, B O, and Muthén, L. K. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol*, *61*(2), 290–300. https://doi.org/10.15288/jsa.2000.61.290

Muthén, Bengt O, and Muthén, L. K. (2000). The development of heavy drinking and alcohol-related problems from ages 18 to 37 in a U.S. national sample. *Journal of Studies on Alcohol*, *61*(2), 290–300. https://doi.org/10.15288/jsa.2000.61.290

NCVER. (2013a). *Longitudinal Surveys of Australian Youth (LSAY): 1995 Cohort user guide, Technical Report 49*. Retrieved from https://www.lsay.edu.au/__data/assets/pdf_file/0031/181876/LSAY_TechnicalReport_49_2122.pdf

NCVER. (2013b). *Longitudinal Surveys of Australian Youth (LSAY): 1998 Cohort user guide, Technical Report 53*. Retrieved from https://www.lsay.edu.au/__data/assets/pdf_file/0016/181501/LSAY_TechnicalReport_53_2199.pdf

NCVER. (2018). *Longitudinal Surveys of Australian Youth (LSAY): 2009 Cohort user guide, Technical Paper 74*. Retrieved from https://www.lsay.edu.au/__data/assets/pdf_file/0017/181511/LSAY_2009_cohort_user_2547.pdf

Nesselroade, J. R., and Baltes, P. B. (1979). *Longitudinal research in the study of behavior and development*. Retrieved from https://books.google.co.uk/books/about/Longitudinal_research_in_the_study_of_be.html?id=5ZhpAAAAMAAJ and redir_esc=y

Nicholson, J., Sanson, A., Rempel, L., Smart, D., and Patton, G. (2002). Longitudinal studies of children and youth: Implications for future studies. *Children's Health and Development: New Research Directions for Australia*, 38–59.

Orth, U., Trzesniewski, K. H., and Robins, R. W. (2010). Self-esteem development from young adulthood to old age: A cohort-sequential longitudinal study. *Journal of Personality and Social Psychology*, *98*(4), 645–658. https://doi.org/10.1037/a0018769

Pearce, A., Hope, S., Lewis, H., and Law, C. (2014). Pearce, A., Hope, S., Lewis, H., and Law, C. (2014). Family structure and socio-emotional wellbeing in the early years: a life course approach. Longitudinal and life course studies, 5(3), 263-282. *Longitudinal and Life Course Studies*, *5*(3), 263–282.

Prinzie, P., Onghena, P., and Hellinckx, W. (2006). A cohort-sequential multivariate latent growth curve analysis of normative CBCL aggressive and delinquent problem behavior: Associations with harsh discipline and gender. *International Journal of Behavioral Development*, *30*(5), 444–459. https://doi.org/10.1177/0165025406071901

Rabe-Hesketh, S., and Skrondal, A. (2012). *Multilevel and longitudinal modeling using Stata* (3rd ed.). Texas: STATA press.

Raudenbush, S. W., and Chan, W.-S. (1992). Growth Curve Analysis in Accelerated Longitudinal Designs. *Journal of Research in Crime and Delinquency*, *29*(4), 387–411.

Raudenbush, S. W., and Chan, W.-S. (1993). Application of a hierarchical linear model to the study of adolescent deviance in an overlapping cohort design. *Journal of Consulting and Clinical Psychology*, *61*(6), 941–951.

Rawana, J. S., and Ames, M. E. (2012). Protective Predictors of Alcohol Use Trajectories Among Canadian Aboriginal Youth. *Journal of Youth and Adolescence*, *41*(2). Retrieved from https://scholar.google.co.uk/scholar?hl=en and as_sdt=0%2C5 and q=Rawana+%26+Ames+%282012%29 and btnG=

Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith, and S. C. Rawlings (Eds.), *Methodological issues in aging research*. New York: Springer.

Sanson, A. (2002). *Children's health and development: New research directions for Australia.* (Text No. 8). Retrieved from Australian Institute of Family Studies website: https://aifs.gov.au/publications/childrens-health-and-development/3-longitudinal-studies-children-and-youth-implications-future-studies

Schaie, K. Warner. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, *64*(2), 92–107. https://doi.org/10.1037/h0022371

Schoon, I., Barnes, M., Brown, V., Parsons, S., Ross, A., and Vignoles, A. (2012). *Intergenerational transmission of worklessness: Evidence from the Millennium Cohort and the Longitudinal Study of Young People In England* (No. DFE-RR234). Retrieved from Department for Education website: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/183328/DFE-RR234.pdf

Statistics Canada. (2019a). *Microdata User Guide. National Longitudinal Survey of Children and Youth (NLSCY), Cycle 8. September 2008 to July 2009.* Retrieved from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey and Id=56797

Statistics Canada. (2019b). Survey of Young Canadians (SYC). Retrieved 17 May 2019, from http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey and Id=81301#a1

Stoel, R. D., Van Den Wittenboer, G., and Hox, J. (2003). Analyzing Longitudinal Data using Multilevel Regression and Latent Growth Curve Analysis. - PsycNET. *Metodologia de Las Ciencias Del Comportamiento*, *5*, 21–42.

Strohschein, L., and Gauthier, A. H. (2018). Poverty dynamics, parenting, and child mental health in Canada. *Society and Mental Health*, *8*(3), 231–247.

Tonry, M., Ohlin, L. E., and Farrington, D. P. (1991). *Human Development and Criminal Behavior: New Ways of Advancing Knowledge*. New York: Springer-Verlag.

Tampubolon, G. and Maharani. A. (2018). Trajectories of allostatic load among older Americans and Britons: longitudinal cohort studies. *BMC Geriatrics* , 18 (1) 255.

Ward, C., Lewis, S., and Coleman, T. (2007). Prevalence of maternal smoking and environmental tobacco smoke exposure during pregnancy and impact on birth weight: retrospective study using Millennium Cohort. *BMC Public Health*, *7*(1). https://doi.org/10.1186/1471-2458-7-81

Watt, H. M. G. (2008). A latent growth curve modeling approach using an accelerated longitudinal design: the ontogeny of boys' and girls' talent perceptions and intrinsic values through adolescence. *Educational Research and Evaluation*, *14*(4), 287–304. https://doi.org/10.1080/13803610802249316

Weber, D. (2016). Differences in physical aging measured by walking speed: evidence from the English Longitudinal Study of Ageing. *BMC Geriatrics*, 28, 16-31.

Zaninotto, P., Batty, G.D., Allerhand, M. and Deary, I.J. (2018). Cognitive function trajectories and their determinants in older people: 8 years of follow-up in the English Longitudinal Study of Ageing. *Journal of Epidemiology Community Health*, 72, 685-694.