The Lifestyle During Pregnancy Study - Brief introduction to analysis of data

Maiken Ina Siegismund Kjaersgaard & Erik Thorlund Parner
Department of Public Health, Section of Biostatistics, Aarhus University, Denmark

Version: July 23 2012

Table of contents

Introduction

This note presents an introduction to the basic variables and statistical analyses of the Lifestyle During Pregnancy Study (LDPS). The LDPS studies the associations between alcohol intake and intellectual ability (IQ), attention (TEACh-5), executive functioning (BRIEF), social skills (SDQ), and motor development. The emphasis in the note is on individuals' sampling probability and on imputing missing values by multiple imputation. For questions regarding the specific outcomes and decisions made in the analyses hereof the reader is encouraged to read specific papers mentioned, and contact the corresponding authors. The overall study design and population are described in detail in:

Kesmodel US, Underbjerg M, Kilburn TR et al. Lifestyle during pregnancy: Neurodevelopmental effects at 5 years of age. The design and implementation of a prospective follow-up study. Scand J Public Health 2010; 38(2):208-19.

The different outcomes are studied in:

Eriksen HLF, Mortensen EL, Kilburn TR et al. The effects of low to moderate prenatal alcohol exposure in early pregnancy on IQ in 5-year-old children. Accepted for publication in BJOG.

Kesmodel US, Eriksen HLF, Underbjerg M et al. The effect of alcohol binge drinking in early pregnancy on child general intelligence. Accepted for publication in BJOG.

Underbjerg M, Kesmodel US, Landrø NI et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on selective and sustained attention in five-year-old children. Accepted for publication in BJOG.

Skogerbø Å, Kesmodel US, Wimberley T et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on executive function in five-year-old children. Accepted for publication in BJOG.

Skogerbø Å, Kesmodel US, Denny CH et al. The effects of low to moderate alcohol consumption and binge drinking in early pregnancy on behavior in five-year-old children. In preparation.

Bay B, Støvring H, Wimberley T et al. Low to moderate alcohol intake during pregnancy and risk of psychomotor deficits. Alcohol Clin Exp Res 2012; 36(5):807-14.

Kesmodel US, Bay B, Støvring H et al. Does binge drinking in early pregnancy increase the risk of psychomotor deficits? In preparation.

Please note that this documentation does not describe the study on the association between alcohol intake and the Sternberg test of speed of information processing. Manuscript on Sternberg by Kilburn TR et al. is in preparation. Also, the study on multiple outcomes is not covered by this documentation, and the interested reader is referred to the publication:

Kesmodel US, Bertrand J, Støvring H et al. The effect of different alcohol drinking patterns in early to mid pregnancy on the child's intelligence, attention, and executive function. Accepted for publication in BJOG.

Sampling strategy

The Lifestyle During Pregnancy Study was designed to study two distinct alcohol drinking patterns: Average intake among four groups (0; 1–4; 5–8; and 9+ drinks per week) and binge drinking (yes or no). To study the interaction between average alcohol intake and binge drinking, seven main categories were defined according to either average drinking levels *during pregnancy* (categories 1–5) or *pre-pregnancy* drinking levels (categories 6–7). Each of these seven categories was further stratified on the basis of binge drinking patterns: whether or not binge drinking occurred and, if so, during which weeks during the pregnancy the binge episodes occurred. See Kesmodel et al. (Scand J Public Health, 2010) for details. The sample was stratified on both average alcohol intake and binge drinking to ensure adequate representation in every stratum. Each stratum is then sampled as an independent subpopulation, out of which individuals were randomly sampled. A sample where individuals are not sampled with the same probability is generally called a survey sample.

The dataset

The LDPS dataset is called `LDPS_data`. This dataset was created by Associate Professor, PhD Henrik Støvring and M.Sc. Theresa Wimberley from the cleaned datasets provided by Associate Professor, PhD Jakob Grove. Henrik Støvring has provided the Stata do-file `Analysedata_alle` (see the file Analysedata_alle.pdf) which shows how the various source datasets were combined into the analysis dataset. The only changes made to the dataset provided by Henrik Støvring were a later update of the sampling fractions (see the section on sampling fractions) and the computation of the exposure variables for average alcohol intake (see the section on exposure). Please note that the dataset `LDPS_data` does not include twins (see Figure 1 and Figure 2) even though sampling fractions are calculated on the actual number of invited participants including twins since some twins (n=11) actually were invited.

The dataset contains a variable for the alcohol category, `kategori`, to which each individual was sampled. Furthermore, an indicator variable, `excl_bef`, describing whether an individual was sampled on alcohol intake during pregnancy (categories 1–5) or on pre-pregnancy drinking (categories 6–7), is given. This variable can be used to include or exclude observations according to whether the analyses should be on women sampled on alcohol intake during pregnancy or on women sampled on pre-pregnancy drinking, or both.

Another variable of interest is the variable containing information on participation, `particind_cat`. This variable indicates whether a woman participated in the LDPS, did not participate in the LDPS, or returned the questionnaire but without being tested. Only a subgroup of participating women was invited to participate in the tests on motor development. This information can be found in the variable `ind_motor`.

Sampling fractions

Definition and computation

In survey sampling, the standard strategy according to the statistical literature is to calculate the sampling fractions for each stratum and use these as inverse probability weights in a weighted analysis and computing the precision of the estimates by a statistical technique named robust variance estimation (Kish, 1965).

In practice, survey researchers sample without replacement but calculate the sampling fractions associated with their samples as if they were sampling with replacement because the latter calculations are much easier. Survey research typically involves relatively small samples from huge populations, in which case sampling with and without replacement are equivalent for all practical purposes.

Data for the LDPS were sampled from the Danish National Birth Cohort (DNBC) without replacement. The sampling was done by a data manager, and the data were subsequently handed over to the statisticians for analysis. The number of women in each stratum as well as the number of invited participants from each stratum obtained from these data are given in Table 1 depicting the sampling scheme for the LDPS. If the sampling fractions are calculated as if sampling were done without replacement we would obtain an individual sampling fraction for each woman in each stratum. As an example we will look at category 1. The first woman sampled in this category would then have had the sampling fraction 1/39,004, the second woman would have had the sampling fraction 1/(39,004-1), the third woman would have had the sampling fraction 1/(39,004-2), and so on.

As it is, the sampling fractions will be calculated as if data were sampled with replacement simply by dividing the number sampled in each category 1-5 by the number identified in the DNBC in a given category 1-5, i.e. if we look at category 1 the sampling fraction for each individual in category 1 is calculated as 579/39,004=0.0148446. The sampling fractions are given in Table 1.

The categories 1-5 are mutually exclusive, i.e. if a woman was sampled for, say category 1a, she could not have been sampled for, say category 2c. Likewise, the categories 6-7 are mutually exclusive. Only, the categories 1-5 and 6-7 are not mutually exclusive, and since the categories 6-7 are sampled after the categories 1-5 have been sampled the calculations of the sampling fractions for the categories 6-7 need to take this into account making the calculation slightly more substantial for individuals in the categories 6-7. The sampling fractions for the women sampled for categories 6-7 cannot be summarized by a single sampling fraction, as was the case for the categories 1-5, but rather there will be a sampling fraction for each individual sampled depending on which category 1-5 the individual could initially have been sampled for and the category 6 or 7 to which the individual actually was sampled.

To obtain the sampling fractions in category 6 and 7 we need first to identify the category to which each individual could have been sampled (if possible) when sampling for the categories 1-5 was done. The corresponding sampling fractions are subtracted from 1 resulting in individual probabilities of not being sampled for categories 1-5.

Next we will calculate the probability of being sampled for category 6 or 7 given that the women have not been sampled previously for one of the categories 1-5. Note that these sampling probabilities are *conditional* on the women not being sampled for categories 1-5, which is sufficient for the comparison of category 6 and 7. Thus the conditional sampling fractions cannot be used for combining data from categories 1-5 with data from categories 6-7. First, the number of women in the strata 6 and 7 are identified, and then these numbers are corrected to only contain women not already sampled resulting in 10,367 and 974 women for the categories 6 and 7, respectively. The sampling fractions for category 6 and 7 in Table 1 are therefore not the final sampling fractions, but rather the probability of being sampled for category 6 or 7 given that the ones considered for sampling have not been sampled previously for one of the categories 1-5. Multiplying these conditional probabilities by the individual probabilities of not being sampled for categories 1-5 leaves us with the individual sampling fractions to be used. As an example we will look at an individual sampled for category 7 that could have been sampled for category 2c, but was not. The sampling fraction for category 2c is given by 0.1379071, and the probability of being sampled for category 7 given that the ones considered for sampling have not been sampled previously for one of the categories 1-5 is given by 0.150924. The individual will therefore have a sampling fraction given by (1-0.1379071)*0.150924=0.13011051.

Please note that sampling was done on the DNBC including twins. Sampling fractions are calculated on the actual number of invited participants including twins since some twins actually got invited, but in the analysis in the LDPS these twins are excluded. The numbers corrected for twins that are used in the analyses are given in Table 1.

Also, the data include women sampled on fish oil and thus not sampled on alcohol consumption. The women sampled on fish oil were added to the categories 1-5 according to their alcohol consumption pattern if possible. The remaining women were added to the categories 6-7 if possible. One woman could not be added to any category because of missing information on alcohol consumption. See Figure 1 (or Figure 2).

How to incorporate in analyses

The LDPS dataset `LDPS_data` contains a variable named `sampfrac` containing the sampling fractions calculated as described above.

Sampling weights or pweights are weights that denote the inverse of the probability that the observation is included because of the sampling design, i.e. the inverse of the sampling fraction, and can be incorporated in a regression analysis by adding `[pweight=1/sampfrac]` to the regression command line.

Exposure

Average alcohol intake during pregnancy (0; 1–4; 5–8; and 9+ drinks per week) is created using the variable `int1_undergraviditet` and is contained in the variable `av_during`. Average alcohol intake before pregnancy (0; 15-21; and 22+ drinks per week) is created using the variable `int1_foergraviditet` and is contained in the variable `av_before`. Please note that whereas `av_during` has no missing values `av_before` has many missings since the information on alcohol consumption before pregnancy for women in the LDPS were missing for many of the women in the DNBC from which the LDPS dataset is sampled. Women sampled on pre-pregnancy intake have no missing information on alcohol intake before pregnancy though.

As for binge drinking in pregnancy (yes or no), the information is contained in the variable `binge`. The number of binge drinking episodes (0; 1; 2; and 3+) is contained in the variable `nbinge`, and the timing of binge drinking episodes (no binge drinking; 1-2; 3-4; 5-8; and 9+; multiple episodes) is contained in the variable `binge_timing`.

Some women had missing information on timing of binge drinking episode(s) or invalid information on binge drinking. In the LDPS the decision was made to recode the binge timing variable, `binge_timing`, as week 9+ if an observation was sampled to either category 1d or category 2d when `binge_timing` was missing (n=4 and n=5, respectively). This is not done in dataset `LDPS_data` and must be done by the researcher if deemed appropriate. As mentioned, some women had invalid binge information. Some women with missing `binge_timing` information sampled to the categories 1, 2, 4 or 5a reported a number of binge episodes > 0 in the variable `a145a`, which is inconsistent with these women being sampled to one of the categories 1, 2, 4 or 5a. In the LDPS these women (n=18) were subsequently dropped before analysis using binge as exposure (see Figure 1).

Outcome

This note focuses on imputing missing values using multiple imputation as was done in the LDPS, and the following description of generating outcomes builds upon that imputing missing values is needed. Of course it is also possible to compute outcomes without doing multiple imputation and consequently do complete-case analyses. It is up to the researcher to do what is deemed appropriate.

Computation of IQ outcome

Intelligence is assessed using the Wechsler Primary and Preschool Scales of Intelligence – Revised (WPPSI-R). This revised version of the WPPSI includes five verbal sub tests and five performance (non-verbal) sub tests, from which an overall verbal IQ, an overall performance IQ and the total IQ can be derived. In the LDPS, only three verbal sub tests and three performance sub tests were administered during the test sessions:

The three verbal sub tests are:

- arithmetic (`aritmeti`)
- information (`informat`)
- vocabulary (`vocabula`)

The three performance sub tests are:

- block design (`block_as`)
- geometric design (`figure_d`)
- object assembly (`object_a`)

The six sub test variables mentioned above contain raw scores. Scaled scores on the six sub tests (`aritmeti2`, `informat2`, `vocabula2`, `block_as2`, `figure_d2`, `object_a2`) are derived using Swedish norms. These norms are available for use in the data file wppsi.csv. Children are tested sometime during the first 4 months after their fifth birthday, and different scaled scores are derived for children aged 5 years and 1-3 months and 5 years and 4 months since we expect differences in IQ at that age level to occur even during a time period of only 4 months, and a period of 3 months is often seen in literature.

Two composite IQ measures are derived as the sum over the three verbal scaled scores (`aritmeti2`, `informat2`, `vocabula2`) and the sum over the three performance scaled scores (`block_as2`, `figure_d2`, `object_a2`), respectively. In both cases, if one out of the three scaled scores is missing, the composite IQ measure are calculated as the sum over the two scaled scores not missing multiplied by 3/2, thereby resulting in an imputed value for the sum (note that the missing scaled score and the corresponding missing raw score will still be missing). Otherwise, multiple imputation using linear regression on the six raw scores is performed, and consequently the two composite IQ measures are derived after imputation of the missing raw scores and the subsequent conversion of the imputed raw scores to scaled scores. To simplify we fit a simple linear regression of the norm table's scaled scores on its raw scores, which enables us to derive (approximate) scaled scores for the imputed raw scores without having to look up every single imputed raw score in the norm table. Furthermore, another composite IQ measure is derived as the sum over the two aforementioned composite IQ scores, i.e. if none of the six scaled scores are missing it is simply the sum over all six scaled scores.

The three composite IQ measures above are then probated by multiplying by 5/3 (WPPSI is derived for five tests but we only have three), and scaled scores on the resulting corrected sums are derived using Swedish norms (the data file wppsi.csv) resulting in the three outcomes for IQ: the overall verbal IQ (VIQ), the overall performance IQ (PIQ) and the total IQ, also called the full scale IQ (FSIQ). These are furthermore dichotomized using the sample mean minus one SD as cut-off for below average IQ.

The three outcomes for IQ (VIQ, PIQ and FSIQ) are already computed in the LDPS dataset for women not requiring imputation (`verbaliq`, `performiq` and `fulliq`, respectively). These outcomes will need to be computed after imputation for women with missing information on either one of these though, following the steps described above.

Computation of TEACh-5 outcome

Attention is assessed using the Test of Everyday Attention for Children at Five (TEACh-5), from which several sub test scores can be derived. Two sub tests assessing selective attention (Great Balloon Hunt and Hide and Seek II) and two sub tests assessing sustained attention (Barking and Draw-a-Line) have been selected, from which a composite overall mean attention score is derived. In addition, a composite mean attention score is calculated for selective attention as well as for sustained attention.

The four sub test scores are:

- Barking (`barking_slow`): The number of correct answers to the question how many barks sounded in each of the six 'slowly presented' sound clips. The distribution is negatively skewed. We will however consider it an approximately symmetric distribution.
- Draw-a-Line (`logdraw`): Time used to trace a line as slowly as possible without stopping and without lifting the pen. Time is log-transformed to obtain approximate normality.
- Hide and Seek II (`logHS2_present_score3`): The task is to report whether a bark from a dog was absent or present in a total of 14 sound clips, and the level of performance was measured as the mean reaction time to giving a correct answer if a bark is present divided by the total number of correct answers. This ratio is subsequently divided by the total number of sound clips. The Hide and Seek II compound score is log-transformed to obtain approximate normality.
- Great Balloon Hunt (`c2_plus_c3`): The sum of the number of balloons marked in a sheet with only balloons and a sheet with balloons and visual distractions. The sub test score follows approximately a normal distribution.

Constructing the four sub test scores is performed as follows:

- The Barking sub test score, `barking_slow`, is constructed as the sum over the number of correct answers to the six 'slowly presented' sound clips given by the variables `wpp_a5_1_3`, wpp_a5_1_4, wpp_a5_1_5, wpp_a5_1_7, wpp_a5_1_8 and wpp_a5_1_10 with the correct answers 3, 4, 5, 4, 5 and 6 barks, respectively, for the children completing the Barking sub test indicated by the variable chk_a5_hund_2 taking on a value different from 1.

- The Draw-a-Line sub test score, logdraw, is constructed as the logarithm to the variable wpp_a5_4_1 containing the time used to trace the line.
- The Hide and Seek II compound score, logHS2_present_score3, is constructed as mean reaction time to giving a correct answer if a bark is present divided by the total number of correct answers which is divided by the total number of sound clips for children completing the Hide and Seek II sub test indicated by the variable chk_a5_gemaud_2 taking on a value different from 1. The mean reaction time to giving a correct answer if a bark is present is constructed as the indicator functions for correct answers if a bark is present, wpp_a5_6_5=="v", wpp_a5_6_7=="v", wpp_a5_6_12=="v", wpp_a5_6_14=="v", wpp_a5_6_30=="v", wpp_a5_6_31=="v", wpp_a5_6_32=="v", wpp_a5_6_33=="v" and wpp_a5_6_34=="v", where "v" indicates a correct answer as opposed to "x" indicating a wrong answer, times the corresponding time to the answer subtracted the stimuli time, wpp_a5_6_1-8, wpp_a5_6_2-2, wpp_a5_6_3-7, wpp_a5_6_4-2, wpp_a5_6_25-3, wpp_a5_6_26-7, wpp_a5_6_27-3, wpp_a5_6_28-8 and wpp_a5_6_29-8, respectively, and divided by the total number of correct answers if a bark is present, which is the sum of the indicator functions mentioned above.

  The total number of correct answers is constructed as the sum of the indicator functions for a correct answer, wpp_a5_6_5=="v", wpp_a5_6_7=="v", wpp_a5_6_12=="v", wpp_a5_6_14=="v", wpp_a5_6_30=="v", wpp_a5_6_31=="v", wpp_a5_6_32=="v", wpp_a5_6_33=="v", wpp_a5_6_34=="v", wpp_a5_6_15=="v", wpp_a5_6_16=="v", wpp_a5_6_17=="v", wpp_a5_6_35=="v" and wpp_a5_6_36=="v".

- The Great Balloon Hunt sub test score, c2_plus_c3, is constructed as the sum of the variables wpp_a5_0_3 and wpp_a5_0_4 containing the number of balloons in the sheet with only balloons and the sheet with balloons and visual distractions, respectively.

The four sub test scores (barking_slow, logdraw, logHS2_present_score3 and c2_plus_c3) are already computed in the LDPS dataset for women not requiring imputation. After imputing missing values for the four sub test scores, these are standardised to a mean of 0 and a SD of 1. In particular, for Hide and Seek II the standardised score is reversed to ensure that higher score reflects better performance.

A composite overall mean attention score will be derived as the mean of the four standardised sub test scores standardised to a mean of 0 and a SD of 1. Similarly, the composite selective mean attention score and the composite sustained mean attention score are derived as the mean of the two standardised selective attention sub test scores standardised to a mean of 0 and a SD of 1 and the mean of the two standardised sustained attention sub test scores standardised to a mean of 0 and a SD of 1, respectively.

The 3 outcomes for TEACh-5 are thus the ‘overall attention mean’, the ‘selective attention mean’ and the ‘sustained attention mean’.

Computation of BRIEF outcome

The Behavior Rating Inventory of Executive Function (BRIEF) is a questionnaire for parents and teachers assessing how the child manages and behaves in everyday life. Forms of the BRIEF are composed of eight non-overlapping clinical scales measuring several aspects of executive functioning: inhibit, shift, emotional control, initiate, working memory, plan/organize, organization of materials, and monitor.

For teachers the eight clinical scales are contained in:

- inhibit (`briefo_inhibit`)
- shift (`briefo_shift`)
- emotional control (`briefo_emotional`)
- initiate (`briefo_initiate`)
- working memory (`briefo_memory`)
- plan/organize (`briefo_plan`)
- organization of materials (`briefo_organization`)
- monitor (`briefo_monitor`)

For parents the eight clinical scales are contained in:

- inhibit (`brieff_inhibit`)
- shift (`brieff_shift`)
- emotional control (`brieff_emotional`)
- initiate (`brieff_initiate`)
- working memory (`brieff_memory`)
- plan/organize (`brieff_plan`)
- organization of materials (`brieff_organization`)
- monitor (`brieff_monitor`)

These scales form two broader indexes, the Behavioral Regulation Index (BRI) and the Metacognition Index (MI), and yield an overall summary score, the Global Executive Composite (GEC). The BRI includes the scales inhibit, shift and emotional control. The MI includes initiate, working memory, plan/organize, organization of materials and monitor.

The eight clinical scales are not following a normal distribution. A normalizing T-score transformation for the observed BRIEF scores is computed (`briefo_inhibit_t` - `briefo_monitor_t` and `brieff_inhibit_t` - `brieff_monitor_t`), with higher scores indicating more executive function difficulties. This is done by ranking each score and computing the quantiles as the rank divided by the total number of observations plus one for each score, and afterwards taking the inverse cumulative distribution function of the standard normal distribution on the obtained quantiles and multiplying the resulting z-score by 10 and adding 50 to the answer.

After imputing missing values for the eight transformed clinical scales the two indexes BRI and MI and the overall summary score GEC are calculated as the mean over the BRI scales, the MI scales and all eight scales, respectively.

Computation of SDQ outcome

The Strengths and Difficulties Questionnaire (SDQ) scores are available in a parents version and a teachers version. The parents version and the teachers version will be analyzed separately.

The five SDQ sub scores (teachers and parents) are:

- emotional symptoms score (`sdq_t_emo` and `sdq_p_emo`)
- conduct problems score (`sdq_t_cond` and `sdq_p_cond`)
- hyperactivity score (`sdq_t_hyp` and `sdq_p_hyp`)
- peer problems score (`sdq_t_peer` and `sdq_p_peer`)
- prosocial score (`sdq_t_prosoc` and `sdq_p_prosoc`)

A total difficulties score is calculated as an average of the first four sub scores if all four sub scores are non-missing, and missing otherwise. Note that most of the missing scores correspond to individuals with missing values on all sub scores.

As the sub scores are categorical variables with a skew distribution (many 0's in case of the first four sub scores and many 10's in case of prosocial score), the outcomes are dichotomized/categorized in the following way:

- The total difficulties score (categorical variable): Cut-offs at the 80^{th} and 90^{th} percentile (lower score is better), resulting in the three groups: normal, borderline and abnormal.
- The four difficulties sub scores (dichotomous variables): Cut-off at the 90^{th} percentile, resulting in the two groups: normal and abnormal.
- The prosocial score (categorical variable): Cut-offs at the 10^{th} and 25^{th} percentile (higher score is better), resulting in the three groups: abnormal, borderline and normal.
- The prosocial score (dichotomous variable): Cut-off at the 10^{th} percentile, resulting in the two groups: abnormal and normal.

After imputing missing values three binary scores are generated indicating abnormal vs. normal, borderline vs. normal, and (abnormal+borderline) vs. normal in case of the total difficulties score and the categorized prosocial score.

Computation of motor function outcome

Motor function is assessed using the performance test of the Movement Assessment Battery for Children (MABC). The performance test includes eight tests covering manual dexterity, ball skills, and static and dynamic balance.

The eight tests are:

- Posting coins in a bank (`abc_pp_11`)
- Threading beads (`abc_ps_5`)

- Drawing a line into a trail (abc_cs_6)
- Rolling a ball into a goal (abc_tb_4)
- Catching a bean bag (abc_ga_3)
- Standing on one leg (abc_1b_11)
- Jumping over a cord (abc_hs_4)
- Walking heels raised on a line (abc_hs_25)

Every task is scored from 0-5 according to the speed and number of correctly executed components of the task. Thus the total score on the performance test is between 0-40, with a lower score indicating better performance. The total score is already computed in the LDPS dataset for women not requiring imputation (m_abcsco).

The total motor impairment score (TIS) can be interpreted by using percentile norm tables for the specific age group. In the manual, the 5th and the 15th percentile, derived from American standardisation, are suggested as cut-off scores resulting in the following three groups indicating a definitive motor problem, a borderline degree of difficulties, and no difficulties, respectively:

- Abnormal: TIS ≤ 5th percentile implies 17 ≤ TIS ≤ 40
- Borderline: 5th percentile < TIS ≤ 15th percentile implies 10.5 ≤ TIS < 17
- Normal: TIS > 15th percentile implies 0 ≤ TIS < 10.5

The categorized total impairment score is already computed in the LDPS dataset for women not requiring imputation (m_perc).

Furthermore, the eight tests are divided into three sub types of motor function tests according to test for fine motor function, gross motor function, and balance, and the corresponding composite scores are derived. The fine motor function composite score includes the tests posting coins in a bank (abc_pp_11), threading beads (abc_ps_5) and drawing a line into a trail (abc_cs_6). The gross motor function composite score includes rolling a ball into a goal (abc_tb_4) and catching a bean bag (abc_ga_3). The balance composite score includes standing on one leg (abc_1b_11), jumping over a cord (abc_hs_4) and walking heels raised on a line (abc_hs_25).

Covariates

The publications each use a number of the following covariates recorded in the dataset `LDPS_data` – the corresponding variable names are specified in the brackets following the covariate:

Maternal age (in years) – continuous, normal (`malder`)

Parity – categorical: 0, 1, ≥2 (`paritet`)

Maternal pre-pregnancy BMI – continuous, skewed normal (`bmi_before_int1`)

Maternal marital status – dichotomous: single either at the prenatal interview or at follow-up/married or cohabitating (`marital`)

Parental education (total duration in years averaged for both parents or, if information on the father was unavailable, maternal only) – continuous, skewed normal (`eduindex`)

Family/home index – dichotomous: normal/suboptimal as determined by the presence of two or more of the following adverse conditions: living with only one biological parent, changes in primary care givers, daycare more than 8 hours/day before age 3, ≥14 days of separation from parents, irregular breakfast, maternal depression, maternal and paternal alcohol intake above the official recommendations from the Danish National Board of Health at the time of the data collection (`home_dic`)

Maternal IQ – continuous, normal (`n_ik_me3`)

Maternal smoking in pregnancy – dichotomous: yes/no (`presmoke`)

Parental postnatal smoking – dichotomous: yes if at least one of the parents smoked in the home and no otherwise (`postsmoke_p`)

Child's gender – dichotomous: male/female (`gender`)

Child's age at testing –continuous, skewed normal (`alder_vtest`) or categorical: 5 years 1 month, 5 years 2 months, 5 years 3 months, 5 years 4 months (`agecat`)

Child's birth weight (in grams) – continuous, normal (`vagt`)

Child's gestational age (in days) – continuous, skewed normal (`ga_dage`)

Child's health status – dichotomous: presence/absence of major medical conditions or regular use of prescription medications (`healthindex`)

Child's hearing abilities – dichotomous: normal/impaired (`hearing`)

Child's vision abilities – dichotomous: normal/impaired (`vision`)

Child's physical activity – dichotomous: yes/no (`nosports`)

Testing psychologist – categorical, 8 categories (`test_psyc1`)

Furthermore, depending on the exposure being analyzed other exposures may be considered as a covariate:

Maternal binge drinking in pregnancy – dichotomous: yes/no (`binge`)

Maternal average alcohol intake during pregnancy – categorical: 0, 1–4, 5–8, ≥9 (`av_during`)

Maternal average alcohol intake before pregnancy – categorical: 0, 15-21, ≥22 (`av_before`)

Imputation of missing values

The method of multiple imputation – why do imputation?

Missing data is a well-recognized problem in datasets. In medicine, for example, it is common for observations to be missing in a sporadic way for different covariates, and a complete-case analysis may therefore omit as many as, say, half of the available cases. One approach for dealing with incomplete covariates is multiple imputation (Rubin, 1987). Multiple imputation is a technique in which each missing value is replaced by m>1 simulated values, so that m plausible versions of the complete data is created, each of which are analyzed by standard complete-data methods. The m imputation datasets reflect the uncertainty about the true values of the missing data. The results of the m analyses are then combined to a single set of estimates. In general, multiple imputation techniques require that missing observations are missing at random (MAR). MAR means that what caused the data to be missing does not depend upon the missing data itself, i.e. a missing value could, for instance, be due to an accidental omission of an answer on a questionnaire. The missing values are allowed to depend on observed variables though. Not missing at random is data that are missing for a specific reason, e.g. if a question on a questionnaire has been skipped deliberately by the participant. Doing multiple imputation should be done in a rich model with many explanatory variables. On the other hand, the number of parameters must not be too large so as not to violate the usual rule of thumb of 15 events per variable (Peduzzi, 1996).

Multiple imputation strategies

In the LDPS missing values are imputed based on the following two strategies: A dedicated model for imputations, for which variables are modeled from a dedicated set of variables considered to be the most predictive, and a black-box strategy, for which a rich set of variables are used to predict missing values. Only results of the dedicated imputation strategy are reported in the articles.

Multiple imputation was implemented by the use of the `-ice-` command in Stata 11, where m=100 imputed datasets were created using weighted regressions. Please note that you may have to download some additional Stata programs including `-ice-` and you can do this with the `-findit-` command. Also note that `-ice-` has been updated several times since it was originally released so if you have an older version of `-ice-` you may need to download the most recent version. To determine which version of `-ice-` you have, and to ensure that you have the current version of `-ice-`, you can type:

```
which ice
ssc describe ice
ssc install ice, replace
```

Stata's program `-ice-` was written by Patrick Royston, and he has published a number of articles in Stata Journal, introducing `-ice-` and documenting improvements made to it (Royston 2004, 2005 a+b, 2007, 2009).

An example of multiple imputation

As an example we will now go through a code for doing multiple imputation in the case of the strengths and difficulties questionnaire (SDQ). The codes presented below are meant as examples and will need recoding to fit the setup of the problem being investigated, but the principles will be the same for other outcomes than the SDQ. The outcome scores to be imputed are the 10 dichotomized sub scores in `sdq_t_emo_dic-sdq_t_prosoc_dic` and `sdq_p_emo_dic-sdq_p_prosoc_dic` and the 4 categorized scores in `sdq_t_totaldif_cat`, `sdq_t_prosoc_cat`, `sdq_p_totaldif_cat` and `sdq_p_prosoc_cat`.

First we need to take a look at the list of confounders that will be used in the analyses of the SDQ so as to impute missing values of these confounders along with outcome scores to be imputed. The confounders used in the analyses of the SDQ outcomes are specified as follows:

<u>Core confounders</u>: Parental education* (`eduindex`), maternal IQ* (`n_ik_me3`), maternal smoking in pregnancy (`presmoke`), child's age at testing (`alder_vtest`) and child's gender (`gender`).

<u>Potential confounders</u>: Maternal age (`malder`), parity (`paritet`), maternal marital status* (`marital`), family/home index* (`home_dic`), parental postnatal smoking* (`postsmoke_p`), maternal pre-pregnancy BMI* (`bmi_before_int1`), maternal average alcohol intake during pregnancy (`av_during)`, maternal binge drinking in pregnancy (`binge`) and child's health status (`healthindex`).

<u>Potential mediators</u>: Child's birth weight* (`vagt`) and child's gestational age* (`ga_dage`).

The variables which require imputation (the outcome scores and the confounders marked with *) are predicted using an appropriate regression type, i.e. for the listed variables the corresponding regression type is used when these variables are outcome in the multiple imputation:

<u>Choice of regression types</u>:

- Logistic regression: The 10 SDQ dichotomized sub scores, maternal marital status, family/home index and parental postnatal smoking.
- Ordered logistic regression: The 4 SDQ categorized scores.
- Linear regression: Parental education, maternal IQ, maternal pre-pregnancy maternal BMI, child's birth weight and child's gestational age.

The variables which require imputation are predicted by all the variables to be used in the analyses as either outcome or as a covariate using the appropriate regression type given above. The code is as follows:

Example code to black-box imputation

```
************************
* Black-box imputation *
************************
set seed 891895
ice sdq_t_emo_dic-sdq_t_prosoc_dic sdq_p_emo_dic-sdq_p_prosoc_dic ///
    sdq_t_totaldif_cat sdq_t_prosoc_cat sdq_p_totaldif_cat sdq_p_prosoc_cat ///
    eduindex n_ik_me3 marital home_dic postsmoke_p bmi_before_int1 vagt ga_dage ///
    /*non-missing:*/ presmoke alder_vtest gender ///
```

```
    /*non-missing:*/ malder i.paritet i.av_during binge healthindex ///
    [pw=1/sampfrac], m(100) cycles(20) clear ///
        cmd(sdq_t_emo_dic-sdq_t_prosoc_dic sdq_p_emo_dic-sdq_p_prosoc_dic ///
            marital home_dic postsmoke_p : logit, ///
            sdq_t_totaldif_cat sdq_t_prosoc_cat ///
            sdq_p_totaldif_cat sdq_p_prosoc_cat : ologit, ///
            eduindex n_ik_me3 bmi_before_int1 vagt ga_dage : regress)
mi import ice, auto clear
```

As already mentioned the black-box strategy uses all variables to predict missing values. The structure of the code is that we write -`ice`- followed by the list of variables to be used in the analyses as either outcome or as a covariate. The -`ice`- command will then use all listed variables to predict missing values. That is the SDQ scores variables will be predicted by each other together with the covariates, and the covariates will be predicted by each other together with the SDQ scores variables. Dependant variables will, as default, be included in regressions as continuous variables even if they are categorical. The variables for parity (`paritet`) and maternal average alcohol intake during pregnancy (`av_during`) are categorical variables, and since they do not have missing values, we can add `i.` before the variable names (i.e. `i.paritet` and `i.av_during`) to instruct -`ice`- to replace all instances of `paritet` and `av_during` with a set of dummy variables. For reproducibility you want to set the seed for the random number generator before the -`ice`- command by writing `set seed` followed by some number. This covers the first six lines of code.

We specify -`ice`- to do weighted regressions by adding `[pw=1/sampfrac]` (see the section on sampling fractions) to the -`ice`- command line before turning our attention to the options following after the comma sign. We need to specify the number of imputed datasets to be created, in this case 100 as seen by `m(100)`. For each imputed dataset imputation is first done by considering the complete-case data alone after which the imputed values are recalculated using both the observed and the imputed values leading to further changes in the imputed values. The imputed values are recalculated a number of times, and, eventually, this process should converge. The option `cycles` specifies how many iterations of this process should be performed before producing each of the final imputed datasets. The number of cycles to be carried out was in this case set to 20 as seen by `cycles(20)`. The option `clear` allows the imputed data to reside in memory without yet having been manually saved to a file using Stata's -`save`- command.

Next we need to specify the regression commands to be used for each variable when it becomes the dependent variable. The choice of regression types for the variables needing imputation was stated above and is implemented in the lines 8-12 by using the option `cmd( )`. In the brackets we have specified, that the 10 SDQ dichotomized sub scores, maternal marital status, family/home index and parental postnatal smoking are imputed using logistic regression by writing `sdq_t_emo_dic-sdq_t_prosoc_dic sdq_p_emo_dic-sdq_p_prosoc_dic marital home_dic postsmoke_p : logit`. The 4 SDQ categorized scores are imputed using ordered logistic regression by `sdq_t_totaldif_cat sdq_t_prosoc_cat sdq_p_totaldif_cat sdq_p_prosoc_cat : ologit`. Finally, the covariates parental education, maternal IQ, maternal pre-pregnancy maternal BMI, child's birth weight and child's gestational age are imputed using linear regression by `eduindex n_ik_me3 bmi_before_int1 vagt ga_dage : regress`.

The last line, `mi import ice`, converts the data in memory to mi data.

Example code to dedicated imputation

The variables which require imputation (the outcome scores and the confounders marked with *) are predicted using an appropriate regression type. When doing a dedicated imputation we also need an appropriate choice of predictors. When looking at exposure to alcohol intake during pregnancy on the SDQ scores the following choices were made.

Choice of predictors:

- The 5 SDQ dichotomized sub scores (parents) are predicted by each other together with the 5 SDQ dichotomized sub scores (teachers), maternal IQ, maternal smoking in pregnancy, parental education, child's gender and child's age at testing.
- The 5 SDQ dichotomized sub scores (teachers) are predicted by each other together with the 5 SDQ dichotomized sub scores (parents), child's gender and child's age at testing.
- The 2 SDQ categorized scores (parents) are predicted by the 5 SDQ dichotomized sub scores (parents), the 5 SDQ dichotomized sub scores (teachers), maternal IQ, maternal smoking in pregnancy, parental education, child's gender and child's age at testing.
- The 2 SDQ categorized scores (teachers) are predicted by the 5 SDQ dichotomized sub scores (parents), the 5 SDQ dichotomized sub scores (teachers), child's gender and child's age at testing.
- Maternal marital status, maternal IQ, maternal pre-pregnancy BMI, family/home index, parental education and parental postnatal smoking are predicted by each other together with maternal age and maternal smoking in pregnancy.
- Child's birth weight and child's gestational age are predicted by each other together with child's gender, maternal smoking in pregnancy, maternal pre-pregnancy BMI, maternal binge drinking in pregnancy and maternal average alcohol intake during pregnancy.

The code for the dedicated imputation is as follows:

```
************************
* Dedicated imputation *
************************
set seed 891895
ice sdq_t_emo_dic-sdq_t_prosoc_dic sdq_p_emo_dic-sdq_p_prosoc_dic ///
    sdq_t_totaldif_cat sdq_t_prosoc_cat sdq_p_totaldif_cat sdq_p_prosoc_cat ///
    eduindex n_ik_me3 marital home_dic postsmoke_p bmi_before_int1 vagt ga_dage ///
    /*non-missing:*/ presmoke alder_vtest gender ///
    /*non-missing:*/ malder i.paritet i.av_during binge healthindex ///
    [pw=1/sampfrac], m(100) cycles(20) clear ///
        cmd(sdq_t_emo_dic-sdq_t_prosoc_dic sdq_p_emo_dic-sdq_p_prosoc_dic ///
            marital home_dic postsmoke_p : logit, ///
            sdq_t_totaldif_cat sdq_t_prosoc_cat ///
            sdq_p_totaldif_cat sdq_p_prosoc_cat : ologit, ///
            eduindex n_ik_me3 bmi_before_int1 vagt ga_dage : regress) ///
         eq(sdq_p_emo_dic :      sdq_p_cond_dic sdq_p_hyp_dic sdq_p_peer_dic ///
                                 sdq_p_prosoc_dic sdq_t_emo_dic sdq_t_cond_dic ///
                                 sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                                 n_ik_me3 presmoke eduindex gender alder_vtest, ///
            sdq_p_cond_dic :     sdq_p_emo_dic sdq_p_hyp_dic sdq_p_peer_dic ///
                                 sdq_p_prosoc_dic sdq_t_emo_dic sdq_t_cond_dic ///
                                 sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                                 n_ik_me3 presmoke eduindex gender alder_vtest, ///
            sdq_p_hyp_dic :      sdq_p_emo_dic sdq_p_cond_dic sdq_p_peer_dic ///
                                 sdq_p_prosoc_dic sdq_t_emo_dic sdq_t_cond_dic ///
                                 sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                                 n_ik_me3 presmoke eduindex gender alder_vtest, ///
```

```
         sdq_p_peer_dic :     sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_prosoc_dic sdq_t_emo_dic sdq_t_cond_dic ///
                              sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                              n_ik_me3 presmoke eduindex gender alder_vtest, ///
         sdq_p_prosoc_dic :   sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_t_emo_dic sdq_t_cond_dic ///
                              sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                              n_ik_me3 presmoke eduindex gender alder_vtest, ///
         sdq_t_emo_dic :      sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_cond_dic ///
                              sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         sdq_t_cond_dic :     sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_hyp_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         sdq_t_hyp_dic :      sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_peer_dic sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         sdq_t_peer_dic :     sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         sdq_t_prosoc_dic :   sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_peer_dic ///
                              gender alder_vtest, ///
         sdq_p_totaldif_cat : sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_peer_dic ///
                              sdq_t_prosoc_dic ///
                              n_ik_me3 presmoke eduindex gender alder_vtest, ///
         sdq_p_prosoc_cat :   sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_peer_dic ///
                              sdq_t_prosoc_dic ///
                              n_ik_me3 presmoke eduindex gender alder_vtest, ///
         sdq_t_totaldif_cat : sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_peer_dic ///
                              sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         sdq_t_prosoc_cat :   sdq_p_emo_dic sdq_p_cond_dic sdq_p_hyp_dic ///
                              sdq_p_peer_dic sdq_p_prosoc_dic sdq_t_emo_dic ///
                              sdq_t_cond_dic sdq_t_hyp_dic sdq_t_peer_dic ///
                              sdq_t_prosoc_dic ///
                              gender alder_vtest, ///
         marital :            n_ik_me3 bmi_before_int1 home_dic eduindex ///
                              postsmoke_p presmoke malder, ///
         n_ik_me3 :           marital bmi_before_int1 home_dic eduindex ///
                              postsmoke_p presmoke malder, ///
         bmi_before_int1 :    marital n_ik_me3 home_dic eduindex postsmoke_p ///
                              presmoke malder, ///
         home_dic :           marital n_ik_me3 bmi_before_int1 eduindex ///
                              postsmoke_p presmoke malder, ///
         eduindex :           marital n_ik_me3 bmi_before_int1 home_dic ///
                              postsmoke_p presmoke malder, ///
         postsmoke_p :        marital n_ik_me3 bmi_before_int1 home_dic ///
                              eduindex presmoke malder, ///
         ga_dage :            vagt gender presmoke i.av_during binge ///
                              bmi_before_int1, ///
         vagt :               ga_dage gender presmoke i.av_during binge ///
                              bmi_before_int1)
mi import ice, auto clear
```

The code of the dedicated imputation is identical to the code of the black-box imputation except for the equations specified in the `eq( )` option. In the brackets we have specified which variables should be used to predict a specific variable that has missing values. As an example look at the last equation for child's birth weight (`vagt`). The child's birth weight is predicted using child's gestational age (`ga_dage`), child's gender (`gender`), maternal smoking in pregnancy (`presmoke`), maternal binge drinking in pregnancy (`binge`), maternal average alcohol intake during pregnancy (`av_during`) and maternal pre-pregnancy BMI (`bmi_before_int1`) by writing `vagt : ga_dage gender presmoke i.av_during binge bmi_before_int1`.

This concludes our example on doing multiple imputation in the case of SDQ. The principles are the same for other outcomes than SDQ.

Analyses

In the following, separate analyses are performed for the effect of average alcohol intake (continuous, coded as 0, 1–4, 5–8, ≥9 drinks per week), binge drinking (dichotomous, yes or no), number of binge episodes (continuous, coded as 0, 1, 2, ≥3 episodes), and timing of binge drinking (none, gestational weeks 1-2, 3-4, 5-8, ≥9, multiple timings). It is assumed that missing values have been imputed by multiple imputation (see the section on multiple imputation). As for confounders of the outcomes adjusted for in the analyses described below, the reader is referred to the corresponding publication.

All statistical analyses were conducted with Stata 11 and weighted by sampling probabilities with robust variance estimation. The analyses were performed as 'complete case analysis' and 'multiple imputation analysis', i.e. after multiple imputation.

When doing multiple imputation, m>1 imputation datasets are created. These datasets can be identified by the variable `_mi_m`, where the original dataset, i.e. the dataset including missing values, will have `_mi_m==0`. When wanting to do the complete case analysis, we only consider this dataset, thereby ignoring the m imputation dataset, and analyse it by standard complete-data methods.

As an example we will now go through a code for doing logistic regression in the case of the strengths and difficulties questionnaire (SDQ). Assume that the exposure of interest is the binary binge drinking variable `binge` and that the outcome of interest is the binary score for the total difficulties score indicating abnormal vs. normal. For the parents version let us call the latter `sdq_p_abnormal`. Furthermore, the model should include a number of confounders, here parental education (`eduindex`), maternal IQ (`n_ik_me3`), maternal smoking in pregnancy (`presmoke`), child's age at testing (`alder_vtest`) and child's gender (`gender`). The sampling weights are incorporated into the regression analysis by adding `[pweights=1/sampfrac]` to the regression command line. The code is as follows:

Example code to complete-case analysis

```
**************************
* Complete case analysis *
**************************
keep if _mi_m == 0
logit sdq_p_abnormal binge ///
      eduindex n_ik_me3 presmoke alder_vtest gender [pw=1/sampfrac], vce(robust) or
```

Analyzing the imputation datasets is done by performing the analysis on each imputation dataset and afterwards combining the results into a single set of estimates. This is done by the -`mi estimate`-command by simply adding this in front of the regression command in the following way:

Example code to multiple imputation analysis

```
********************************
* Multiple imputation analysis *
********************************
mi estimate, esampvaryok or: logit sdq_p_abnormal binge ///
      eduindex n_ik_me3 presmoke alder_vtest gender [pw=1/sampfrac], vce(robust)
```

To the -`mi estimate`- command you may add a number of options. One of them is the option `esampvaryok`, which allows estimation when estimation sample varies across samples. Another option is the option `or`, which tells Stata to display coefficients in exponentiated form and is nothing new per se, but the ordering of the code is slightly different compared to the complete-case code. For the multiple imputation analysis the option `or` needs to be put as an option to the -`mi estimate`- command, whereas the standard complete-data methods require this as an option to the regression command.

This concludes our example on doing analyses on imputation datasets in the case of SDQ.

Analyses of IQ outcome

The associations between alcohol exposures and the continuous FSIQ, VIQ and PIQ outcome scores for IQ were estimated using multiple linear regression adjusted for a number of confounders.

Also, raw scores of each individual IQ sub test (`aritmeti`, `informat`, `vocabula`, `block_as`, `figure_d`, `object_a`) were analyzed with linear regression models adjusting for a number of confounders. Furthermore, potential interactions with alcohol exposures average alcohol intake and binge drinking were assessed for child's gender, parental education, maternal smoking during pregnancy, and binge drinking episodes and average alcohol intake, respectively.

Additionally, the three continuous FSIQ, VIQ and PIQ outcome scores for IQ were dichotomized using the sample mean minus one SD as cut-off for subnormal scores on FSIQ, VIQ and PIQ. Logistic regression models adjusting for a number of confounders were used in the analyses of these dichotomized outcome variables using the category of IQ above the cut-off as reference group.

Analyses of TEACh-5 outcome

The associations between alcohol exposures and the continuous selective attention, sustained attention and overall attention outcome scores for TEACh-5 were estimated using multiple linear regression adjusted for a number of confounders. Standardized scores of each individual TEACh-5 sub test (`barking_slow`, `logdraw`, `logHS2_present_score3` and `c2_plus_c3`) were analyzed with linear regression models adjusting for a number of confounders. Furthermore, potential interactions between average alcohol intake and binge drinking as well as interactions of the alcohol exposures with child's gender, parental education and maternal smoking were assessed.

Additionally, the three continuous attention scores were dichotomized using the sample mean minus one SD as cut-off for subnormal scores. Logistic regression models adjusting for a number of confounders were used in the analyses of these dichotomized outcome variables using the category of attention scores above the cut-off as reference group.

Analyses of BRIEF outcome

The associations between alcohol exposures and the continuous BRI, MI and GEC outcome scores for BRIEF were estimated using multiple linear regression adjusted for a number of confounders. A normalizing T-score transformation for the BRIEF raw scores (`briefo_inhibit_t` - `briefo_monitor_t` and `brieff_inhibit_t` - `brieff_monitor_t`) is computed, and the resulting transformed scores are analyzed with linear regression models adjusting for a number of confounders. Furthermore, potential interactions between average alcohol intake and binge drinking as well as interactions of the alcohol exposures with child's gender, parental education and maternal smoking were assessed.

Additionally, the three continuous BRI, MI and GEC outcome scores were dichotomized using the sample mean minus 1 SD and 1.5 SD, respectively, as cut-off for subnormal scores. Logistic regression models adjusting for a number of confounders were used in the analyses of these dichotomized outcome variables using the category below the cut-off as reference group.

Analyses of SDQ outcome

The main analysis was a logistic regression on the dichotomized total difficulties score (abnormal+borderline) vs. normal as well as on the dichotomized prosocial score (abnormal+borderline) vs. normal. A supplementary analysis was conducted in the same way for the dichotomized total difficulties scores abnormal vs. normal and borderline vs. normal as well as for the dichotomized prosocial scores abnormal vs. normal and borderline vs. normal. Similarly, each of the five SDQ sub scores (parents and teachers) is analysed using logistic regression.

Supplementary analyses for interactions between average alcohol (binge drinking) and the dichotomized variables gender, education, presmoke and binge drinking (average alcohol), and the corresponding strata specific analyses were conducted. If necessary the number of binge episodes groups “2” and ”3+” were collapsed into a “2+” group and the timing of binge episodes groups “5-8” and ”9+” were collapsed into a “5+” group, otherwise they were reported as not available.

Analyses of motor function outcome

The associations between alcohol exposures and the continuous TIS were estimated using multiple linear regression adjusted for a number of confounders. The TIS is categorized into three groups indicating a definitive motor problem, a borderline degree of difficulties, and no difficulties, respectively, and the resulting variable is analyzed using multiple logistic regression adjusted for a number of confounders with the category of normal performance as reference. Furthermore, potential interactions between average alcohol intake and binge drinking as well as interactions of the alcohol exposures with child's gender, parental education and maternal smoking were assessed.

Additionally, the three continuous composite scores for fine motor function, gross motor function and balance motor function were assessed using multiple linear regression adjusted for a number of confounders.

References

Kish L. Survey sampling. Wiley, 1965.

Peduzzi P, Concato J, Kemper E et al. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49(12):1373-9.

Royston P. Multiple imputation of missing values. Stata Journal 2004; 4(3):227-41.

Royston P. Multiple imputation of missing values: Update. Stata Journal 2005; 5(2):188-201.

Royston P. Multiple imputation of missing values: Update of ice. Stata Journal 2005; 5(4):527-36.

Royston P. Multiple imputation of missing values: Further update of ice, with an emphasis on interval censoring. Stata Journal 2007; 7(4):445-64.

Royston P. Multiple imputation of missing values: Further update of ice, with an emphasis on categorical variables. Stata Journal 2009; 9(3):466-77.

Rubin DB. Multiple imputation for nonresponse in surveys. John Wiley & Sons, 1987.