MELC: Blueprint for Version 3

Modelling CHDS, health service use, 0-13 yrs

Contents

[1. Framework: Version 2 3](#_Toc314836129)

[2. Framework: Version 3 5](#_Toc314836130)

[3. Version 3 versus Version 2 6](#_Toc314836131)

[4. Simulation Order 7](#_Toc314836132)

[5. Janet’s Blurb: Notes On Decision Tree/Typology Of Family/Making Things Consistent 8](#_Toc314836133)

[5.1 Definitions of derived variables mentioned 8](#_Toc314836134)

[5.2 Inclusion of some lagged x’s where time characteristic of outcome dictates- 5 outcomes 9](#_Toc314836135)

[5.3 Relationship and consistency among the ‘parental presence’ indicator variables in the data 10](#_Toc314836136)

[5.4 Modelling and simulating a change in the parents, even when z1single was the same from one year to the next 12](#_Toc314836137)

[6. The Pseudo-Models 13](#_Toc314836138)

[6.1 Step 1 13](#_Toc314836139)

[6.1.1 Single Parent 13](#_Toc314836140)

[6.1.2 Calculate Family Type Variables 13](#_Toc314836141)

[6.1.3 Family Type Flow Diagram 16](#_Toc314836142)

[6.1.4 Number of Children and Household Size 17](#_Toc314836143)

[6.1.5 Change of Parents 18](#_Toc314836144)

[6.1.6 Changes in Residence 19](#_Toc314836145)

[6.1.7 Calculate More Family Type Variables 19](#_Toc314836146)

[6.1.8 Welfare 19](#_Toc314836147)

[6.1.9 Mother’s Hours Worked 20](#_Toc314836148)

[6.1.10 Father’s Hours Worked 20](#_Toc314836149)

[6.2 Step 2 21](#_Toc314836150)

[6.2.1 Accomodation Type 21](#_Toc314836151)

[6.2.2 Owned or Rented 21](#_Toc314836152)

[6.2.3 Over-crowding 21](#_Toc314836153)

[6.2.4 Mother Current Smoking 21](#_Toc314836154)

[6.2.5 Father Current Smoking 21](#_Toc314836155)

[6.3 Step 3 22](#_Toc314836156)

[6.3.1 Total Number of GP Visits 22](#_Toc314836157)

[6.3.2 Number of Hospital Admissions 22](#_Toc314836158)

[6.3.3 Number of Outpatient Attendances 22](#_Toc314836159)

[6.4 Step 4: Justice Outcomes 22](#_Toc314836160)

[6.5 Step5: Education Outcomes 24](#_Toc314836161)

[7. Special Case Models 26](#_Toc314836162)

[Special Case 1: The Starting Value Is Not Provided In the Base File and Needs to Be Generated In the Simulation 26](#_Toc314836163)

[8. Possible scenarios 26](#_Toc314836164)

[9. Ways to test scenarios? 27](#_Toc314836165)

# Framework: Version 2



Text



# Framework: Version 3



Need to add into the above Justice and Education in the final outcomes box.



Need to add after Health service use, Justice and Education

# Version 3 versus Version 2

**Pseudo-models:**

|  |  |
| --- | --- |
| **Pseudo-models** | **Version 3** |
| **Family-household characteristics** | Similar to v2 |
| **Employment** | Reduced set of potential predictors |
| **Material circumstances:** | Similar to v2 |
| **Psychosocial factors:** | Reduced set of potential predictors |
| **Behavioural factors:** | Increased set of potential predictors |
| **Health service use**: | Similar to v2 |

**V3: Continuous/count variables that could be re-specified?**

|  |  |  |
| --- | --- | --- |
| **Outcome** | **Version 2** | **Version 3** |
| Mother’s age (at child’s birth) |  | Continuous |
| Father’s age (at child’s birth) | 6 categories | 6 categories |
| Breastfeeding (months)  Birthweight (g),  Gestational age (weeks),  Pregnancy-mother-cigs-per-day,  Pregnancy-Mother-alcohol-drinks-per-day | ? | continuous |
| **intermediate outcomes** | |  |
| Household-size  Children-number | <4, >=4 (below/above average)  <=2, >2 (below/above average) | continuous |
| Mother-hours-worked  Father-hours-worked | 0, >=1  0, >=1 | continuous |
| Bedrooms-number | <3, >=3 (below/above average) | Using dichotomous overcrowding |
| Change-of-residence-number | 0, >=1 | continuous |
| Maternal smoking (cigs/day)  Paternal smoking (cigs/day) | 0, >=1  0, >=1 | continuous |

Most outcomes will be change scores in version three and we will use the normal or logistic distribution for all deviation scores .This means that the continuous form of the variable can be used as predictors in the model building and in the simulation.

For models where we need to simulate the initial value (e.g. conduct and Burt-reading score) we will still need to find a suitable distribution for the outcome variable.

# Simulation Order

**STEP 1**

* Family / household characteristics
* Employment

**STEP 2**

* Material circumstances
* Psychosocial factors
* Behavioural factors

**STEP 3**

* Health service use outcomes

**STEP 4**

* Justice outcomes

**STEP 5**

* Education outcomes

# Janet’s Blurb: Notes on Decision Tree/Typology of Family/Making Things Consistent

## Definitions of derived variables mentioned

**Sptype** – in dataset this is derived from the change from z1single\_previous to z1single:

1='became partnered'

2='broke up'

3='same status as b4';

**Typeofchange** – in dataset this is derived from the change from previous parental ages variables (whether 99 or not) to current parental ages variables (whether 99 or not). For the purpose of creating and using typeofchange variable in the simulation, I put the -2 category in as a number 5, as -2 is really a rare subset of number 5.

-**2**='diffkindparent in singlep fam'

**0**='both father and mother left'

**1**='got new mother/ mother back'

**2**='got new father/ father back'

**3**='mother left'

**4**='father left'

**5**='have same number of parents'

**Mumgroup**

**0=** no mum

**1=** birthmother

**2=** not birthmother not same mum as last year

**3=** not birthmother but same mum as last year

**Dadgroup**

**0=** no dad

**1=** birthfather

**2=** not birthfather not same dad as last year

**3=** not birthfather but same dad as last year

**Fage\_imputed**

Already provided in the basefile, this is a continuous version of **fage**, the father’s age at the birth of the child. It is derived via the program ‘imputing birth fathers age for oliver - sep011.sas’ which uses, depending on the circumstances, a combination of father’s age at year 1; mother age at birth of child; and average ages of fathers in the unbounded groups, to impute father’s continuous age at birth of child.

**Methn**

If r1stmethn=1 then methn=3;

If r1stmethn=2 then methn=2;

If r1stmethn=3 then methn=1;

**Meduc**

If r1stmeduc=1 then meduc=3;

If r1stmeduc=2 then meduc=2;

If r1stmeduc=3 then meduc=1;

## Inclusion of some lagged x’s where time characteristic of outcome dictates- 5 outcomes

If outcome is an ‘at time of survey’ variable, then time variant predictors should be current ones.

If outcome is an ‘over last year’ variable, then time variant predictors should be previous ones.

This enables chpar outcome to be modelled before mhrswrk, and fhrswrk (as well as before msmoke and fsmoke as it already is) – now not necessary as the typology of family circumstances (the parental age creation section) does not depend on chpar anymore , like I had originally had it when I first started thinking about it)

Have changed other models accordingly (putting lagged x’s in), in order to be consistent, however I am aware of issues with this meaning that we are potentially sometimes using information from up to two years earlier – this only happens however, for the final health outcomes –, as if we did so here, there would be a cascade effect, where, for example, we use Num-changes-of-residence(t-1), but Num-changes-of-residence(t-1) was itself based on for eg welfare(t-1) – really welfare(t-2) at the time we are at when we are modelling the health outcome . So we would end up with health outcomes at time t being linked to welfare at time (t-2)! But is this really an issue – when the same model also already has current values of all the prior step variables (e.g including welfare(t) ) in there as potential predictors.

Another concern here would be that we may run out of simulated values for these cascaded (t-2) variables. This is not the case, as for example, when we are modelling our first year in the simulation– that is t=2, (t-1) covariates may be year= t-1 = 2-1 = year 1 variables, that don’t actually need to be modelled as they are already present in the basefile.

When we are modelling year 3 (t=3) health outcomes, its (t-1) covariates would be year2 variables. These year 2 variables would have been imputed via models that may have with respect to the current year – (t-2) variables in them. Here, (t-2)=1, and again, we have these available for us in the basefile.

So perhaps we should use lagged x’s where appropriate (i.e where the outcome is a ‘over last year’ variable).

## Relationship and consistency among the ‘parental presence’ indicator variables in the data

In terms of creating a fage\_years and mage\_years variables at each time period, in the data itself there is a near 1-1 relationship between whether there is a mother/father figure present in the home (i.e the parental age variables) in sequential years and whether z1single is 0 or 1 in the same years (z1single seems to be mostly derived from the age variables). There are only 10 tall records where they do not match up (fage\_years indicating a new father in a year, but z1single variables indicating they had the same single parent status from the previous year to the next) . Looking at individual years (and not the change from one year to the next) there is a perfect 1-1 relationship between the parental age variables and the z1single variable.

However there is not a 1-1 relationship between chpar and change in z1single, nor between fage\_years/mage\_years and z1chpar – there were major discrepancies (see below for an example of some of the investigations). Therefore, we cannot use chpar to impute the parental ages afterall.

*There are some major discrepancies here in the percentage of each* ***typeofchange*** *that have* ***z1chpar*** *for that time equal to 1*

*(they should all be 100%):*

*Typeofchange Yr2%chpar2=1 Yr3%chpar3=1 Yr4%chpar4=1 Yr5%chpar5=1*

*got new mother 100 (5/5) 100 (5/5)*

*got new father 89 67 (60/90) 84(105/125) 70(70/100)*

*mother left 100 100(10/10) 80(20/25)*

*father left 91 91(145/160) 100(150/150) 65(55/85)*

*The definition in survey states that z1chpar should include parental separation, reconciliation, remarriage, entry of a step parent,*

*placement with foster parents, parental death, or any other change of custodial parents, so fage\_years or mage\_years change to 99*

*(i.e no father or mother figure in the home - got new mum, got new dad, mu left, dad left - you think would be a chpar=1) .*

*So I do not understand why the above percentages aren't 100%.*

*Perhaps a parent is working away from home - that's why there is 'no mum/dad figure' all of a sudden, but with z1chpar a no?*

*However according to data dictionary parental age variables being 99*

*" implies there was no mother/father (figure) in the home i.e. the family was a solo mother/father family at the time of the interview."*

*So does solo= singleparent? or just they weren't home? I think maybe sometimes ‘just they weren’t home’ as on questionnaire it says 'details of all person living in the household' before asking about mum and dad’s age.*

*Or perhaps a change in parents means they changed then changed back again?*

Also table of sptype vs z1chpar shows they don’t tie up:

The only indication that there was any tie up at all, was when looking into the subset that 98 to 99% of those with chpar=0 have parental ages that also indicate that there was no change (i.e <=two years of last year’s age) is there a tie up.

Therefore decided to use z1single togther with the parental age variables to model the typology of family circumstances(i.e what parent was there, who left, who was new etc), and these tie up, and also z1single is the first outcome in the steps so that the typology can be used early on in the simulation. I will ask that sptype be placed in as a confounder for z1chpar, though, so that some consistency can be achieved between z1chpar and sptype (as much as what was present in the data).

## Modelling and simulating a change in the parents, even when z1single was the same from one year to the next

If there is no change in single parent status, there may still have been a change in one or other or both of the parents, affecting the parental age variables of course.

I have concentrated on modelling if parental age was 99 or not (i.e presence or absence of parent) and whether there had been a change or not (‘changewasmum’ and ‘changewasdad’ variables). It was sometimes not possible to model the new age of parent if they changed, as I would have liked to have based this on the previous parent’s age, using the difference in parental ages in the dataset (i.e in the simulation – current age of new parent=last parents age +/- some value), and this would have resulted in potentially negative ages though!

As code #12 and #13 show, most ages for changed parents, where the same age as the parent that they replaced, and this has been reflected in the pseudo code.

Also please not that the changewasmum/wasdad variables are there only to denote a change in the corresponding parental age variable. There may be cases in the survey were there was a change in parent, but the ages are the same (+0,1,2), and so we cannot tell here if it was a different person or not. The z1chpar variable does not help here, as there is not a proper tie-up between when there IS a change in parental age, to this variable (i.e of records with a change (beyond +0,1,2 years) in parental age: 49% had z1chpar=0, 50.94% had z1chpar=1 – **see CODE #11**) indicating that the z1chpar variable would be unrealiable for this purpose (i.e the purpose of seeing if parent at time of survey was different than last year).

The changewasmum/dad variable IS however used as an indication of if the parent actually changed, when informing the derivation of if each parent was ‘birthparent, not birthparent but same as before, not birthparent and not same as before’ – but this is the best we can do.

If indeed, in our virtual world, somebody is really a different parent, but they have a changewasmum/dad indicating that there wasn’t a change, the worse that could happen is the following:

-A person could be allocated as being the birthmother/father when they weren’t – in which case Mother/Father ethnicity at birth would also be used as predictors, but these are likely correlated with any new parents values. Mother/Father age at birth would also be used as a predictor, but the value of this, for the new parent would be the same as the birthparents in this scenario. Previous smoking would also be used as a predictor, if the case of the person also being alloced as being the same person as in the last year and this is where some error may occur in the prediction. However this would be similar to what happened in V2, for this subset of potential individuals.

In summary, the changewasmum/dad variables are essential to improving the tie up between the parental age variables and the predictors for the parental outcome models.

# The Pseudo-Models

Considered whether best to model in a decision tree way via sptype variable(based on z1single change) or via typeofchange variable(more categories, and based on whether parental ages were/are 99 or not). Decided to use typeofchange variable as this gave more categeries that were of importance to the prediction of change in children and change in householdsize.

## Step 1

(Response variables at time t and all potential predictors)

Note: The ‘code#’s’ below refer to the same ‘code#’s’ distributed throughout the 'decision tree investigation for jess.sas' program which give investigations that show how I justified what I am getting Oliver to do in the simulation here.

**Family/household characteristics**

### Single Parent

**If Single-or-2-parent (t-1) = 0** then:

* **Single-or-2-parent (t)** = function(Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth)

**If Single-or-2-parent (t-1) = 1** then:

* **Single-or-2-parent (t)** = function(Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth)

(i.e two separate statistical models here based on subsets of the data where firstly Single-or-2-parent(t-1)=0 and then where Single-or-2-parent(t-1)=1)

### Calculate Family Type Variables

* **Onemorep(t**): if Single-or-2-parent (t) =0 and Single-or-2-parent(t-1)=1 then onemorep(t)=1; else onemorep(t)=0;
* **Onelessp(t**): if Single-or-2-parent (t) =1 and Single-or-2-parent(t-1)=0 then onelessp(t)=1; else onelessp(t)=0;
* **Samestatus(t):** if (Single-or-2-parent (t) =1 and Single-or-2-parent(t-1)=1) or (Single-or-2-parent (t) =0 and Single-or-2-parent(t-1)=0) then samestatus(t)=1; else samestatus(t)=0;
* Age (including 99 or not – presence or absence or not) of parents – and typeofchange(t), changewasmum, changewasdad variables totally created in this part – see below for flow diagram of this process too.
  + **If onemorep**(t)=1 (**see CODE #2** - 82 tall records in dataset where sptype=1 and typeofchange variable not missing – 2.4% or 2 records new mother, 97.56% or 80 records new father) and **fage\_years(t-1)** =99 then **fage\_years**(t)=function(**mage\_years(t)**); **mage\_years(t)=mage\_years(t-1**)+1**; typeofchange(t)**=2 – know for sure new dad or old dad back again; **changewasdad(t)**=1;

[Can base **fage\_years**(t)=function(**mage\_years(t)**) statistical model on the 80 tall records in data where sptype=1(i.e one more parent) and typeofchange=2 – i.e got new dad – different families in the main. See **code #3** for subset of data for this model]

* + **Else if** **onemorep**(t)=1 and **mage\_years(t-1)**=99 then

**mage\_years** =mage+age -4; typeofchange(t)=1 – know for sure new mum or old mum back again; changewasmum(t)=1;

[**see CODE#4** assume about 4 years younger than birth mother, as only 2 records in data where this happened- turned out to be two children with experience of new mother – one with new mother 4yrs younger than what birth mother would be, and the other 5yrs younger]

Fage\_years(t)=fage\_years(t-1)+1

* + **Else if onelessp**(t)=1 then do:
    - Take random uniform number between [0, 1]
    - **If** random number 1<=0.0738 - mum left– make **mage\_years(t)**=99; **typeofchange(t)**=3 –know for sure mum left; **changewasmum(t)**=1; **fage\_years(t)**=fage\_years(t-1)+1;
    - **Else** - know for sure father left– make **fage\_years(t)=**99; **typeofchange(t)**=4 –know for sure dad left; **changewasdad(t)**=1; **mage\_years(t)**=mage\_years(t-1)+1;
    - **End**;

[**see CODE#1** -of 122 records with sptype=2 that had non-missing **typeofchange** variable in the dataset- 7.38% or 2 records mum left, 96.2% or 113 records father left – this was based on the typeofchange variable which was based on the parental age variables ]

* + **Else if samestatus(t)=1** then do;
    - **typeofchange(t)**=5 – know for sure same number of parents this year as last year;
  + **If single mother family i.e if fage\_years(t-1)=99** then do;
    - If random number3<=0.0064 (see CODE #14) then do-make birth father return and mother leave
      * Make mage\_years(t)=99
      * Make fage\_years(t)= fage\_imputed + age. (see CODE#15)
      * Changewasmum(t)=1
      * Changewasdad(t)=1
      * End;
    - **Else if random number3>0.0064 then do; -assign if new mum or not**
      * **If random number4>0.0128 (see CODE#16 – adding %s for those were mumdiff not 0,1,2) then do;**
        + Same mum
        + Make **mage\_years(t)= Mage\_years(t-1)+1**
        + Make **fage\_years(t)=99**
        + **End;**
      * **Else if random number4<=0.0128 then do;**
        + New mum
        + **Changewasmum(t)=1**
        + **DifferenceInMumAge(t)=**

according to the 3 age differences in the data (see **code#17**):

-39 if mage(t-1)>=55 (=39+16 youngest age in data);

41 if mage(t-1)<=19 (=60 oldest in data set – 41)

Otherwise random chance [-1,3]

* + - * + Make **mage(t)=mage(t-1) + DifferenceInAge(t)**
        + Make **fage(t)=99**
        + **End;**
    - **End (of single mother sub statements);**
  + **If single father family i.e if mage(t-1)=99 then do;**
    - Make **fage(t)=fage(t-1) + 1 (**see **code #18 – all are consistent with same father – so just add one year here)**
    - Make **mage(t)=99**
    - **End  (of single sub father statements);**
  + **If two parent family i.e if mage(t-1) ne 99 and fage(t-1) ne 99 then do;**
    - **If randomnumber2<=0.0045 (code #11) then do-make both parents change**
      * **Changewasmum(t)=1;**
      * **changewasdad(t)=1;**
      * **Mage(t)=mage(t-1) +1**
      * **Fage(t)=fage(t-1) +1**
      * See **code #12 and #13** for how most new parents here are same age
      * **End;**
    - **Else if 0.0045<randomnumber2<=0.0098 then do-mum change only**
      * **Changewasmum(t)=1;**
      * **Mage(t)=mage(t-1) +1** (See **code #12 and #13** for how most new parents here are same age)
      * **Fage(t)=fage(t-1) +1**
      * **End;**
    - **Else if 0.0098<randomnumber2<=0.0272 then do-dad change only**
      * **Changewasdad(t)=1;**
      * **Mage(t)=mage(t-1) +1**
      * **Fage(t)=fage(t-1) +1** (See **code #12 and #13** for how most new parents here are same age)
      * **End;**
    - **Else if randomnumber2>0.0272 then do -neither change**
      * **Mage(t)=mage(t-1) +1**
      * **Fage(t)=fage(t-1) +1**
      * **End;**
    - **End (of 2 parent family sub statements)**
  + **End (of same status statements)**

Need to compare z1chpar\*sptype and z1chpar\*typeofchange distributions between the data and the simulation, particularly to check that not too many samestatus records had z1chpar, for unchanging single parent families, and for unchanging 2 parent families.

Note: whether a ‘onemorep=1’ child gets a new mum or a new dad is already determined by whether the previous mage\_years or fage\_years was 99 in the simulation – need to validate typeofchange distribution in simulation (created above based on combination of z1single and previous parental age variables) against the data (where typeofchange is based solely on parental age variables)?

### Family Type Flow Diagram



### Number of Children and Household Size

**Change-Children-num (t)** = round(function (typeofchange + Children-num(t-1) + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth))/\*although 28.89% of typeofchange is missing in dataset due to parental ages current and previous being missing – need to pay attention to validation of simulation here –however in v2 model weren’t able to use 27.18% of the data due to missing data anyway – **see code#7**\*/

*Tweaking* Change-Children-num *if out of range (see ‘working for method 4’ in ‘tweaking householdsize and kids to stop impossible values.docx) :*

if Change-Children-num <1- Children-num\_previous then make Change-Children-num=1- Children-num\_previous

Children-num(t)= Children-num(t-1) + Change-Children-num(t)

*Tweaking* Children-num *if out of range (based on what range was in 1-5 data)*

If Children-num(t)>10 then Children-num(t)=10

**Prop\_Additional\_household-size** **(t) =** function (typeofchange + householdsize(t-1) + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth)/\*although 28.89% of typeofchange is missing in dataset due to parental ages current and previous being missing– need to pay particular attention to validation of simulation here–however in v2 model weren’t able to use 27.18% of the data due to missing data anyway – see code#7 \*/

Additional\_household-size (t)=[ householdsize(t-1) –oneless(t)(1 or 0) +onemorep(t)(1 or 0) + Change-Children-num (t) ] \* Prop\_Additional\_household-size (t)

*Tweaking Additional\_household-size if out of range (see ‘working for method 4’ in ‘tweaking householdsize and kids to stop impossible values.docx) :*

If **z1single(t)**=1 then **numparents(t)**=1;

If **z1single(t)**=0 then **numparents(t)**=2;

**Distlimit(t)*=*** [Children-num(t) +numparents(t)] – [ Household-size(t-1) – onelessp(t)(1 or 0) +onemorep(t)(1 or 0) + Change-Children-num (t) ]

If additional\_household-size< distlimit(t) then additional\_householdsize(t)=distlimit(t).

**Household-size** **(t)** = round[ Household-size(t-1) – onelessp(t) (1 or 0) +onemorep(t) (1 or 0) + Change-Children-num (t) + additional\_household-Size]

*Tweaking household-size if out of range (based on what range was in 1-5 data):*

If household-size>14 then household-size=14

**Psychosocial factors**

***based on previous year’s as well as current year’s information***

### Change of Parents

**Change-of-parents (t)** = function(Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + **sptype(t-1))**

With sptype in here being compulsory as this enables consistency in its tie-up to sptype

Separate models depending on whether changed in the previous year.

(Separate models for each sptype??).

[See code#10 – table/graphs of sptype\*z1chpar where sptype=(onemorep, onelessp, samestatus) – seems highly predictive!]

Validation - check %(sptype \* chpar) against numbers in data especially if not doing a separate model for each sptype.

### Changes in Residence

**Change-score-Num-changes-of-residence (t)** = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Children-num(t-1) + Household-size(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1)+**typeofchange(t-1)**

Not sure on the sptype(t) covariate here – check

### Calculate More Family Type Variables

If **changewasdad(t)** ne 1 (for t=1 to T) and **fage\_years(t)** ne 99 and **group(fage\_years(1))=fage)** [checking out year 0 to 1 step that isn’t simulated] then **birthfather**=1; else **birthfather**=0;

If **changewasmum(t)** ne 1 (for t=1 to T) and **mage\_years(t)** ne 99 and **abs(mage\_years(1)**- **mage)**<2[checking out year 0 to 1 step that isn’t simulated] then **birthmother**=1; else **birthmother**=0;

If **changewasdad(t)** ne 1 and **fage\_years(t)** ne 99then **sameDadAsb4**=1; else **sameDadAsb4**=0;

If **changewasmum(t)** ne 1 and **mage\_years(t)** ne 99 then **sameMumAsb4**=1; else **sameMumasb4**=0;

If (**mage\_years**==99) then **mumgroup**=0

If (**birthmother**==1) then **mumgroup**=1

If ((**birthmother**==0)&(**sameMumAsb4**==0)) then **mumgroup**=2

If ((**birthmother**==0)&(**sameMumAsb4**==1)) then **mumgroup**=3

If (**fage\_years**==99) then **dadgroup**=0

If (**birthfather**==1) then **dadgroup**=1

If ((**birthfather**==0)&(**sameDadAsb4**==0)) then **dadgroup**=2

If ((**birthfather**==0)&(**sameDadAsb4**==1)) then **dadgroup**=3

**Employment**

### Welfare

**Change-score-Welfare (t)** = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education+ Father-education+ SES-birth + Single-or-2-parent-birth

### Mother’s Hours Worked

**If mumgroup=0** (mage\_years=99) then change-score-**Mother-hours-worked** (t)=0

Else if mumgroup=1 (birthmother=1) then change-score-Mother-hours-worked (t) = Mother-age-current + Child- age + Child-gender + Child-ethnicity + Mother-age-birth + Mother-ethnicity + Mother-education + SES-birth + Single-or-2-parent-birth

[5206 records to make model with]

Else **if mumgroup=2** (birthmother=0 and sameMumAsb4=0) then change-score-Mother-hours-worked (t) = Mother-age-current + Child- age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth

[14 records to make model with]

Else **if mumgroup=3** (birthmother=0 and sameMumAsb4=1) then change-score-Mother-hours-worked (t) = Mother-age-current + Child- age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth

[40 records to make model with]

### Father’s Hours Worked

**If** **dadgroup=0** (fage\_years=99) then **Father-hours-worked (t)=0**

Else **if dadgroup=1** (birthfather=1) then change-score-Father-hours-worked (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + Father-age-birth + Father-ethnicity + Father-education + SES-birth + Single-or-2-parent-birth

[4727 records to make model with]

Else **if dadgroup=2** (birthfather=0 and sameDadAsb4=0) then change-score-Father-hours-worked (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth

[188 records to make model with]

Else if **dadgroup=3** (birthfather=0 and sameDadAsb4=1) then change-score-Father-hours-worked (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth

[197 records to make model with]

## Step 2

**Material circumstances**

### Accomodation Type

**Accommodation-type (t)** = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Household-size + Children-num + Welfare + Mother-hours-worked + Father-hours-worked

### Owned or Rented

**Owned-rented (t)** = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Household-size + Children-num + Welfare + Mother-hours-worked + Father-hours-worked

### Over-crowding

**Overcrowding (t)** = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Household-size + Children-num + Welfare + Mother-hours-worked + Father-hours-worked

**Behavioural factors**

### Mother Current Smoking

**If mumgroup=0** (mage\_years=99) then change-score-**Mother-smoking-current(t)**=0;

Else **if mumgroup=1** (birthmother=1) then change-score-Mother-smoking-current (t) = Mother-age-current + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Mother-ethnicity + Mother-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num+ Household-size+ Welfare + Mother-hours-worked + Father-hours-worked

Else if **mumgroup=2** (birthmother=0 and sameMumAsb4=0) then change-score-Mother-smoking-current (t) = Mother-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num+ Household-size+ Welfare + Mother-hours-worked + Father-hours-worked

Else **if mumgroup=3** (birthmother=0 and sameMumAsb4=1) then change-score-Mother-smoking-current (t) = Mother-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num+ Household-size+ Welfare + Mother-hours-worked + Father-hours-worked

### Father Current Smoking

If **dadgroup=0** (fage\_years=99) then change-score-**Father-smoking-current (t)**=0;

Else **if dadgroup=1** (birthfather=1) then change-score**-**Father-smoking-current (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + Father-age-birth + Father-ethnicity + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num + Household-size + Welfare + Mother-hours-worked + Father-hours-worked

Else **if dadgroup=2** (birthfather=0 and sameDadAsb4=0) then change-score-Father-smoking-current (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num + Household-size + Welfare + Mother-hours-worked + Father-hours-worked

Else **if dadgroup=2** (birthfather=0 and sameDadAsb4=1) then change-score-Father-smoking-current (t) = Father-age-current + Child-age + Child-gender + Child-ethnicity + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current + Children-num + Household-size + Welfare + Mother-hours-worked + Father-hours-worked

## Step 3

**Health service use**

### Total Number of GP Visits

**Num-GP-visits (t)** = Birth-order + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Over-crowding(t-1) + Change-of-parents(t-1) + Number-of- changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy

### Number of Hospital Admissions

**Num-Hospital-admissions (t)** = Birthorder + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Overcrowding(t-1) + Change-of-parents(t-1) + Num-changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy

### Number of Outpatient Attendances

**Num-Outpatient-attendances (t)** = Birthorder + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Overcrowding(t-1) + Change-of-parents(t-1) + Num- changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy

## Step 4: Justice Outcomes

***General list of predictors:***

* **(change-score?)conduct-problems (t)** = Birth-order + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t) + Household-size(t) + Children-num(t) + Welfare(t) + Mother-hours-worked(t) + Father-hours-worked(t) + Accommodation-type(t-1) + Owned-rented(t) + Over-crowding(t) + Change-of-parents(t) + Number-of- changes-of-residence(t) + Mother-smoking-current(t) + Father-smoking-current(t) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + Burt-reading-score(t)
  + *Should be ts (as opposed to t-1s) because the conduct variable is effectively a at time of questionnaire variable (rather than an over the past year question). The wording of the question was “at present” or “within the last month or so”.*
* The below info is what we will have in the final model. What we have currently in the simulation is the models from using the LDV approach. Year 6 does not have an LDV, years 7 to 10 do. Only years 9 and 10 have burt previous in the model.
* Currently we do not have data on the intermediate outcomes so instead of using the deviation score in the models (or the actual value for year 6) we are doing cross-sectional/between models using the mean of the intermediate variables across years 1 to 5.

***Year 6:***

* **conduct-problems (t)** = Birth-order + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Over-crowding(t-1) + Change-of-parents(t-1) + Number-of- changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + Burt-reading-score(t-1)
* This will be a cross-sectional model as only will only have data on the outcome for one year (for the year 6 model).
* At year 6 (first year conduct is available) we do not use deviation score as the outcome but just the outcome itself. So in the simulation we simulate the actual value of conduct at year 6
* Burt reading score is first measured at age 9 so not included here

**Years 7 & 8:**

* **change-score-conduct-problems (t)** = Birth-order\*age + Child-age? + Child-gender\*age + Child-ethnicity\*age + Mother-age-birth\*age + Father-age-birth\*age + Mother-education\*age + Father-education\*age + SES-birth\*age + Single-or-2-parent-birth\*age + deviation-score-Single-or-2-parent-current(t-1) + deviation-score-Household-size(t-1) + deviation-score-Children-num(t-1) + deviation-score-Welfare(t-1) + deviation-score-Mother-hours-worked(t-1) + deviation-score-Father-hours-worked(t-1) + deviation-score-Accommodation-type(t-1) + deviation-score-Owned-rented(t-1) + deviation-score-Over-crowding(t-1) + deviation-score-Change-of-parents(t-1) + deviation-score-Number-of-changes-of-residence(t-1) + deviation-score-Mother-smoking-current(t-1) + deviation-score-Father-smoking-current(t-1) + Birth-weight\*age + Gestational-age\*age + Breast-feeding\*age + Mother-smoking-pregnancy\*age + Mother-drinking-pregnancy\*age
* For years 7 to 10 we do use the deviation score as the outcome (this is a fixed-effects model) and in the simulation we simulate the change from the previous time and add or subtract this to/from the simulated value at year 6 (or year 7).
* Burt reading score is first measured at age 9 so not included here

**Years 9 & 10:**

* **change-score-conduct-problems (t)** = Birth-order\*age + Child-age? + Child-gender\*age + Child-ethnicity\*age + Mother-age-birth\*age + Father-age-birth\*age + Mother-education\*age + Father-education\*age + SES-birth\*age + Single-or-2-parent-birth\*age + deviation-score-Single-or-2-parent-current(t-1) + deviation-score-Household-size(t-1) + deviation-score-Children-num(t-1) + deviation-score-Welfare(t-1) + deviation-score-Mother-hours-worked(t-1) + deviation-score-Father-hours-worked(t-1) + deviation-score-Accommodation-type(t-1) + deviation-score-Owned-rented(t-1) + deviation-score-Over-crowding(t-1) + deviation-score-Change-of-parents(t-1) + deviation-score-Number-of-changes-of-residence(t-1) + deviation-score-Mother-smoking-current(t-1) + deviation-score-Father-smoking-current(t-1) + Birth-weight\*age + Gestational-age\*age + Breast-feeding\*age + Mother-smoking-pregnancy\*age + Mother-drinking-pregnancy\*age + deviation-score-Burt-reading-score(t-1)
* For years 9 and 10 we do use the deviation score as the outcome (this is a fixed-effects model) and in the simulation we simulate the change from the previous time and add or subtract this to/from the simulated value at year 8 (or year 9).

## Step5: Education Outcomes

***General list of predictors:***

* **(change-score-)Burt-reading-score(t)** = Birth-order + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Over-crowding(t-1) + Change-of-parents(t-1) + Number-of- changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + conduct(t?)
  + *Should there be ts or t-1s above?*
* The below info is what we will have in the final model. What we have currently in the simulation is the models from using the LDV approach. Year 8 does not have an LDV, years 9 to 13 do.
* Currently we do not have data on the intermediate outcomes so instead of using the deviation score in the models (or the actual value for year 8) we are doing cross-sectional/between models using the mean of the intermediate variables across years 1 to 5.

***Year 8:***

* **Burt-reading-score(t)** = Birth-order + Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + Single-or-2-parent-current(t-1) + Household-size(t-1) + Children-num(t-1) + Welfare(t-1) + Mother-hours-worked(t-1) + Father-hours-worked(t-1) + Accommodation-type(t-1) + Owned-rented(t-1) + Over-crowding(t-1) + Change-of-parents(t-1) + Number-of- changes-of-residence(t-1) + Mother-smoking-current(t-1) + Father-smoking-current(t-1) + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + conduct(t?)
* *ts or t-1s?*
* This will be a cross-sectional model as only will only have data on the outcome for one year (for the year 8 model).
* At year 8 (first year Burt is available) we do not use deviation score as the outcome but just the outcome itself. So in the simulation we simulate the actual value of Burt at year 8.

**Years 9 & 10:**

* **Deviation-score-Burt-reading-score(t)** = Birth-order\*age + Child-age? + Child-gender\*age + Child-ethnicity\*age + Mother-age-birth\*age + Father-age-birth\*age + Mother-education\*age + Father-education\*age + SES-birth\*age + Single-or-2-parent-birth\*age + deviation-score-Single-or-2-parent-current(t-1) + deviation-score-Household-size(t-1) + deviation-score-Children-num(t-1) + deviation-score-Welfare(t-1) + deviation-score-Mother-hours-worked(t-1) + deviation-score-Father-hours-worked(t-1) + deviation-score-Accommodation-type(t-1) + deviation-score-Owned-rented(t-1) + deviation-score-Over-crowding(t-1) + deviation-score-Change-of-parents(t-1) + deviation-score-Number-of-changes-of-residence(t-1) + deviation-score-Mother-smoking-current(t-1) + deviation-score-Father-smoking-current(t-1) + Birth-weight\*age + Gestational-age\*age + Breast-feeding\*age + Mother-smoking-pregnancy\*age + Mother-drinking-pregnancy\*age + deviation-score-conduct(t?)
* For years 9 to 13 we do use the deviation score as the outcome (this is a fixed-effects model) and in the simulation we simulate the change from the previous time and add or subtract this to/from the simulated value at year 8 (or year 9).

**Years 11 to 13:**

* **Deviation-score-Burt-reading-score(t)** = Birth-order\*age + Child-age? + Child-gender\*age + Child-ethnicity\*age + Mother-age-birth\*age + Father-age-birth\*age + Mother-education\*age + Father-education\*age + SES-birth\*age + Single-or-2-parent-birth\*age + deviation-score-Single-or-2-parent-current(t-1) + deviation-score-Household-size(t-1) + deviation-score-Children-num(t-1) + deviation-score-Welfare(t-1) + deviation-score-Mother-hours-worked(t-1) + deviation-score-Father-hours-worked(t-1) + deviation-score-Accommodation-type(t-1) + deviation-score-Owned-rented(t-1) + deviation-score-Over-crowding(t-1) + deviation-score-Change-of-parents(t-1) + deviation-score-Number-of-changes-of-residence(t-1) + deviation-score-Mother-smoking-current(t-1) + deviation-score-Father-smoking-current(t-1) + Birth-weight\*age + Gestational-age\*age + Breast-feeding\*age + Mother-smoking-pregnancy\*age + Mother-drinking-pregnancy\*age + mean-conduct-over-years-6-to-10\*age
* For years 11 to 13 we do use the deviation score as the outcome (this is a fixed-effects model) and in the simulation we simulate the change from the previous time and add or subtract this to/from the simulated value at year 10 (or year 11 or 12).
* Conduct becomes a static variable for years 11 to 13 because conduct was only measured up to age 10.

# Special Case Models

For most outcomes, the model will be a fixed effects model with the deviation score as the outcome. However, there are some special cases where this blanket idea of using a fixed effects model with a deviation score as the outcome variable will not be suitable. These special cases and the modeling approach solution are given below.

## Special Case 1: The Starting Value Is Not Provided In the Base File and Needs to Be Generated In the Simulation

This occurs for variables that do not start at birth or year 1 (like conduct problems and Burt reading score). For these models we create a cross-sectional model using one year of the data. The outcome variable is the variable in its original form (rather than a deviation score) and this is a classical linear model including direct effects for time-invariant variables.

# Possible scenarios

* **Outcomes (final)**
  + GP visits: total (= preventive + total morbidity); preventive; total morbidity; respiratory
  + hospital admissions: total
  + hospital outpatient attendances (incl. acute and arranged): total
* **Levers** *(P=Perinatal, S=Structural, I=Intermediate)*
  + reduce smoking in pregnancy (smoking or number of cigarettes)
  + reduce drinking in pregnancy (drinking or number of drinks)
  + prevent low birth weight
  + promote breast-feeding
  + improve family socio-economic status (SES at birth of child)
  + improve parental educational level (at birth of child)
  + reduce proportion of (very young) single mothers (at birth of child)
  + increase support for single parents (given single-parent status of family impacts on health)
  + reduce level of unemployment
  + increase proportion of working mothers or hours worked (mother’s hours worked)
  + improve home ownership level (home owned/rented)
  + reduce overcrowding (ratio of number of bedrooms to household size)
  + reduce parental smoking (especially mother’s smoking)

**Do any of these levers work on any of the relevant outcomes?**

# Ways to test scenarios?

Current Way:

* Change value of a variable where change is most likely and thus change the distribution of that variable, e.g. identify the children with mothers most likely to have reduced or stopped smoking and assign them as having done so. Individual records are altered with only the variable of interest changing.

Other ways:

* Re-weight data on a variable so that the distribution of that variable changes, e.g. the proportion of children with a single-parent is weighted up (also used for alignment to external benchmarks). The records that are weighted up (e.g. for children with a single parent) are essentially replicated in total.
* Change a probability, e.g. of belonging to a single-parent family (also used for sensitivity analysis).
* Change a coefficient of interest (without changing others) in equation governing probability, e.g. of belonging to a single-parent family (not recommended).
* The above techniques could be applied to a combination of 2 or more variables simultaneously.
* Scaling the aggregate outcome is also used for alignment to external benchmarks.