**MEL- C PSEUDO CODE**

**Variables to be given in the basefile (each record represents one child):**

***A0***– four digit code for each child

Child characteristics at birth:

***Age0* =0, *z1gender, r1stchildethn*** (based on prioritisation of parents ethnicity)***, bw\_kg, ga, z1nicu, pregsmk, pregalc2, z1twin***

Child characteristics end of year 1:

***BREAST***

Parent characteristics at birth:

***mage, fage, r1stmethn, r1stfethn, r1stmeduc, r1stfeduc, SESBTH, z1single0***

Initial states for modeled variables (i.e year 1 values):

***z1Single1, householdsize1, kids1, welfare1, mhrswrk1, fhrswrk1, z1accom1, z1homeown1, z1overcrowd1, z1chpar1, chres1, gptotvisit1, hadmtot1, houtptot1,*** m***smoke1, fsmoke1***

***Age1=1***

**Simulation Process**

1. Create the following ‘indicator’ variables, using the following logic:

If r1stMeduc=1 then r1stMeducLvl1=1 ; else if r1stMeduc not in( .) then r1stmeducLvl1=0;

If r1stMeduc=2 then r1stMeducLvl2=1 ; else if r1stMeduc not in( .) then r1stmeducLvl2=0;

If r1stMeduc=3 then r1stMeducLvl3=1 ; else if r1stMeduc not in( .) then r1stmeducLvl3=0;

If R1stchildethn=1 then R1stchildethnLvl1=1 ; else if r1stchildethn not in( .) then r1stchildethnLvl1=0;

If R1stchildethn=2 then R1stchildethnLvl2=1 ; else if r1stchildethn not in( .) then r1stchildethnLvl2=0;

If R1stchildethn=3 then R1stchildethnLvl3=1 ; else if r1stchildethn not in( .) then r1stchildethnLvl3=0;

If Fage=1 then FageLvl1=1 ; else if fage not in(.) then fageLvl1=0;

If Fage=2 then FageLvl2=1 ; else if fage not in(.) then fageLvl2=0;

If Fage=3 then FageLvl3=1 ; else if fage not in(.) then fageLvl3=0;

If Fage=4 then FageLvl4=1 ; else if fage not in(.) then fageLvl4=0;

If Fage=5 then FageLvl5=1 ; else if fage not in(.) then fageLvl5=0;

If Fage=6 then FageLvl6=1 ; else if fage not in(.) then fageLvl6=0;

If R1stmethn=1 then R1stmethnLvl1=1 ; else if R1stmethn not in(.) then r1stmethnLvl1=0;

If R1stmethn=2 then R1stmethnLvl2=1 ; else if R1stmethn not in(.) then r1stmethnLvl2=0;

If R1stmethn=3 then R1stmethnLvl3=1 ; else if R1stmethn not in(.) then r1stmethnLvl3=0;

If R1stfethn=1 then R1stfethnLvl1=1 ; else if R1stfethn not in(.) then r1stfethnLvl1=0;

If R1stfethn=2 then R1stfethnLvl2=1 ; else if R1stfethn not in(.) then r1stfethnLvl2=0;

If R1stfethn=3 then R1stfethnLvl3=1 ; else if R1stfethn not in(.) then r1stfethnLvl3=0;

If R1stfeduc=1 then R1stfeducLvl1=1 ; else if r1stfeduc not in(.) then r1stfeducLvl1=0;

If R1stfeduc=2 then R1stfeducLvl2=1 ; else if r1stfeduc not in(.) then r1stfeducLvl2=0;

If R1stfeduc=3 then R1stfeducLvl3=1 ; else if r1stfeduc not in(.) then r1stfeducLvl3=0;

If R1stmeduc=1 then R1stmeducLvl1=1 ; else if R1stmeduc not in(.) then r1stmeducLvl1=0;

If R1stmeduc=2 then R1stmeducLvl2=1 ; else if R1stmeduc not in(.) then r1stmeducLvl2=0;

If R1stmeduc=3 then R1stmeducLvl3=1 ; else if R1stmeduc not in(.) then r1stmeducLvl3=0;

If Sesbth=1 then SesbthLvl1=1 ; else if sesbth not in(.) then sesbthLvl1=0;

If Sesbth=2 then SesbthLvl2=1 ; else if sesbth not in(.) then sesbthLvl2=0;

If Sesbth=3 then SesbthLvl3=1 ; else if sesbth not in(.) then sesbthLvl3=0;

if z1msmoke=**1** or z1fsmoke=**1** then z1smoke=**1**;

else if z1msmoke=**0** and z1fsmoke=**0** then z1smoke=**0**;

if breast=**1** and z1smoke=**1** then breastsmoklvl1=**1**;else breastsmoklvl1=**0**;

if breast=**1** and z1smoke=**0** then breastsmoklvl2=**1**;else breastsmoklvl2=**0**;

if breast=**0** and z1smoke=**1** then breastsmoklvl3=**1**;else breastsmoklvl3=**0**;

if breast=**0** and z1smoke=**0** then breastsmoklvl4=**1**;else breastsmoklvl4=**0**;

Then for each year (X) going from 2- 13 *(note ‘function’ below refers to listed variables being multiplied by ‘betas’ given by separate excel sheets)*:

***AgeX*=*Age(X-1) +1*** (e.g ***age1***= ***age0*** +1)

***Age\_minus1X***= ***ageX*** -1

**Version 3: pseudo-models - STEP 1**(Response variables at time t and all potential predictors)

***Family / household characteristics***

1. Imputing **single/two parent status**
   1. If *z1single(t-1)* = 0 then:

* *logit\_z1singleChangeX* = function(*ageX + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3+ MAGEX + MAGEX^2 + fage + fage^2+ r1stmethnLvl1 + r1stmethnLvl2 + r1stmethnLvl3 + r1stfethnLvl1 + r1stfethnLvl2 + r1stfethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + sesbthLvl1 + sesbthLvl2 + sesbthLvl3 + z1single0*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1singlePrev0.csv**
* *prob\_z1singleChangeX* = exp(logit\_z1singleChangeX)/ [1+ exp(*logit\_z1singleChangeX*)]
* *z1singleChangeX* = rbinom(1, 1, prob\_z1singleChangeX)
* *z1singleX = z1singleX(t-1)* ***+*** *z1singleChangeX* **(add the change)**
  1. If *z1single(t-1)* = 1 then:
* *logit\_z1singleChangeX* = function(*ageX + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3+ MAGEX + MAGEX^2 + fage + fage^2+ r1stmethnLvl1 + r1stmethnLvl2 + r1stmethnLvl3 + r1stfethnLvl1 + r1stfethnLvl2 + r1stfethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + sesbthLvl1 + sesbthLvl2 + sesbthLvl3 + z1single0*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1singlePrev1.csv**
* *prob\_z1singleChangeX* = exp(*logit\_z1singleChangeX*)/ [1 + exp(*logit\_z1singleChangeX*)]
* *z1singleChangeX* = rbinom(1, 1, prob\_z1singleChangeX)
* *z1singleX = z1singleX(t-1)* ***-*** *z1singleChangeX* **(subtract the change)**

1. **Change-Children-numX**
2. *Mean\_Change\_Children\_numX* = function (*typeofchangeLvl1 + typeofchangeLvl2 + typeofchangeLvl3 + typeofchangeLvl4 + typeofchangeLvl5 + age\*typeofchangeLvl1 + age\*typeofchangeLvl2 + age\*typeofchangeLvl3 + age\*typeofchangeLvl4 + age\*typeofchangeLvl5 + kids\_previous + age + MAGE + MAGE^2 + r1stfeducLvl1 + r1stfeducLvl2 +r1stfeducLvl3 + z1single0 + age\*z1single0*)
3. *Change\_Children\_numX* is a (rounded) random draw from a normal distribution with mean = *Mean\_Change\_Children\_numX*and sd = *kidschsigma.* E.g.: round(rnorm(1, mean = Mean\_Change\_Children\_numX, sd = kidschsigma) )
   * *kidschsigma* = sqrt(*Scale*) where *Scale* = 0.1624
4. Coefficients and *Scale* can be found in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **kidschbetas.csv**
5. if *Change-Children\_numX* < (1- *Children-num(X-1)*) then make *Change-Children\_numX* = (1- *Children-num(X-1)*)
6. *Children\_numX* = *Children\_num(X-1)* + *Change-Children\_numX*
7. *Children\_numX*: If *Children\_numX* > 10 then *Children\_numX* = 10
8. **Additional\_household-sizeX**
   1. *Mean\_Prop\_Additional\_household\_sizeX*= function(*typeofchangeLvl1 + typeofchangeLvl2 + typeofchangeLvl3 + typeofchangeLvl4 + typeofchangeLvl5 + age\*typeofchangeLvl1 + age\*typeofchangeLvl2 + age\*typeofchangeLvl3 + age\*typeofchangeLvl4 + age\*typeofchangeLvl5 + householdsize\_previous age\*householdsize\_previous + age + age^2 + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + age\*r1stchildethnLvl1 + age\*r1stchildethnLvl2 + age\*r1stchildethnLvl3 + MAGE + age\*MAGE + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3*)
   2. *Prop\_Additional\_household\_sizeX* is a random draw from a normal distribution with mean = *Mean\_Prop\_Additional\_household\_sizeX*and sd **=** *phssigma.* E.g.:rnorm(1, mean = Mean\_Prop\_Additional\_household-sizeX, sd = phssigma)
      * *Phssigma* = sqrt(*Scale*) where *Scale* = 0.0211
   3. Coefficients and *Scale* can be found in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **paddhsbetas.csv**
   4. *Additional\_household\_sizeX* **=** [*householdsize(X-1)*– *onelessX* (1 or 0) + *onemorepX* (1 or 0) *+ Change-Children\_numX*] \**Prop\_Additional\_household\_size*
   5. If *z1singleX*= 1 then *numparentsX*= 1; If *z1singleX* = 0 then *numparentsX*= 2;
   6. *DistlimitX* *=* [*Children\_numX* + *numparentsX*] – [*Household\_size(X-1)* – *onelesspX* (1 or 0) + *onemorepX* (1 or 0) + *Change\_Children\_numX*]
   7. If *additional\_household\_sizeX* < *distlimitX* then *additional\_householdsizeX* = *distlimitX*
   8. *Household\_size(t)* = round[*Household\_size(X-1)* – *onelesspX* (1 or 0) + *onemorepX* (1 or 0) + *Change\_Children\_numX* + *additional\_household\_SizeX*]
   9. *Household\_sizeX*: If *household\_sizeX* > 14 then *household\_sizeX* = 14

***Psychosocial factors– based on previous year’s as well as current year’s information***

1. ***Change-of-parents (t)*** :
   1. *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))***
   2. If *zchpar(t-1)* = 0 then:

* *logit \_z1chparChangeX* = function(*sptypeX + sptypeX\*ageX + ageX + z1gender + MAGEX + z1single\_previousX + ageX\*z1single\_previousX + welfare\_previousX + mhrswrk\_previousX)*
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1chparPrev0.csv**
* *prob\_z1chparChangeX* = exp(*logit\_z1chparChangeX*)/ [1+ exp(*logit\_z1chparChangeX*)]
* *z1chparX* = rbinom(1, 1, prob\_z1chparChangeX)
  1. If *zchpar(t-1)* = 1 then:
* *logit\_z1chparChangeX* = function(*sptypeX + z1gender + MAGEX)*
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1chparPrev1.csv**
* *prob\_z1chparChangeX* = exp(*logit\_z1chparChangeX*)/ [1+ exp(*logit\_z1chparChangeX*)]
* *z1chparX* = rbinom(1, 1, prob\_z1chparChangeX)

1. **Number of -changes-of-residence**
   1. *meanChangeChresX* = function(*SPTYPE\_PREVIOUS + age + chres\_previous + chres\_previous^2 + age^2 + z1gender + MAGE + MAGE^2 + age\*MAGE + age\*MAGE^2 + z1single\_previous + age\*z1single\_previous + householdsize\_previous + kids\_previous + welfare\_previous + mhrswrk\_previous*)
      * find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **chres.csv**
   2. *simChangeChresX* = rnorm(1, mean=meanChangeChresX, sd= sqrt(0.3819))
      * 0.3819 is the “Scale” in chres.csv
   3. *preChresX = chresX(t-1)* ***+*** *simChangeChresX*
   4. *chresX* = round(*preChresX*) (round to nearest whole number)
   5. if *chresX* < 0 then *chresX* = 0 (if negative just make 0)

***Employment***

1. **Welfare** 
   1. If *welfare(t-1)* = 0 then:

* *logit\_welfareChangeX* = function(*ageX + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + MAGEX + MAGEX^2 + fage + fage^2+ r1stmethnLvl1 + r1stmethnLvl2 + r1stmethnLvl3 + r1stfethnLvl1 + r1stfethnLvl2 + r1stfethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + sesbthLvl1 + sesbthLvl2 + sesbthLvl3 + z1single0*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **welfarePrev0.csv**
* *prob\_welfareChangeX* = exp(*logit\_welfareChangeX*)/ [1+ exp(*logit\_welfareChangeX*)]
* *welfareChangeX* = rbinom(1, 1, prob\_welfareChangeX)
* *welfareX = welfareX(t-1)* ***+*** *welfareChangeX* **(add the change)**
  1. If *welfare(t-1)* = 1 then:
* *logit\_welfareChangeX* = function(*ageX + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3+ MAGEX + MAGEX^2+ fage + fage^2+ r1stmethnLvl1 + r1stmethnLvl2 + r1stmethnLvl3 + r1stfethnLvl1 + r1stfethnLvl2 + r1stfethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + sesbthLvl1 + sesbthLvl2 + sesbthLvl3 + z1single0*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **welfarePrev1.csv**
* *prob\_welfareChangeX* = exp(*logit\_welfareChangeX*)/ [1+ exp(*logit\_welfareChangeX*)]
* welfareChangeX = rbinom(1, 1, prob\_welfareChangeX)
* *welfareX = welfareX(t-1)* ***-*** *welfareChangeX* **(subtract the change)**

1. **Mother Hours Worked**
   1. If *mumgroup*=0
      * then *mhrswrkX*=0
   2. If *mumgroup* = 1
      * then *meanChangeMhrswrkX* = function(*age + mhrswrk\_previous + age\*mhrswrk + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + age\*r1stmeducLvl1 + age\*r1stmeducLvl2 + age\*r1stmeducLvl3 + z1single0*)
      * Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **mhrswrkMumgroup1.csv**
      * *simChangeMhrswrkX* = rnorm(1, mean=meanChangeMhrswrkX, sd= sqrt(115.84))
        + 115.84 is the “Scale” in mhrswrkMumgroup1.csv
   3. If *mumgroup* = 2
      * then *meanChangeMhrswrkX* = 14.8333 (*intercept only model)*
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **mhrswrkMumgroup2.csv**
      * *simChangeMhrswrkX* = rnorm(1, mean=meanChangeMhrswrkX, sd= sqrt(289.21))
        + 289.21 is the “Scale” in mhrswrkMumgroup2.csv
   4. If *mumgroup* = 3
      * then *meanChangeMhrswrkX* = function(*mhrswrk\_previous*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **mhrswrkMumgroup3.csv**
      * *simChangeMhrswrkX* = rnorm(1, mean=meanChangeMhrswrkX, sd= sqrt(154.68))
        + 154.68 is the “Scale” in mhrswrkMumgroup3.csv
   5. preMhrswrkX = mhrswrkX(t-1) + simChangeMhrswrkX
   6. *mhrswrkX* = round(*preMhrswrkX*) (round to nearest whole number)
   7. If *mhrswrk* < *zero.twk* then *mhrswrk* = 0
      * See below on how to get the value of *zero.twk*.
      * The same value of *zero.twk* used is throughout all years (i.e. it is only calculated once using the year 1 basedata and the year 2 simulated data
   8. **Generic Tweak**
      * **Calculated after the year two mhrswrk data has been simulated**
      * obq = round(quantile(mhrswrk1, prob=(1:100)/100, na.rm=T), 2)
        + *mhrswrk1* is basefile data
      * For multiple iterations:
        + simqmat <- matrix(nrow=n, ncol=100)
        + simqmat[i,] <- quantile(preMhrswrk2, prob=(1:100)/100, na.rm=T)
          - *preMhrswrk2* is a vector of all simulated untweaked values of *mhrswrk* for year 2
        + simq = apply(simqmat, 2, mean)
      * for one iteration at a time (the way I think it actually is):
        + simq <- quantile(preMhrswrk2, prob=(1:100)/100, na.rm=T)
          - *preMhrswrk2* is a vector of all simulated untweaked values of *mhrswrk* for year 2
      * zero.twk = round(simq[max(which(obq==0))])
2. **Father Hours Worked**
   1. If *dadgroup*=0
      * then *fhrswrkX*=0
   2. If *dadgroup* = 1
      * then *meanChangeFhrswrkX* = function(*age + fhrswrk\_previous + age\*fhrswrk + age\*age + fage\_years + fage\_years^2 + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3 + age\*SESBTHLvl1 + age\*SESBTHLvl2 + age\*SESBTHLvl3*)
      * Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fhrswrkDadgroup1.csv**
      * *simChangeFhrswrkX* = rnorm(1, mean=meanChangeFhrswrkX, sd= sqrt(115.49))
        + 115.49 is the “Scale” in mhrswrkMumgroup1.csv
   3. If *dadgroup* = 2
      * then *meanChangeFhrswrkX* = function(*age*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fhrswrkDadgroup2.csv**
      * *simChangeFhrswrkX* = rnorm(1, mean=meanChangeFhrswrkX, sd= sqrt(640.82))
        + 640.82 is the “Scale” in fhrswrkDadgroup2.csv
   4. If *dadgroup* = 3
      * then *meanChangeFhrswrkX* = function(*fhrswrk\_previous + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3 + zsingle0*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fhrswrkDadgroup3.csv**
      * *simChangeFhrswrkX* = rnorm(1, mean=meanChangeFhrswrkX, sd= sqrt(140.84))
        + 140.84 is the “Scale” in fhrswrkDadgroup3.csv
   5. preFhrswrkX = fhrswrkX(t-1) + simChangeFhrswrkX
   6. *preFhrswrkRndX* = round(*preFhrswrkX*) (round to nearest whole number)
   7. apply tweak:
      * if *preFhrswrkRndX* is <= *zero.twk* then *fhrswrkX* = 0
      * if *preFhrswrkRndX* is >= *min40* and <= *max40* then *fhrswrkX*=40
        + instructions to calculate *zero.twk*, *min40* and *max40* given below. Values change each year (unlike for *mhrswrk*)
      * if *preFhrswrkRndX* > 100 then *fhrswrkX*=100
   8. **Generic Tweak:**
      * **Calculated at end of year 2 and adjusted after each following year**
      * **At end of year 2:**
      * obq = round(quantile(fhrswrk(t-1), prob=(1:100)/100, na.rm=T), 2)
        + above is not quite proper R code (the *fhrswrk(t-1)* bit)
        + *fhrswrk(t-1)* is the vector of *fhrswrk* from the previous year. So in this case it will be the *fhrswrk* values in the basefile for year 1
        + E.g. if *preFhrswrk* has just been simulated for year 3 then *fhrswrk(t-1)* is the vector of simulated values for year 2
      * For multiple iterations:
        + simqmat <- matrix(nrow=n, ncol=100)
        + simqmat[i,] <- quantile(preFhrswrkRnd, prob=(1:100)/100, na.rm=T)
        + simq = apply(simqmat, 2, mean)
        + preFhrswrkRnd is just those values for year 2
      * for one iteration at a time (the way I think it actually is)
        + simq <- quantile(preFhrswrkRnd, prob=(1:100)/100, na.rm=T)
        + preFhrswrkRnd is just those values for year 2
      * y2zero.twk = round(simq[max(which(obq<2))])
      * y2min40 = round(simq[min(which(obq==40))])
      * y2max40 = round(simq[max(which(obq==40))])
      * for year 2
        + *zero.twk = y2zero.twk*
        + *min40 = y2min40*
        + *max40 = y2max40*
      * **At end of other years:**
      * *zero.twk* = round(*y2zero.twk* + .2937\*(*ageX* – 2))
      * *min40* = round(*y2min40* + .373\*(*ageX* – 2))
      * *max40* = round(*y2max40* - .0909\*(*ageX* – 2))

**Version 3: pseudo-models - STEP 2**

***Material circumstances***

1. **Accomodation**
   1. If *z1accom(t-1)* = 0 then:

* *logit \_accomChangeX* = function(*ageX + z1gender + r1stchildethnLvl2 + r1stchildethnLvl3+ MAGEX + ageX\*MAGEX + r1stmeducLvl2 + r1stmeducLvl3 + sesbthLvl2 + sesbthLvl3 + z1single0 + householdsizeX + ageX\*householdsizeX + kidsX + welfareX + mhrsrwkX*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1accomPrev0.csv**
* *prob\_accomChangeX* = exp(*logit\_accomChangeX*)/ [1+ exp(*logit\_accomChangeX*)]
* *accomChangeX* = rbinom(1, 1, prob\_accomChangeX)
* *z1accomX = z1accomX(t-1)* ***+*** *accomChangeX* **(add the change)**
  1. If *z1accom(t-1)* = 1 then:
* *logit \_accomChangeX* = function(*r1stchildethnLvl2 + r1stchildethnLvl3 + z1singleX + householdsizeX + kidsX)*
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1accomPrev1.csv**
* *prob\_accomChangeX* = exp(*logit\_accomChangeX*)/ [1+ exp(*logit\_accomChangeX*)]
* *accomChangeX* = rbinom(1, 1, prob\_accomChangeX)
* *z1accomX = z1accomX(t-1)* ***-*** *accomChangeX* **(subtract the change)**

1. **Home-own**
   1. If *z1homeown(t-1)* = 0 then:

* *logit \_homeownChangeX* = function(*ageX + z1gender + MAGEX + r1stmeducLvl2 + r1stmeducLvl3 + z1singleX + householdsizeX + kidsX + welfareX + mhrsrwkX + fhrsrwkX + ageX\*fhrswrkX*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1homeownPrev0.csv**
* *prob\_homeownChangeX* = exp(*logit\_homeownChangeX*) / (1 + exp(*logit\_homeownChangeX*))
* *homeownChangeX* = rbinom(1, 1, prob\_homeownChangeX)
* *z1homeownX = z1homeownX(t-1)* ***+*** *homeownChangeX* **(add the change)**
  1. If *z1homeown (t-1)* = 1 then:
* *logit \_homeownChangeX* = function(*ageX + r1stchildethnLvl2 + r1stchildethnLvl3 + SESBTHLvl2 + SESBTHLvl3 + z1singleX + kidsX +ageX\*kidsX + welfareX*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1**homeown**Prev1.csv**
* *prob\_homeownChangeX* = exp(*logit\_homeownChangeX*) / (1 + exp(*logit\_homeownChangeX*))
* *homeownChangeX* = rbinom(1, 1, prob\_homeownChangeX)
* *z1homeownX = z1homeownX(t-1)* ***-*** *homeownChangeX* **(subtract the change)**

1. **Overcrowding**
   1. If *z1overcrowd(t-1)* = 0 then:

* *logit \_overcrowdChangeX* = function(*ageX + z1gender + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3+ MAGEX + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3+ age\*r1stmeducLvl1 + age\*r1stmeducLvl2 + age\*r1stmeducLvl3 + householdsizeX + ageX\*householdsizeX + kidsX + agex\*kidsX + welfareX*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1overcrowdPrev0.csv**
* *prob\_overcrowdChangeX* = exp(*logit\_overcrowdChangeX*) / (1 + exp(*logit\_overcrowdChangeX*))
* *overcrowdChangeX* = rbinom(1, 1, prob\_overcrowdChangeX)
* *z1overcrowdX = z1overcrowdX(t-1)* ***+*** *overcrowdChangeX*  **(add the change)**
  1. If *z1overcrowd(t-1)* = 1 then:
* *logit \_overcrowdChangeX* = function(*ageX + ageX^2 + householdsizeX + welfareX*)
* Find coefficients in S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **z1overcrowdPrev1.csv**
* *prob\_overcrowdChangeX* = exp(*logit\_overcrowdChangeX*) / (1 + exp(*logit\_overcrowdChangeX*))
* *overcrowdChangeX* = rbinom(1, 1, prob\_overcrowdChangeX)
* *z1overcrowdX = z1overcrowdX(t-1)* ***-*** *overcrowdChangeX* **(subtract the change)**

***Behavioural factors***

1. **Mother Smoking**
   1. Same method as for *mhrswrk*
   2. If *mumgroup*=0
      * then *msmokeX*=0
   3. If *mumgroup* = 1
      * then *meanChangeMsmokeX* = function(*msmoke\_previous + mage\_years + mage\_years^2 + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3 + kids + welfare + mhrswrk*)
      * Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **msmokeMumgroup1.csv**
      * *simChangeMsmokeX* = rnorm(1, mean=meanChangeMsmokeX, sd= sqrt(16.9314))
        + 16.9314 is the “Scale” in msmokeMumgroup1.csv
   4. If *mumgroup* = 2
      * then *meanChangeMsmokeX* = function (*age + age^2 + mage\_years + z1single0 + z1single + householdsize + age\*householdsize + kids + age\*kids + mhrswrk*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **msmokeMumgroup2.csv**
      * *simChangeMsmokeX* = rnorm(1, mean=meanChangeMsmokeX, sd= sqrt(3.0621))
        + 3.0621 is the “Scale” in msmokeMumgroup2.csv
   5. If *mumgroup* = 3
      * then *meanChangeMsmokeX* = function(*msmoke\_previous, z1gender*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **msmokeMumgroup3.csv**
      * *simChangeMsmokeX* = rnorm(1, mean=meanChangeMsmokeX, sd= sqrt(48.2458))
        + 48.2458 is the “Scale” in msmokeMumgroup3.csv
   6. preMsmokeX = msmokeX(t-1) + simChangeMsmokeX
   7. *msmokeX* = round(*preMsmokeX*) (round to nearest whole number)
   8. If *msmoke* < *zero.twk* then *msmoke* = 0
      * See below on how to get the value of *zero.twk*.
      * The same value of *zero.twk* used is throughout all years (i.e. it is only calculated once using the year 1 basedata and the year 2 simulated data
   9. **Generic Tweak**
      * **Calculated after the year two *msmoke* data has been simulated**
      * obq = round(quantile(msmoke1, prob=(1:100)/100, na.rm=T), 2)
        + *msmoke1* is basefile data
      * For multiple iterations:
        + simqmat <- matrix(nrow=n, ncol=100)
        + simqmat[i,] <- quantile(preMsmoke2, prob=(1:100)/100, na.rm=T)
          - *preMsmoke2* is a vector of all simulated untweaked values of *msmoke* for year 2
        + simq = apply(simqmat, 2, mean)
      * for one iteration at a time (the way I think it actually is):
        + simq <- quantile(preMsmoke2, prob=(1:100)/100, na.rm=T)
          - *preMsmoke2* is a vector of all simulated untweaked values of *msmoke* for year 2
      * zero.twk = round(simq[max(which(obq==0))])
2. **Father Smoking**
   1. Same method as for *msmoke* except slightly different zero tweak (add 1 – see last line of this *fsmoke* section)
   2. If *dadgroup*=0
      * then *fsmokeX*=0
   3. If *dadgroup* = 1
      * then *meanChangeFsmokeX* = function(*fsmoke\_previous + fage\_years + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3 + z1single0 + welfare*)
      * Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fsmokeDadgroup1.csv**
      * *simChangeFsmokeX* = rnorm(1, mean=meanChangeFsmokeX, sd= sqrt(27.5158))
        + 27.5158 is the “Scale” in fsmokeDadgroup1.csv
   4. If *dadgroup* = 2
      * then *meanChangeFsmokeX* = function (*age* + *age*^2)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fsmokeDadgroup2.csv**
      * *simChangeFsmokeX* = rnorm(1, mean=meanChangeFsmokeX, sd= sqrt(128.71))
        + 128.71 is the “Scale” in msmokeMumgroup2.csv
   5. If *dadgroup* = 3
      * then *meanChangeFsmokeX* = function(*fsmoke\_previous, fage\_years, SESBTHLvl1, SESBTHLvl2, SESBTHLvl3, welfare*)
      * Find intercept and scale in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **fsmokeDadgroup3.csv**
      * simChangeFsmokeX = rnorm(1, mean=meanChangeFsmokeX, sd= sqrt(23.883))
        + 23.883 is the “Scale” in fsmokeDadgroup3.csv
   6. preFsmokeX = fsmokeX(t-1) + simChangeFsmokeX
   7. *fsmokeX* = round(*preFsmokeX*) (round to nearest whole number)
   8. If *fsmoke* < *zero.twk* then *fsmoke* = 0
      * See below on how to get the value of *zero.twk.*
      * The **same value of *zero.twk* used is throughout all years** (i.e. it is only calculated once using the year 1 basedata and the year 2 simulated data
   9. **Generic Tweak**
      * **Calculated after the year two *fsmoke* data has been simulated**
      * obq = round(quantile(fsmoke1, prob=(1:100)/100, na.rm=T), 2)
        + *fsmoke1* is basefile data
      * For multiple iterations:
        + simqmat <- matrix(nrow=n, ncol=100)
        + simqmat[i,] <- quantile(preFsmoke2, prob=(1:100)/100, na.rm=T)
          - *preFsmoke2* is a vector of all simulated untweaked values of *fsmoke* for year 2
        + simq = apply(simqmat, 2, mean)
      * for one iteration at a time (the way I think it actually is):
        + simq <- quantile(preFsmoke2, prob=(1:100)/100, na.rm=T)
          - *preFsmoke2* is a vector of all simulated untweaked values of *fsmoke* for year 2
      * zero.twk = round(simq[max(which(obq==0))]) + 1

**Version 3: pseudo-models - STEP 3**

***Health service use***

1. **GP Total Visits**

*Gptotvis* only goes up to age 10 in the data so **only simulate up to age 10**

There are three separate models: one for age 2, one for ages 3 to 6 and one for ages 7 to 10.

* Age 2:
  + *LogGPTotVisX* is an intercept only model
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **gptotvisAge2.csv**
* Ages 3 to 6:
  + *LogGPTotVisX* = function(*age + bthorder + age\*bthorder + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + fage + kids\_previous + kids\_previous*^2 + *z1homeown\_previous + z1chpar\_previous + pregsmk + pregsmk*^2)
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **gptotvisAges3\_6.csv**
* Ages 7 to 10
  + *LogGPTotVisX* = function(*age + bthorder + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + fage + SESBTHLvl1 + SESBTHLvl2 + SESBTHLvl3 + kids\_previous + msmoke\_previous + ga + pregalc + pregalc^2*)
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **gptotvisAges7\_10.csv**
* *meanGPTotVisX* = exp(*LogGPTotVisX*)
* *gptotvisX* is a random draw from a Negative Binomial distribution with mean *meanGPTotVisX*, and size=1/*Alpha*
* use **rnbinom(1, mu=meanGPTotVisX, size=1/Alpha)** in R, where *Alpha* is ‘Scale’ in the above csv files and is different for each of the three models. For the age 2 model *Alpha* is .1441, for the ages 3 to 6 model *Alpha* is .5012 and for the ages 7 to 10 model *Alpha* is .8589)

(Num-GP-visits sub-outcomes: respiratory, total morbidity, preventive still to do)

1. **Total Number of Hospital Admissions**

hadmtot only goes up to age 10 in the data so **only simulate up to age 10**

* *Log\_hadmtotX =* function(*bthorder + age + age\*age + z1gender + MAGE + age\* MAGE + sesbthLvl1 + sesbthLvl2 + sesbthLvl3 + householdsize\_previousX + age\*householdsize\_previousX + kids\_previousX + age\*kids\_previousX + welfare\_previousX + chres\_previousX + age\* chres\_previousX + bwkg[[1]](#footnote-1) + age\*bwkg*)
  + - Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **hadmtot.csv**
* *mean\_hadmtotX =* exp(*Log\_hadmtotX*)
* *hadmtotX* is a random draw from a Poisson distribution with mean *mean\_hadmtotX* (use rpois(1, mean\_hadmtotX) in R)

1. **Total Number of Outpatient Attendances**

houtptot only goes up to age 10 in the data so **only simulate up to age 10**

* *LogHoutptotX =* function*(age + age^2 + r1stmeducLvl1 + r1stmeducLvl2 + r1stmeducLvl3 + welfare\_previous + fsmoke\_previous + BREAST + BREAST^2)*
  + - Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **houtptot.csv**
* *meanHoutptotX* = exp(*LogHoutptotX*)
* *houtptotX* is a random draw from a Negative Binomial distribution with mean *meanHoutptotX*, and size=1/*Alpha*
* use **rnbinom(1, mu=meanHoutptotX, size=1/Alpha)** in R, where *Alpha* is given from SAS output)

1. **GP Preventative Visits**

*Gpprev* only goes up to age 7 in the data so **only simulate up to age 7**

There are three separate models: one for age 2, one for ages 3 and4 and one for ages 5 to 7.

* Age 2:
  + *meanGPPrevX* = function(bthorder + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + MAGE + MAGE^2 + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + z1overcrowd\_previous)
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\**gpprevAge2Norm.csv**
  + *PreGPPrevX* is a random draw from a normal distribution with mean *meanGPPrevX* and standard deviationsqrt(*Scale*)in the csv file (sqrt(0.2679)).
  + Usernorm(1, mean=meanGPPrevX, sd=sqrt(.2679)).
  + *GPPrevX* is *PreGPPrevX* rounded in a modified manner
    - Rounding rules:
    - *simval2* is *PreGPPrev*

if (PreGPPrev[k]<.7) {

GPPrev[k] <- 0

}

if ((PreGPPrev [k]>=.7)&(PreGPPrev[k]<1.3)) {

GPPrev[k] <- 1

}

if ((PreGPPrev[k]>=1.3)&(PreGPPrev[k]<2.7)) {

GPPrev[k] <- 2

}

if ((PreGPPrev[k]>=2.7)&(PreGPPrev[k]<4.5)) {

GPPrev[k] <- 3

}

* Ages 3 and 4:
  + *LogGPPrevX* = function(*age + bthorder + age\*bthorder + r1stfeducLvl1 + r1stfeducLvl2 + r1stfeducLvl3 + householdsize\_previous + kids\_previous + fhrswrk\_previous + fhrswrk\_previous^2 + age\*fhrswrk\_previous* + *z1homeown\_previous + age\* z1homeown\_previous + fsmoke\_previous + age\*fsmoke\_previous + BREAST*)
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\**gpprevAges3\_4Pois.csv**
  + *mean\_gpprevX =* exp(*LogGPPrevX*)
  + *gpprevX* is a random draw from a Poisson distribution with mean *mean\_gpprevX* (use rpois(1, mean\_gpprevX) in R)
* Ages 5 to 7
  + *LogitGPPrevX* = function(*age + age^2 + z1accom\_previous*)
  + Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\**gpprevAges5\_7Logistic.csv**
  + *probGPPrevX* = exp(*LogitGPPrevX*)/(1 + exp(*LogitGPTotVis*))
  + *gpprevX* is a random draw from a Binomial distribution with probability *probGPPrevX*
  + use **rbinom(1, size=1, prob=probGPPrevX)** in R

1. **GP Visits for a Morbidity**

gpmorb only goes up to age 7 in the data so **only simulate up to age 7**

* *LogGPMorbX =* function*(age + r1stchildethnLvl1 + r1stchildethnLvl2 + r1stchildethnLvl3 + age\*r1stchildethnLvl1 + age\*r1stchildethnLvl2 +age\*r1stchildethnLvl3 + bthorder + age\*bthorder + fage + fage^2 + kids\_previous + z1homeown\_previous + ga + BREAST + pregsmk + pregsmk^2 + age\*pregsmk + pregalc)*
  + - Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **gpmorb.csv**
* *meanGPMorbX* = exp(*LogGPMorbX*)
* *gpmorbX* is a random draw from a Negative Binomial distribution with mean *meanGPMorbX*, and size=1/*Alpha*
* use **rnbinom(1, mu=meanGPMorbX, size=1/Alpha)** in R, where *Alpha* is given from SAS output and in this case is 0.5729.

1. **GP Visits for a Respiratory Condition**

gpresp only goes up to age 10 in the data so **only simulate up to age 10**

* *LogGPRespX =* function*(age + age^2 + bthorder + age\*bthorder + fage + fage^2 + SESBTHLvl1 + SESBTHLvl2 + age\*SEBTHLvl1 + age\*SESBTHLvl2 + r1stmeducLvl1 + r1stmeducLvl2 + age\*r1stmeducLvl1 + age\*r1stmeducLvl2 + kids\_previous + z1accom\_previous + z1homeown\_previous + z1chpar\_previous + msmoke\_previous + age\*msmoke\_previous + fsmoke\_previous + bwkg + pregsmk + pregsmk^2 + pregalc)*
  + - Find coefficients in: S:\Symonds Group\soc\Sociology Research Group\Projects\FRST - MEL-C\Health service use model version3\modelCoefficients\ **gpresp.csv**
* *meanGPRespX* = exp(*LogGPRespX*)
* *gprespX* is a random draw from a Negative Binomial distribution with mean *meanGPRespX*, and size=1/*Alpha*
* *Alpha = g0 + g1\*age*
* *g0* = 0.384 and *g1* = 0.1567 and can be found in gpresp.csv
* use **rnbinom(1, mu=meanGPRespX, size=1/Alpha)** in R
* So a different *Alpha* is used for each age. At the start of the simulation for each year you can calculate the *Alpha* for that year. Each child in a given year will have the same *Alpha*.

From version 2:

**STEP 7 (*Health-service-use)*:**

**CONDUCT**

1. **Conduct**
   1. Create variables at end of year 5 simulation: (each child will get a value for each of the below variables)
      * **Meansingle** = 1 - mean(z1single1, z1single2, z1single3, z1single4, z1single5)
      * **Meanmhrswrk** = 15.6\*mean(z1mhrswrk1, z1mhrswrk2, z1mhrswrk3, z1mhrswrk4, z1mhrswrk5)
      * **meanfhrswrk** = 45.3\*mean(z1fhrswrk1, z1fhrswrk2, z1fhrswrk3, z1fhrswrk4, z1fhrswrk5)
      * **catsumchpar** = sum(z1chpar1,..., z1chpar5)
        + if catsumchpar>2 catsumchpar=2
      * **meanchres** = 1.6\*mean(z1chres1, ..., z1chres5)
   2. **Age 6** 
      * **use Cond6.csv**
      * This model also have squared terms (not sure how you deal with these):
        + **meanmhrswrk\*meanmhrswrk**
        + **meanfhrswrk\*meanfhrswrk**
      * **log\_condt(t) =** Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + meansingle + meanhhsize + meankids + meanWelfare + mean-Mother-hours-worked + mean-Mother-hours-worked\*mean-Mother-hours-worked + mean-Father-hours-worked + mean-Father-hours-worked\*mean-Father-hours-worked + mean-z1Accomm + mean-Owned-rented + meanz1overcrowd + mean-Change-of-parents + mean-Num-changes-of-residence + mean-Mother-smoking-current + mean-Father-smoking-current + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + punish + interact
      * **mean\_condt(t)** = exp(log\_condt(t))
      * ***condtX*** is a random draw from a negative binomial distribution with mean ***mean\_condtX***, and size=1/***alpha*** (use **rnbinom(1, mu=mean\_condtX, size=1/alpha)** in R, where ***alpha*** is given from SAS output)
      * if condtX > 14 then condtX=14
      * **condX = condtX + 9**
   3. **Ages 7 and 8:** 
      * use Cond7\_8.csv
      * **Condt\_previous** is the condt value (not the cond value) from the previous year
      * Create variable:
        + **condt\_previous\*age** = condt\_previous\*age
      * **log\_condt(t) = condt\_previous +** Child-age + condt\_previous\*age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + meansingle + meanhhsize + meankids + meanWelfare + mean-Mother-hours-worked + mean-Father-hours-worked + mean-z1Accomm + mean-Owned-rented + meanz1overcrowd + mean-Change-of-parents + mean-Num-changes-of-residence + mean-Mother-smoking-current + mean-Father-smoking-current + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-drinking-pregnancy + punish + interact
      * **mean\_condt(t)** = exp(log\_condt(t))
      * ***condtX*** is a random draw from a negative binomial distribution with mean ***mean\_condtX***, and size=1/***alpha*** (use **rnbinom(1, mu=mean\_condtX, size=1/alpha)** in R, where ***alpha*** is given from SAS output)
      * if condtX > 30 then condtX=30
      * **condX = condtX + 20**
   4. **Ages 9 and 10**:
      * use Cond9\_10.csv
      * **Condt\_previous** is the condt value (not the cond value) from the previous year
      * Create variable: **pregsmk\*pregsmk** = pregsmk\*pregsmk
      * **log\_condt(t) = condt\_previous +** Child-age + condt\_previous\*age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + meansingle + meanhhsize + meankids + meanWelfare + mean-Mother-hours-worked + mean-Father-hours-worked + mean-z1Accomm + mean-Owned-rented + meanz1overcrowd + mean-Change-of-parents + mean-Num-changes-of-residence + mean-Mother-smoking-current + mean-Father-smoking-current + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-smoking-pregnancy\* Mother-smoking-pregnancy + Mother-drinking-pregnancy + punish + interact + burt\_previous
      * **mean\_condt(t)** = exp(log\_condt(t))
      * ***condtX*** is a random draw from a negative binomial distribution with mean ***mean\_condtX***, and size=1/***alpha*** (use **rnbinom(1, mu=mean\_condtX, size=1/alpha)** in R, where ***alpha*** is given from SAS output)
      * if condtX > 35 then condtX=35
      * **condX = condtX + 20**

**EDUCATION**

1. **Burt Reading Score**
   1. Note – currently Burt is the only variable that goes up to age 13
   2. **Age 8**
      * Use Burt8.csv
      * Create variables at anytime after end of year 5 simulation (but before this step): (each child will get a value for each of the below variables)
      * **meankids** = 1.67 + .348\*mean(z1kids1, z1kids2, z1kids3, z1kids4, z1kids5)
      * **meanwelfare** = mean(welfare1, …, welfare5)
      * This model has cond and **cond\*cond** terms. For cond you will use the cond value generated for year 8 which has already been generated by the simulation. cond\*cond is the squared which you may need to create.
      * Burt Reading Score(t) = Child-age + Child-gender + Child-ethnicity + Mother-age-birth + Father-age-birth + Mother-education + Father-education + SES-birth + Single-or-2-parent-birth + meansingle + meanhhsize + meankids + meanWelfare + mean-Mother-hours-worked + mean-Father-hours-worked + mean-z1Accomm + mean-Owned-rented + meanz1overcrowd + mean-Change-of-parents + mean-Num-changes-of-residence + mean-Mother-smoking-current + mean-Father-smoking-current + Birth-weight + Gestational-age + Breast-feeding + Mother-smoking-pregnancy + Mother-smoking-pregnancy-squared + Mother-drinking-pregnancy + punish + interact + conduct
      * ***burtX*** is a random draw from a normal distribution with mean ***mean\_burtX***, and sd= 15.06858 (use **rnorm(1, mean=mean\_burtX, sd=15.06858)** in R)
      * if burtX > 99 then burtX=99
      * if burtX<0 then burtX=0
   3. **Ages 9 and 10**
      * Use Burt9\_10.csv
      * This model has coefficients for:
        + **burt\_previous\*burt\_previous and**
        + **age\*burt\_previous.**
      * Burt Reading Score(t) = Burt-reading-score(t-1) + Burt-reading-score(t-1)\* Burt-reading-score(t-1) + Child-age + child-age\*burt-reading-score(t-1) + … + conduct (at current age)
      * ***burtX*** is a random draw from a normal distribution with mean ***mean\_burtX***, and sd= 7.587072 (use **rnorm(1, mean=mean\_burtX, sd=** **7.587072)** in R)
      * if burtX > 108 then burtX=108
      * if burtX<0 then burtX=0
   4. **Ages 11 to 13**
      * Use Burt11\_13.csv
      * This model has a coefficient for:
        + **burt\_previous\*burt\_previous**
      * Burt Reading Score(t) = Burt-reading-score(t-1) + Burt-reading-score(t-1) \* Burt-reading-score(t-1) + Child-age + … + meanconduct (but was not significant)
      * ***burtX*** is a random draw from a normal distribution with mean ***mean\_burtX***, and sd= 6.512851 (use **rnorm(1, mean=mean\_burtX, sd=** **6.512851)** in R)
      * if burtX > 110 then burtX=110
      * if burtX<0 then burtX=0

**STEP 8 *(collating and presenting of outputs/ aggregate numbers):***

***GptotvisX, hadmtotX, HoutptotX***

Mean, and distribution of number of visits

across the years

* by gender and ethnicity

1. Bwkg is birthweight in kilograms and can be calculated from bw by dividing bw by 1000. [↑](#footnote-ref-1)