### Supplementary material

Supplementary Table 1. Parent reported Strengths and Difficulties Questionnaire (SDQ) items used to measure MHC in the LSAC and MCS.

<table>
<thead>
<tr>
<th>SDQ subscale</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosocial behaviours</td>
<td>Considerate of other people's feelings</td>
</tr>
<tr>
<td></td>
<td>Shares readily with others</td>
</tr>
<tr>
<td></td>
<td>Helpful if someone is hurt, upset or feeling ill</td>
</tr>
<tr>
<td></td>
<td>Kind to younger children</td>
</tr>
<tr>
<td></td>
<td>Often volunteers to help others (parents, teachers, other children)</td>
</tr>
<tr>
<td>Conduct problems</td>
<td>Generally obedient, usually does what adults request</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>Sees tasks through to the end</td>
</tr>
<tr>
<td></td>
<td>Thinks things out before acting</td>
</tr>
</tbody>
</table>

Note. Parents reported on a 3-point scale, 1=not true; 2=somewhat true; 3=certainly true. There were minor wording changes to items between child and youth waves.

Supplementary Table 2. Number and proportion (%) of participants with missing data on study variables in the LSAC (N=4983) and MCS (N=18,296).

<table>
<thead>
<tr>
<th>Variable</th>
<th>LSAC n(%)</th>
<th>MCS n(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mental health competence (SDQ)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-5 years</td>
<td>20 (0.4%)</td>
<td>8039 (43.9%)</td>
</tr>
<tr>
<td>6-7 years</td>
<td>683 (13.7%)</td>
<td>8069 (44.1%)</td>
</tr>
<tr>
<td>10-11 years</td>
<td>869 (17.4%)</td>
<td>6958 (38.0%)</td>
</tr>
<tr>
<td>14-15 years</td>
<td>1617 (32.5%)</td>
<td>8865 (48.5%)</td>
</tr>
<tr>
<td>Sociodemographic characteristics (4-5 years)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child’s sex</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
</tr>
<tr>
<td>Parental education</td>
<td>63 (1.3%)</td>
<td>3043 (16.6%)</td>
</tr>
<tr>
<td>Parent unemployment</td>
<td>17 (0.3%)</td>
<td>3835 (21.0%)</td>
</tr>
<tr>
<td>Young maternal age at child’s birth</td>
<td>2 (0.0%)</td>
<td>58 (0.3%)</td>
</tr>
</tbody>
</table>
Supplementary Table 3. Sociodemographic characteristics of participants in the LSAC (N=4,983) and MCS (N=18,296).

<table>
<thead>
<tr>
<th></th>
<th>LSAC (N = 4,983)</th>
<th></th>
<th></th>
<th>MCS (N = 18,296)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
<td>95% CI</td>
<td>%</td>
<td>95% CI</td>
</tr>
<tr>
<td><strong>Child’s sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>51.2</td>
<td>49.8–52.7</td>
<td>51.4</td>
<td>50.5–52.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>48.8</td>
<td>47.3–50.2</td>
<td>48.6</td>
<td>47.8–49.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parental education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher (degree level)</td>
<td>27.2</td>
<td>24.7–29.6</td>
<td>28.7</td>
<td>26.4–31.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower (below degree)</td>
<td>72.8</td>
<td>70.4–75.3</td>
<td>71.3</td>
<td>69.0–73.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Parent unemployment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Working parent</td>
<td>87.0</td>
<td>85.7–88.3</td>
<td>84.7</td>
<td>83.7–85.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neither parent in paid employment</td>
<td>13.0</td>
<td>11.7–14.3</td>
<td>15.3</td>
<td>14.2–16.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Young maternal age at child’s birth</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Over 23 years of age</td>
<td>96.0</td>
<td>95.3–96.8</td>
<td>79.1</td>
<td>77.8–80.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Equal to or less than 23 years</td>
<td>4.0</td>
<td>3.2–4.7</td>
<td>20.9</td>
<td>19.5–22.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Lone parent</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Partnered</td>
<td>84.9</td>
<td>83.7–86.2</td>
<td>81.5</td>
<td>80.5–82.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lone parent</td>
<td>15.1</td>
<td>13.8–16.3</td>
<td>18.5</td>
<td>17.4–19.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Language background</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>84.6</td>
<td>82.5–86.7</td>
<td>97.2</td>
<td>96.6–97.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-English</td>
<td>15.4</td>
<td>13.3–17.5</td>
<td>2.8</td>
<td>2.1–3.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Table 4. Proportion (with 95% CI) of children with high MHC from 4-5 to 14-15 years according to country and gender. Estimates corresponding to Figure 1.

<table>
<thead>
<tr>
<th>Age</th>
<th>Male Australia</th>
<th>Females Australia</th>
<th>All Australia</th>
<th>Male UK</th>
<th>Females UK</th>
<th>All UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% 95% CI</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
<td>% 95% CI</td>
</tr>
<tr>
<td>10-11 yrs</td>
<td>15.45 13.78</td>
<td>17.12 30.23</td>
<td>17.12 31.25</td>
<td>22.66 21.28</td>
<td>24.03 22.60</td>
<td>23.92 23.08</td>
</tr>
</tbody>
</table>

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).

Supplementary Table 5. Proportion of children with high mental health competence from 4-5 to 14-15 years according to country and indicators of social disadvantage. Estimates corresponding to Figure 2.

<table>
<thead>
<tr>
<th>Group comparison</th>
<th>4-5 years</th>
<th>6-7 years</th>
<th>10-11 years</th>
<th>14-15 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff</td>
<td>95% CI</td>
<td>% diff</td>
<td>95% CI</td>
</tr>
<tr>
<td>Lower (versus higher) education</td>
<td>-5.7</td>
<td>-8.21</td>
<td>-3.19</td>
<td>-4.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed (vs employed)</td>
<td>-2.87</td>
<td>-5.86</td>
<td>0.12</td>
<td>-2.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger (vs older) mother</td>
<td>-7.08</td>
<td>-11.04</td>
<td>-3.13</td>
<td>-4.24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group comparison</th>
<th>4-5 years</th>
<th>6-7 years</th>
<th>10-11 years</th>
<th>14-15 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% diff</td>
<td>95% CI</td>
<td>% diff</td>
<td>95% CI</td>
</tr>
<tr>
<td>Lower (versus higher) education</td>
<td>-7.93</td>
<td>-10.05</td>
<td>-5.81</td>
<td>-7.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unemployed (vs employed)</td>
<td>-5.88</td>
<td>-8.26</td>
<td>-3.5</td>
<td>-6.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger (vs older) mother</td>
<td>-5.11</td>
<td>-7.24</td>
<td>-2.98</td>
<td>-6.52</td>
</tr>
</tbody>
</table>
Supplementary Table 6. Proportion of children in each trajectory group according to country and indicators of social disadvantage. Estimates corresponding to Figure 4.

<table>
<thead>
<tr>
<th>Trajectory group</th>
<th>Australia</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Female (vs male)</td>
<td>Lower (vs higher) education</td>
</tr>
<tr>
<td></td>
<td>% diff</td>
<td>95% CI</td>
</tr>
<tr>
<td>Low baseline, decreasing</td>
<td>-5.78</td>
<td>-7.00</td>
</tr>
<tr>
<td>Mid baseline, decreasing</td>
<td>-2.47</td>
<td>-3.66</td>
</tr>
<tr>
<td>High baseline, decreasing</td>
<td>1.68</td>
<td>0.29</td>
</tr>
<tr>
<td>Low baseline, stable</td>
<td>-5.22</td>
<td>-6.84</td>
</tr>
<tr>
<td>Mid baseline, stable</td>
<td>0.02</td>
<td>-1.28</td>
</tr>
<tr>
<td>Low baseline, increasing</td>
<td>-3.78</td>
<td>-5.32</td>
</tr>
<tr>
<td>Mid baseline, increasing</td>
<td>4.43</td>
<td>3.13</td>
</tr>
<tr>
<td>High baseline, increasing</td>
<td>3.85</td>
<td>3.00</td>
</tr>
</tbody>
</table>

BMJ Publishing Group Limited (BMJ) disclaims all liability and responsibility arising from any reliance placed on this supplemental material which has been supplied by the author(s).
Supplementary Figure 1. Mean (95% CI) continuous MHC score from ages 4 to 15 years, according to country and gender.
Approach to accounting for survey design in each cohort

Proportions of children with high MHC from 4-15 years of age
To account for stratification in the survey design, clustering within the primary sampling unit, and survey weights, the svyset stata command was used. For example, this was implemented as mi svyset [pweight=zweight] zpsun, strata(zstrn) singleunit(scaled), where the primary sampling unit (zpsun) refers to the original variables SPTN00 and pcodes in the MCS and LSAC respectively; strata (zstrn) to PTTYPE2 and stratum; and the wave 1 weighting variable (zweight) to AOVWT2 and cweight. We note that the Wave 1 weighting variable was applicable for estimates at all waves included as subsequent sample attrition was accounted for using multiple imputation. Estimates were made on a wave specific basis thus no additional adjustment for repeated measurements was required.[1]

Approach to deriving trajectory groups to explore within individual change over time in MHC from 4-15 years

Additional notes on method
Cut points at the 33rd and 66th percentile of the random slopes do not impose a direction or magnitude on the slope. For example, all tertiles may comprise of increasing slopes. However, in these data, the cut points can be interpreted as stated.

Assignment of each individual to one of the nine pre-specified trajectory groups, and estimates of the distribution of the nine trajectory groups across gender and the indicators of social disadvantage, were performed for each imputation, with results combined using Rubin’s rules.

To estimate the mean intercept and change over time, and corresponding confidence interval, in each of the nine trajectory groups, we fit a random effects model between age and MHC score. Change over time was assumed to follow a linear relationship with age.

Stata code for trajectory models
Code is provided below to illustrate the trajectory modelling approach and inform approaches of other researchers interested; it should not be applied without careful consideration of and validation within different data and contexts. For brevity, data cleaning and imputation steps are not presented.

*****************************************************************************
* data dictionary
*****************************************************************************
* ahedpn_hi:    Education level present in the household
* ahempd:    Neither parent in paid employment
* alonpn:    Lone parent
* zagepn:   Mothers age at child's birth
* zsexpn:   Child’s sex
* zcounn:   Country
* mhcpd:   Mental health competency score
* agep:    Wave 1 population weight
* ID:    Unique identifier for each child
* wave:    Wave identifier
*****************************************************************************
* DO FILE 1 - classify trajectories
*****************************************************************************
* Author:  Sarah Arnup
* Purpose: This do file classifies each imputed trajectory into one of nine groups based on intercept and slope tertile,
as obtained from a mixed model, with random effect
for intercept and slope.

Date: 27 Nov 2019
Edited: 22 Jun 2020

version 16.0
clear all

*** open data
*** data in long format

* manage data

* change mi format to flong for mixed modelling
mi convert flong, clear

* trajectory modelling

* set up variables needed to extract trajectory groups
forval y=1/9 {
    gen b_slope`y' = .
gen se`y' = .
gen intercept`y' = .
}

foreach var of varlist ahedpn_hi ahempd alonpn zagepn zsexpn zcounn {
    gen mean_`var'`y' = .
gen var_`var'`y' = .
}

gen baseline = .
gen slope = .
gen groups = .

* obtain number of imputations
mi query
local M = r(M)

* iterate through imputations
forvalues i = 1/`M' {
    * for each imputation, obtain random effect and slope for each individual
    * include pweights at the level of the individual, not observation
    mi xeq `i': meglm mhcpd agep_childage || ID: wave, covariance(unstructured)
pweight(zweight)

    * obtain estimates of random effects
    predict b* if _mi_m==`i', reffects

    * obtain 33rd and 66th centile of random intercept
centile bl if _mi_m==`i', centile(33)
local p33 = r(c_1)
centile bl if _mi_m==`i', centile(66)
local p66 = r(c_1)

    * create baseline categories for intercept based on tertile
gen bl = 1 if bl < `p33' & _mi_m==`i'
replace bl = 2 if bl < `p66' & bl==. & _mi_m==`i'
replace bl = 3 if bl < . & bl==. & _mi_m==`i'
replace baseline = bl if _mi_m==`i'

    * obtain 33rd and 66th centile of random slope
centile b2 if _mi_m==`i', centile(33)
local p33 = r(c_1)
centile b2 if _mi_m==`i', centile(66)
local p66 = r(c_1)

    * create slope categories for slope based on tertile
gen s = 1 if b2 < `p33' & _mi_m==`i'
replace s = 2 if b2 < `p66' & s==. & _mi_m==`i'
replace s = 3 if b2 < . & s==. & _mi_m==`i'
replace slope = s if _mi_m==`i'

    * create 9 groupings of baseline and slope tertiles
gen g = group(baseline slope) if _mi_m==`i'
replace groups = g if _mi_m==`i'
* drop variables needed for next iteration
  drop b? b1 s g
}
compress
save "imputed trajectories_re_groups", replace
exit
********************************************************************************

* DO FILE 2 - obtain estimate of slope and intercept for each trajectory group
********************************************************************************

********************************************************************************

* Author: Sarah Arnup
* Purpose: Obtain estimate of slope and intercept for each group, with
  variance obtained using Rubin's rules.
* Date: 27 Nov 2019
* Edited: 22 June 2020
********************************************************************************

* Notes:
  * Rubin's Rules: $V_T = V_W + \frac{1}{m} \sum (V_i)$
  * $V_W = \frac{1}{m} \sum (V_i - \bar{Y}^2)$
  * $d_f = (m-1) \left( \frac{1}{V_W} + \frac{1}{(1+1/m) V_B} \right)^2$
  * Student's T-distribution with $v$ degrees of freedom.
  * CI $\bar{Y} = t_d f, (1-\alpha/2) \sqrt{V_T}$
  * increase in variance due to missing data
    $\sqrt{V_T/V_W} - 1$
********************************************************************************

version 16.0
clear all
* open data
use "imputed trajectories_re_groups"
* we wish to obtain an estimate of the slope in each group
* fit linear regression to each of nine groups
********************************************************************************

* obtain number of imputations
mi query
local M = r(M)
* iterate through imputations
forvalues i = 1/`M' {
  * iterate through each of the 9 groups
  forval y=1/9 {
    * fit a linear regression, accounting for repeated measurements within
      individual
    regress mhcpd agep_childage if groups == `y' & _mi_m==`i', vce(cluster ID) noheader
    * extract intercept, beta and se for each regression
    replace b_slope`y' = _b[agep_childage] if _mi_m==`i' & groups == `y'
    replace se`y' = _se[agep_childage] if _mi_m==`i' & groups == `y'
    replace intercept`y' = _b[_cons] if _mi_m==`i' & groups == `y'
  }
  * Obtain variances using Rubins Rules for estimate of slope in each group
  bysort ID wave: will sum each observation over the 50 runs
  * run number is given by _mi_m
  * drop observed data
  drop if _mi_m == 0
  * create indicators
...gen one = 1
bysort ID wave: egen check = total(one) // this should total the number of imputations, i.e. 50

* create mean of mhc over all imputations
bysort ID wave: egen y_imputed = mean(mhcpd) // needed for figure, isn’t used further

* for each imputation (_mi_m) the regression parameters will be constant within each group
* we want to create a complete dataset with regression parameters for each group
forval y=1/9 {
bysort _mi_m (b_slope`y''): replace b_slope`y' = b_slope`y'[1]
bysort _mi_m (intercept`y''): replace intercept`y' = intercept`y'[1]
bysort _mi_m (se`y''): replace se`y' = se`y'[1]
}

* create the within and between and total variance, and d.f.
forval y=1/9 {
* create the variance of the estimate of slope
gen var`y' = (se`y'[1])^2
* this is V_W: within imputation variance for each observation
bysort ID wave: egen V_W`y'= mean(var`y'})

* create mean regression parameters for each observation for each group
bysort ID wave: egen mean_slope`y' = mean(b_slope`y')
bysort ID wave: egen mean_intercept`y' = mean(intercept`y') // this isn’t used

* create difference between mean and observation for slope for each observation
gen between`y' = (b_slope`y' - mean_slope`y'[1])^2

* calculate V_B: between imputation variance
bysort ID wave: egen sum_between`y' = total(between`y')
gen V_B`y' = sum_between`y' / (check-1)

* calculate total variance and dfs
gen df`y' = (check-1) * (1 + V_W`y' / (1+(1/check)))* V_B`y'
compress
save "imputed trajectories_trajectories", replace
exit

******************************************************************************
* DO FILE 3 - Obtain estimate of the probability of being in each group,
* given socio-demographic factor
******************************************************************************

******************************************************************************
* Author: Sarah Arnup
* Purpose: Obtain estimate of the probability of being in each group,
* given socio-demographic factor, with variance obtained using
* Rubin’s rules as per do-file 2.
* Date: 27 Nov 2019
* Edited: 22 June 2020
******************************************************************************

version 16.0
clear all

t * open data
use "imputed trajectories_re_groups"

t * tidy up data

* drop observed data
drop if _mi_m ==0

* drop unneeded variables
drop mhcpd mhcpd_hi
drop *slope* se* intercept* baseline slope agep_childage
compress

* create some indicator variables
gen one = 1
bysort ID wave: egen check = total(one)

* estimate probabilities for each group using multinomial logistic regression
* and margins
********************************************************************************

* obtain number of imputations
mi query
local M = r(M)

* estimate proportion
forvalues i = 1/50 {
    * repeat for each imputation run
    preserve
    keep if _mi_m==`i'
    * repeat for each socio-demographic factor
    foreach var of varlist ahedpn_hi ahempd alonpn zagepn zsexpn {
        * run multinomial logistic regression
        mlogit groups i.`var'
        * run margins to estimate effect of factor on probability
        margins, dydx(`var')
        * extract table of estimates
        mat table1 = r(table)
        * run margins to obtain probabilities for each level of factor
        in each group
        margins `var', post
        * extract table of estimates
        mat table2 = r(table)
        * extract estimates
        forval y = 1/9 {
            * extract effect estimate and variance
            local a = 9 + `y'
            replace mean_`var'`y' = table1[1,`a']
            replace var_`var'`y' = table1[2,`a']
            replace var_`var'`y' = (var_`var'`y')^2
            * extract mean and variance of probability for each
            group
            local a0 = (2*`y') - 1
            local a1 = 2*`y'
            gen mean_`var'0_`y' = table2[1,`a0']
            gen mean_`var'1_`y' = table2[1,`a1']
            gen var_`var'0_`y' = table2[2,`a0']
            gen var_`var'1_`y' = table2[2,`a1']
            replace var_`var'0_`y' = (var_`var'0_`y')^2
            replace var_`var'1_`y' = (var_`var'1_`y')^2
        }
    }
    compress
    * save separate data file for each imputation
    save "imputed_trajectories_proportions_`i'"., replace
    restore
}

clear all

* append all imputation data files together
********************************************************************************
use "imputed_trajectories_proportions_1"
forval i=2/50 {
    append using "imputed_trajectories_proportions_`i'"
}
save "imputed_trajectories_proportions_1_50", replace
* calculate estimate and variance for each socio-demographic factor
* (follows the same methodology as do-file 2)
*** REPEAT for each of the indicators: ahedpn_hi ahempd alonpn zagepn zsexpn zcounn
* open appended data
use "imputed_trajectories_proportions_1_50", clear all
* unset data
mi unset, asis
* calculate estimate and variance as per do-file 2
foreach var of varlist ahedpn_hi {
    forval y=1/9 {
        no di `y'
        bysort _mi_m (mean_`var'`y'): replace mean_`var'`y' = mean_`var'`y'[1]
        bysort _mi_m (var_`var'`y'): replace var_`var'`y' = var_`var'`y'[1]
        bysort ID wave: egen mean_`var'`y' = mean(mean_`var'`y')
        bysort ID wave: egen mean_`var'`y' = mean(var_`var'`y')
        gen between_`var'`y' = (mean_`var'`y' - mean_mean_`var'`y')
        bysort ID wave: egen sum_between_`var'`y' = total(between_`var'`y')
        gen V_B`var'`y' = sum_between_`var'`y' / (check -1)
        gen V_T`var'`y' = mean_var_`var'`y' + (1+(1/check))*V_B`var'`y'
        foreach x=0/1 {
            bysort _mi_m (mean_`var'`x'_`y'): replace mean_`var'`x'_`y' = mean_`var'`x'_`y'[1]
            bysort _mi_m (var_`var'`x'_`y'): replace var_`var'`x'_`y' = var_`var'`x'_`y'[1]
            bysort ID wave: egen mean_`var'`x'_`y' = mean(mean_`var'`x'_`y')
            bysort ID wave: egen mean_`var'`x'_`y' = mean(var_`var'`x'_`y')
            gen between_`var'`x'_`y' = (mean_`var'`x'_`y' - mean_mean_`var'`x'_`y')
            bysort ID wave: egen sum_between_`var'`x'_`y' = total(between_`var'`x'_`y')
            gen V_B`var'`x'_`y' = sum_between_`var'`x'_`y' / (check -1)
            gen V_T`var'`x'_`y' = mean_var_`var'`x'_`y' + (1+(1/check))*V_B`var'`x'_`y'
        }
    }
}
* tidy up data
* calculate confidence intervals
rename mean_ahedpn_hi mean
rename V_Tahedpn_hi V_T
rename dfahedpn_hi df
rename mean_ahegpn_hi0 mean0
rename V_Tahedpn_hi0 V_T0
rename dfahedpn_hi0 df0
rename mean_ahegpn_hi1 mean1
rename V_Tahedpn_hi1 V_T1
rename dfahedpn_hi1 df1
gen ci = invt(df,0.975) * sqrt(V_T)
gen ci = invt(df0,0.975) * sqrt(V_T0)
gen ci = invt(df1,0.975) * sqrt(V_T1)
gen ci = mean - ci
* label trajectories
rename wave trajectory
label define traj 1 "Low baseline, decreasing" ///
  2 "Mid baseline, decreasing" ///
  3 "High baseline, decreasing" ///
  4 "Low baseline, stable" ///
  5 "Mid baseline, stable" ///
  6 "High baseline, stable" ///
  7 "Low baseline, increasing" ///
  8 "Mid baseline, increasing" ///
  9 "High baseline, increasing"
label values trajectory traj

order name trajectory mean ci_l u mean0 ci_l0 u0 mean1 ci_l1 u1
save "MHCproportions_shedpn_hi", replace
exit

********************************************************************************
* DO FILE 4 - construct trajectories for each group based on intercept and slope
********************************************************************************

* formula for line
  y1 = mean_b_slope1 * agep_childage + mean_intercept1
  95% CI t_0 = 1-alpha/2 * sqrt(V_T1)
  95% CI invt(df1,0.975) * sqrt(V_T1)

version 16.0
clear all
* open data
use "imputed_trajectories"
* keep first imputation as example 'raw' trajectories
keep if _mi_m==1
* create some dummy IDs for mean trajectory per group
count
local temp = _N + 2
set obs `temp'
replace ID = 1 if _n==_N-1
replace ID = 2 if _n==_N
* replicate one for each of 9 groups
gen replicate = 1 if ID ==1|ID==2
expand 9 if replicate==1
gsort -ID
count
replace groups = (_N - _n) + 1 in -9/-1
replace groups = (_N - _n) - 8 in -18/-10
* create variables needed for each line
forval y=1/9 {
gen y`y'=.  
gen ci_y`y' = .  
gen ci_l`y' = .  
gen ci_u`y' = .
}
* set age at 3 for first point
replace agep_childage = 4*12 if ID==1
forval y=1/9 {
  replace y`y' = mean_slope`y'[1] * agep_childage + mean_intercept`y'[1] if ID==1 &
  groups==`y'
}
* set age at 16 for second point
replace agep_childage = 16*12 if ID==2
forval y=1/9 {
  replace y`y' = mean_slope`y'[1] * agep_childage + mean_intercept`y'[1] if ID==2 &
  groups==`y'
}
* calculate CI (very narrow)
forval y=1/9 {
replace ci`y' = invt(df`y'[1], 0.975) * sqrt(V_T`y'[1]) if (ID==1 | ID==2) & groups==`y'
replace ci_l`y' = y`y' - ci`y' if (ID==1 | ID==2) & groups==`y'
replace ci_u`y' = y`y' + ci`y' if (ID==1 | ID==2) & groups==`y'
}

* y_imputed contains mean trajectory per groupings

exit
Supplemental references