**Adrenarche\_Gonadarche.R\_README**

Robert Toms

Sections:

1. Purpose of Code
2. Researcher-Facing Explanation of Code
3. Explanation of Variables
4. Methods Section Summary
5. References

PURPOSE OF CODE:

The purpose of this code is to determine age at Adrenarche, Gonadarche and general pubertal onset in females.

RESEARCHER-FACING EXPLANATION OF CODE:

1. Pubertal Development data from ABCD 6.0 were imported (ph\_y\_pds.tsv), as were biological sex data from the ABCD static variables spreadsheet (ab\_g\_stc.tsv).
2. Clean and Prepare Data
   1. Irrelevant columns were removed, remaining columns were renamed for readability, and sex data were added to pubertal data.
   2. “I don’t know” (999) and “Refuse to Answer” (777) responses to Pubertal Development Scale (PDS) items (prefix: ph\_y\_pds\_, items: 001, 002, 003, f\_001, f\_002, m\_001, m\_002) were set to NA, in line with the best practices of Herting et al, 2021.
   3. Female-only data were split off.
3. Manually Calculate PDS Values
   1. Background: General PDS (Pubertal Development Score) values for women are computed by averaging 5 pubertal datapoints with ranges of 1-4 (ph\_y\_pds\_001, ph\_y\_pds\_002, ph\_y\_pds\_003, ph\_y\_pds\_\_f\_001, ph\_y\_pds\_\_f\_002—this last one refers to menarche status, where Premenarche = 0 and PostMenarche = 1, these are manually reset to 1 and 4, respectively). The ABCD 6.0 dataset contains pre-calculated PDS values, with a minimum requirement of 4 of the 5 datapoints for an average to be calculated for a given timepoint. However, previous ABCD datasets did not have precalculated PDS values, thus the extant best practices involve manually calculating them, with a minimum requirement of 2 of 5 datapoints increasing across time for inclusion into analysis (Larson, Chaku, & Moussa-Tooks, 2025). Because of this difference, both will be utilized, and the results will be compared.
   2. The menarche variable (Premenarche = 0 and PostMenarche = 1), were manually reset to (Premenarche = 1 and PostMenarche = 4), so as not to skew the PDS average calculations, because the rest of the variables range from 1-4.
   3. Calculate PDS Values:
      1. PDS values were calculated by adding the values across all 5 variables, and dividing by the number of variables with non-NA values. This is saved to “manual\_pds”.
      2. Adrenarche PDS is calculated by averaging the PDS values for progress in body hair growth (body\_hair) and skin changes (skin\_pimples)
      3. Gonadarche PDS is calculated by averaging the PDS values for thelarche and menarche
4. Calculate Slopes of each Pubertal Category
   1. Best practices from (Larson, Chaku, & Moussa-Tooks, 2025) require a minimum of 2 of 5 pubertal categories increasing in score across timepoints.
   2. A slope-finding function is defined called “find\_slope”.
      1. Takes in a dataframe and the column in it that will act as the “y” value
      2. Keeps only the x and y values
      3. Gets rid of NA values, and returns NA in the column if there isn’t enough information to calculate the slope (< 2 timepoints)
      4. Calculates a linear model
      5. Pulls the slope value from the linear model
   3. “find\_slope” function is run on the 5 Pubertal categories, and the number of positive (increasing) slopes is counted for each pubertal onset group (general, adrenarche, and gonadarche) and saved to “num\_increasing”, “num\_adrenarche\_increasing”, and “num\_gonadarche\_increasing”, respectively.
   4. These “\*\_increasing” values are then counted to tally exclusions pertaining the best-practices version of the model (at least 2 items increasing).
5. Calculate Age at Adrenarche, Gonadarche, & General Pubertal Onset
   1. Explanation:
      1. Age of onset in each category is considered to be when PDS = 2.5.
      2. According to (Larson, Chaku, & Moussa-Tooks, 2025) and (Beltz et al., 2014), Linear mixed-effects modelling can be used to determine age at pubertal onset, using the following equation:
         1. PDS = b0i + ( b1i \* ageit ) + rit
            1. PDS = PDS value
            2. b0i = intercept for an individual *i*
            3. b1i = slope for an individual *i*
            4. ageit = age for an individual *i* at a timepoint *t*
            5. rit = residual for an individual *i* at a timepoint *t*
      3. Because we are computing onsets individually, residuals will be zero. Thus, to find age, we isolate age and substitute 2.5 for PDS, leaving us with this equation:
         1. Age = (2.5 – intercept) / slope
   2. An intercept-finding function is defined called “find\_intercept”.
      1. Takes in a dataframe and the column in it that will act as the “y” value
      2. Keeps only the x and y values
      3. Gets rid of NA values, and returns NA in the column if there isn’t enough information to calculate the slope (< 2 timepoints)
      4. Calculates a linear model
      5. Pulls the intercept value from the linear model
   3. Slopes and Intercepts are calculated using “find\_slope” and “find\_intercept”.
   4. The equation [ Age = (2.5 – intercept) / slope ] is used to determine age at pubertal onset in each of the onset types (adrenarche, gonadarche, general).
6. Clean and Organize for Export
   1. Onset ages, PDS items, and PDS item counts are isolated and combined into a single data frame “pubertal\_timing”.

#### FINAL NOTES

1. Some ages are exorbitantly incorrect – adrenarche at age 2760741055120846.5 for example. This is due to near horizontal slope values, when computing ages. Cases like these, and negative/outlier ages will need to be systematically excluded.

Explanation of Variables:

|  |  |
| --- | --- |
| participant\_id | ABCD subject identifier |
| session\_id | ABCD visit identifier |
| ab\_g\_stc\_\_cohort\_sex | Biological Sex indicator, 1 = Male, 2 = Female |
| sex | ab\_g\_stc\_\_cohort\_sex, renamed, set “M” = Male, “F” = Female |
| ph\_y\_pds\_dtt | Interview datetime |
| ph\_y\_pds\_age | Age at interview (in years) |
| ph\_y\_pds\_001 | Growth in Height: Has not yet begun to spurt (spurt means grow faster than usual) = 1; Has barely started = 2; Is definitely underway (has definitely started) = 3; Seems complete = 4; I don't know = 999; Decline to answer = 777 |
| ph\_y\_pds\_002 | Body Hair Growth: Has not yet begun to grow = 1; Has barely started = 2; Is definitely underway (has definitely started) = 3; Seems complete = 4; I don't know = 999; Decline to answer = 777 |
| ph\_y\_pds\_003 | Skin changes/Pimples: skin has not yet started changing = 1; skin has barely started changing = 2; skin changes are definitely underway (has definitely started) = 3; skin changes seem complete = 4; I don't know = 999; Decline to answer = 777 |
| ph\_y\_pds\_\_f\_001 | Breasts begun to grow? : have not yet started growing = 1; have barely started growing = 2; breast growth is definitely underway (has definitely started) = 3; breast growth seems complete = 4; I don't know = 999; Decline to answer = 777 |
| ph\_y\_pds\_\_f\_002 | Begun to Menstruate? : Yes = 1; No = 0; I don't know = 999; Refuse to answer = 777 |
| ph\_y\_pds\_\_f\_mean | PDS value : mean of ph\_y\_pds 001, 002, 003 and f\_001 & f\_002, when f\_002 reset to Yes = 4, No = 1 |
| interview\_datetime | ph\_y\_pds\_dtt, renamed |
| interview\_age\_yrs | ph\_y\_pds\_age, renamed |
| growth\_spurt | ph\_y\_pds\_001, renamed |
| body\_hair | ph\_y\_pds\_002, renamed |
| skin\_pimples | ph\_y\_pds\_003, renamed |
| Thelarche | ph\_y\_pds\_\_f\_001, renamed |
| menarche | ph\_y\_pds\_\_f\_002, renamed, reset to Yes = 4, No = 1 |
| ABCD\_PDS | ph\_y\_pds\_\_f\_mean, renamed |
| manual\_pds | Mean of available PDS items: growth\_spurt, body\_hair, skin\_pimples, thelarche, menarche |
| num\_items | Number of PDS items (growth\_spurt, body\_hair, skin\_pimples, thelarche, menarche) with non-NA data |
| adrenarche\_pds | Mean of adrenal axis-related PDS items (body\_hair, skin\_pimples) |
| gonadarche\_pds | Mean of gonadal axis-related PDS items (thelarche, menarche) |
| find\_slope | Function. Finds the slope of a variable’s progression as age increases |
| growth\_spurt\_slope | Slope of a subject’s growth\_spurt PDS values vs. age in years |
| body\_hair\_slope | Slope of a subject’s body\_hair PDS values vs. age in years |
| skin\_pimples\_slope | Slope of a subject’s skin\_pimples PDS values vs. age in years |
| thelarche\_slope | Slope of a subject’s thelarche PDS values vs. age in years |
| menarche\_slope | Slope of a subject’s menarche PDS values vs. age in years |
| num\_general\_increasing | Number of slopes increasing (out of 5) from: growth\_spurt\_slope, body\_hair\_slope, skin\_pimples\_slope, thelarche\_slope, menarche\_slope |
| num\_adrenarche\_increasing | Number of slopes increasing (out of 2) from: body\_hair\_slope, skin\_pimples\_slope |
| num\_gonadarche\_increasing | Number of slopes increasing (out of 2) from: thelarche\_slope, menarche\_slope |
| find\_intercept | Function. Finds the intercept of a variable’s progression as age increases |
| ABCD\_slope | Slope of ABCD\_PDS values vs. age in years |
| ABCD\_intercept | Intercept of ABCD\_PDS values vs. age in years |
| manual\_slope | Slope of manual\_pds values vs. age in years |
| manual\_intercept | Intercept of manual\_pds values vs. age in years |
| adrenarche\_slope | Slope of adrenarche\_pds values vs. age in years |
| adrenarche\_intercept | Intercept of adrenarche\_pds values vs. age in years |
| gonadarche\_slope | Slope of gonadarche\_pds values vs. age in years |
| gonadarche\_intercept | Intercept of gonadarche\_pds values vs. age in years |
| ABCD\_pubertal\_age | Age at General pubertal onset, using precalculated ABCD PDS values |
| manual\_pubertal\_age | Age at General pubertal onset, using manually calculated PDS values *a la* Larson, Chaku, & Moussa-Tooks, 2025 |
| adrenarche\_age | Age of Adrenarche |
| gonadarche\_age | Age of Gonadarche |
| adrenarche\_tempo | Increase in Adrenarche PDS score per year |
| gonadarche\_tempo | Increase in Gonadarche PDS score per year |
| manual\_pubertal\_tempo | Increase in manually calculated PDS scores per year |
| ABCD\_pubertal\_tempo | Increase in pre-calculated ABCD PDS scores per year |
| ABCD\_Manual\_Flag\_Y1N0 | Flag, Yes = 1 No = 0, if precalculated ABCD PDS score does not exist, but manually calculated PDS score does. This is due to differences in criteria for calculation (ABCD = 4 minimum, manual = no minimum) |
| Exclude\_Adren\_Y1N0 | Exclusion flag for age at adrenarche, flagged if has less than 2 adrenal-axis values increasing as age increases (out of 2)  Yes = 1, No = 0 |
| Exclude\_Gonad\_Y1N0 | Exclusion flag for age at adrenarche, flagged if has less than 2 gonadal-axis values increasing as age increases (out of 2)  Yes = 1, No = 0 |
| Exclude\_General\_Y1N0 | Exclusion flag for age at adrenarche, flagged if has less than 2 pubertal category values increasing as age increases (out of 5)  Yes = 1, No = 0 |

Methods Section Summary:

**Main:**

To calculate pubertal onsets and tempos—Adrenarche, Gonadarche, and General Pubertal Onset (to contextualize axis-specific effects)—we utilized extant best practices (Larson, Chaku, & Moussa-Tooks, 2025; Beltz et al., 2014) to derive estimates of pubertal timing via linear mixed effects modelling. The model can be represented by this equation:

PDS = b0i + ( b1i \* ageit ) + rit

where PDS = Pubertal Development Score value, b0i = intercept for an individual *i*, b1i = slope for an individual *i*, ageit = age for an individual *i* at a timepoint *t*, and rit = residual for an individual *i* at a timepoint *t* (Beltz et al., 2014).

Pubertal onset is considered to have occurred when PDS = 2.5, and pubertal tempo is conceptualized as the rate of change in PDS score relative to age. Methods for calculating each onset’s PDS values are described in the Supplement.

**Supplement:**

Adrenarche PDS values were calculated using the mean of ph\_y\_pds\_002 (body hair growth) and ph\_y\_pds\_003 (skin changes/pimples) values. Those lacking increases in at least one of these values across time were flagged for exclusion, per (Larson, Chaku, & Moussa-Tooks, 2025).

Gonadarche PDS values were calculated using the mean of ph\_y\_pds\_\_f\_001 (breast development) and ph\_y\_pds\_\_f\_002 (first menses) values. Those lacking increases in at least one of these values across time were flagged for exclusion, per (Larson, Chaku, & Moussa-Tooks, 2025).

General PDS (Pubertal Development Score) values for women are computed by averaging 5 pubertal datapoints with ranges of 1-4 (ph\_y\_pds\_001, ph\_y\_pds\_002, ph\_y\_pds\_003, ph\_y\_pds\_\_f\_001, ph\_y\_pds\_\_f\_002)—this last one refers to menarche status, where PreMenarche = 0 and PostMenarche = 1, these are manually reset to 1 and 4, respectively). The rest of the variables (growth spurt in height, new body hair, skin changes, breast development, respectively) are continuous Likert-style ratings ranging from 1 to 4 (1 = has not begun, 2 = has barely begun, 3 = is definitely underway, 4 = seems complete). The ABCD 6.0 dataset contains pre-calculated PDS values, with a minimum requirement of 4 of the 5 datapoints for an average to be calculated for a given timepoint. However, previous ABCD datasets did not have precalculated PDS values, thus the extant best practices involve manually calculating them, with a minimum requirement of 2 of 5 datapoints increasing across time for inclusion into analysis (Larson, Chaku, & Moussa-Tooks, 2025). Because of this difference, both will be utilized, and the results will be compared.

According to (Larson, Chaku, & Moussa-Tooks, 2025) and (Beltz et al., 2014), Linear mixed-effects modelling can be used to determine age at pubertal onset, using the following equation:

PDS = b0i + ( b1i \* ageit ) + rit

where PDS = Pubertal Development Score value, b0i = intercept for an individual *I*, b1i = slope for an individual *I*, ageit = age for an individual *i* at a timepoint *t*, and rit = residual for an individual *i* at a timepoint *t*. Age of onset in each category is considered to be when PDS = 2.5, and because we are computing onsets individually, residuals will be zero. Thus, to find age, we isolate age and substitute 2.5 for PDS, leaving us with this equation, that can be applied to each subject’s data individually:

Age = (2.5 – intercept) / slope

Intercepts and slopes were calculated using R version 4.2.1 (*stats* package), extracting intercept and slope values from linear models regressing each subject’s onset-type-specific PDS scores against age. The extracted slope values are equivalent to our conceptualizations of pubertal tempo, and were used as such.

References:

Beltz, A. M., Corley, R. P., Bricker, J. B., Wadsworth, S. J., & Berenbaum, S. A. (2014). Modeling pubertal timing and tempo and examining links to behavior problems. Developmental psychology, 50(12), 2715–2726. <https://doi.org/10.1037/a0038096>

Herting, M. M., Uban, K. A., Gonzalez, M. R., Baker, F. C., Kan, E. C., Thompson, W. K., Granger, D. A., Albaugh, M. D., Anokhin, A. P., Bagot, K. S., Banich, M. T., Barch, D. M., Baskin-Sommers, A., Breslin, F. J., Casey, B. J., Chaarani, B., Chang, L., Clark, D. B., Cloak, C. C., Constable, R. T., … Sowell, E. R. (2021). Correspondence Between Perceived Pubertal Development and Hormone Levels in 9-10 Year-Olds From the Adolescent Brain Cognitive Development Study. Frontiers in endocrinology, 11, 549928. https://doi.org/10.3389/fendo.2020.549928

Larson, E., Chaku, N., & Moussa-Tooks, A. (2025, August 3). Early Pubertal Development is a Risk Factor for Psychotic-Like Experiences in Boys and Girls. https://doi.org/10.31234/osf.io/249pt\_v1