Father’s Analysis - Data Codebook

This codebook follows the structure of the R file 02\_wrangle-data.R in that variables are described in the same order as the corresponding subsections of the R file. However, variables coded but not used in the analysis are not described here. In the last section we discuss missing data imputation.

## Outcome variable - Intimate Partner Violence

We are primarily interested in reports of Intimate Partner Violence at Year 1.

**Question D6**: *Now, think about how (FATHER) behaves towards you. For each statement I read, please tell me how often he behaves this way.*

Answer options: 1 - OFTEN; 2 - SOMETIMES; 3 - NEVER

Statements:

1. He is fair and willing to compromise when you have a disagreement
2. He expresses affections or love for you
3. He insults or criticizes you or your ideas
4. He encourages or helps you to do things that are important to you
5. He tries to keep you from seeing or talking with your friends or family
6. He tries to prevent you from going to work or school
7. He withholds money, makes you ask for money, or takes your money
8. He slaps or kicks you
9. He hits you with his fist or an object that could hurt you
10. He tries to make you have sex or do sexual things you don’t want to do
11. He listens to you when you need someone to talk to
12. He really understands your hurts and joys

**Question D8** are the same as D6, but with all items in the past tense. It is for parents who are no longer together at follow-up, and **specifically asks about the last month of the relationship**.

To construct our IPV outcome variable, we score the items as such:

* Positive items (1, 2, 4, 11, and 12) are coded as Often = 0, Sometimes = 1, and Never = 2.
* Negative items are coded as Often = 2, Sometimes = 1, and Never = 0.

All scores are summed and the total score is divided by 24 to create an IPV index from 0 to 1. For analyses with binary IPV, an index of 0 was taken as 0 and an index greater than 0 was taken as 1. If there are any missing values for the individual items, the total score counts as missing. For the imputed dataset, any missing values are first imputed and only them summarized into an IPV index.

If the woman was still with the father at year 1, the variables ipv\_prop and ipv\_binary capture IPV from the father at that point in time; if not, we take IPV from the father in the last month of their relationship (question D8). Women who did not report a relationship that would be amenable to IPV score were excluded from the dataset (see sample delineation document).

Three subscales were constructed from these items: - Emotional: items 1,2,3,4,11,12 - Physical: items 8,9,10 - Controlling: items 5,6,7 And the respective indices were calculated by dividing by 12, 6, and 6, respectively.

## Predictors / Covariates

[All measured at baseline, unless otherwise indicated]

### Race

**Question H3**:*Which of these categories best describes your race?*

Answer options: 1 — White, 2 — Black, African-American, 3 — Asian or Pacific Islander, 4 — American Indian, Eskimo, Aleut, 5 — Other, not specified, -2 — DON’T KNOW

We create the categorical variable f\_race with levels “White” and “Black”, after filtering out those who didn’t identify with either. “White” is the reference category.

### Employment

**Question J1**: *Last week, did you do any regular work for pay?*

Answer options: 1 — Yes, 2 — No

We create the categorical variable f\_employment with levels “Unemployed” and “Employed”. “Unemployed” is the reference category.

### Age

Age of the father at baseline. From the baseline guide:

*Age is recorded from the Core Baseline survey for mother, father and child and can be retrieved through the constructed variables:* ***cm1age*** *(mother’s age at the interview),* ***cf1age*** *(father’s age at the interview),* ***cm1b\_age*** *and* ***cf1b\_age*** *for the child’s age at the mother and father interview, respectively.* (p. 20)

f\_age is always a positive integer, and it is treated as a continuous variable.

### Education

Father’s level of education at baseline.

**Question I1:** *Now I’d like to ask some questions about your education and work experience. What is the highest grade or year of regular school that you have completed?*

Answer options: 1 — No formal schooling, 2 — 8th grade or less, 3 — Some high school (Grades 9, 10, 11, & 12), 4 — High school diploma (Completed 12th grade), 5 — G.E.D., 6 — Some college or 2 year degree, 7 — Technical or trade school, 8 — Bachelor’s degree, 9 — Graduate or professional school

Variable f\_education is a categorical variable, where original answers 1-3 are coded as “Below HS”, and 4-9 as “HS and above”, and all missing values as NA. The reference category is “Below HS”.

### Alcohol use

Father’s alcohol use.

**Question G2:** *In the past three months, about how often did you drink alcoholic beverages?*

Answer options: 1 — NEARLY EVERY DAY, 2 — SEVERAL TIMES A WEEK, 3 — SEVERAL TIMES A MONTH, 4 — LESS THAN ONCE A MONTH, 5 — NEVER

Variable f\_alcohol is a categorical variable with 3 levels: “Never”, “<1 per month”, and “>1 per month” (original values 1-3) with all missing values as NAs and “Never” as the reference level.

### Drug use

Father’s drug use.

**Question G3:** *In the past three months, about how often did you use drugs such as marijuana, crack cocaine, or heroin?*

Answer options: 1 — NEARLY EVERY DAY, 2 — SEVERAL TIMES A WEEK, 3 — SEVERAL TIMES A MONTH, 4 — LESS THAN ONCE A MONTH, 5 — NEVER

Variable f\_drugs is a categorical variable with 3 levels: “Never”, “<1 per month”, and “>1 per month” (original values 1-3) with all missing values as NAs and “Never” as the reference level.

### Depression

Fathers were asked, for 12 items measuring depressive symptoms, how many days of the past seven days they experienced the symptom (scale from 0 to 7, items G9a through G9l). For those who had valid answers to all twelve items, we took the arithmetic mean, which became the variable f\_depression. Otherwise NA.

### Incarceration

Our measure of whether the father had ever experienced incarceration was constructed out of mother and father reports at baseline and year one.

Baseline father survey: 6 items measuring different things (e.g., reasons for not being in a relationship, not living together, or not looking for a job when unemployed) listed “incarceration” as one of the reasons (items B4c, B9a5, B9b5, B10a5, B10b5, J3b). We created temporary variable f\_incarceration\_f\_baseline based on these items - if the father responded with “incarceration” to any of the 6 items, the variable took the value “Experienced incarceration”, and otherwise “No incarceration”. If responses were missing for all six items, the temporary variable took the value NA.

Year 1 father survey: 2 items (H22 and H24) ask whether the father had ever spent time in a juvenile or adult correctional facility, with one answer option being “Yes, jail”. We created temporary variable f\_incarceration\_f\_year\_one based on these items - if the father responded with “Yes, jail” to any of the 2 items, the variable took the value “Experienced incarceration”, and otherwise “No incarceration”. If responses were missing for both items, the temporary variable took the value NA.

Baseline mother survey: 7 items measuring different things (e.g., what was father doing last week, reasons having ended the relationship) listed “incarceration” or “incarcerated” as one of the options (items I6, I11, B4c, B9a5, B9b5, B10a5, B10b5). We created temporary variable f\_incarceration\_m\_baseline based on these items - if the mother responded with “incarceration” or “incarcerated” to any of the 7 items, the variable took the value “Experienced incarceration”, and otherwise “No incarceration”. If responses were missing for all six items, the temporary variable took the value NA.

Year 1 mother survey: 2 items were used. Item A9d if equal to 1 if the mother reports that their marriage ended due to the father’s incarceration, and item C36 asks whether the father has spent any time in jail or prison. We created temporary variable f\_incarceration\_m\_year\_one: if either item indicates that the father has ever been incarcerated, it takes the value “Experienced incarceration”, otherwise “No incarceration”. If responses were missing for both items, the temporary variable took the value NA.

Finally, we took these four temporary variables and summarised them into the variable f\_incarceration with the follow algorithm: if any of the 4 temporary variables had the value “Experienced incarceration”, then our new variable took that value; if all 4 temporary variables had missing values, then our new variable took the value NA; otherwise it took the value “No incarceration”.

### Number of children

Fathers were asked whether they had any other biological children and if so, how many (items A6 and A6a). We created the variable f\_children by pulling information from both items. It is a numeric variable consisting of positive integers (and 0).

### Poverty

f\_poverty is a constructed variable available within the dataset (originally named cf1inpov). From the baseline guide: “cm1inpov and cf1inpov indicate the poverty ratio. The poverty ratio is the ratio of total household income, as defined in c\*1hhinc, to the official poverty thresholds, designated by the U.S. Census Bureau.”

### Home ownership

**Question F2:** *Is the home or apartment where you currently reside…..*

Answer options: 1 — Owned or being bought by someone in your family?, 2 — Rented?

We created f\_home, a categorical variable with levels “Owned” (reference level) and “Rented”.

## Data imputation

Missing data were imputed using the missForest::missForest() command. In order for correct imputation to be possible, IPV variables were turned into factor – this way missForest would only impute possible values. Variables were later turned into numeric for analysis. The IPV index (as well as the subscores for physical, emotional, and controlling IPV) was only calculated after imputation.