How did three papers use the PSID-CDS dataset?

| **Category** | [**Parenting with Patience**](https://www.dropbox.com/scl/fi/ac6c7cvxz2u6e3nexvgd7/Parenting-with-Patience_JPE_forthcoming.pdf?rlkey=hzurn1mo9mg6h37x5gvdugrjs&dl=0) | [**Child Skill Production**](https://www.dropbox.com/scl/fi/r2omkz3pqcgwij3zllcnh/ChildSkillProduction_JPEForthcoming.pdf?rlkey=kab0n807xiiq9m8s08pj1bbme&dl=0) | [**Bruno JMP**](https://www.dropbox.com/scl/fi/p4b483y2fnpidumpiq9zq/Bruno_jmpJan2025_Parenting-Decisions-and-Child-Skill-Development.pdf?rlkey=oq9cjzfik3w0pyth4rmvnkr1l&dl=0) |
| --- | --- | --- | --- |
| 1. **Cognitive Skill -CDS** | Letter Word (LW) | LW, Applied Problems (AP) | LW |
| 1. **Noncognitive Skill -CDS** | Another dataset | – | Behavior Problem Index (BPI-E) |
| 1. **Parental Time Use -CDS** | Mother/father/joint active time | Mother/father active time | Mother/father/joint active time |
| 1. **Child Time Use - CDS** | Self-investment time, school time | – | Study/self time |
| 1. **Goods/Expenditure -CDS** | (Modeled, external CEX data) | Child care & goods expenditures | – |
| 1. **Parenting Style -CDS** | Allowance/CCTs | – | Rules & style classification |
| 1. **Demographic variables – CDS?** |  |  |  |
| 1. **Household Controls -Family data** | Wages, income, education, family size | Wages, income, structure | Wages, income, education |

This note is based on the name of variables in 1997

1. **Cognitive Skill**

| **Source file** | **Variable name** | **Description** | **Used in papers** |
| --- | --- | --- | --- |
|  | Q3AGE  Q3A3 | "AGE OF CHILD 97  "WHAT GRADE IN SCHOOL 97" |  |
| CHILD97\_codebook.pdf | Q3LW\_SS, Q3AP\_SS, Q3PC\_SS, Q3BMA\_SS, Q3BRE\_SS | **Woodcock–Johnson** standardized scores (Letter–Word, Applied Problems, Passage Comprehension, Broad Math, Broad Reading) | *Parenting with Patience*, *Child Skill Production*, *Bruno JMP* |
| CHILD97\_codebook.pdf | Q3LWW, Q3PCPR, Q3PCW, Q3BRPR, Q3BRW, Q3CALPR, Q3CALW, Q3APPR, Q3APW, Q3BMPR, Q3BMW | Letter Word W Score, Passage Comprehension Percentile Rank, Passage Comprehension W Score, Broad Reading Summation Percentile Rank, Broad Reading Summation W Score, Calculation Percentile Rank, Calculation W Score, Applied Problems Percentile Rank, Applied Problems W Score, Broad Math Summation Percentile Rank, Broad Math Summation W Score | It is second-order issue |
| PCG97\_CHLD\_codebook.pdf | Q1PCSS, Q1PCPR, Q1PCW | Parent comprehension summary scores (for adult literacy controls) | Might use as parental skill control? |

**2. Noncognitive Skill**

| Source file | Variable name | Description | Comments |
| --- | --- | --- | --- |
| EMSTEACH97\_codebook.pdf | Q9A6A – Q9A6H | Teacher ratings of behavior: e.g., *sudden mood swings*, *feels no love*, *impulsive*, *restless*, *irritable*, etc. | Can form the Behavior Problems Index (BPI) or *externalizing* measure (used in *Bruno JMP*) |
| EMSTEACH97\_codebook.pdf | Q9A5A–Q9A5H | Teacher ratings of attention, communication, and cooperation | Often used as *noncognitive / socioemotional* composite |
| The data below, HB\_CPROV97\_codebook.pdf and PCG97\_CHLD\_codebook.pdf, have more data  At the moment second-order issue  **No Need to find them for now** | | | |
| HB\_CPROV97\_codebook.pdf | Q11A11A – Q11A11E | Home-based caregiver’s reports of mood, discipline, aggression, etc. | Supplementary source of BPI-type measures |
| PCG97\_CHLD\_codebook.pdf | Q1A2 – Q1A5C,  Q1A6 | Health and socio-emotional attributes | Used in robustness controls |

**3. Parental Time Use**

**4. Child Time Use**

As we discussed, we need to consider all variables and construct the mother/father and children's time using the whole dataset before doing the merge

**5. Goods/Expenditure**

**As pointed out in the JPE paper, only the 2002 and 2007 data have data for expenditure.**

**I mainly worked on 1997 and did not study this. Please check the file for expenditure and list variables yourself**

**Refer to page 27 of the paper** [**Child Skill Production**](https://www.dropbox.com/scl/fi/r2omkz3pqcgwij3zllcnh/ChildSkillProduction_JPEForthcoming.pdf?rlkey=kab0n807xiiq9m8s08pj1bbme&dl=0)

**Child Care Expenditure: In all 3 PCG interviews, respondents for children older than age 5 answer questions about current child care arrangements, costs, and time spent in each arrangement. For children younger than age 5, a retrospective history of arrangements is collected, from which we take all arrangements that are reported as ongoing. We construct a measure of weekly expenditures from these answers. We also create a second measure from total household expenditures on child care from the main interview, divided by the number of children ages 12 or younger. Both measures are strongly correlated and have similar sample averages. Therefore, we use the average of these measures when both are reported, while we use the available measure when only one is reported. Among families with children ages 5–12 who report spending on child care, 60% report that their main arrangement is a before/after-school program, center-based care, or a non-relative family care provider. Only about 7% report non-relative care within the family’s home (e.g., babysitters).**

Goods Expenditures: **In the 2002 and 2007 PCG interviews, respondents report annual expenditures for the child on school supplies and toys. They also answer questions on child participation in private lessons, sports, tutoring, or community groups, and report the costs of these activities. Our measure of market goods expenditures sums over the weekly values of spending on these goods and services.**

**6. Parenting Style**

Parent participation and monitoring, Authoritative and other parenting style proxies

|  | PCG97\_HH\_codebook.pdf | **Variable Code** | **Variable Description** |
| --- | --- | --- | --- |
|  |  | Q2A29A | BEING PARENT IS HARD 97 |
|  |  | Q2A29B | FEEL TRAPPED AS PARENT 97 |
|  |  | Q2A29C | CHILDREN ARE WORK 97 |
|  |  | Q2A29D | FEEL TIRED RAISING FAMILY 97 |
|  |  | Q2A30 | ALL CHILDREN UNDER 3 97 |
|  | Q2A22\_1-4 is used for | Q2A22\_1 | MOST IMPORT QUALITY 97 ( |
|  | Constructing parenting | Q2A22\_2 | 2ND IMPORT QUALITY 97 |
|  | Style | Q2A22\_3 | 3RD IMPORT QUALITY 97 |
|  |  | Q2A22\_4 | 4TH IMPORT QUALITY 97 |
|  |  | Q2A23 | CHILD LEARN TOLLERANCE-IMPORT 97 |
|  |  | Q2A24 | AMOUNT OF RULES 97 |
|  |  | Q2A25 | RULES STRICTLY ENFORCED 97 |
|  |  | Q2A26A | # HRS TV ON / DAY 97 |
|  |  | Q2A31A | LIMITS ON TV 97 |
|  |  | Q2A31B | LIMITS ON KINDS OF TV 97 |
|  |  | Q2A31C | NO TV DURING MEALS 97 |
|  |  | Q2A31D | LIMITS ON BED TIME 97 |
|  |  | Q2A31E | LIMITS ON SWEETS 97 |
|  |  | Q2A31F | TRY TO CONTROL WHO CHILD IS WITH 97 |
|  |  | Q2A31G | CONTROL CHILD DOES AFTER SCHL 97 |
|  |  | Q2A31H | SET HOMEWORK TIME 97 |
|  |  | Q2A31I | DISCUSS RULES WITH CHILD 97 |

| **File** | **Variables** |  |
| --- | --- | --- |
| PCG97\_CHLD\_codebook.pdf | Q1E12–Q1E15 | Parental monetary incentives (allowance and conditional allowance) |
| PDTEACH97\_codebook.pdf | Q10A9A–Q10A9G | Parental involvement index  Teacher report of parental engagement (conferences, open house, volunteering) |

**7. Demographic file in CDS**

Second order importance - so in this stage can be ignored if we don’t know where to get the data

Are there any useful demographic variables in the CDS dataset? We can collect this data in the family dataset. But in case there is a useful one, we should consider

For example, it is useful to know who the primary caregiver is. Because he/she answers the questions related to parenting style

**7. Household Controls**

I am not sure if the three papers that we have use the individual file or the family file. But it seems the family file has data that we want (the names of variables are based on data for 1997)

ER10002 "1997 Family Interview Number"

ER12084 "1997 Longitudinal Family Weight"

ER12224 "1997 Family Cross-Sectional Weight"

ER10008 "# in Family Unit"

ER10012 "# Children in Family Unit"

ER11927 "M24 WTR CHLDRN OUT FU-HD"

ER11928 "M25 # CHLDRN OUT OF US-H"

ER11929 "M26 # US CHLD OUT FU-HD

ER10013 "Age of Youngest Child"

ER12223S "# Born Only to Head in 1996"

ER12223T "# Born Only to Wife in 1996"

ER12223U "# Born to Head and Wife Jointly in 1996"

ER10009 "Age of Head"

ER10010 "Sex of Head"

ER10011 "Age of Wife"

ER12222 "Completed Education - Head"

ER12223 "Completed Education - Wife"

ER12217 "Head Wage Rate - 1996"

ER12218 "Wife Wage Rate - 1996"

ER12079 "Total family income"

ER12069 "Taxable income of head and wife"

ER12071 "Transfer income of head and wife "

ER12073 "Taxable income of other family unit members-OFUMs"

ER12075 "Transfer income of OFUMs "

ER12077 "Social Security income "

ER12080 "LABOR INCOME-HEAD"

ER12082 "LABOR INCOME-wife "

ER12085 "B9-9A MAIN OCCUPATION: 3 DIGIT (HD-E)"

ER12116 "D9-9A MAIN OCCUPATION:3 DIGIT (WF-E)"

ER12214 "WF LABOR INCOME FROM BUSINESS-1996"

ER12171 "Head Weekly Work Hours - 1996"

ER12174 "Head annual Total Work Hours - 1996"

ER12182 "Wife Weekly Work Hours - 1996"

ER12185 "Wife annual Total Work Hours - 1996"

ER10016 "Head Marital Status"

ER12223A "Marital Status - Generated"

ER12223B "Change in Marital Status"

ER12223C "COUPLE STATUS OF HEAD "

ER12221 "1997 PSID State of Residence Code"

ER10007 "Year of Current Interview"

ER11045 "HOUSEWORK HRS-WIFE an average week"

ER11046 "HOUSEWORK HOURS-HEAD an average week"

ER11048 "COST CHILD CARE "

ER11067 "F14 BUY FOOD NOT FD STMP"

ER11068 "F15 $SPENT NOT FDSTMP AMT"

ER11076 "F19 AMT SPENT ON FOOD"

ER11077 "F19 $ SPENT ON FOOD PER"

XXXXXXX weight variables for family

Request to check in family data

We need parents' wages to calculate the opportunity cost of parental time and to calibrate the household’s income constraint. I picked the wage data (check above). If you find any other measure of wage or household income, please add them to the list

**Request to check in family data and CDS data**

**For example, Is there any data for the intended number of children? I couldn’t find in 1997 file, but I read in some other years it is available.**

**If you come across any other variables that can be useful for our question, please add them to the list – Thank you for all your hard work**

**First stage: Provide summary statistics and graphs**

In the first stage, we aim to produce a set of summary statistics and descriptive plots using the CDS-PSID dataset.

To guide this stage, it may be helpful to refer to examples from previous studies (three papers that I attached in my last email). Specifically, Section 4 of Bruno’s paper, Sections 3.1 and 3.4 of Del Boca et al., and Sections 2 and 5 of Caucutt et al. In some cases, the appendices of these papers also contain useful data presentations. These sections can serve as illustrative examples for how to construct, summarise, and visualise the data in our study. You may use other sections from the paper too.

Here is the detailed steps on what we need to to do. Please feel free to reach out if you have any questions => I update the note on Dropbox as well

1. First, confirm the sample:
   1. I think the same families appear across the 1997, 2002, and 2007 CDS datasets, though there may be some attrition and new additions over time. It would be best to confirm this in the first stage of the analysis.
   2. I think CDS randomly chose households in 1997, and they also randomly picked a maximum of two kids in each household
   3. Is there any change in the sample of parents and kids?
   4. In 1997, I had around 25 observations where kids assigned different primary caregivers, and for one household, we had two primary caregivers. Probably we should know the size of these observations in each year and delete them while we construct summary statistics.
   5. Think about other issues in the data and document them.
2. Second, variables from the family dataset
   1. **Number of children** in the family unit.
   2. Age for both the **reference person** and **spouse/partner**
   3. Any information about the age of the kids (e.g., age of the younger kids or older kids)
   4. Information about household size and whether kids live in the household
   5. **Number of intended children** (if available)
   6. **Employment status** of both the **reference person** (formerly “head”) and **spouse/partner** (formerly “second person”)
   7. **Occupation and industry classification** for both the **reference person** and **spouse/partner**, if employed
   8. **Education level** for both the **reference person** and **spouse/partner**
   9. Marital status
   10. Income level for both the **reference person** and **spouse/partner**.
   11. Hours of work for both the **reference person** and **spouse/partner**
   12. Wage per hour (or other variable for wage) for both the **reference person** and **spouse/partner**,
   13. Household income level
   14. Total annual family-level spending on key durable and non-durable categories.
   15. Main categories of spending: Food, housing, transport, health, education, childcare, Miscellaneous (clothing, recreation, etc.), and so forth
   16. For parental labour supply, wage and income we can check and Del Boca et al, page 35-36 – section 3.1.2 and other papers
   17. If I forgot any other variable please add them to the list
3. Variables from CDS
   1. Parenting style (Bruno’s paper, page 19)
   2. Time use data
      1. Kids' time spent on activity and parents spending time on activity (Bruno’s paper, page 20, and Del Boca et al, page 33-35 – section 3.1.1, Caucutt et al’s paper section 5). (1) Active time with the mother, (2) active time with the father, and (3) the child’s productive self-investment time, Note that there are differences between papers here. This difference goes back mainly to how the active time of the mother and father is calculated.
   3. This is a second order and I don’t think we have time before presentation=> Kids' measure of cognitive and non-cognitive (Bruno’s paper, pages 18-19 and Del Boca et al, page 37 - section 3.1.3)
      1. Kids below 3 do not have data on this (e.g. LW score)

Recommended Figure/Descriptive Statistics Order for Paper

1. Sample and summary Statistics
   * We have a presentation on early November. So it is better to focus on one year before that (we may start with 2002 or 2007, which has data for expenditure, and in the next step, confirm for other years.
   * In the next steps, we can even study Change over Time (Panel Comparison). We can examine whether fertility differences across parenting styles evolve between waves (e.g., 1997 → 2002 → 2007).
2. Fertility vs. parenting style (main stylised fact)
3. Labour supply mechanisms
4. Time-use mechanisms (more important)
   * Here break down for age of the kids can be important (cited papers did that)
5. Child outcome correlations (I don’t think we have time to finalize this before the presentation and it is fine

We need summary statistics

We also need to have some graphs. Here are some examples:

1. Distribution of Fertility by Parenting Style

* For example, displays median, quartiles, and variability of the number of children within each parenting style. Helps identify differences in central tendency and dispersion

1. Average Fertility and Confidence Intervals

* Bar chart with confidence intervals: Each bar represents the mean number of children for a parenting style, with vertical lines showing 95% confidence intervals.

1. Joint Fertility–Parental Characteristics Patterns

* Explore how parenting style interacts with parental characteristics linked to fertility decisions. One panel per parenting style showing the relationship between fertility and another continuous variable (e.g., household income or parents education (e.g. collage above and below)).