FRAGILE FAMILIES AND CHILD WELLBEING STUDY
FALL 2018 PUBLIC DATA RE-RELEASE MEMO

The Fall 2018 re-release of the Fragile Families and Child Wellbeing Study (FFCWS) Public Data improves the data, documentation, and resources of the FFCWS in the following ways:

1. Provides new variable search resources at metadata.fragilefamilies.princeton.edu
2. Revises data file structure for improved ease of use
3. Standardizes variable naming system across core and in-home data
4. Standardizes negative codes for missing data
5. Restructures and revise documentation according to the changes above
6. Includes newly released data and weights
7. Provides updates for previously released data

These updates have been completed in response to feedback from our data users. We thank you for your feedback and hope that these updates improve your experience working with FFCWS data. Please review this memo for further details on these updates and send any remaining questions you may have ffdata@princeton.edu. Please download the new data files described in this memo from the [Princeton University Office of Population Data Archive](). The remainder of this memo will provide further detail regarding each of these updates.

1. Variable Search-Metadata Resources

Based on feedback from FFCWS data users, including participants in the Fragile Families Challenge,[1] we have developed new metadata resources to improve data users' experience of working with FFCWS data. These include a web Application Programming Interface (API) that provides direct access to the metadata as well as three different front-ends to the API: a website, an R package, and a Python package. To use these resources, please visit metadata.fragilefamilies.princeton.edu. For further detail regarding the preparation of these fields, please see our [forthcoming paper in *Socius*]().

2. Revised Data File Structure

All FFCWS public data are now provided in two different options of file structure:

(1) one file with all waves and components of public data (called FF_allwaves_2018) or (2) a set of six files, one for all components at each wave of data (FF_wave1_2018, FF_wave2_2018, etc.).

Table 1 shows the list of files previously available to public data users along with the name of the file that now contains those corresponding data in the re-release.

---

[1] See http://www.fragilefamilieschallenge.org/ for more information regarding the Fragile Families Challenge.

Table 1. Old and New File Names

Old File Description	Old File Name	New File*
Core-merged file	ff_pub_merge2	Split by wave across: *FF_wave1_2018-FF_wave4_2018*
Nine Year Core and More	ff_y9_pub1	*FF_wave5_2018*
Nine Year Weights	ff_y9_pubweights082013	*FF_wave5_2018*
Fifteen Year Core	ff_Y15_pub	*FF_wave6_2018*
Three-Year In-Home	inhome3yr inhome3yrcalendar ff_attachment_variables_pub	*FF_wave3_2018*
Five-Year In-Home	inhome5yr2011 inhome5yrcalendar	*FF_wave4_2018*
Three-Year Child Care Provider	ffchildcareprovider	*FF_wave3_2018*
Five-Year Teacher Survey	ff_kteachersurvey_fnlpub	*FF_wave4_2018*

* All data are also available in *FF_allwaves_2018.*

3. Standardization of Variable Names

In the previous version of the public data, a variable naming convention was used across the Core Surveys: a prefix of one-letter and one number indicated the survey and wave in which that question was asked (ex. m1 = mother baseline, f5 = father Year 9). Additionally, a "c" for "constructed" was added in front of the one letter prefix if a variable was constructed by the study staff as a short-cut for researchers. However, this convention was not consistently applied for Year 3 and Year 5 In-Home and other supplemental files. In this re-release, we applied these variable naming rules across all surveys. For a full list of old variable names and new variable names, see the relevant columns included in the metadata resources.[2] Table 2 provides a summary of all prefix changes.

Table 2. Old and New Variable Prefixes

old prefix		new prefix
kind_	=	t4
c[2-5][city|nat] wt	=	q[2-5][city|nat]wt
ffcc_centsurvey_	=	d3
ffcc_centobs_	=	e3
ffcc_famsurvey_	=	r3
ffcc_famobs_	=	s3
ffcc_pof_	=	u3
hv5_	=	[c|blank][h][5]

---

[2] Use "Old Name" in the metadata Advanced Search to identify name changes to a particular variable.

previously no prefix		new prefix
In-Home PCG Survey Years 3 & 5		[c | blank][p][3 | 4]
In-Home Observations Years 3 & 5		[c | blank][o][3 | 4]
In-Home Activity Workbook Years 3 & 5		[c | blank][h][3 | 4]
Attachment variables (Year 3 only)		ch3att
Employment calendar		ch[3 | 4]emp
Child care calendar		ch[3 | 4]cc

4. Standardization of Negative Codes for Missing Data

Previously, the FFCWS Core data files used standard negative codes for missing data. However, these negative codes were not applied consistently to Year 3 and Year 5 In-Home and other supplemental data files.[3] We recoded the missing values to match the standard below:

Value	Value Label
-9	-9 Not in wave
-8	-8 Out of range
-7	-7 N/A
-6	-6 Skip
-5	-5 Not asked
-4	-4 Multiple ans
-3	-3 Missing
-2	-2 Don't know
-1	-1 Refuse

5. Restructuring of Documentation

We have also restructured our PDF documentation. There is now one User's Guide for each wave of data. Each User's Guide now contains the information which was previously found in the "Guide to the Public Use Files", "Scales Documentation", and "In-Home User's Guide." One TXT format codebook for each wave is also provided.

6. Newly Released Data

There are now **national and city weights for the PCG, Teen, and Home Visit at Year 15**, constructed using a similar process to previous waves.

We have constructed two new variables to measure **PCG Anxiety at Year 15**, cp6gad: (PCG meets anxious criteria in last year (CIDI)) and cp6gad_9y: (Constructed - PCG meets anxious criteria since last interview (CIDI)). For more information, please see the Year 15 Guide to the Public Use Files, page 42.

---

[3] For example, in the survey of kindergarten teachers, some variables had data coded to 98 for "not ascertained" and 99 for "don't know", with a system missing if the respondent was not in the survey. These variables were marked with missing data code 14, and then our program recoded all 98 "not ascertained" to -3 "-3 Missing" and 99 "don't know" to -2 "-2 Don't know". The system missing codes were changed to -9 "-9 Not in wave".

Previously, there were only constructed variables for education at baseline and Year 9. We have added equivalent variables at Years 1, 3 and 5 (cm2edu, cm3edu, cm4edu, cf2edu, cf3edu, cf4edu) so researchers can more easily capture **parent's educational attainment** across waves.

Finally, we added a flag (cm1med) for **whether a mother's hospital medical records was abstracted** at baseline.

Three measures from the In-Home Study, which were previously available by request, are now included in the public data files, including 7 variables related to gross motor skills, 9 variables regarding sustained attention, and a series of variables regarding parent-child affect. Further documentation of these measures is available in the Year 3 and Year 5 User Guides.

7. Updates to the Data

Recoding of "Circle All that Apply" and String Variables

In the previous release, there were variables in the Child Care Provider Survey (d3) and Family Care Provider Survey (r3) that used the number from the questionnaire as the value for a "Yes" response, as opposed to the standard "1" used throughout the data. For example, in A28 from the Child Care Provider Survey, "Yes" is coded to two for d3a28_2, three for d3a28_3, etc. In order to match the standard of the data, we recoded the "Yes" responses to 1 for a series of these child care variables.[4]

Additionally, in the previous release, there were variables from the In-Home Survey components (p3, o3, p4, o4) and the surveys listed above that contained the exact quote given in response to an "Other (Specify)" response to a multiple-choice question. In order to make these variables easier to use, we have back coded these responses into the applicable existing answer option, whenever possible. If there were over 10 respondents that specified similar "Other" responses, but there was not an applicable answer option, we added a new answer option. This led to the creation of these new variables.[5] We dropped 185 string variables during this process. For a full list of these variables, please email FFData@princeton.edu.

Constructed Variables Updates

cm1bsex: We updated gender in cm1bsex for 26 focal children.[6]

cm5_bmomstat: Three mothers were incorrectly coded as having completed the Year 9 interview while incarcerated. We corrected these cases to "64 Complete by Telephone."[7]

---

[4] These variables include: d3a10a_[2-11], d3a26_[0-4][9], d3a27_[0-3][9], d3a28_[0-6][8-9], d3a33a_[2-14], d3c5a_[2-14], r3a6_a[2-14], d3c6_a[2-9], r3a7a_[2-9], r3a8_[2-9], r3a9_[2-9], r3a10_a[2-9], d3c13a_[0-7][9] , r3a17a_[0-7][9], r3a18a_[0-7][9], d3g9_[2-5][9], r3b13b_[0][2-5][8-9], r3f13_[2-5][9], r3f25[b-d].

[5] These variables include: d3a28_101, d3a15_[d,e], d3a23a_[1-7], d3g9_101, t4g4_[101-106], r3f13_101, d3c13a_10, d3c14a_101, r3a18a_101, r3b26_a, r3b31a.

[6] Cm1bsex was corrected to male for idnums 0788, 1006, 1155, 4207, 4389, 1780, 4785; and recoded to female for idnums 4582, 3108, 3776, 0993, 0689, 4129, 0968, 0076, 4671, 3340, 0874,1524, 0706, 0997, 1010, 2717, 3478, 4437and 4591.

cp5pcgrel: Previously, in order to determine the relationship of the primary caregiver to the child at Year 9, data users needed to use pcg5idstat and n5a1. We've generated cp5pcgrel using these variables and recoded to -9 "not in wave" cases who don't have a PCG or a nonparental caregiver interview. The same variable name convention now exists at Years 3 and 5, and also at Year 15.[8] For example, cp3pcgrel can be found at Year 3. pcg5idstat and n5a1 remain in the data.

cf1edu: In earlier releases, the coding for cf1edu at baseline relied primarily on mother report, and then father report filled in the gap when mother report was incomplete. This variable was modified to be primarily based on father's report. Further, we compared father report to mother report when a father entered the study after the baseline wave. When a father reported a lower educational attainment (when he first participated at a later wave) than the mother reported at baseline, we adjusted his baseline educational attainment to match his report. As a result of these changes, there are 875 fathers in a different education category, approximately half of which are in a higher category and half are in a lower category. We have retained the original variable from earlier releases but renamed it cm1fedu.

cm5edu, and cf5edu: At Year 9, due to adjustments in coding, cm5edu has 51 mothers with lower education attainment and cf5edu has 268 fathers with lower educational attainment.

child's BMI at Year 9: Previously, child's age at assessment was rounded down to the 1st of the month. We revised our age at assessment variable to the best practice of using the day of the month at birth and assessment, resulting in some changes in values for hv5_waz, hv5_wap, hv5_bmiz, hv5_bmip, hv5_haz, and hv5_hap. Nearly all cases are within 1 percentile of the previous value, with a handful of cases with a larger difference. A result of this change is an adjustment in who is flagged as an outlier by the CDC; thus, hv5_cflag has been adjusted.   Variables ch5waz (previously hv5_waz), ch5bmiz (hv5_bmiz), and ch5haz (hv5_haz) were also recoded to have missing cases as system missings. Previously, valid negative values such as "-3 Missing" were included within the values for these variables.

Survey Variable Updates
m4i23q1, m4i23q2, f4i23q1, f4i23q2: The two pilot cities were not asked these questions so their non-responses were changed from "-6 Skip" to "-5 Not Asked".

Dropped Variables
The following variables were dropped from the data file: incitysm, f5cs, ihostat, m2g5b_a1, p6e33_91, m5d2f_flag, fathid*, mothid*, p5childgen_wrong, p6box_*, ih*_resprel, *datacoll, and *inhom.   As mentioned above, we also dropped about 100 string variables after recoding the responses as described in the above section, "Recoding String variables".

---

[7] This affected idnums 0165, 2359, and 4222.
[8] We also renamed ch4pcgrel to cp4pcgrel to match this convention.