

Pathways to Desistance

Gainful Activity

This includes logic changes for all versions of the interview

June 10, 2011

Last Updated September 6, 2011

GAINFUL ACTIVITY	3
<u>SECTION 1: EXPLANATION OF VARIABLES</u>	<u>4</u>
ITEMS AT THE MONTHLY LEVEL	4
Regular Attendance.....	4
Working at least part time.....	5
Gainful Activity	6
ITEM AT THE RECALL LEVEL	7
Proportion of community months gainfully active	7
CUMULATIVE 1, 2, 3, 4, 5, 6 AND 7-YEAR VARIABLES	7
<u>SECTION 2: DESCRIPTIVE INFORMATION REGARDING THE RECALL PERIOD</u>	<u>9</u>
COMMUNITY VS. INSTITUTION MONTH MARKER	9
NUMBER OF DAYS COVERED IN EACH MONTH OF THE RECALL PERIOD.....	10
CALENDAR MONTH AND YEAR MAPPED TO A RECALL PERIOD MONTH	10
SUBJECT AGE AT EACH MONTH	10
<u>SECTION 3: VARIABLES IN THE LINEAR FORMAT</u>	<u>11</u>
CONVERSION OF RECALL MONTHLY DATA TO LINEAR MONTHS	11
EXPLANATION OF LINEAR VARIABLES.....	12
VARIABLES AVAILABLE IN THE LINEAR FORMAT	12
Gainful Activity	12
Community vs. Institution Month Marker	12
Subject Age	12
Number of Days Covered, and Calendar Months.....	12
Recall period month mapped to linear month number	12
<u>SECTION 4: DATA ISSUES</u>	<u>14</u>
OVERLAPPING MONTHS	14
Introduction.....	14
Impact on S#NMonths and calendar data	14
GAP MONTHS.....	15
Introduction.....	15
<u>APPENDIX A: GAINFUL ACTIVITY SUMMARY TABLE.....</u>	<u>16</u>

Gainful Activity

Introduction

This is a complex document meant to serve as a reference guide for individuals thinking about how to structure the information about gainful activity that is obtained in the interview. Our intent is to provide a map of the skip patterns and the meanings of variables produced in the transformation of the data from its original question form to its final SPSS data file.

You need to know a few things to use this document effectively:

1. It is laid out in sections that provide different views of the data collected.
 - The first section provides detailed information regarding the specific created by staff at the coordinating center.
 - The second section provides basic information related to each month covered in the recall period, such as the number of days covered in each month and the calendar month and year associated with each month.
 - The third section provides a detailed explanation of the conversion of the recall level monthly data to the linear format.

A summary table of the variables available is included at the end of this document (See Appendix A).

Please contact the Study Coordinator, Carol Schubert (412-647-4760; schubertca@upmc.edu) for clarification regarding any aspect of this document or specific variables.

Section 1: Explanation of Variables

Items at the monthly level

Each timepoint interview allows for a maximum of eight or 14 months in the recall period, depending on the follow-up wave (eight months was the maximum for timepoints 6-36 and 14 was the maximum for timepoints 48-84). This means that there is a corresponding variable in the dataset reflecting events occurring in each of those months through a maximum of 14 months (the outer limit of any of the possible months covered). If, however, the recall period did not include the maximum number of months (as is most often the case), there will be variables with no data. For example, Subject 1 has a recall period of five months for follow-up 12 so this means he/she will have data in five monthly variables but not the remaining nine. Subject 2, has seven months in the recall period for follow-up 12 so he/she will have data in all but seven of the monthly variables. The recall length is set by programming code based on the current date in relationship to the date of the previous interview (see “Interview Information” under Measures for a more detailed description of how the length of the recall period is determined).

Regular Attendance

This is a marker indicating whether the subject regularly attended school in the community for each month of the recall period. Regular attendance is defined as attending a community school and missing fewer than five days of school that month. The subject receives credit for attendance both if they attended a community school during a community month (a community month is defined later, [here](#)), or if they attended community school during an institution month without missing five or more days.

The attendance determination is made using variables from the school calendar, specifically those listed below. For a complete explanation of these items, please refer to the School Calendar Documentation and codebook under “Calendars”.

- Enrolled in school by month
- Number of days suspended or expelled by month
- Whether they missed five or more days of school that month for various reasons:
 - Physical illness
 - Mental health reasons/emotional problems
 - Afraid to go
 - Refusing to go/truancy
 - Any other reason

S#Gainful_Attendance_M01 to S#Gainful_Attendance_M14

“Did the subject regularly attend school in the community during month x?”

1 = Regular attendance. This is assigned in the following two instances:

- a) -- The month is a community month, and
-- The subject attended community school without missing five or more days
- b) -- The month is a facility month, and
-- The subject attended community school without missing five or more days.

0 = Irregular attendance. This is assigned in the following four instances:

- a) -- The month is a community month, and
-- The subject attended community school and missed five or more days

- b) -- The month is a community month, and
-- The subject was not enrolled in school that month
- c) -- The month is a community month, and
-- School was not in session that month
- d) -- The month is a facility month, and
-- The subject attended community school and missed five or more days

-97 = NA (Facility month). This is assigned in the following instance:

- a) -- The month is an institution month, and
-- The subject was not enrolled in school, or they attended a facility school

-88 = Cannot compute: Conflicting information. This is assigned in the following instance:

- a) -- The month is a community month, and
-- The subject reported being enrolled in a facility school

-77 = Cannot compute: Data missing. This is assigned in the following instance:

- a) -- The month is a community month, and
-- The subject's enrollment status is unknown

Working at least part time

This is a marker indicating whether the subject worked at least part time in the community for each month of the recall period. Working part time is defined as working in the community for at least two weeks of the month, at 20 or more hours/week that month.

The working part time determination is made using variables from the work calendar, specifically those listed below. For a complete explanation of these items, please refer to the "Making and Spending Money Calendar" documentation and codebook under "Calendars".

- Had a community based job that month
- Number of hours worked per week

S#Gainful_Work_M01 to S#Gainful_Work_M14

"Did the subject work at least part time in the community during month x?"

1 = Worked part time. This is assigned in the following two instances:

- a) -- The month is a community month, and
-- The subject had a community/under-the-table job for at least two weeks where they worked at least 20 hours/week
- b) -- The month is a facility month, and
-- The subject had a community/under-the-table job for at least two weeks where they worked at least 20 hours/week

0 = Worked less than part time. This is assigned in the following three instances:

- a) -- The month is a community month, and
-- The subject had a community/under-the-table job but worked less than two weeks at 20 hours/week

- b) -- The month is a community month, and
-- The subject did not work at all
- c) -- The month is a facility month, and
-- The subject had a community/under-the-table job but worked less than two weeks at 20 hours/week

-97 = NA (Facility month). This is assigned in the following instance:

- a) -- The month is an institution month, and
-- The subject did not hold any community jobs

-88 = Cannot compute: Conflicting information. This is assigned in the following instance:

- a) -- The month is a community month, and
-- The subject reported having facility jobs only

-77 = Cannot compute: Data missing. This is assigned in the following instance:

- a) -- The month is a community month, and
-- The subject's work status is unknown

Gainful Activity

This variable combines the school and work variables described above. A month is considered to be a gainfully active month if the subject either attended community school without missing five or more days of school, and/or if they worked at least two weeks of the month at 20 or more hours per week.

S#GainfulActivity_Comm_M01 to S#GainfulActivity_Comm_M14

“Gainfully active in the community this month”

1 = Gainfully active in the community. This is assigned in the following instances:

- a) -- The month is a community month, and
-- The subject regularly attended community school (attendance value = 1), **or**
-- The subject worked a community job at least part time (worked value = 1)

0 = Not gainfully active in the community. This is assigned in the following instances:

- a) -- The month is a community month, and
-- The subject did not regularly attend community school (attendance value = 0), **and**
-- The subject worked less than part time in the community (work value = 0)

-97 = NA (facility month). This is assigned in the following instances:

- a) -- The month is an institution month
-- The subject did not work part time in the community

-88 = Cannot compute: Conflicting information. This is assigned in the following instances:

- a) -- School attendance is unknown due to conflicting information (value = 88), and
-- The subject worked less than part time in the community (work value = 0)
- b) -- The subject did not attend school regularly (attendance value = 0), and
-- Work status is unknown due to conflicting information (work value = 88)

- c) -- School attendance is unknown due to conflicting information (value = 88), and
 -- Work status is unknown due to conflicting information (value = -88)

-102 = Cannot compute: Missing data. This is assigned in the following instance:

- a) -- School attendance is unknown due to missing data (value = -77), and
 -- Work status is unknown due to missing data (value = -77)

-103 = Cannot compute: Missing data. This is assigned in the following instances

- a) -- The subject did not regularly attend school that month (value = 0)
 -- Work status is unknown due to missing data (value = -77)

-104 = Cannot compute: Missing data. This is assigned in the following instances

- a) -- School attendance is unknown due to missing data (value = -77), and
 -- The subject did not work at least part time that month (value = 0)

Item at the recall level

Proportion of community months gainfully active

This variable represents the proportion of the community months the subject was gainfully active. A score is only computed if more than half of the community months have valid data.

S#PropMths_GainfulComm	“Proportion of community months gainfully active” (-97) NA: Zero community months in recall period (-101) NA: Not enough valid data present to receive a computed score (more than half of the community months have invalid/missing data)
------------------------	---

Note that this proportion score cannot be reproduced by summing the number of community months (S#CommunityMonth_M##) and the number of months gainfully active (S#GainfulActivity_Comm_M##) for two reasons:

- 1) The restrictions placed on the amount of data required to be present to receive a proportion score.
- 2) Gainful activity is not limited to community months. There are cases where the subject worked part time or regularly attended school in the community during an institution month. A month is marked as an institution month if they spent eight or more days in a secure facility that month, so these are generally cases where the month is counted as an institution month but the subject spent the majority of the month in the community (8 or 9 days in a facility).

Cumulative 1, 2, 3, 4, 5, 6 and 7-year variables

The proportion of time gainfully active in the community is also computed for each yearly interval. The rules implemented for the recall level proportion is also used for the cumulative scores. Each timepoint has to have valid data for more than half of the community months in the recall period in order for it to be included in the final cumulative computations. For instance, for a recall period containing six community months, the case has to have valid data for at least four of those months for that timepoint to be considered in the calculation. If the recall period proportion score has a value of -101 (cannot compute due to insufficient data), that timepoint is considered missing.

In addition, each case is required to have a certain number of completed timepoints in order to receive a score. The denominator for each yearly interval is set to the total number of community months with valid/non-missing data rather than the total number of community months.

- For 1-year cumulative scores, must have 1 of 2 timepoints complete (includes follow06 – follow12)
- For 2-year cumulative scores, must have 3 of 4 timepoints complete (includes follow06 – follow18)
- For 3-year cumulative scores, must have 4 of 6 timepoints complete (includes follow06 – follow36)
- For 4-year cumulative scores, must have 5 of 7 timepoints complete (includes follow06 – follow48)
- For 5-year cumulative scores, must have 6 of 8 timepoints complete (includes follow06 – follow60)
- For 6-year cumulative scores, must have 7 of 9 timepoints complete (includes follow06 – follow72)
- For 7-year cumulative scores, must have 8 of 10 timepoints complete (includes follow06 – follow84)

GainfulActivity_Year1	“Cumulative 1-year: Proportion of community months gainfully active”
GainfulActivity_Year2	“Cumulative 2-year: Proportion of community months gainfully active”
GainfulActivity_Year3	“Cumulative 3-year: Proportion of community months gainfully active”
GainfulActivity_Year4	“Cumulative 4-year: Proportion of community months gainfully active”
GainfulActivity_Year5	“Cumulative 5-year: Proportion of community months gainfully active”
GainfulActivity_Year6	“Cumulative 6-year: Proportion of community months gainfully active”
GainfulActivity_Year7	“Cumulative 7-year: Proportion of community months gainfully active”

(-97) NA: Zero community months to this point

(-201) Cannot compute: Not enough timepoints with valid data to receive a computed score

Section 2: Descriptive Information Regarding the Recall Period

As a standard practice, the specific calendar information will be accompanied by three sets of variables which describe the recall period. This information is important for the user to consider when attempting to use data characterizing the recall period (e.g. measures) in conjunction with the monthly-level calendar data. In addition, this information is useful if the user is viewing events from a developmental perspective.

Community vs. Institution Month Marker

This variable indicates whether each month of the recall period is a community or institution month, where a community month is defined as a month where the subject spent eight or fewer days in a facility. The following facility types are considered an “institution” for this purpose because these types of places do not generally allow a client to “come and go” and they provide educational services to the youth where appropriate:

- **Drug and alcohol facility or detoxification unit.** These are facilities where the primary focus is the provision of substance use treatment. Detoxification and longer-term substance use treatment programs are both included, with the vast majority being longer-term treatment facilities.
- **Psychiatric hospital or unit.** These are settings which provide inpatient acute care to evaluate and stabilize individuals with mental health or behavioral problems.
- **Jail or prison.** These settings have incarceration as the main goal, and are almost all adult settings. Jails, usually locally run, hold people until trial for relatively short sentences after trial. Prisons are usually state-run and hold people for a longer term after trial. Federal adult detention centers and Immigration and Naturalization Services (INS) detention centers are also included in this category, although these constitute a very small portion.
- **Detention.** These are juvenile facilities where adolescents await their adjudication hearing or more permanent placement after adjudication and disposition.
- **Pennsylvania Youth Development Centers (YDC) / Arizona Department of Juvenile Corrections (ADJC).** These facilities are state-run, secure juvenile facilities, formerly characterized as “state training schools”. They provide secure custody, education, and treatment to committed youth.
- **Contracted residential treatment (general).** These settings provide residential care within a structured environment. A range of services may be offered, usually centered on a specific model of intervention (e.g., peer culture, physical challenge), and there may be varying amounts of security and access to the community.
- **Contracted residential treatment (mental health).** These settings have an integrated residential program of therapies and activities. The primary focus of treatment is on the youth’s mental health needs, and the facility targets mentally ill adolescents.

Any stays in the following setting types are not considered “institutions” for this purpose because the youth in these settings have access to the community (i.e. they may attend the local schools, may hold a job in the community):

- **Foster home.** These are defined as placements with non-kin ordered by child welfare or dependency court.
- **Shelter.** These facilities provide short-term, non-secure, temporary out-of-home care.
- **Other.** This includes any residential setting not captured by the above categories such as residential military-style high school or YMCA.

S#CommunityMonth_M01 to S#CommunityMonth_M14 “Community or institution month”

(1) Community month

(0) Institution month

Number of days covered in each month of the recall period

This value represents the number of days in each month that is represented in the calendar data (i.e., if Month03 covers February, NDays for that month will = 28). The values for the last month of the recall period are reduced, however, if the interview was actually conducted in that month. For example, if the current interview is conducted on 02/18/05, the value for the last month of the recall period is 18. If the interview is being conducted before the 15th of the month, the last month covered is actually the **previous** month. However, if the interview is being conducted after the 15th of the month, then that recall period covers all days of the month up through the day of the interview.

S#NDays01 to S#NDays14 “Number of days covered in month x”

Calendar month and year mapped to a recall period month

These variables (labeled “RealDate”) provide the calendar month and year associated with each month covered in the recall period (e.g. Month 3 of the 18th month follow-up interview is equal to July 2003 for one subject but it could be referencing December 2006 for another subject).

S#RealDate01 to S#RealDate14 “Calendar month and year linked S#M#”

Subject age at each month

These variables plot the subject’s age at each month of the recall period. Subject age at the time of the interview is also available (see the “Demographics” measure codebook for more information). Two sets of variables are available for each month, 1) the subject’s age truncated to a whole number, and 2) the subject’s age as a continuous variable. Each is computed as the calendar date for the month (from the S#RealDate## variables above) minus the subject’s date of birth.

S#SubjAge_M01 through S#SubjAge_M14 “Subject age during month x (truncated0”

S#CTSubjAge_M01 through S#CTSubjAge_M14 “Subject age during month x (continuous)”

For 328 cases, the subject’s age at the baseline was greater than the age in the first month of follow06. This happened when the calendar month for the date of the baseline interview and the calendar month for S1M01 are the same. The RealDate variables are in the Year/Month/Day format, and are set to the first of each month (e.g., S1M01 = January 1, 2007; S1M02 = February 1, 2007; S1M03 = March 1, 2007, etc.), so the baseline age will be slightly higher than the age at S1M01 when this is the same month. For these cases, we have set the age at S1M01 equal to the age at the baseline interview.

Also, note that the subject’s age at the time of the follow-up interviews and their age in the final month of the recall period will not always be the same, because the last month of the recall period is not always set to the month in which the interview was conducted. If the interview was conducted before the 15th of the month, the final month of the recall period will be set (through programming code) to the month prior (e.g., if conducted on November 14th, the last month will be set to October). If the interview was conducted after the 15th of the month, the age in the final month and the age at the recall level will be the same.

Section 3: Variables in the linear format

Conversion of recall monthly data to linear months

The monthly Pathways data, in its raw form, is not suited for some kinds of analytic approaches (e.g., trajectory analysis). Each time point interview allows for a maximum of eight or 14 months in the recall period, depending on the follow-up wave (eight months was the maximum for time points 6-36 and 14 was the maximum for time points 48-84). This means that there is a corresponding variable in the dataset reflecting events occurring in each of those months through a maximum of 14 months (the outer limit of any of the possible months covered). If, however, the recall period did not include the maximum number of months (as is most often the case), there will be variables with no data. For example, subject 1 has a recall period of five months for follow-up 12 so this means he/she will have data in five monthly variables but not in the remaining nine. Subject 2, has seven months in the recall period for follow-up 12 so he/she will have data in all but seven of the monthly variables. The recall length is set by programming code based on the current date in relationship to the date of the previous interview (see “Interview Information” under Measures for a more detailed description of how the length of the recall period is determined). However, the programming code did permit the interviewer to “reset” by hand the length of the recall period. This was done infrequently, but in some instances it created a situation in which we obtained two reports of the same month. For example, follow06 covered months January to June and the interviewer resets the follow12 recall period to start with June (leading to two different reports for the month of June). A series of data cleaning decisions (described in Section 4) were implemented to correct these situations but we note them here because they are relevant to the conversion of the data to linear months.

The “linear months” data reformats the variables so that each variable is a sequential representation of life event data for each month of the research participant’s life from the baseline interview forward. Thus, “linear month 8” actually represents eight calendar months from the baseline and “linear month 16” is actually 16 months past the baseline interview. In this format, several issues are resolved. Specifically:

1. **Months not covered in the recall period are removed.** In the “recall period by-month” format (as outlined in Section 1 above), there are place-markers for months not covered in the recall period (e.g., if S#NMonths = 5, variables for months 6 – 14 are still included but will be blank for that case). The actual number of months covered in the recall period will vary from case to case, with the majority having fewer than 14 months covered.

In the linear format, variables outside of the range of months covered in the recall period are eliminated, so that only months with a subject self-report remain.

2. **Overlapping months are removed.** Situations where there were two reports for the same calendar month are corrected, and the duplicate month is removed entirely (click [here](#) for more information about overlapping months).
3. **Gap months are accounted for.** In some cases, a subject report is missing entirely for a particular calendar month. For instance, the end of the follow06 recall period may cover January 2000, and the beginning of follow12 covers March, resulting in an absence of a report for February. In this case, the missing calendar month (February) is included in the sequence of linear data but is noted as being a “gap” month (e.g., Linear month #6 = January; Linear month #7 = February (“gap month”); Linear month #8 = March, etc.).

It’s important to note that while a place-holder for the gap month is included in the linear sequence, there will not be any subject self-report data for that month (click [here](#) for more information about gap months).

4. **Missed timepoint interviews.** Place-holders for calendar months that are missing a subject self-report due to a missed timepoint interview are also included in the linear sequence. As with gap months, all self-report data from the subject will be missing for these months.

Explanation of Linear Variables

Variables available in the linear format

All items in this calendar that are available in the “by month – by recall period” format are also available in the linear format.

Gainful Activity

L##Community_Attendance	“Did the subject regularly attend school in the community”
L##Community_Work	“Did the subject work at least part time this month”
L##GainfulActivity_Comm	“Gainfully active in the community this month”

Community vs. Institution Month Marker

L01CommunityMonth to L87CommunityMonth	“Community month marker (<8 days in D/A, Psych, JailPrison, Detention, YDC, CR, and CR-MH)”
--	---

Subject Age

L01SubjAge to L87SubjAge	“Subject age (truncated)”
L01CTSubjAge to L87CTSubjAge	“Subject age (continuous)”

Number of Days Covered, and Calendar Months

L01NDays to L87NDays	“Number of days covered this month”
L01RealDate to L87RealDate	“Calendar month and year linked to linear month”

Recall period month mapped to linear month number

In addition, a “map” is provided to link the linear month back to the recall period and month in which the information was originally collected. This is important to know when recall-level data is being used in conjunction with the monthly event calendar data.

For example, a trajectory analysis of employment at the monthly level is conducted. The investigator observes a distinct period of time when employment declines (or increases) and wants to understand what else is going on in the life of the participant on that trajectory during that particular period in time. He/she can look at recall-level data for that time period (e.g., symptoms, relationship quality) to draw some inferences.

L01TpMo to L87TpMo	“Recall period month (s#m#) mapped to linear month number ##”
--------------------	---

These variables link the recall period month (s#m#) to each linear month (e.g., whether linear month number 26 is S5M1 or S5M4, etc.). A code and label is inserted into each month to indicate why data is missing for that particular calendar month. There are three important codes:

- **Months missing due to the timepoint interview being missed.**
 - These are noted with a series of five dashes (-----) in the TpMo variables (dashes are used instead of a numeric code because these are string variables).
 - In all other linear variables, a sysmis code of -201 (Month missing – timepoint interview missed) is assigned.

- **Months missing due to a gap in months between recall periods** (e.g., the end of s1 covered January, but the beginning of s2 started with March)
 - These are noted as “-GAP-” in the TpMo variables.
 - In all other linear variables, a sysmis code of -699 (Month missing – gap in months covered between recall periods) is assigned.

- **Months outside of the range of linear months covered by the case through the 84month interview.** Most cases do not have exactly 84 months covered through the follow84 timepoint. The maximum number of months covered by any case is 87, therefore all cases are allowed for that number of months, but most have fewer than that maximum.
 - Months outside of the range for a particular case are noted as ‘nirp’ (not in recall period) in the TpMo variables.
 - In all other linear variables, a sysmis code of -555 (Month outside the range of linear months covered by the case) is assigned.

Section 4: Data Issues

Overlapping Months

Introduction

The number of months covered in each recall period (S#NMonths) is usually automatically determined by the interview programming. Follow-up interviews for a particular timepoint must be conducted within a specific period of time, ranging from six weeks prior to the exact target date to eight weeks after the target date. The last month covered in the recall period is determined based on whether the interview was actually conducted in that month. If the interview was conducted after the 15th of the month, the last month of the recall period was set to the month in which it was conducted; if completed prior to the 15th, the last month was set to the month prior. For example, the last month of the recall period is set to February if the interview is conducted on or after February 15th. If the interview was conducted between February 1st and 14th, the last month of the recall period was set to January.

However, interviewers were able to override the programming that set the time frame for reporting, and in some situations they created two reports for the same month, collected at two different interviews. For instance, the follow30 timepoint may have a seven month recall period covering January through June, while the follow48 timepoint covered May through December, resulting in months May and June being covered in both timepoints.

Through follow84, there are 221 cases where the same calendar month was covered in two timepoint interviews. Data cleaning was carried out by staff at the coordinating center to remove data for the duplicate month, so that the end result is a single report for that calendar month. As a general rule, the recall period month that was closest to the time of the event for which information was being collected was kept, and the other month was removed. For example:

- The date of the follow06 interview is July 20th, 2000.
- S1M06 covers the month of July. In this recall period, the subject reports they were arrested and charged this month.
- The date of the follow12 interview is December 2000.
- S2M01 covers the month of July. Here, the subject does not report an arrest this month.

The date in S2M01 was removed. We believe the report of an arrest in S1M06 because the interview is being conducted in July, and the subject is reporting an arrest in that month. We believe the report of an arrest in S1M06 because this is closer to the time of the event (reporting an arrest in July while being interviewed in July) than the follow12 report (where the subject is being interviewed in December and trying to report on something that happened six months prior).

Impact on S#NMonths and calendar data

S#NMonths “Number of months covered in the recall period”

This variable holds the number of months covered in the recall period and is important for all data plotted by month (refer to the “Interview Information” measure codebook for more specific information about this variable) and is impacted by the overlapping months issue.

The number of months removed as a result of overlap cleaning is deducted from the interview-generated value of S#NMonths, so that the **overlap corrected** value of S#NMonths represents the actual number of months for which there is a valid subject report for that recall period.

In the majority of cases, data within overlapping months at the beginning of the recall period were removed (e.g., when S1M06 and S2M01 overlap, S2M01 is removed). A system missing code is inserted for that month to indicate that the month overlaps with the one in the prior recall period, and S2M01 will no longer hold any data. In these cases, the subject will have data in recall months that fall outside of the range of months indicated in S#NMonths. In other words, if a case originally had a five month recall period, and the first two months at the beginning are removed as a result of overlap, the new value of S#NMonths will be 3, but data will still be present in recall months 4 and 5.

For example:

- The original value of S#NMonths for follow12 = 4.
- S1M05 and S2M01 cover the month of May; all data for S2M01 are removed.
- The **overlap corrected** value of S#NMonths for follow12 = 3. Data is still present in S2M04.
- All self-reported subject data for S2M01 are removed. A system missing code indicates that that month overlaps with the one in the prior recall period.

	S1M01	S1M02	S1M03	S1M04	S1M05 (KEEP)	S2M01 (REMOVE)	S2M02	S2M03	S2M04
Calendar Month	January	February	March	April	May	May	June	July	August
Arrested Calendar (before cleaning)	No Arrest	No Arrest	No Arrest	No Arrest	Arrest	No Arrest	No Arrest	No Arrest	No Arrest
Arrested Calendar (after cleaning)	No Arrest	No Arrest	No Arrest	No Arrest	Arrest	Overlap month – refer to other timepoint for data	No Arrest	No Arrest	No Arrest

Gap Months

Introduction

In some situations, there is a gap in the months covered between recall periods. For instance, the follow30 timepoint may cover the calendar months January through July, while the follow36 recall period may cover calendar months October through December. As a result, self-reported subject data is not obtained for the months of August and September.

Through follow84, there are 125 cases that have at least one “gap” month between recall periods, with 31 of these cases having two or more missing months.

Appendix A: Gainful Activity Summary Table

Description of Variable	Variable Name	Version Change	Page Number
By recall period			
Proportion of community months gainfully active	S#PropMths_GainfulComm		7
By month			
<i>Gainful Activity</i>			
Marker for being gainfully active in the community that month	S#GainfulActivity_Comm_M##		6
Marker for regularly attending community school that month	S#GainfulActivity_Attendance_M##		4
Marker for working part-time in the community that month	S#GainfulActivity_Work_M##		5
<i>Data characterizing the recall period</i>			
Subject age at each month (truncated)	S#SubjAge_M##		10
Subject age at each month (continuous)	S#CTSubjAge_M##		10
Community vs. Institution month marker	S#CommunityMonth_M##		9
Number of days covered in each month	S#NDays##		10
Calendar month linked to each s#m#	S#RealDate##		10
Cumulative variables (1yr, 2yr, 3yr, 4yr, 5yr, 6yr, 7yr)			
Proportion of community months gainfully active	GainfulActivity_Year#		8
By linear month			
<i>Gainful Activity</i>			
Marker for being gainfully active in the community that month	L##GainfulActivity_Comm		12
Marker for regularly attending community school that month	L##GainfulActivity_Attendance		12
Marker for working part-time in the community that month	L##GainfulActivity_Work		12
<i>Data characterizing the recall period</i>			
Subject age at each month (truncated)	L##SubjAge		12
Subject age at each month (continuous)	L##CTSubjAge		12
Community vs. Institution month marker	L##CommunityMonth		12
Number of days covered in each month	L##NDays		12
Calendar month linked to each s#m#	L##RealDate		12
Additional sections supplement this calendar. Refer to the codebook section for each listing for more information			
Interview Information – contains variables that describe basic information related to the interview, such as interview completion status, interview date, version, and number of months and days covered by the recall period. This can be found under the “Interview Information” section of the Measures codebook.			

School Calendar – Contains information regarding school enrollment and attendance for each month of the recall period.

Making and Spending Money Calendar – contains information regarding employment, including whether they had a job in each month, the length of each job, and the hours worked.