Chapter 2: File Contents

Grantee welfare outcomes files should include all administrative and survey data used in analysis and reporting. To assist grantees with identifying and reporting specific data elements in these files, the Work Group developed a set of recommendations for commonly reported administrative data items, examples of grantee-specific administrative data items, and a list of some auxiliary data elements grantees may choose to report in their files.

The Group's recommendations and guidelines for administrative and survey data were developed to provide grantees with a model for reporting and labeling their data. However, for some grantees that have already produced data files for reporting and analysis, these recommendations and guidelines may require an additional level of effort to implement.

The following chapter provides further information on the Group's recommendations regarding administrative and survey data file contents and variable labeling. Ideally, grantees should follow these recommendations; however, where this proves to be a significant additional burden or is impossible to implement within existing data agreements, grantees may choose to adopt a different approach. If a different approach is selected, it is incumbent on grantees to clearly document this approach.

A. Administrative Data Files

Tip 2.1 contains a checklist identifying minimum contents of administrative data files.

Grantee administrative data files should include all data elements reported in their administrative data reports. At a minimum, the data contained in grantees' administrative data files should be sufficient for researchers to replicate results in the administrative data reports (see Tip 2.1). Additionally, case-level data should be able to be linked to data in other administrative or survey data files.

Common data elements are listed in Figure 2.1.

Grantee-specific data elements are listed in Figure 2.3.

Auxiliary data elements are listed in Figure 2.3.

Generally speaking, grantee administrative data fall into one of two categories:

- **Common data elements** analyzed by most grantees in their administrative data reports. The Group identified 18 such elements (see Figure 2.1).

- **Grantee-specific data elements** or data elements that were included in individual grantees' research, but not found across all grantees (see Figure 2.2).

In addition to examples of the common and grantee-specific data elements, the Work Group also identified a brief list of **auxiliary data** on the characteristics of former recipients and other welfare outcomes that would be useful for further research (see Figure 2.3).

Grantees are not required to include auxiliary data in their files; however, if possible, reporting such additional elements is recommended.

✓ **Tip 2.1:** Checklist for Producing Administrative Data Files

At a minimum, grantees' administrative data files should include the data necessary to replicate results in their administrative data reports. These data may include:

- ❑ Common and grantee-specific data elements.
- ❑ New variables constructed for analysis and included in administrative data reports.
- ❑ The raw data used to construct variables for analysis, except where several raw data elements were used to construct binary flags for benefit receipt during a given month or quarter.
- ❑ Variables that identify subgroups or other characteristics used in report cross-tabulations.
- ❑ Record identifiers that allow case-level data to be related or linked to survey or administrative data contained in other files.

The Work Group also recommends that grantees include common and auxiliary data elements that were not included in grantees' analysis and reporting, if this inclusion does not present a significant additional burden.

Common Data Elements

Figure 2.1 presents common data element definitions and reporting guidelines.

Appendix B contains additional detail in support of these common data items.

The Work Group identified 18 administrative data items that were reported by most grantees in their welfare outcomes research. For these items, the Work Group has developed a uniform set of definitions and reporting guidelines to ensure that these data are reported in a standardized manner. Figure 2.1 provides a brief overview of these items; Appendix B provides a more detailed description and the recommended reporting requirements for these items that may be used when constructing the file.

Although much of Figure 2.1 and Appendix B refer to **the "exit" month, this term should be understood as the "base" month or quarter in which selection into the study population occurred.** For example, the selection event could be the month or quarter in which an individual left TANF or applied for or was diverted from TANF assistance.

While it is preferred that grantees report all items and use the recommended standardized approach, **alternative approaches may be used** where it is impossible to follow these recommendations or they present a significant burden to grantees. Where another approach is used, **grantees should provide a description of the alternative in their documentation.** Grantees may also omit some common data elements if including them would present an additional burden. For example, a grantee may report only one or two of the three common data elements related to Medicaid enrollment.

Figure 2.1: Recommended "Common" Data Elements

File Field	Description	Reported Value	Minimum Time Period to Be Reported	Variable Name(s)
Quarterly Earnings	Earnings information extracted from quarterly Unemployment Insurance data.	Dollar value for total earnings during a given calendar quarter	Four quarters before and four quarters after the selection event.	Variable name prefix: EARN Examples for leavers studies: <u>Prior to TANF "exit":</u> EARNBQ01 through EARNBQ99 <u>Quarter of TANF "exit":</u> EARNAQ00 <u>After TANF "exit":</u> EARNAQ01 through EARNAQ99
TANF Benefit Receipt or Authorization	A binary variable that indicates whether a casehead received or was authorized to receive a TANF payment during a given month.	0 = Casehead did not receive or was not authorized to receive a TANF payment in a given month 1 = Casehead received or was authorized to receive a TANF payment in a given month	Twelve months before and twelve months after the selection event.	Variable name prefix: TNFR Examples for leavers studies: <u>Prior to TANF "exit":</u> TNFRBM01 through TNFRBM99 <u>Month of TANF "exit":</u> TNFRAM00 <u>After TANF "exit":</u> TNFRAM01 through TNFRAM99
TANF Benefit Amount	The TANF benefit amount paid to a casehead during a given month.	Dollar value of benefit	Twelve months before and twelve months after the selection event.	Variable name prefix: TNFB Examples for leavers studies: <u>Prior to TANF "exit":</u> TNFBBM01 through TNFBBM99 <u>Month of TANF "exit":</u> TNFBAM00 <u>After TANF "exit":</u> TNFBAM01 through TNFBAM99

File Field	Description	Reported Value	Minimum Time Period to Be Reported	Variable Name(s)
Selection Event Date	The date used for selection into the study cohort. For example, for leavers studies, this would be the month/year in which a casehead last received benefits; or first month/year in which a casehead did not receive benefits. For applicant studies, this would be the month/year in which an individual applied for TANF assistance.	MMYYYY MM = Numeric Month YYYY = Numeric Year	N/A	Leavers Studies: EXITDATE Applicant/Diversion Studies: APPLDATE
Casehead Medicaid Enrollment	A binary variable noting whether the casehead was enrolled in the Medicaid program during a given month.	0=Casehead was not enrolled in the Medicaid program during a given month 1=Casehead was enrolled in the Medicaid program during a given month	Twelve months before and twelve months after the selection event.	Variable name prefix: MEDR Examples for leavers studies: <u>Prior to TANF "exit":</u> MEDRBM01 through MEDRBM99 <u>Month of TANF "exit":</u> MEDRAM00 <u>After TANF "exit":</u> MEDRAM01 through MEDRAM99
Medicaid Enrollment for Any Member of the Former Case Unit	A binary variable that indicates whether any member of the former case unit was enrolled in the Medicaid program during a given month.	0=No member of the case unit was enrolled in the Medicaid program during a given month 1=At least one other member of the case unit was enrolled in the Medicaid program during a given month (e.g., casehead, child, or other adult)	Twelve months before and twelve months after the selection event.	Variable name prefix: MEDT Examples for leavers studies: <u>Prior to TANF "exit":</u> MEDTBM01 through MEDTBM99 <u>Month of TANF "exit":</u> MEDTAM00 <u>After TANF "exit":</u> MEDTAM01 through MEDTAM99

File Field	Description	Reported Value	Minimum Time Period to Be Reported	Variable Name(s)
Medicaid Enrollment for Any Child Who Was a Member of the Former Case Unit	A binary variable that indicates whether a child who was a member of the former case unit was enrolled in the Medicaid program during a given month.	0=No child was enrolled in the Medicaid program during a given month 1=At least one child was enrolled in the Medicaid program during a given month	Twelve months before and twelve months after the selection event.	Variable name prefix: MEDC Examples for leavers studies: Prior to TANF "exit": MEDCBM01 through MEDCBM99 Month of TANF "exit": MEDCAM00 After TANF "exit": MEDCA01 through MEDCA99
Food Stamp Receipt	A binary variable that indicates whether the casehead received or was authorized to receive food stamps during a given month.	0=Casehead did not receive or was not authorized to receive food stamps during given month 1=Casehead received or was authorized to receive food stamps during given month	Twelve months before and twelve months after the selection event.	Variable name prefix: FOOD Examples for leavers studies: Prior to TANF "exit": FOODBM01 through FOODBM99 Month of TANF "exit": FOODAM00 After TANF "exit": FOODAM01 through FOODAM99

File Field	Description	Reported Value	Minimum Time Period to Be Reported	Variable Name(s)
TANF Case Composition	Grantees should report at least two of the following: 1. Number of certified adults on TANF case. 2. Number of certified children on TANF case. 3. Total number of certified individuals on TANF case.	Number of certified adults, children, or total number of individuals on TANF case.	Twelve months before and twelve months after the selection event. For example: the month of exit, and other months up to twelve months before and after exit (where data are available).	Variable name prefix for adults: ADLT Variable name prefix for children: CHLD Variable name prefix for total number: TOTC Examples for leavers studies: <u>Prior to TANF "exit":</u> i.e., ADLTBM01 i.e., CHLDBM01 i.e., TOTCBM01 <u>Month of TANF "exit":</u> i.e., ADLTAM00 i.e., CHLDAM00 i.e., TOTCAM00 <u>After TANF "exit":</u> i.e., ADLTAM01 i.e., CHLDAM01 i.e., TOTCAM01
Casehead Date of Birth	The month and year in which the casehead was born.	MMYYYY MM = Numeric Month YYYY = Numeric Year	N/A	CASE_AGE
Casehead Gender	Casehead gender.	1=Male 2=Female	N/A	CASE_SEX
Case Type	The case unit's designation as a single, two, or no parent household.	1=Single parent household 2=Two parent household 3=No parent on case/ child only case 4=Other	N/A	CASETYPE
Race/Ethnicity	Casehead Race/ethnicity.	Grantees may use state-specific definitions and codes	N/A	CASERACE
Age of Oldest Child	Age of oldest child on case at time of selection into the study.	Age at TANF "exit" or other selection event	N/A	CHLDAGE2

File Field	Description	Reported Value	Minimum Time Period to Be Reported	Variable Name(s)
Age of Youngest Child	Age of youngest child on case at time of selection into the study.	Age at TANF “exit” or other selection event	N/A	CHLDAGE1
Geographic Identifier	County in which casehead resided at time of selection into the study.	State/County FIPS Code	N/A	COUNTYCD
Survey Sample	A binary variable that indicates whether a case was included in a survey research cohort.	0=Not included in survey sample 1=Included in survey sample	N/A	SURVFLAG
Marital Status	Casehead marital status at time of selection into the study.	Grantees may use state-specific definitions and codes	N/A	MARISTAT

Note: In all cases, the term “exit” month should be understood to mean the critical selection event that marks entry into the study population, whether that be a study population of leavers, divertees, applicants, or entrants to the TANF program.

Grantee-Specific and Auxiliary Data Elements

Each welfare outcomes grantee has selected **additional data for analysis and reporting** that extend beyond those identified in the list of common elements. These data should also be included in grantee administrative data files, and are considered **grantee-specific data elements.**

The nature of grantee-specific data requires each grantee to **define the data being reported, including its values, the time period(s) covered, and the assigned variable names.** When completing this task, grantees should follow the Work Group's recommended approach (see Chapter 1). To assist grantees in applying this approach, the Group identified a list of grantee-specific data elements to serve as examples for applying these standards. A summary of these examples is provided in Figure 2.2.

Figure 2.2 summarizes examples of reporting standards for grantee-specific data.

Figure 2.3 provides an overview and examples of auxiliary data items.

The Work Group also identified a brief list of example auxiliary data items that would be helpful for grantees to include in their files. While most grantees did not use these data in their analysis and reports, these additional data elements represent an opportunity to improve future welfare outcomes research. Figure 2.3 provides a brief overview of these items and examples of such items. Grantees are encouraged to report these auxiliary data items in their administrative data files, to the extent possible.

Figure 2.2: Examples of “Grantee-specific” Data Elements

File Field	Description	Reported Value	Recommended Time Period to Be Reported	Variable Name(s)
Number of Jobs Held	Number of jobs held by an individual in a particular quarter as indicated by the number of EIN’s listed in Unemployment Insurance wage records.	Number	Four quarters before and four quarters after selection event.	Variable name prefix: JOBS Examples for leavers studies: Prior to TANF “exit”: JOBSBQ01 through JOBSBQ99 Month of TANF “exit”: JOBSAQ00 After TANF “exit”: JOBSAQ01 through JOBSAQ99
Reported Wages	Information on casehead wages obtained from a source other than UI wage data. For example, wage information that is available on State TANF or food stamp files.	Dollar value	Twelve months after and twelve months before selection event.	Variable name prefix: WAGE Examples for leavers studies: Prior to TANF “exit”: WAGEBM01 through WAGEBM99 Month of TANF “exit”: WAGEAM00 After TANF “exit”: WAGEAM01 through WAGEAM99

File Field	Description	Reported Value	Recommended Time Period to Be Reported	Variable Name(s)
Length TANF Spell Before Selection Event	The number of consecutive months the casehead received TANF assistance immediately prior to selection event. For TANF leavers studies, this should be the number of consecutive months immediately prior to exit. For diversion/applicant studies, this should be the number of consecutive months a casehead received assistance for last episode. If there is no past history of TANF benefit receipt, the value would be 0.	Number of Months	N/A	SPELLLENT
Childcare Subsidy	The childcare subsidy amount received by a casehead in a given month.	Dollar value	Twelve months before and twelve months after selection event.	Variable name prefix: CSUB Examples for leavers studies: <u>Prior to TANF "exit":</u> CSUBBM01 through CSUBBM99 <u>Month of TANF "exit":</u> CSUBAM00 <u>After TANF "exit":</u> CSUBAM01 through CSUBAM99
Reason for Case Closure	A state-specific notation for the reason a casehead's TANF was closed at time of exit.	Individual state codes should be used and clearly documented.	N/A	CCLOSURE

Note: In all cases, the term "exit" month should be understood to mean the critical selection event that marks entry into the study population, whether that be a study population of leavers, divertees, applicants, or entrants to the TANF program.

Figure 2.3: Examples of Auxiliary Administrative Data Elements and How They May Be Defined

File Field	Description	Reported Value	Recommended Time Period to Be Reported	Variable Name(s)
Urban/Rural	A binary variable noting whether the casehead resided in an urban or rural area at time of the selection event. File documentation should clearly note how urban and rural areas have been defined.	0=Resided in rural area 1=Resided in urban area	N/A	URBRURAL
Geographic Identifier	A geographic identifying variable at the smallest possible unit (i.e., zip code or census tract) noting casehead location at time of selection event.	Preferred codes include: Zip Code or Census tract	N/A	LOCATION
Number of Children Under 6 Years of Age	Number of children on cash assistance case who were under the age of 6 at selection event.	Number	N/A	CHUNDER6
Age of Youngest Child Not on Case	Age of youngest child not on cash assistance case at time at selection event.	Number	N/A	YOUNGEST
Language Barrier	A binary variable that indicates whether the casehead spoke English or had a language barrier at time of selection event.	0=No language barrier 1=Experienced language barrier	N/A	LANGUAGE
Educational Attainment	A state-specific notation as to the highest level of education a casehead had achieved at time of selection event.	Individual state codes should be used and clearly documented	N/A	EDUCATION

File Field	Description	Reported Value	Recommended Time Period to Be Reported	Variable Name(s)
Child Support Receipt	A binary variable noting whether casehead received child support payments during the month of the selection event.	0=Did not receive child support payments 1=Received child support payments	N/A	CHLDSUPP
Subsidized or Public Housing	A binary variable noting whether a casehead received a housing subsidy or lived in publicly funded housing at the selection event.	0=Did not receive housing subsidy or live in publicly funded housing 1= Received housing subsidy or lived in publicly funded housing	N/A	PHOUSING

Note: In all cases, the term "exit" month should be understood to mean the critical selection event that marks entry into the study population, whether that be a study population of leavers, divertees, applicants, or entrants to the TANF program.

Coding Missing Administrative Data

The Work Group agreed that it is important that researchers using the welfare outcomes data files understand and correctly analyze **missing data**. For example, flags for missing data listed as "blank" or "0" in administrative data files mean very different things across data elements and, as a result, may be easily misunderstood or misinterpreted by outside users in their analyses.

Figure 2.4 provides definitions and examples of recommended codes for missing administrative data.

To mitigate possible interpretation problems caused by missing data, the Work Group recommended a **standard approach to labeling missing administrative data** in grantee files. An overview of this approach is provided in Figure 2.4. The Work Group recommends that grantees **use negative numbers between the values of one and five to flag missing data.** While it is impossible to identify all of the circumstances that may result in missing values, the Group did identify some common occurrences and assigned reserved codes for these situations.

Where recommended codes are insufficient to describe certain types of missing data, **grantees should construct additional codes using values of "-4" or "-5"** and **provide clear documentation for how these codes were applied.**

Tip 2.2 and Appendix B provide examples of how codes for missing administrative data may be applied.

Tip 2.2 provides additional information as to how recommended codes may be applied to missing Unemployment Insurance wage data.

Examples of how these codes may be applied to the recommended common data elements are also provided in Appendix B.

Figure 2.4: Reserved Codes for Missing Administrative Data

Reserved Code	Description	Definition
-1	Match File Unavailable for Cohort or Subgroup	Data are missing for a whole cohort or subgroup because the administrative data file was unavailable for data matches (e.g., data are missing due to time lags or other reasons).
-2	No Match Attempted for Individual Case	Data are missing for certain cases because the individual TANF case was excluded from the data match (e.g., necessary identifying information such as Social Security Number in the TANF record are missing).
-3	Data were Missing in Linked File	While the case itself appears in the match file, the specific data element was not reported for a particular case (e.g., there was incomplete information or missing values for a particular case).

✓ Tip 2.2: Applying the Reserved Codes for Missing Administrative Data

Using UI wage data as an example, the recommended missing administrative data codes may be applied as follows:

- **"No Match File Available" or "-1"**

Grantees should use a "-1" to label missing UI data when information for certain cases is absent because the match file was unavailable for data matches with a subset of cases. For example, UI data are missing for a whole cohort or subgroup because of time lags between the file becoming available and including the data in the analyses, OR because the UI file does not exist for some other reason.

- **"No Match in Data File" or "-2"**

Grantees should use a "-2" to label missing administrative data when information for certain cases is absent because the individual TANF case was excluded from the data match. For example, a match between the TANF records and UI data could occur for a particular case where the TANF data are missing a Social Security Number.

- **"Data Were Not Reported" or "-3"**

Grantees should use a "-3" to label missing administrative data in situations where the case appears in the match file, but the specific data element was not reported for a particular case. That is, there was incomplete information or missing data, such as evidence of benefit receipt but missing data on the benefit amount for a particular month. However, it is important to note that in UI earnings records, "missing" earnings information is generally is assumed to mean 0 earnings, not missing data. Accordingly, for the UI example, a "-3" missing code would probably not be used.

2. Survey Data Files

Grantees should produce **data files for survey data** collected as a part of their welfare outcomes grant, including all data collected by a specific questionnaire as well as some basic survey administration variables that researchers will need to correctly interpret the data. The following sections provide more specific information on the recommended contents for grantee survey data files.

Survey Sample Cases to Be Included in Grantee Files

Figure 2.5 provides instructions for including a "respondent flag" in survey data files.

The Work Group recommends that grantee survey data files include all cases, or records, that were sampled for a specific survey. That is, both survey respondents and non-respondents should be included in the survey data file. A binary "respondent flag" should be included to identify the two groups (see also Figure 2.5). Including all sampled cases will allow researchers to accurately examine differences between survey respondents and non-respondents.

Survey Data Items to Be Reported by Grantees

A checklist for producing survey data files is provided in Tip 2.3.

The Work Group recommends that grantees follow the basic premise that **all survey data collected should be included in grantee survey data files** (see Tip 2.3). However, the Group also noted several important **exceptions** to this premise:

- Verbatim, or "open-ended," question responses that have not been coded do not need to be included. Coded verbatim responses should be included in the file.

- Grantees may substitute constructed variables for raw survey data in cases where the same survey question was administered several times to different survey sample subgroups. For example, the same income question administered to different subgroups at different points in the survey may be aggregated into a single measure or variable.

- Obvious personal identifiers such as name, address, phone number, social security number, and case identification number should not be included in the file.

Similar to administrative data files, grantees should also **include variables constructed for survey data analysis and reporting.** Supporting information as to how these variables were created and their corresponding values and labels should be clearly presented in the survey data file documentation.

In some cases, grantees may feel that data from a particular survey question or bank of questions are **“bad” or “unreliable”** based on their subsequent analysis. These data should still be included in the welfare outcomes survey data file, and grantee **concerns regarding limitations should be clearly described** in the survey data file’s documentation.

✓ **Tip 2.3:** Checklist for Producing Survey Data Files

What goes into a survey data file?

- ❑ All survey data collected by the questionnaire.
- ❑ All sampled cases – not just those with whom a survey interview was conducted.
- ❑ All variables constructed for analyses that were included in grantee survey data reports; raw data used to construct variables should also be included in the file and corresponding documentation should indicate how variables were constructed.
- ❑ Survey data weights.
- ❑ Survey administration variables that describe the survey process (see Figure 2.4).
- ❑ Flags that indicate where values have been imputed.
- ❑ A unique record identifier for each case. This identifier should be constructed using the approach outlined in Chapter 1 and should allow survey data to be linked to administrative data included in other welfare outcomes files.

What does not need to go into a survey data file?

- ❑ Verbatim or open-ended survey question responses that have not been coded.
- ❑ Obvious personal identifiers, such as name, address, phone number, social security number, and case identification number.
- ❑ Data collected by computer-assisted interviewing systems for sample management and data edit checks (e.g., interviewer call attempt message fields, responses from verification questions, “soft edits”).
- ❑ Survey data collected during questionnaire pretesting, unless these data are included in grantee survey data analysis and reports.

Survey Administration Items to be Reported by Grantees

Figure 2.5 gives a complete description of the eight survey administration variables for survey data files.

To correctly interpret survey data, researchers must understand the nature of the survey implementation process. The Work Group recommends that grantees **report eight survey administration variables in the survey data files**. A complete description of these items is provided in Figure 2.5. Grantees should not report survey administration items that do not apply to their survey effort (e.g., pretest flag or sample weights).

Figure 2.5: Recommended Survey Administration Items

File Field	Description	Reported Value	Variable Name
Unique Identifier	Each casehead included in the survey data file should be assigned a unique identifier that allows case-level data to be linked to other files.	Identifier constructed using recommended approach outlined in Chapter 1.	MASTERID
Final Survey Disposition	The final survey disposition for a particular case should be indicated.	Grantee-specific codes may be used. These codes should correspond with the response rate calculation methodology outlined in the survey file's documentation.	DISPOSIT
Respondent Flag	A binary flag noting cases that are included as "completed interviews" in the numerator of the reported response rate.	0 = Not included in response rate numerator as completed interview 1 = Included in response rate numerator as completed interview	RESPOND
Sample Weight	Weights calculated to account for survey sample design effects.	Sample weight value	SRWEIGHT (If file has multiple weights, variable names should be noted as: SRWEIGT1; SRWEIGT2; Etc.)
Interview Date	The date on which a survey interview was completed.	MMDDYYYY MM = Numeric month DD = Numeric day YYYY = Numeric year	RESPDATE
Number of Telephone Attempts	The number of telephone interviewer attempts to reach a sample member for a survey interview. Survey data file documentation should clearly note how "telephone interview attempt" was defined.	Number	ATTEMPTS
Interview Mode	A flag indicating by which mode a survey interview was conducted.	1 = CATI 2 = CAPI 3 = Phone with pencil and paper 4 = In-person with pencil and paper 5 = In-person with cellular telephone 6 = Mail survey	INTVMODE

File Field	Description	Reported Value	Variable Name
Pretest Sample Flag	In cases where grantees include data from a survey pretest in their final analysis, a binary flag should be included that notes which cases are from the pretest and which cases are from the final survey fielding.	0= Non-pretest 1 = Pretest data	PRETEST

Protecting Respondent Confidentiality in Survey Data Files

Chapter 5 provides background information on confidentiality and restricted file access.

Work Group members felt that even with proposed plans for researcher-restricted access to survey data files (see Chapter 5), some grantees may feel the need to provide additional confidentiality protections in their data by excluding certain data elements. In these cases, the Group recommends that grantees consider **"top coding"** or **collapsing categories** (e.g., age), as an alternative to variable deletion. However, **if grantees must delete variables to meet confidentiality requirements, the deletions should be clearly noted in their file documentation.**

- **Top Coding**

While Work Group members did not feel that standards for top coding specific variables were necessary, they did recommend a general standard of having **no more than 5% of high- or low-range values top coded**. This strategy eliminates "extreme" high or low values from the data file.

If grantees choose to top code, these data should be included in the survey data file and the **file's documentation should clearly note where and when top coding has been applied.**

- **Collapsing Categories**

Collapsing values for continuous survey data variables into categories may reduce the possibility of identifying individual respondents. For example, children's ages may be collapsed into categories such as 0-2 years, 3-5 years, etc.

Similar to top coding, if grantees choose to collapse categories, these data should be included in the **survey data file** and the file's documentation should **clearly note where and when categories have been collapsed.**

Recommended Strategies for Reporting Special Types of Survey-related Data

Some grantee survey analysis efforts have employed **advanced statistical techniques** to account for sample design and survey non-response. Where these techniques have been applied, it is essential that grantees **clearly document these strategies** and include values or other data elements that will allow future researchers to replicate reported survey results.

The Group noted that the two most popular techniques employed by grantees were **imputation** and survey **sample weights**. Specific recommendations for reporting on these strategies were developed and are described in the following paragraphs.

- **Imputed Data**

Some grantees may choose to impute certain survey data values. In cases where imputed values are reported, **an additional data element or "flag" variable should be created that indicates which records contain imputed data**. The flag should be coded as follows:

0 = Non-imputed data
1 = Imputed data

A separate flag variable should be constructed for each survey data element that has been imputed. For example, if survey data indicating monthly household income and the number of jobs held since TANF exit includes imputed values, two flag variables should be created – imputed income and imputed number of jobs – that indicate which values have been imputed.

Grantee survey data file documentation should include a detailed account of how imputation techniques were applied.

- **Survey Weights**

Figure 2.5 shows guidelines for reporting survey administration items, including sample weights.

All survey data weights should be included in grantee data files and file documentation should include a detailed description of how these weights were calculated. Where grantees have calculated more than one survey weight for a particular record, each weight should be included in the file as a separate variable (see Figure 2.5).

Coding Missing Survey Data

Figure 2.6 shows reserved codes for missing survey data.

Similar to the reserved codes for missing administrative data, the Group recommends that grantees use a set of reserved codes when labeling **missing survey data.** For missing survey data, grantees should use **negative numbers between the values of six and nine.** The recommended reserved codes are presented in Figure 2.6. Grantees should construct additional missing codes where necessary and provide clear documentation for how these codes were applied.

Figure 2.6: Reserved Codes for Missing Survey Data

Reserved Code	Description	Definition
-6	Invalid Missing	Data are missing because a survey question was inadvertently not administered to a respondent. For example, a survey skip pattern was violated or incorrectly applied.
-7	Don’t Know	Data are missing because a respondent answered a question with a “don’t know” response.
-8	Refused	Data are missing because a respondent refused to answer a question.
-9	Valid Missing	Data are missing because survey question was not administered to respondent. For example, the question was correctly skipped.