Appendix O.

Excerpt From WIC PC2018 Final Data Cleaning Plan

Contents

Tables

WIC PC2018 Final Data Cleaning Plan: Chapter 3. Data Cleaning Plan by Phase

This section provides details on the initial data checks, the diagnostic evaluation of the data, and the steps taken to clean and compile the data.

A. Phase I: Initial Data Checks

For PC2018, State agencies will be asked to submit data on all ***WIC participants who were certified to receive WIC benefits in April 2018.*** This includes all WIC participants who were certified, regardless of whether they receive benefits. For example, they include partially breastfeeding women more than 6 months postpartum, even if they receive no food packages, or others who are eligible to receive vouchers but who do not use them. In contrast, for administrative purposes, FNS separately measures participation based on the number of certified individuals who claimed their food instruments each month.[1]

This first phase of the data cleaning process is to verify whether the data have been submitted in the correct file format and are readable in SAS. If the file can be read into SAS, analysts verify (1) whether it has the correct number of variables (depending on whether just MDS or MDS and SDS items were submitted), and (2) whether it has approximately the correct number of records. The file is checked to ensure the data submissions contain information for the reference month of April. The file is also checked to ensure that essential data items, such as certification category or food package codes, have values for most or all participants. If the file meets these conditions, it is diagnostically evaluated in Phase II. If those conditions are not met, the State agency is asked to correct and resubmit its data file. Acceptable data are converted into a SAS dataset containing both MDS and SDS data for the diagnostic evaluation.

B. Phase II: Diagnostic Evaluation

Once the State agency submits a readable file, the data are submitted to the diagnostic SAS program. Substantial resources were invested in the development of these diagnostic tables, and Insight staff will continue to build on these diagnostics to ensure State agency data anomalies are caught and addressed early in the reporting process. This program creates standardized output, consisting of more than 150 tables for each State agency, to identify problems with State agency submissions. The tables facilitate thorough assessment of the data by project analysts.

There are several different ways the data are evaluated. First, each variable is checked to ensure values submitted are within the appropriate ranges and contain valid data. The tables resulting from this analysis include distributions of all MDS and SDS data items. Second, some variables are cross-tabulated with one another to ensure the data are internally consistent. The tables resulting from this analysis

[1] For each PC report, a WIC participant is defined as a person who was certified to receive WIC benefits in April of the reference year, including individuals who did not claim a food instrument in April. In accordance with WIC regulations, this definition includes fully breastfeeding infants who were certified for WIC benefits but were not prescribed a food package, as well as partially breastfeeding women who were not prescribed a food package but whose infants were prescribed a food package. In contrast, for administrative purposes, FNS measures participation based on the number of certified individuals who claimed their FIs each month.

include cross-tabulations of related items such as certification category and risk priority level. Third, an additional test compares distributions in the current data collection year to those of the previously collected wave of data to check for consistency. The tables also highlight any errors; for example, the invalid values are put into an error category and displayed alongside the correct data. These three types of analysis are described separately in this document for the purpose of conceptual clarity, but they are often performed simultaneously. The details of each analysis are described below.

Once the diagnostic tables have been evaluated, the project analyst drafts a Diagnostic Memorandum for the State agency's review. This memo notes each potential problem with the data and suggests possible causes and resolutions. The memo asks the State agency to explain the cause of the data problem(s), and if possible, to resubmit corrected data (see appendix A for a sample memo). Next, a teleconference is held with most State agencies to clarify State agency data practices, ensure correct interpretation of the data, and suggest corrections (if necessary). If a problem can be solved with a limited amount of programming, such as a date submitted in the wrong format, Insight will correct this problem in Phase III as described below. Otherwise, the State agency is asked to resubmit the data within 2 weeks of the call.

This step is repeated for each updated submission so that each submission receives a full diagnostic evaluation to ensure that new errors have not been produced. Once a State agency's file meets quality standards (on or before the final data submission deadline of September 15, 2018), it is ready for the cleaning and compilation stage. Below, we describe the three main checks that occur in the diagnostic evaluation phase.

1. Does Each Variable Have the Appropriate Range of Values?

The first type of diagnostic test ensures the data for each individual variable appear correct. For some variables, this means confirming there are entries for each respondent. For all variables, it means ensuring the values are in the appropriate formats and ranges. Table 3.1 illustrates the range checks that will be assessed in this first stage.

Table 3.1. Variable Range Checks

Variable	Checks
MDS Variables	
State Agency ID	• Must be 100 percent reported • Check caseload against administrative data • Compare ID to WIC local agency directory (LAD) list
Local Agency	• Must be 100 percent reported • Make sure all agencies in WIC LAD list are represented
Site ID	• State agencies with sites in WIC LAD should have sites in WIC PC data; compare submission to WIC LAD
Participant ID	• Must be 100 percent reported
Birth Date	• No birth dates after April 30, 2018 • Few birth dates before 1978 • Few birth dates between April 2006 and March 2013

Variable	Checks
Race/Ethnicity	• Must be 100 percent reported • Confirm that reported values for the race variable match the coding scheme indicated by the data transmittal worksheet • Check that multiple race combinations appear • Check that multiple race/ethnicity combinations appear
Certification Category	• Must be 100 percent reported • Children should be approximately 50 percent of participants • Infants should be approximately 25 percent of participants • Should be more pregnant women than breastfeeding or postpartum women • Check against administrative data
Date of Delivery	• Valid dates: February 2018–February 2019
Weeks Gestation	• Not more than 5 percent more than 40
Certification Date	• Must be 100 percent reported • No certification dates after April 2018 • Few certification dates in March 2017 or earlier
Sex	• Even distribution between male and female
Risk Priority Code	• Must be 100 percent reported • Values 1–7
TANF Participation	• No more than 5 percent missing • Check for very low participation rates; national averages from the last PC data collection are used as benchmarks
SNAP Participation	• No more than 5 percent missing • Check for very low participation rates; national averages from the last PC data collection are used as benchmarks
Medicaid Participation	• No more than 5 percent missing • Check for very low participation rates; national averages from the last PC data collection are used as benchmarks
Migrant Status	• No more than 5 percent missing
Economic Unit Size	• Must be 100 percent reported • Zero is not a valid value
Income	• Check for reasonable values; should be very few less than $51 and very few high values
Income Reporting Period	• Values 1–5
Income Range	• Values should range from 1 to 63
Nutritional Risk #1	• Check for invalid nutritional risk codes • Should be at least 98 percent reported
Nutritional Risk #2	• Check for invalid nutritional risk codes • Should be at least 50 percent reported for women and infants • Should be at least 25 percent reported for children
Nutritional Risk #3	• Check for invalid nutritional risk codes • Should be at least 20 percent reported for women and infants • Should be at least 10 percent reported for children
Nutritional Risk #4 Through #10	• Check for invalid nutritional risk codes

Variable	Checks
Hemoglobin	• Check range: typically 7.0–20.0 • Confirm that values do not clump around whole numbers
Hematocrit	• Check range: typically 23.0–60.0
Blood-Test Date	• No dates after April 2018 • Few dates before February 2017
Weight (pounds)	• Should be at least 95 percent reported (if weight is reported in pounds) • Check range for reasonable values
Weight (quarter pounds)	• Check values (0–3 are valid)
Weight (grams)	• Should be at least 95 percent reported (if weight is reported in grams) • Check range for reasonable values
Height (inches)	• Should be at least 95 percent reported (if height is reported in inches) • Range should be 17–72
Height (eighth inches)	• Range should be 0–7
Height (centimeters)	• Should be at least 95 percent reported (if height is reported in centimeters) • Check range for reasonable values
Date of Height and Weight Measurement	• Few dates before April 2017 • Few dates after April 2018
Date Breastfeeding Data Collected	• Few dates before May 2017 • No dates after April 2018
Food Item (1–14)[a]	• For State agencies reporting food packages in item/quantity format, item 1 should have less than 5 percent missing • Compare to food package translation material; should be able to translate most frequently reported food items
Food Quantity (1–14)	• For State agencies reporting food packages in item/quantity format, quantity 1 should have less than 5 percent missing • Maximum quantity is 806
Food Package Code 1–14	• For State agencies reporting food package codes, packages should have less than 5 percent missing (if applicable) • Compare to food package translation material; should be able to translate most frequently reported food packages
Food Package Type	• Should be at least 95 percent reported • Values 1–28 • Should be at least some participants assigned to Food Package Type 28
SDS Variables	
Date of First WIC Certification	• No dates after April 2018
Education Level	• Range should be 0–18
Number in Household on WIC	• Range should be 1–20
Date Previous Pregnancy Ended	• Check for reasonable values • Should have few if any in January 2018
Total Number of Pregnancies	• Range should be 1–20 • 0 is not valid
Total Number of Live Births	• Range should be 0–20 • 0 is valid
Prepregnancy Weight (pounds)	• Check for reasonable values; most between 60 and 500

Variable	Checks
Prepregnancy Weight (quarter pounds)	• Range should be 0–3
Prepregnancy Weight (grams)	• Check for reasonable values; most between 30,000 and 250,000
Weight Gain During Pregnancy (pounds)	• Check for reasonable values; most between -20 and 50
Weight Gain During Pregnancy (quarter pounds)	• Range should be 0–3
Weight Gain During Pregnancy (grams)	• Check for reasonable values; most between 10,000 and 25,000
Birth Weight (pounds)	• Check for reasonable values; most between 2 and 12
Birth Weight (ounces)	• Range should be 0–15
Birth Weight (grams)	• Check for reasonable values; most between 900 and 5,500
Length at Birth (inches)	• Check for reasonable values; most between 17 and 23
Length at Birth (eighth inches)	• Range should be 0–7
Length at Birth (centimeters)	• Check for reasonable values; most between 25 and 61
Participation in FDPIR	• Range should be 1–2 • Must be reported for 5 percent of caseload

[a] The standard data submission file has room for 14 food items and the corresponding quantities. In some cases, a State agency may list more than 14 items in its food package. In these cases, Insight will work with the State agency to develop an alternative file structure for submissions to allow all of the food items and quantities the agency submits to be tested for data quality.

2. Are the Data Internally Consistent?

The next level of diagnostic analysis compares two or more variables to ensure the data across multiple values are internally consistent (e.g., birth dates should be consistent with certification categories). Table 3.2 provides details on the internal logic comparisons made at this stage in the diagnostic analysis.

Table 3.2. Internal Consistency Checks

Variable	Checks
Site ID/Local Agency	• In State agencies that report site IDs, check against the WIC LAD to confirm the site IDs appear under the correct local agency
Sex/Certification Category	• 100 percent reported for infants and children • Does not need to be reported for women
Age/Certification Category	• Confirm children are in the correct certification categories for their age (fewer than 5 percent of children should be younger than 12 months, fewer than 1 percent of children should be 9–10 months, and fewer than 1 percent of children should be younger than 8 months) • Confirm no infants older than 12 months • Confirm most women are 18–34 years old
Date of Delivery/Certification Category	• Date of delivery should be reported for pregnant women only
Weeks Gestation/Certification Category	• Weeks gestation should be reported for pregnant women only
Date of Delivery/Weeks Gestation/Certification Category	• Either date of delivery or weeks gestation should be reported for all pregnant women

Variable	Checks
Risk Priority Code/Certification Category	Check for valid risk priority codes by certification category: • Pregnant women: 1, 4, 7 • Breastfeeding women: 1, 2, 4, 7 • Postpartum women: 3, 4, 5, 6, 7 • Infants: 1, 2, 4, 7 • Children: 3, 5, 7
Economic Unit Size/Certification Category	• Pregnant women can be in a household size = 1 • Infants and children in foster care may be in a household size = 1, but check with State agency if these are a large percentage of cases
Income/Income Reporting Period	• Respondents who report income should also report income period (weekly, biweekly, monthly, twice monthly, annually)
Income/Income Range	• Confirm that either income or income range is reported for most respondents; no more than 10 percent missing both
Income/Economic Unit Size	• Confirm that higher incomes tend to correlate with higher economic unit sizes • Check for clusters of incomes and economic unit sizes
Income Range/Economic Unit Size	• Check for clusters of income ranges and economic unit sizes
Income/Income Range/Certification Category	• Confirm that either income or income range is reported for most respondents; should be less than 10 percent missing by certification category
Program Participation/Income/Income Range	• If income is missing for a large percentage of participants, check that participants without reported income have reported participation in TANF, SNAP, or Medicaid
Poverty Level (Income/Number in Economic Unit)	• Fewer than 3% of participants should have income greater than 185% of the poverty level.
Nutritional Risk 1/Certification Category	Check for valid nutritional risks by certification category: • Infant of a WIC-eligible mother or mother at risk during pregnancy—reported for infants • Low birth weight/premature birth—reported for infants and children • Pregnancy-induced conditions; delivery of low birth weight/premature infant; prior stillbirth, fetal or neonatal death; and general obstetrical risks—reported for women • Breastfeeding mother/infant dyad—reported for breastfeeding women and infants • Should be less than 2 percent missing in each certification category
Nutritional Risk 2/Certification Category	Check for valid nutritional risks by certification category: • Infant of a WIC-eligible mother or mother at risk during pregnancy—reported for infants • Low birth weight/premature birth—reported for infants and children • Pregnancy-induced conditions; delivery of low birth weight/premature infant; prior stillbirth, fetal or neonatal death; and general obstetrical risks—reported for women • Breastfeeding mother/infant dyad—reported for breastfeeding women and infants • Should be less than 50 percent missing in each certification category

Variable	Checks
Nutritional Risk 3/Certification Category	Check for valid nutritional risks by certification category: • Infant of a WIC-eligible mother or mother at risk during pregnancy—reported for infants • Low birth weight/premature birth—reported for infants and children • Pregnancy-induced conditions; delivery of low birth weight/premature infant; prior stillbirth, fetal or neonatal death; and general obstetrical risks—reported for women • Breastfeeding mother/infant dyad—reported for breastfeeding women and infants • Should be less than 80 percent missing in each certification category
Nutritional Risk 4/Certification Category (repeat for 5–10)	Check for valid nutritional risks by certification category: • Infant of a WIC-eligible mother or mother at risk during pregnancy—reported for infants • Low birth weight/premature birth—reported for infants and children • Pregnancy-induced conditions; delivery of low birth weight/premature infant; prior stillbirth, fetal or neonatal death; and general obstetrical risks—reported for women • Breastfeeding mother/infant dyad—reported for breastfeeding women and infants
Nutrition Risks 1–10	• Participants may report between 0 and 10 nutritional risks • 0 nutritional risks reported should be for less than 2 percent of the caseload
Hemoglobin/Hematocrit	• One of these should be reported
Hemoglobin/Hematocrit/ Certification Category	• Confirm that women do not have more than 5 percent missing • Confirm that children do not have more than 22 percent missing • Infants should be 100 percent missing
Blood-Test Date/Hemoglobin/Hematocrit	• No more than 5 percent of respondents with hemoglobin or hematocrit measures should be missing blood-test date
Blood-Test Date/Certification Date/ Certification Category	• For pregnant women, 95 percent should have blood-test date between 3 months before reported certification date to 9 months after reported certification date • For postpartum and breastfeeding women, no more than 5 percent should have blood-test date more than 9 months after certification date • For postpartum and breastfeeding women, no more than 5 percent should have blood-test date more than 3 months before certification date • For children, no more than 5 percent should have blood-test date more than 15 months after certification date • For children, no more than 5 percent should have blood-test date more than 3 months before certification date
Weight/Certification Category	• No more than 5 percent missing in each certification category • Less than 5 percent should weigh less than 100 pounds
Weight (pounds)/Weight (grams)	• Weight in grams or pounds should be reported
Height/Certification Category	• No more than 5 percent missing in each certification category • At least 75 percent should be between 5’ and 5’11”
Height (inches)/Height (centimeters)	• Height in inches or centimeters should be reported

Variable	Checks
Height/Weight/Certification Category	• Infants: Most should be between 5 and 40 pounds and 16 to 30 inches • Children: Most should be between 21 and 60 pounds and 21 to 48 inches • Breastfeeding and postpartum women (using weight and height converted to body mass index [BMI] measure): Check for outliers (normal is 19.8–26.0; many will be overweight)
Height/Age/Height Measurement Date/Certification Category	• Infants and children aged 0–23 months: Most should be in the normal range of height by age percentiles (Note: Normal range is defined by World Health Organization [WHO] anthropometric curves) • Children aged 24–60 months: Most should be in the normal range of height by age percentiles (Note: Normal range is defined by Centers for Disease Control and Prevention [CDC] anthropometric curves) • Less than 5 percent bad values (biologically implausible or missing data) for infants and for children
Weight/Height/Age/Height Measurement Date/Certification Category	• Infants and children aged 0–23 months: Most should be in the normal range of weight to height percentiles (Note: Normal range is defined by WHO anthropomorphic curves) • Children aged 24–60 months: Most should be in the normal range of weight to height percentiles (Note: Normal range is defined by CDC anthropomorphic curves) • Less than 5 percent bad values (biologically implausible or missing data) for infants and for children
Weight/Age/Height Measurement Date/Certification Category	• Infants and children aged 0–23 months: Most should be in the normal range of weight by age percentiles (Note: Normal range is defined by WHO anthropomorphic curves) • Infants and children aged 24–60 months: Most should be in the normal range of weight by age percentiles (Note: Normal range is defined by CDC anthropomorphic curves) • Less than 5 percent bad values (biologically implausible or missing data) for infants and for children
Height and Weight Measurement Date/Age/Certification Category	• Infants and children: Check that age at height and weight measurement date is appropriate for certification category; more than 5 percent out of range indicates a potential problem
Currently Breastfed/Certification Category	• Infants and children: Should be asked, but 100 percent response not required • Women: Should be missing
Ever Breastfed/Certification Category	• Infants and children: Should be asked, but 100 percent response not required • Women: Should be missing
Ever Breastfed/Currently Breastfed/Certification Category	• Infants and children only: Should have data for Ever Breastfed only if Currently Breastfed = no
Ever Breastfed/Currently Breastfed/Age	Infants 6–13 months old as of April 30, 2018: • Should have few missing Currently Breastfed data • Should have data for Ever Breastfed only if Currently Breastfed = no
Currently Breastfed/Age	• Infants 6–13 months old: Should be at least 85% reported
Length of Time Breastfed/Age/Currently Breastfeeding	• Infants 6–13 months old: If Currently Breastfed = yes, then Length of Time Breastfed should be missing

Variable	Checks
Length of Time Breastfed/Age/Ever Breastfed	Infants 6–13 months old: • If Ever Breastfed = yes, then Length of Time Breastfed should have data • If Ever Breastfed = yes, then Length of Time Breastfed should have some, but not many, zeroes • If Ever Breastfed = no, then Length of Time Breastfed should be missing
Date Breastfeeding Data Collected/Age	• Infants 6–13 months old: Should have few missing
Date Breastfeeding Data Collected/Age Breastfeeding Data Collected/Currently Breastfeeding	Infants 6–13 months old: • Should not be many currently breastfeeding infants with data collected from 0– to 4 months old • High percentage of zeroes may indicate birth date was used instead of breastfeeding collection date
Food Package Type/Certification Category	• Check for correct assignment of food package type by certification category and age; no more than 15 percent out of range for each certification category and age group • Should be more participants assigned to other types than to medical types 11–22 • Check for a high percentage of breastfeeding women with missing food package type data in conjunction with no reports of food package type 28
Missing or Out of Range Food Package Type/Certification Category	• States: No more than 5 percent missing/out of range • ITOs: No more than 10 percent missing/out of range • Check each certification category as well as overall total
Medical Food Package Types/Certification Category	• Pregnant Women: at least 95 percent nonmedical • Breastfeeding Women: at least 97 percent nonmedical • Postpartum Women: at least 98 percent nonmedical • Infants: at least 67 percent nonmedical • Children: at least 95 percent nonmedical • Overall Total: at least 87 percent nonmedical • There should be less than 100 percent nonmedical food package types overall
Food Package Number/Certification Category	• No more than 5 percent mismatch • Check each certification category and overall total
Date of First WIC Certification/Certification Category	• Must be reported for at least 5 percent of each certification category
Date of First WIC Certification/Date of Current Certification	• Date of current certification must not occur before date of first certification
Education Level/Certification Category	• Must be reported for at least 5 percent of each certification category
Number in Household on WIC/Economic Unit Size	• Number in household on WIC must be equal to or smaller than Economic Unit Size
Number in Household on WIC/Certification Category	• Must be at least 5 percent reported for each certification category
Date Previous Pregnancy Ended/Certification Category	• Should be reported for pregnant women only • Must be at least 5 percent reported for pregnant women
Total Number of Pregnancies/Certification Category	• Should be reported for pregnant women only • Must be at least 5 percent reported for pregnant women

Variable	Checks
Total Number of Live Births/Certification Category	• Should be reported for pregnant women only • Must be at least 5 percent reported for pregnant women
Prepregnancy Weight (pounds)/Certification Category	• Should be reported for pregnant women only • Must be at least 5 percent reported for pregnant women (if prepregnancy weight is reported in pounds)
Prepregnancy Weight (quarter pounds)/Certification Category	• Should be reported for pregnant women only
Prepregnancy Weight (grams)/Certification Category	• Should be reported for pregnant women only • Must be at least 5 percent reported for pregnant women (if prepregnancy weight is reported in grams)
Weight Gain During Pregnancy (pounds)/Certification Category	• Should be reported for breastfeeding and postpartum women only • Must be at least 5 percent reported for breastfeeding and postpartum women (if weight gain is reported in pounds)
Weight Gain During Pregnancy (quarter pounds)/Certification Category	• Should be reported for breastfeeding and postpartum women only
Weight Gain During Pregnancy (grams)/Certification Category	• Should be reported for breastfeeding and postpartum women only • Must be at least 5 percent reported for breastfeeding and postpartum women (if weight gain is reported in grams)
Birth Weight (pounds)/Certification Category	• Should be reported for infants and children only • Must be at least 5 percent reported for infants and children (if birth weight is reported in pounds)
Birth Weight (ounces)/Certification Category	• Should be reported for infants and children only
Birth Weight (grams)/Certification Category	• Should be reported for infants and children only • Must be at least 5 percent reported for infants and children (if birth weight is reported in grams)
Length at Birth (inches)/Certification Category	• Should be reported for infants and children only • Must be at least 5 percent reported for infants and children (if length at birth is reported in inches)
Length at Birth (eighth inches)/Certification Category	• Should be reported for infants and children only
Length at Birth (centimeters)/Certification Category	• Should be reported for infants and children only • Must be at least 5 percent reported for infants and children (if length at birth is reported in centimeters)
Participation in Food Distribution on Indian Reservations Program/Certification Category	• Should be reported for all certification categories

3. Do the Data Compare Appropriately to Previous WIC PC Distributions?

The diagnostic program also compares the current PC submission to the State agency's previous PC submission. This comparison across years provides a check against data that may be reported incorrectly but are still within the range of acceptable values. For example, if a State agency reports that 15 percent of its WIC participants are of Hispanic ethnicity in one submission or "wave," but that value changes to 75 percent in the next wave, this indicates the race/ethnicity variables may be reported incorrectly.

Changes greater than +/- 5 percentage points from the previous wave will be investigated. Table 3.3 provides details on the historical data checks made at this stage in the diagnostic analysis.

Table 3.3. Historical Data Checks

Variable	Checks
Number of Participants	• No changes in distribution greater than +/- 5 percent from the previous wave
Race	• No changes in distribution greater than +/- 5 percent from the previous wave
Race/Ethnicity	• For certain State agencies that historically overreported American Indians in conjunction with Hispanic ethnicity, compare percentage of Hispanic and non-Hispanic American Indians with previous wave; no changes greater than +/- 5 percent
Certification Category	• No changes in distribution greater than +/- 5 percent from the previous wave
Risk Priority Code	• No changes in distribution greater than +/- 5 percent from the previous wave
TANF Participation	• No changes in distribution greater than +/- 5 percent from the previous wave
SNAP Participation	• No changes in distribution greater than +/- 5 percent from the previous wave
Medicaid Participation	• No changes in distribution greater than +/- 5 from the previous wave
Migrant Status	• Compare percentages and numbers to previous wave
Food Item	• Compare some large item codes; should be similar item codes (if coding scheme has not changed since PC2016)
Food Package Codes	• Compare some large package codes; should be similar package codes (if coding scheme has not changed since PC2016)

If the State agency continues to experience trouble generating complete and accurate data for any MDS/SDS element after resubmission(s) of data, Insight works with the WIC PC contact at the State agency to identify accurate versus inaccurate data. If Insight is unable to obtain an explanation for questionable data submitted by a State agency, Insight cleans the data during the State-agency-specific editing stage (Phase III).

C. Phase III: Data Cleaning

Once each State agency has submitted data that meet the quality standards of the diagnostic program described above and all data items have met the checks, that State agency data file is cleaned in four separate steps:

1. **State agency-specific cleaning.** The cleaning process has some portions that are specific to the State agency. Some systematic errors discovered during the diagnostic phase may be addressed during this step if the State agency does not have the resources to make the changes internally.
2. **Universal cleaning.** All data are cleaned to universal standards. This step includes setting most outliers to missing and top-coding some data. In cases where a respondent should not be in the universe for a specific item, any data reported for that item are set to missing. This step also

includes standardization of variables with different reporting formats. Standardization of food codes occurs separately, as described in chapter 4.

3. **Analytic variable creation.** Several new variables are created. The new variables are typically created for ease of analysis. For instance, some State agencies report height in inches and eighth inches, while others use centimeters. The new height variable combines those three variables into one variable. Other new variables include anthropometric measures, race/ethnicity variables, and ages at certain measurement dates.
4. **Accuracy checks.** Frequencies and cross-tabulations are created for the final set of variables. These tables are reviewed to ensure the data were cleaned correctly and the derived variables have the correct parameters.

Each of these four steps is described in detail below.

1. State Agency-Specific Cleaning Procedures

The rules for this step of the cleaning are determined in the State-agency-level diagnostic phase described above, and these procedures will be developed and applied for each individual submission. Whether State-agency-specific cleaning is required will be determined after a review of the diagnostics and discussion with the State agency. As mentioned above, every problem identified in a State agency data submission is resolved through a series of communications with the State agency staff. For example, there may be data entry errors (e.g., entries of zero [“0”] for missing income data) or reporting mistakes (e.g., two-digit local agency numbers missing the leading zero). After communicating with the State agency, a State agency-specific program is written to address the specific problems affecting its data. These are handled on a case-by-case basis and therefore cannot be fully detailed here.

2. Universal Cleaning Procedures

Several variables are submitted in different formats across State agencies. These variables are converted into standard formats during universal cleaning. Table 3.4 illustrates these standard editing procedures. See chapter 4 for details on food code standardization.

Table 3.4. Standardization

Variable	Recoding Strategy
Race	Race can be reported by State agencies in two different ways. For State agencies that use three-digit race codes, these codes are converted to six-digit codes.
Race (5 category)	A five-category race variable is created to assign respondents to the mutually exclusive racial/ethnic categories used prior to 2006: White, Black, Asian/Pacific Islander, Hispanic, or American Indian/Alaskan Native. People of multiple races are assigned to one race, with priority given to different races in different State agencies, as described below. The Asian and Native Hawaiian/Pacific Islander groups are combined in this coding structure. • For ITOs, the priority system is as follows: American Indian, Hispanic, Black, Asian/Pacific Islander, White. • For Hawaii, Guam, the Commonwealth of the Northern Mariana Islands, and American Samoa, the priority system is as follows: Asian/Pacific Islander, Hispanic, Black, Asian, American Indian, White. • For all other State agencies, the priority system is as follows: Hispanic, Black, American Indian, Asian/Pacific Islander, White. *Note: This five-category variable permits comparisons of race/ethnicity from before and after 2006.*
Nutritional Risk	For State agencies reporting nutrition risk codes that do not match FNS standard codes, the crosswalks provided by State agencies between the nutrition risk codes in the data and the FNS standard codes are used to translate the data to the FNS standard.
Local Agency	State agencies are asked to submit local agency data using the three-digit code in the WIC LAD, but some may submit their own codes. In these cases, the codes are mapped onto WIC LAD codes. For all States, a character variable with the local agency name is created. *Note: The WIC LAD is divided into 90 crosswalks, 1 for each State agency.*

Once all variables have been standardized, universal cleaning procedures are applied uniformly to all State agencies. Outlying and illogical data are set to missing, or in some cases top coded. Dates with missing days are set to 15, and all dates are converted to SAS formats for ease of manipulation. Consistency edits are also done here—for example, deleting the expected date of delivery if the participant is not a pregnant woman. Table 3.5 illustrates these editing procedures.

Table 3.5. Universal Cleaning Procedures

Variable	Recoding Strategy
Certification Category	• If certification category does not equal 1–5, then the record is dropped.[a]
Certification Date	• If day is missing, set to 15. • If date is April 31, set to April 30. • Converted to SAS date. • If date is before March 1, 2017, or after April 30, 2018, set to missing. • If Birth Date and Certification Date are both present and Birth Date is after Certification Date, set both to missing.

Variable	Recoding Strategy
Birth Date	• If date is after April 30, 2018, record is dropped. • If date is before January 1, 1967, set to missing. • If date is after Certification Date, set both to missing. • For infants and children, if birth date is before January 1, 2010, set to missing. • If day is missing, set to 15. • If date is April 31, set to April 30. • Converted to SAS date. • For infants and children, if age in months at certification date is greater than 60, set to missing. • For women, if age in years at certification date is less than 10, set to missing.
Nutritional Risk	• If a participant has duplicate codes, set to missing. • If the participant is an infant or child with nutritional risks 300, 320, or 330, those are recoded to 700. • If the participant is an infant or child with nutritional risk 310, recoded to 140. • If the participant is a woman with nutritional risk 140, recoded to 310. • If the participant is a woman with nutritional risk 700, recoded to 600.
Date of Height and Weight Measurement	• If day is missing, set to 15. • If date is April 31, set to April 30. • Converted to SAS date. • If before March 1, 2017, or after October 31, 2018, set to missing. • If date of height and weight measurement is before birth date, set to missing.
Date of Delivery	• If day is missing, set to 15. • Converted to SAS date. • If the date is before February 1, 2018, set to missing. • If the date is after February 1, 2019, set to missing. • For all certification categories other than pregnant women, set to missing.
Date Breastfeeding Data Collected	• If day is missing, set to 15. • If date is April 31, set to April 30. • Converted to SAS date. • If date is before birth date, set to missing. • If before February 1, 2017, or after October 31, 2018, set to missing. • For all women, set to missing.
Weeks of Gestation	• If less than or equal to zero, set to missing. • If more than 40,[b] set to missing. • For all certification categories other than pregnant women, set to missing.
Sex	• If not 1 or 2, set to missing. • Set to 2 for all women.
Risk Priority Code	• If not 1–7, set to missing.
TANF Participation	• If not 1–2, set to missing.
SNAP Participation	• If not 1–2, set to missing.
Medicaid Participation	• If not 1–2, set to missing.
Migrant Worker Status	• If not 1–2, set to missing.

Variable	Recoding Strategy
Size of Economic Unit	• If less than 1, set to missing. • If greater than or equal to 95, set to missing. • If greater than 20 and less than 95, set to 20. • If size of economic unit = 1 for a breastfeeding unit, set to missing.
Income	• If -1, 1, 99997, 99998, or 99999, set to missing. • If 0 and income range is present, set to missing. • If 0 and participant also participates in TANF, SNAP, or Medicaid, set to missing. • If income is missing but income range is reported, set income period to 4 and set income to midpoint of the reported income range. • If income is not 0 and income period is missing, set income to missing. • If annual income is greater than $200,000, set to missing.
Income Period	• If not 1–4, set to missing. • If income is 0 and income period is missing, set to 4 (yearly). • If annual income is greater than $200,000, set to missing
Income Range	• If less than 0 or greater than 63, set to missing.
Hemoglobin	• If 0, set to missing. • If more than 20, set to missing.
Blood-Test Date	• If after October 31, 2018, set to missing. • If hematocrit and hemoglobin are both missing, set to missing. • If more than 5 years before April 30, 2018, set to missing. • For children, if blood-test date is more than 450 days before or after certification date, set to missing. • For women and infants, if blood-test date is more than 400 days before or after certification date, set to missing. • If blood-test date is before birth date, set to missing.
Hematocrit	• If more than 55, set to missing. • If less than or equal to 20, set to missing.
Weight (pounds)	• If less than 2, set to missing. • If 997, 998, or 999, set to missing.
Weight (quarter pounds)	• If more than 3, set to missing. • If weight in pounds is missing, set to missing.
Weight (grams)	• If less than 896, set to missing. • If 999997, 999998, or 999999, set to missing.
Height (inches)	• If less than or equal to 10, set to missing. • If 97, 98, or 99, set to missing.
Height (eighth inches)	• If more than 7, set to missing. • If height in inches is missing, set to missing.
Height (centimeters)	• If less than 25.4, set to missing. • If 999.7, 999.8, or 999.9, set to missing.
Currently Breastfed	• If not 1 or 2, set to missing. • For women, set to missing. • If Ever Breastfed is no and Currently Breastfed is blank, then set to 2 (no).

Variable	Recoding Strategy
Ever Breastfed	• If not 1 or 2, set to missing. • For women, set to missing. • If Currently Breastfed is yes, set to missing. • If Ever Breastfed is missing and Length of Time Breastfed is greater than or equal to 1 and Currently Breastfed is no, then Ever Breastfed is set to 1 (yes).
Length f Time Breastfed (weeks)	• If less than 0, set to missing. • If greater than or equal to 60, set to missing. • For women, set to missing. • If Currently Breastfed is yes, set to missing. • If Ever Breastfed is no, set to missing. • If Ever Breastfed is missing, and Length of Time Breastfed is 0, set to missing. • If equal to 0, set to 0.286.
Food Package Type	• If less than 1, set to missing. • If greater than 28, set to missing. • For infants and children, if equal to 20–22 or 25–28, set to missing. • For women, if not equal to 20–22 or 25–28, set to missing.
Date of First WIC Certification	• If day is missing, set to 15. • If after April 30, 2018, set to missing. • If after current certification date, set to missing. • For pregnant women, if before July 1, 2017, set to missing. • For breastfeeding or postpartum women, if before July 1, 2016, set to missing. • For infants and children, if before April 1, 2013, set to missing. • If before birth date, set to missing.
Date Last Pregnancy Ended	• If day is missing, set to 15. • For all certification categories other than pregnant women, set to missing. • If after April 30, 2018, set to missing. • If before January 1, 1970, set to missing. • If equal to or later than [Certification Date-(Weeks Gestation*7)] or [Expected Date of Delivery-280 days], set to missing. • If before birth date, set to missing. • If after March 30, 2018, set to missing.
Number in Household on WIC	• If less than or equal to 0, set to missing. • If greater than 95, set to missing. • If greater than 20 and less than 95, set to 20. • If greater than number in economic unit, set to missing.
Years of Education	• If greater than 20, set to missing. • If less than 0, set to missing.
Total Number of Pregnancies	• If less than 1, set to missing. • If greater than 20, set to missing. • For all certification categories other than pregnant women, set to missing.
Total Number of Live Births	• If less than 0, set to missing. • If greater than 15, set to missing. • For all certification categories other than pregnant women, set to missing.

Variable	Recoding Strategy
Prepregnancy Weight (pounds)	• If less than 0, set to missing. • If greater than 900, set to missing. • For all certification categories other than pregnant women, set to missing.
Prepregnancy Weight (quarter pounds)	• If greater than 3, set to missing. • If prepregnancy weight in pounds is missing, set to missing. • For all certification categories other than pregnant women, set to missing.
Prepregnancy Weight (grams)	• If less than 0, set to missing. • If greater than 408,600, set to missing. • For all certification categories other than pregnant women, set to missing.
Weight Gain During Pregnancy (pounds)	• For all certification categories other than breastfeeding and postpartum women, set to missing. • If less than -20, set to missing. • If greater than 95, set to missing.
Weight Gain During Pregnancy (quarter pounds)	• For all certification categories other than breastfeeding and postpartum women, set to missing. • If weight gain during pregnancy reported in pounds is missing, set to missing. • If greater than 3 or less than -3, set to missing.
Weight Gain During Pregnancy (grams)	• For all certification categories other than breastfeeding and postpartum women, set to missing. • If less than -9072, set to missing. • If greater than 43,091, set to missing.
Birth Weight (pounds)	• For women, set to missing. • If less than 2, set to missing. • If greater than 12, set to missing.
Birth Weight (ounces)	• For women, set to missing. • If greater than 15, set to missing.
Birth Weight (grams)	• For women, set to missing. • If less than 896, set to missing. • If greater than 5,443, set to missing.
Length at Birth (inches)	• For all women, set to missing. • If less than 10, set to missing. • If greater than 24, set to missing.
Length at Birth (eighth inches)	• For all women, set to missing. • If greater than 7, set to missing.
Length at Birth (centimeters)	• For all women, set to missing. • If less than 35.4, set to missing. • If greater than 60.96, set to missing.
Participation in FDPIR	• If not 1–2, set to missing.

[a] Certification Category is never recoded. Any records with values other than 1–5 for Certification Category are dropped; all other variables are edited to match the Certification Category, if necessary.
[b] WIC PC has historically used 40 as the cutoff point.

3. Creation of Derived Variables

Once the original variables have been cleaned, several new variables are created. These variables are typically added to the dataset to aid in analysis. Most of these are simple calculations using MDS items only. Table 3.6 illustrates the new variables created through this process.

Table 3.6. Derived Variables

New Variable	Specifications
Weight (combined)	• For State agencies that report weight in pounds, this variable equals weight in pounds plus weight in quarter pounds. • For State agencies that report weight in grams, this variable equals weight in grams divided by 453.6. • If more than 500, set to missing. • For women, if less than 70, set to missing. • For infants and children, if more than 200, set to missing.
BMI	• Reported for children aged 2–5. Calculated from height and weight. • Calculated using programming code for pediatric anthropometry developed by CDC based on current growth charts.
BMI Percentile	• Reported for children aged 2–5. • The National Center for Health Statistics (NCHS) percentiles for BMI are calculated using a program developed originally by CDC. This program (1) evaluates the raw height and weight measurements on child participants age 24 months and older and (2) identifies biologically implausible values (BIVs) with a flag variable. The program assigns 3 percentiles to each participant. "These percentiles express a child's BMI relative to children in the U.S. who participated in national surveys that were conducted from 1963-65 to 1988-94."[a] (Note: for children aged 2 and older, weight-to-height percentiles are calculated based on BMI.) • Children with height and weight flagged as biologically implausible are reported in PC tables as having "Invalid or unreported" data. The PC data files, however, contain the calculated percentile and BIV flag.
Age in Years at Certification Date	• Calculated for women only. • Age as of the last birthday at the time of certification. • If less than 10, set to missing.
Age in Months at Certification Date	• Calculated for infants and children only. • Number of months between birth date and certification date, counted in months based on the day on which they were born. • If greater than 60, set to missing.

New Variable	Specifications
Age at Height and Weight Measurement	For women: • Age is reported in years as of the last birthday at the time of height/weight measurement. • Age less than 10 years or greater than 51 years is set to missing. For infants and children: • Age is reported in months expressed as the count of monthly birthdays plus additional days expressed as a fraction of a month. (If a child is 24 months and 10 days of age, her age in months would be 24 + 10/30.4375 = 24.33 months old.) For all participants: • Age less than 0 is set to missing. • If date of height and weight measurement is not available, certification date is used instead.
Annual Income	• Calculated from income and income period (income*52 for weekly income, income*12 for monthly income, income*26 for biweekly income, income*1 for annual income). • If more than 200,000, set to missing (also set income and income period to missing).
Height (combined, in inches)	• For State agencies that report height in inches, this variable equals height in inches plus height in eighth inches. • For State agencies that report height in centimeters, this variable equals height in centimeters divided by 2.54. • If 0, set to missing. • If more than 90, set to missing. • For women, if less than 48, set to missing. • For infants and children, if more than 60, set to missing.
Prepregnancy Weight (combined, in pounds)	• For State agencies that report prepregnancy weight in pounds, this variable equals prepregnancy weight in pounds plus prepregnancy weight in quarter pounds. • For State agencies that report prepregnancy weight in grams, this variable equals prepregnancy weight in grams divided by 453.6. • If more than 500, set to missing. • If less than 70, set to missing.
Weight Gain During Pregnancy (Combined, in pounds)	• For State agencies that report weight gain during pregnancy in pounds, this variable equals weight gain during pregnancy in pounds plus weight gain during pregnancy in quarter pounds. • For State agencies that report weight gain during pregnancy in grams, this variable equals weight gain during pregnancy in grams divided by 453.6. • If less than -20, set to missing.
Birth Weight (combined)	• For State agencies that report birth weight in pounds, this variable equals birth weight in pounds plus birth weight in ounces (converted to pounds). • For State agencies that report birth weight in grams, this variable equals birth weight in grams divided by 453.6. • If 0, set to missing.

New Variable	Specifications
Length at Birth (combined)	• For State agencies that report length at birth in inches, this variable equals length at birth in inches plus length at birth in eighth inches . • For State agencies that report length at birth in centimeters, this variable equals length at birth in centimeters divided by 2.54. • If 0, set to missing.
Race/Ethnicity Dichotomous Variables	• A series of six dichotomous race/ethnicity variables (Hispanic, White, Black, Asian, Native Hawaiian/Pacific Islander, American Indian), based on the six-digit race/ethnicity codes.
Single Race Dichotomous Variables	• A series of five dichotomous race variables that indicate whether the respondent belongs only to that racial group (White only, Black only, Asian only, Native Hawaiian/Pacific Islander only, American Indian only), based on the six-digit race/ethnicity codes.
Multiple Race Flag	• A dichotomous flag is created to indicate if the respondent is of multiple races.
Missing Race Flag	• A dichotomous flag is created to indicate if race data are missing.
Total Number of Nutritional Risks	• Sums the number of nutritional risks for each individual.
FNS Region	• FNS region in which participant's local agency is located.
Age in Years at Blood Test	• Calculated for women only. • Age in years at blood-test date. • If blood-test date is missing, certification date is used instead. • If less than 10, set to missing.
Age in Months at Blood Test	• Calculated for infants and children only. • Age in months at the certification date counted in months based on the day on which they were born. • If blood-test date is missing, certification date is used instead. • If greater than 60, set to missing.
Percent of Poverty Level	• This variable compares calculated annual income and size of economic unit to the U.S. Department of Health and Human Services poverty guidelines, as published in the *Federal Register* and used by WIC to determine eligibility. The State-specific measures for Alaska and Hawaii are used. For PC2018, the 2017 guidelines are used.
Weight to Height Percentile Range	• For infants and children aged 0–23 months, percentiles for height and weight are calculated using a program developed originally by WHO. These growth standards reflect the growth of children in environments believed to support optimal growth. • For children aged 24 months or older, z-scores and percentiles for height and weight are calculated using a program developed originally by CDC, based on anthropometric percentiles normed to the U.S. population. (Note: For children 2 years of age and older, weight to-height percentiles are calculated based on BMI.)
Weight to Age Percentile Range	• Same as Weight to Height Percentile Range.
Height to Age Percentile Range	• Same as Weight to Height Percentile Range.
Weight to Height Percentile	• Same as Weight to Height Percentile Range.
Weight to Age Percentile	• Same as Weight to Height Percentile Range.
Height to Age Percentile	• Same as Weight to Height Percentile Range.

New Variable	Specifications
Blood Measures Below FNS Standard	• Using the FNS-issued nutrition risk criteria for hemoglobin and hematocrit values, this variable indicates if (0) blood measures are not below the cutoff values, (1) hemoglobin or hematocrit is below the cutoff value, or (2) blood measures were not reported.
Breastfeeding Age Flag	• Reported for infants 6–13 months old in April 2018. • A flag value of 1 represents a breastfeeding age of 6 months, or 12–13 months; a flag value of 2 represents a breastfeeding age of 7–11 months.
Breastfeeding Status	• This variable indicates if the infant is (1) currently breastfeeding or breastfed at some time, (2) never breastfed, or (3) breastfeeding status not reported.
Breastfeeding Duration, in Weeks	• For currently breastfeeding infants and children, ROUND ((date of breastfeeding response—birth date)/7). • For not currently breastfeeding infants and children, set equal to length of time breastfed.
Trimester of Pregnancy at WIC Certification	• Calculated for pregnant women only. • Calculated from certification date and date of expected delivery, or weeks gestation.
Trimester of Pregnancy at Time of Blood Test	• Calculated for pregnant women only. • Calculated from blood test date and date of expected delivery. • If blood-test date is missing, certification date is used instead.
Food Prescription ID	• This is the linking identifier that is also on the food package data analytic file; it is created by the data cleaning program.

[a] See https://www.cdc.gov/healthyweight/assessing/bmi/childrens_bmi/about_childrens_bmi.html

4. Final Data Check

After the universal cleaning has taken place, an abbreviated set of data frequencies and crosstabs is created from each State agency's clean data file. This output is reviewed to ensure all cleaning has occurred and composite variables were created as intended. This review confirms that all data are within expected ranges. The data are also examined by certification category to ensure the universes are correct for each variable. This step also checks that the ranges for each variable are appropriate within each certification category. Variables in the SDS are checked to ensure that at least 5 percent of participants who should report that data item are represented in the data. The analyst uses a checklist to ensure the data can be considered final. This checklist is provided in appendix B. After all checks have been completed, the file is considered final and ready for tabulations.

D. Phase IV: File Creation

Once the State agency data files have been cleaned, they are used in the creation of the final data files as follows:

- Ninety cleaned State agency census files
- Combined census file of all WIC participants with appended geographic variables (see section E of this chapter for more information on geographic variables)
- Nationally representative sample files (FNS internal file and public use file)
- Race/ethnicity dataset

- Food package data file

The individual State agency census files, the combined census file, and the nationally representative sample files contain participant-level data. Table 3.7 lists the variables included in the national sample public use file. The Race/ethnicity dataset contains data on the race and ethnicity of participants within each local agency, State agency, and region; three accompanying worksheets summarize this data. The food package file is described in chapter 4.

Table 3.7. WIC PC2018 Participant Characteristics File Variables

Variable Name	Description	Variable Type
STATE	State agency	MDS
STATE_NAME	State agency name	Derived
REGION	FNS region	Derived
LOCAL_NAME	Local agency name	Derived
SITE	Site code	MDS
ID_10	Ten-digit local agency identifier	Derived
ID	Case ID	MDS
PARTICID	Participant identifier	Created
BDATE	Birth date	MDS
A_MONTHS	Infant/child age in months	Derived
A_YEARS	Woman's age in years	Derived
ORIGINAL_RACE	Race as submitted	Derived
RACE5	Race recoded into pre-2006 categories	Derived
HISP	Hispanic/Latino origin	Derived
INDIAN	American Indian/Alaska Native	Derived
ASIAN	Asian	Derived
BLACK	Black/African-American	Derived
HIPI	Native Hawaiian/Pacific Islander	Derived
WHITE	White	Derived
INDIAN_ONLY	American Indian/Alaska Native only	Derived
ASIAN_ONLY	Asian only	Derived
BLACK_ONLY	Black only	Derived
HIPI_ONLY	Native Hawaiian/Pacific Islander only	Derived
WHITE_ONLY	White only	Derived
MULTIPLE_RACE	Two or more races	Derived
NO_RACE_REPORTED	Race not reported	Derived
CERT_CAT	Certification category	MDS
EDATE	Expected date of delivery	MDS
GEST	Weeks gestation	MDS
TRIMSTR	Trimester of pregnancy at WIC certification	Derived
CDATE	Date of current certification	MDS
SEX	Sex	MDS
RISK_PRI	Risk priority codes	MDS
TANF	Participation in TANF	MDS

Variable Name	Description	Variable Type
SNAP	Participation in SNAP	MDS
MEDICAID	Participation in Medicaid	MDS
MIGRANT	Status as a migrant worker	MDS
ECO_UNIT	Number in economic unit/family	MDS
INCOME	Economic unit/family income	MDS
INC_PER	Reporting period for INCOME	MDS
INC_ADJN	Annual income ranges	MDS
ANUALINC	Economic unit/family annual income	Derived
POVLEVEL	Percent of Federal poverty level	Derived
NAWD1-NAWD10	Nutritional risk(s) present at certification	MDS
MAX_NAWD	Total number of nutritional risks	Derived
HEMOGLOB	Hemoglobin measure	MDS
HEMACRIT	Hematocrit measure	MDS
R_BELOW	Blood measures below FNS standard	Derived
BLDATE	Date of blood test	MDS
ANEMIA_MONTHS	Infant/child age in months at blood test	Derived
ANEMIA_YEARS	Woman's age in years at blood test	Derived
ANEMIA_TRIMSTR	Trimester of pregnancy at blood test	Derived
WGHT	Weight	Derived
HGHT	Height	Derived
HDATE	Date of height/weight measurements	MDS
A_WTAG	Combined weight-to-age percentile (ranges): WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
A_WTAG_OLD	Weight-to-age percentile (ranges): NCHS standards	Derived
A_HTAG	Combined height-to-age percentile (ranges): WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
A_HTAG_OLD	Height-to-age percentile (ranges): NCHS standards	Derived
A_WTHT	Combined weight-to-height percentile (ranges): WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
A_WTHT_OLD	Weight-to-height percentile (ranges): NCHS standards	Derived
P_WTAG	Weight-to-age percentile: WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
P_WTAG_OLD	Weight-to-age percentile: NCHS standards	Derived
P_HTAG	Height-to-age percentile: WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	
P_HTAG_OLD	Height-to-age percentile: NCHS standards	Derived
P_WTHT	Weight-to-height percentile: WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
P_WTHT_OLD	Weight-to-height percentile: NCHS standards	Derived
BMI	BMI: WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived

Variable Name	Description	Variable Type
BMI_OLD	Body mass index (BMI): NCHS standards	Derived
BMIPCT	BMI percentile: WHO standards for infants and children (0–23 months) and CDC standards for children (24–60 months)	Derived
BMIPCT_OLD	BMI percentile: NCHS standards	Derived
CBRSTFED	Currently breastfeeding indicator	MDS
EBRSTFED	Ever breastfed indicator	MDS
BRSTFED	Length of time breastfed	MDS
AGE_BF_FLAG	Breastfeeding age flag	Derived
BFED	Breastfeeding status	Derived
BFDATE	Date of breastfeeding measurements	MDS
DURWEEKS	Breastfeeding duration, in weeks	Derived
FPACK1-FPACK14	Food package code(s)	MDS
ITEM1-ITEM14	Food item code(s)	MDS
QTY1-QTY14	Food item quantity	MDS
FP_TYPE	Food package type	MDS
SCRIPTID	Food prescription identifier	Created
FDATE	Date of first WIC certification	SDS
EDUC	Years of education	SDS
NUMONWIC	Number in household on WIC	SDS
PDATE	Date previous pregnancy ended	SDS
TOTLPREG	Total number of pregnancies	SDS
TPREGLIV	Total number of live births	SDS
WTHTAGE	Infant/child age in months at time of weight and height measurement	Derived
PREGWHT	Prepregnancy weight in pounds	SDS
GAINWGHT	Weight gain during pregnancy in pounds	SDS
BIRTHWGT	Birth weight in pounds	SDS
BIRTHHGT	Birth length in inches	SDS
FDPIR	Participation in FDPIR	SDS
n	Weighting variable	Created
TARGET_GROUP	Target group for sampling	Derived
SELECT	Percentile for sampling	Derived