

Development of a WIC Participant and Program Characteristics Longitudinal Data Set

The goal of this study was to pilot creating a Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) participant and program characteristics (PC) longitudinal data set with one WIC state agency. This report describes the process of working with one WIC state agency to create a pilot WIC PC longitudinal data set of infant and child participants and the challenges and successes of this effort.

Key Findings

Why

How

Seven of eight WIC state agencies surveyed had “ideal” or “sufficient” capabilities to provide a longitudinal WIC PC data set.

Three had “ideal” capabilities and only one scored “insufficient” for any of the necessary capabilities (Figure 1; see Table 1 for definitions). The ability to provide longitudinal WIC PC data was based on whether a state agency included data elements that could be used to link records across data sets (e.g., participant identification (ID), last and first name), the type and amount of retrospective data available, availability of key WIC PC variables, and whether participant IDs are retained when participants re-enroll.

Table 1. Summary of management information system (MIS) criteria for state agency selection.

MIS Criteria	MIS Capabilities: Ideal	MIS Capabilities: Sufficient	MIS Capabilities: Insufficient
Data elements that could be used to link records across data sets	Consistent participant identification (ID), household ID, and first and last name	Consistent participant ID only	No data elements that could be used to link records across data sets
Type of available data	All instances of updates and changes to each participant's record	All certification and recertification visits but no records from other visit types	Inconsistent or uncertain frequency of records across caseload
Years of available data	5 years of retrospective data	Between 4 and 5 years of retrospective data and capability to provide periodic prospective submissions	Fewer than 5 years of data
Scope of available data	All infants and children enrolled at any time during previous 5 years with an indicator for currently enrolled	All infants and children enrolled at any time during the previous 5 years without an indicator for currently enrolled or records for currently enrolled infants and children only	The state agency can provide records for only a subset of enrolled infants and children
Available variables	Five key supplemental variables (i.e., date of first WIC certification, education level of parent, number in household on WIC, birth weight, and birth length) available for all available years of data	Fewer than five key supplemental variables (i.e., date of first WIC certification, education level of parent, number in household on WIC, birth weight, and birth length) available for all available years of data	N/A

ID persistence	ID retained for infants and children who re-enroll in WIC after a period of nonparticipation	New ID assigned for infants and children who re-enroll in WIC	N/A
----------------	--	---	-----

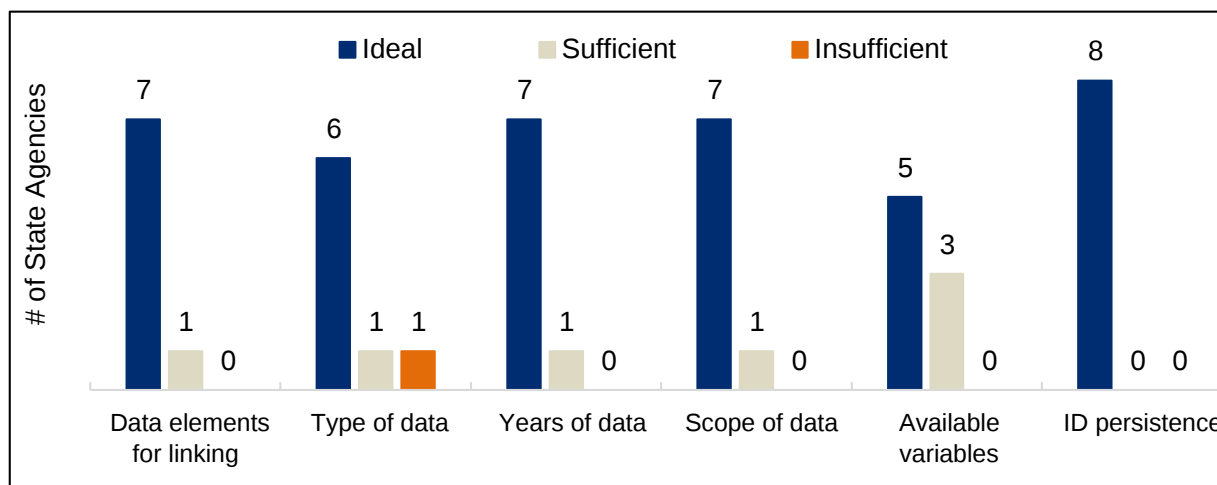


Figure 1. State agency management information systems (MIS) data provision capabilities survey results, N=8 state agencies.

Matching techniques correctly linked nearly all (99%) WIC participant records over time.

- **Deterministic matching** uses participant ID to link records belonging to the same person. It **cannot** link records of participants with more than one ID. Nearly all (99.9%) infants and children had only one participant ID and were able to be linked by participant ID (i.e., through deterministic matching). As with all state agencies, the pilot state agency had processes in place to minimize the number of participants with multiple participant IDs over time in their management information systems (MIS). However, probabilistic matching is still helpful because participants could receive more than one participant ID over time due to data entry mistakes or imperfect data cleaning processes. Although women were not included in this pilot, we expect that probabilistic matching is especially important for women as they are more likely to leave WIC and return if they have multiple pregnancies, and thus may be more likely to be assigned multiple participant IDs.
- **Probabilistic matching** uses other variables to link records likely to belong to the same person. It **can** link records of participants with more than one ID. To inform decisions about

whether records with different IDs belonged to the same participant, the researchers used variables that should not change over time and are available in all state agencies (i.e., first name, last name, date of birth, sex, race, and ethnicity). The researchers then developed a similarity score that used those variables to quantify the similarity between records. During testing, the probabilistic matching similarity score correctly matched records longitudinally for 99.97% of tested cases.

- The matching procedure for this study resulted in a high-quality longitudinal WIC PC data set containing records for all infants and children over a six year period for one state agency. This study suggests that similar results could be achieved with other state agencies.

The pilot longitudinal WIC PC data set was successfully used to determine retention and anemia resolution among WIC participants.

- Analysis from the WIC PC longitudinal data set showed that among infants and children, 85% first enrolled before the age of one and 56% were last certified for WIC benefits before turning 3 years old (Figure 2). Children are eligible for WIC until age 5, yet the pilot data analysis found that most participants drop out by age 3. Future analysis using a longitudinal WIC PC data set could determine factors associated with later enrollment in WIC and dropping out of WIC early.

Age at First Certification

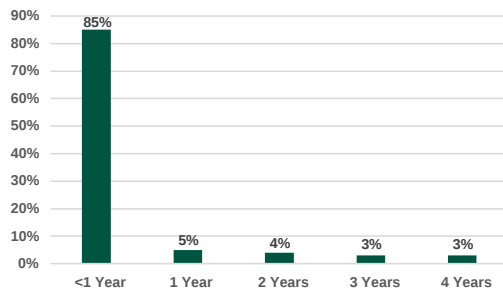


Figure 2a. Age at first certification among infants/children born in 2014.

Age at Last Certification

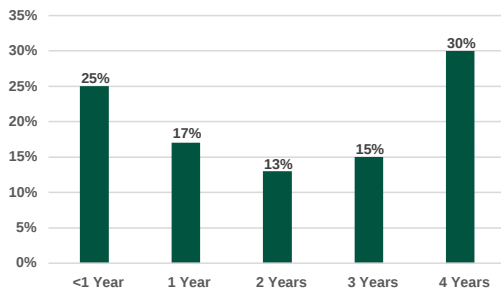


Figure 2b. Age at last certification among infants/children born in 2014.

- The longitudinal WIC PC data were used to determine that 21% of infants and children had a normal hemoglobin level within 12 months of first being identified as anemic. Although further analysis on resolution of anemia status was not conducted for this study, future analyses could allow for a better understanding of which WIC participants are most likely to experience poor health outcomes such as anemia or unhealthy weight.

Why FNS Did This Study

At the time of this study, we had no participant-level longitudinal data sets containing WIC PC data. Since 1992, we have collected WIC PC data every two years. The WIC PC data are a census of WIC participants and describe participant-level information on demographics, income, nutritional risks, anthropometrics, hematology, breastfeeding status, and food package prescriptions during the month of April for each WIC PC year.

Currently, WIC PC data can only be used to identify population trends in the program over time. Because the WIC PC data lack identifiers to link participants over time, individual-level longitudinal analyses such as participant retention cannot be measured. This report describes the process of working with one WIC state agency to create a pilot WIC PC longitudinal data set of infant and child participants and the challenges and successes of this effort.

How FNS Did This Study

Eight state agencies with varying MIS platforms and high quality WIC PC 2020 data completed a survey about their MIS and longitudinal data provision capabilities. The survey assessed these state agencies' MIS capabilities, as measured in six areas, to provide a longitudinal data set.

Next, we chose one state agency and worked with state agency staff and their MIS contractor to extract longitudinal data that included one record per infant or child per week from January 2014 to December 2019.

Once the data were collected, we used two matching approaches to link the records for each infant or child over time. First, we used deterministic matching where the participant ID provided by the state agency was used to link records belonging to the same infant/child. Second, for records that had more than one participant ID, we used probabilistic matching where other identifying variables (e.g., date of birth, first name, and last name) were used to match records that likely belonged to the same participant. Each potential probabilistic match was rated with a similarity score. We reviewed the probabilistic matches and compared their similarity scores to a predetermined threshold to determine whether matches were accepted.

Finally, we conducted example analyses to demonstrate the type of information that could be gained from longitudinal WIC PC data.

Next Steps

- We have taken lessons learned from this pilot (see [Key Findings](#)) to inform the first national collection of longitudinal WIC PC data as part of the 2024 WIC PC data collection. These findings helped shape guidance and technical assistance for state agencies.
- The WIC PC 2024 collection built off recommendations from the pilot to include women in the longitudinal data set, and to link household members via household ID. We will use these data for new analyses to better understand retention and changes in other key measures over time.

Suggested Citation

Beckerman-Hsu, J., Huret, N., & Zvavitch, P. (2025). Development of a WIC Participant and Program Characteristics Longitudinal Data Set. Prepared by Insight Policy Research, Inc., Contract No. GS-10F-0136X, Order No. 12319820F0078. Alexandria, VA: U.S. Department of Agriculture, Food and Nutrition Service, Project Officer: Amanda Reat. Available online at: [www.fns.usda.gov/research/wic/pc-longitudinal-dataset](https://fns.usda.gov/research/wic/pc-longitudinal-dataset).

###

USDA is an equal opportunity provider, employer, and lender.