APPENDIX W
Base Study Sampling Methods

Respondent Universe

The target population for the Base Study was the set of WIC participants ages 0 to 3 months, and the sampling unit was the infant. The Base Study used a longitudinal design, asking the mother or primary caregiver questions (via survey) about their infant multiple times between the infant's birth and second birthday, and collecting health data from WIC administrative records, hospitals, and healthcare providers. In order to gather data starting at birth, we recruited WIC participants for the study at WIC sites during their WIC enrollment appointment. The respondents were either pregnant or enrolling an infant less than 3 months old. The target sample sizes were based on the sample needed to support minimum detectable differences (MDD) between subgroup estimates for infants at 24 months. (See below for further discussion of these calculations.) Since some of the rarer subgroups (e.g., African-American women who are breastfeeding) required more extensive screening to identify the target number of participants, a supplemental sample (above and beyond the "core" sample needed to support most subgroup estimates) was selected. Table B2.3 shows our sampling estimates which are described here. Based on the MDD calculations, the target total number of completed 24-month interviews was 2,758. Assuming response rates for the 24-month interview of 70 percent and 68 percent for the core and supplemental samples, respectively, the target size of the consented and enrolled cohort was 3,991 (2,805 consented core sample enrollees and 1,186 consented supplemental sample enrollees). Factoring in expected consent rates (85% core and 90% supplemental prenatal enrollees), live birth rate (87%), core sample eligibility rates

(98 % prenatal and 80% postnatal enrollees who met screening criteria), and supplemental sample eligibility and subsampling rates (30% prenatal enrollees; 66% postnatal enrollees),[1] the total target number of sampled WIC enrollees was 7,840.

The WIC enrollees were sampled from a stratified, nationally representative sample of 80 WIC sites in 27 State Agencies. In addition to facilitating access to and creating efficient sampling frames for recruiting WIC participants, WIC program representatives in the WIC State and Local Agencies provided important information to the study. Table W.1 presents the estimated population size and the expected number of respondents to have been contacted to provide data for each respondent type. We estimated that there were 2.19 million WIC participants aged 0 to 3 months,[2] and that the final sample size would be 7,840 WIC enrollees. Further, we estimatef the population of respondents for the WIC State and Local Key Informant Interviews to be 12,180, which represents one WIC director and one nutrition coordinator at each of the 90 State Agencies and one local WIC administrator (the most knowledgeable person) at 12,000 services sites;[3] accordingly, we expected a sample size of 107 WIC State and Local administrators (27 WIC State and 80 local administrators) for these interviews. The population of WIC site staff who could complete a Local Staff Online Survey was 36,000 (based on WIC sites having an average of 1.2 staff per 300 WIC participants), and the

1 The eligibility and subsampling rates are combined into one rate: 30% prenatal = 98% eligibility x 30% subsampling ; 66% postnatal = 80% eligibility x 82% subsampling; the subsampling rate are predetermined to target the rarer subgroups to meet precision requirements for estimates for these groups

2 Estimate based on a total of 2.37 million infant participants, 92.4 percent of whom were enrolled by 3 months of age. Source: U.S. Department of Agriculture, Food and Nutrition Service, Office of Research and Analysis, WIC Participant and Program Characteristics 2010, WIC-1-PC, by Patty Connor, Susan Bartlett, Michele Mendelson, Kelly Lawrence, Katherine Wen, et al. Project Officer, Fred Lesnett Alexandria, VA: December 2011.

3 National WIC Association http://www.nwica.org/?q=nwa/1

expected sample size was 800 (10 staff per 80 sites). Finally, the population of data managers was 4.38 million, which reflects the finite number of hospitals, health care providers, and State Agencies that would need to be contacted to obtain data on the population of WIC participants aged 0 to 3 months.[4] The expected sample size was 4,537 data managers, which represents 3,991 hospitals of consented and enrolled participants with live births, 519 health care providers for enrollees for whom hospital records are not available (13% of consented/enrolled participants), and 27 State WIC Agencies.

Table W.1. Estimated population and sample sizes

Respondents	Estimated population size	Expected sample size
WIC Participants Age 0-3 mo. at enrollment (Participant Interviews)	2.19 million	7,840
State and Local WIC administrators (Key Informant Interview)	12,180	107
WIC Site Staff (Local Staff Online Survey)	36,000	800
Data Managers (Health data)	4.38 million	4,537

Sampling Methods

The study used probability sampling methods to select the WIC site sample and the WIC participant sample. We sampled the lowest WIC unit that delivers services to WIC participants, called a "service site". Within each service site we sampled new WIC enrollees within a pre-determined recruitment window.

- **Sampling WIC Service Sites.** As shown in Figure W-1, we used a two-stage sampling approach that used the WIC 2010 Participant Characteristics data (WIC PC 2010) to develop the WIC site sampling frame and a stratified sample design to select the sample of sites. In the first-stage we used a

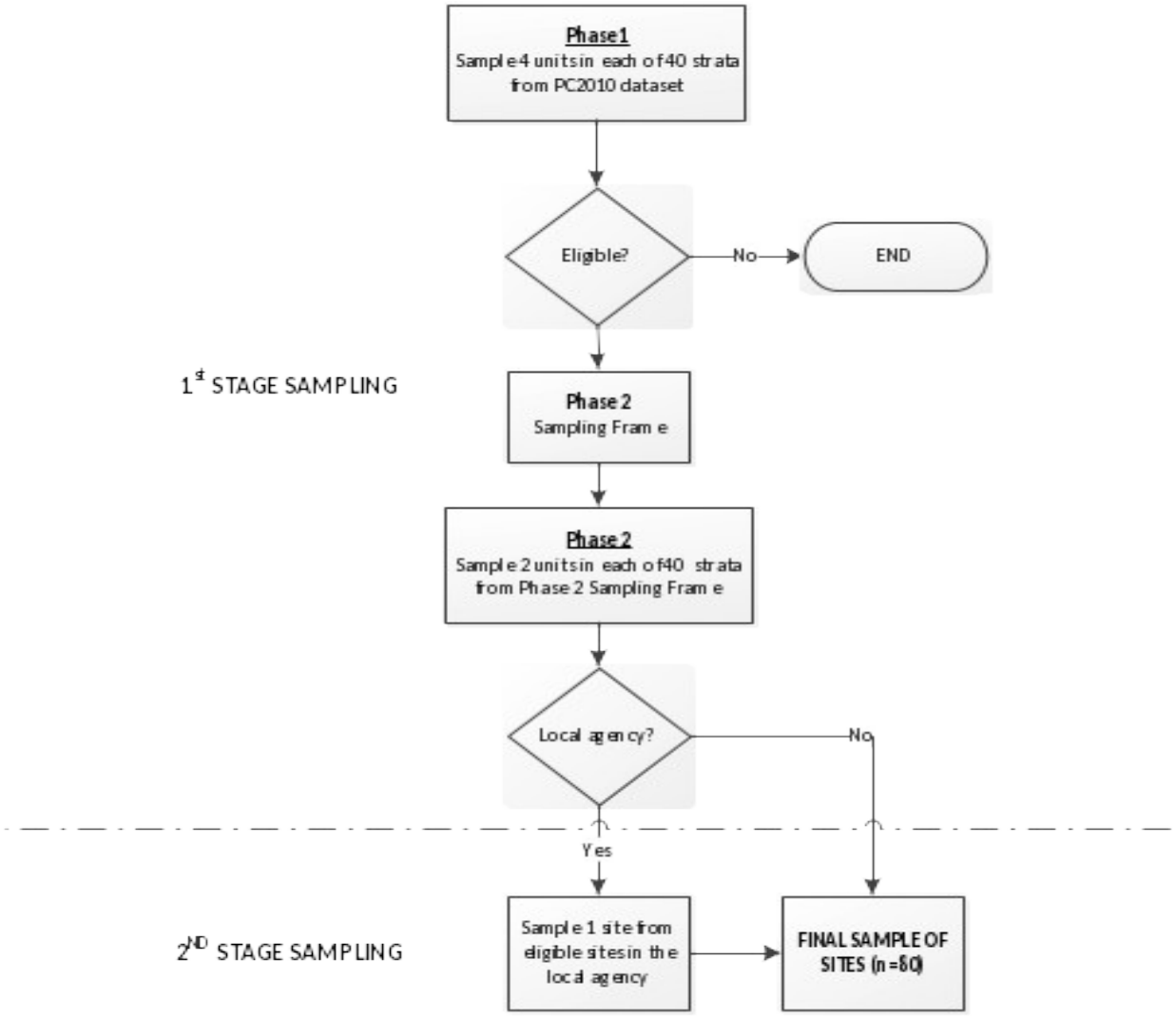
4 2.19 million hospitals, 2.19 million healthcare providers, and 90 State Agencies

group of characteristics to stratify the WIC sites into 40 strata; details of the formation of the 40 strata are given below. Because of uncertainties about the eligibility of the first-stage sampling units, these units were selected in two phases. In the first phase an expected total of 160 sampling units in 42 State Agencies were selected—4 from each of the 40 strata. After the phase 1 selection, we listed the service sites associated with each first-stage sampling unit selected and determined the eligibility of each unit. To be eligible for the study, a site must have had an average minimum daily flow of 1.5 new WIC ITFPS-eligible enrollees per day and must have been expected to remain in operation and enrolling new WIC participants during the WIC ITFPS recruitment period. In the second phase we subsampled eligible first-stage sampling units to arrive at the final sample of 80 first-stage sampling units (2 from each of the 40 strata). In first-stage sampling units that are local agencies with more than one eligible service site, a second stage of sampling was conducted to select one service site. The final sample consisted of 80 eligible service sites. Once the second-stage sampling was complete, recruitment efforts began in earnest. Although due diligence was used to recruit service sites, we anticipated that some sites may be unable or unwilling to cooperate. According to plan, such service sites were replaced by members of a matched sample. This replacement of service sites by matched substitutes is similar to imputation.

- **Sampling WIC participants within a sampled recruiting window.** The WIC participant sample was designed such that the total target number of sampled WIC enrollees (7,840) was spread uniformly across the 80 sampled

sites; that is, the recruitment of study participants was designed so that each site was expected to yield 98 sampled WIC enrollees. An important part of our sampling plan was the concept of recruiting "windows." A recruiting window was a string of consecutive workdays during which we recruited new WIC enrollees at each sampled service site. These windows were expected to vary in length from 7 to 66 workdays. The length of the window was pre-determined, based on typical daily enrollment volumes (obtained from the State following selection of the phase 1 sample of first-stage sampling units) and was calculated in such a way as to yield an expected average of 98 sampled WIC enrollees per site. Early in the site recruitment process, the WIC service site was informed of the length of the recruiting window. The 80 windows were randomly assigned to a spread of starting dates across the 22-week field period for recruiting, with the pool of possible starting dates for a given site determined based on its average daily enrollments and enrollment schedule. Due to the variations in actual WIC enrollments over time, the actual number of sampled enrollees who enrolled in WIC during the specified recruiting window was expected to vary from site to site. Among those who enrolled at each service site during the site's recruiting window, two samples were selected, a core longitudinal and supplemental cross-sectional sample.

Figure W-1. WIC site sampling process



Statistical methodology for stratification and sample selection

The WIC sites were sampled using a stratified probability proportionate to size (PPS) selection procedure. The strata were formed by creating groups of sites that were fairly homogeneous with respect to the following characteristics:

- **Features of the State WIC Agency Plan.** The State WIC Agency plan contents were reviewed to identify a few easily obtained features of the State Agencies' WIC programs, including whether the State Agency:

(1) has a breastfeeding peer counseling program; (2) has trained paraprofessionals to provide nutrition education (vs. requiring that staff who provide nutrition education have professional training or credentials); and (3) provides one can of formula for breastfeeding infants during the first 30 days of life. These features were used to group the WIC State Agency programs into categories.

- **Percent of women who used fully breastfeeding package.** This variable is an estimate of the percentage of women in the first-stage sampling unit who utilized the fully breastfeeding food package during the postpartum period. The PC 2010 data were used to measure food-package selection by first-stage sampling unit, and this rate was computed by taking the ratio of the number of postpartum women who received the fully breastfeeding package during April of 2010 to the total number of postpartum women receiving any food package that same month.
- **Average of children's and mothers' high weight for height rates.** The PC 2010 data were used to estimate the percent of children and the percent of mothers who were "high weight for height"[5] at the first-stage sampling unit level, and these were averaged together to get a measure of risk of being overweight for all participants at the first-stage sampling unit level.

5For children (12 months or older), "high weight for height" is determined based on nutrition risk code 110. For children 24 months and older, it is defined as higher than the 95th percentile of BMI for age. For children 12 to 24 months, it is defined as at risk of being overweight by virtue of having a mother or father who is obese (BMI of 30 or greater). For mothers, the criterion is a pregravid BMI of 25 or higher.

WIC sites were sampled with probabilities proportional to a measure of size (MOS). The MOS was the expected number of eligible enrollees, based on the April 2010 enrollment counts from the WIC PC 2010. The MOS was calculated by summing the total prenatal enrollment and 20 percent of the total enrollment of infants less than 3 months.[6]

Degree of Accuracy Needed for the Purpose Described in the Justification

The sample size requirements for the WIC ITFPS-2 were determined based on power projections and precision requirements. Our primary source of information for these analyses was the baseline interview of ECLS-B (Early Childhood Longitudinal Study – Birth Cohort). We used this survey both to estimate the sizes of key subgroups and to project likely intra-class correlation (ICC). In projecting sample sizes, we focused on the following key outcomes: breastfeeding initiation, breastfeeding at 6 months (with no sub-setting on initiation), and the introduction of solid foods before the age of six months.

The precision requirement was that for key national estimates for the full "currently on WIC" group, a 95% confidence interval should have a half-width of no more than 5 percentage points. Additionally, for subgroup estimates (for key subgroups), a 90% confidence interval should have a half-width of no more than 5 percentage points. The sample should also support detection of minimum detectable differences (MDDs) among the categories of each of the key subgroups

6 The 20 percent figure is based on an estimate from the Early Childhood Longitudinal Study-Birth (ECLS-B) Cohort that 20 percent of infants enrolled in WIC were not enrolled prenatally

with power of at least 0.80 and a significance level of 0.05. The core sample size was determined by the need to meet the precision target on the breastfeeding initiation rate for African-American mothers. The supplemental sample size was driven by the need to provide the same precision on the comparable statistic restricted to African-American mothers who keep their children on WIC for 24 months. Other statistics for the population that keep their children on WIC for 24 months were also at the desired precision limit. Table B2.1 shows minimum detectable differences (MDDs) between subgroups of interest for three critical outcomes using a test size of 0.05 and power of 0.80. They ranged from 5 to 10 percentage points. Based on subgroup differences observed in ECLS-B, it appeared reasonable to expect differences of this magnitude for some but not all of the comparisons. Note that MDDs for upward and downward changes were slightly different. The numbers shown in this table are the average of the MDDs and upward and downward change. These figures used the expected total of the core and supplemental samples.

Table B2.2 shows minimum detectable differences (MDDs) in child obesity and overweight status by timing of maternal WIC enrollment, controlling for maternal weight status. We assumed that controlling for maternal weight status in these analyses would reduce variances by 20 percent. These projections applied to either age 12 months or 24 months. Power to detect these effects was 0.80.

Table W.2. Minimum detectable differences between subgroups of interest (based on both the core longitudinal and supplemental cross-sectional samples)

Comparison	Percent initiating breastfeeding	Percent breastfeeding at six months	Percent introduced solid food prior to 6 months
African-American vs. white	9.7	5.8	8.4
Other vs. white	7.6	7.0	7.8

Hispanic vs. Non-Hispanic	6.0	5.9	6.4
Breastfed 1-3 months vs. never	na	na	7.2
Breastfed 4+ months vs. never	na	5.9	7.8
1st trimester enrollment vs. postnatal	8.6	6.4	7.5
2nd or 3rd trimester vs. postnatal	9.0	6.8	6.7
Mom overweight vs. normal or low	8.2	6.8	7.3
Mom obese vs. normal or low	8.1	5.1	7.0
Under 75% poverty vs. 76 to 129%	8.7	5.8	7.9
Over 130% poverty vs. 76 to 129%	7.5	5.9	7.6

Table W.3. Minimum detectable differences for child obesity and overweight status by timing of WIC enrollment – controlled for maternal weight status – valid at both 12 and 24 months

Timing of maternal WIC enrollment	MDD in percent obese	MDD in percent overweight	MDD in percent obese or overweight
1st trimester enrollment vs. postnatal	3.2	4.4	5.2
2nd or 3rd trimester vs. postnatal	3.3	4.5	5.4

Table W.4 shows our projected sample sizes and response rates at various recruitment and interviewing stages over time. The overall sample size required to obtain the MDDs shown in Tables W.2 and W.3 was a total of 2,758 respondents to the 24-month interview. In order to attain this expected sample size, this target was adjusted to account for assumptions about attrition rates, consent rates, live birth rates, eligibility rates, and subsampling rates for the supplemental sample. These assumptions are shown in the “Rates” columns of Table W.4. For all interviews except the prenatal and 3-month supplemental, the rates shown were computed as the number of completed interviews divided by the cohort size (2,805 for the core sample and 1,186 for the supplemental sample). In general, these rates accounted only for nonresponse to the particular interview (including

attrition). The 1-month and 3-month interviews were the exception; for those interviews, these rates accounted for the fact that only a portion of the sample will have enrolled in WIC in time to be eligible for the 1-month interview. For the supplemental sample, the only enrollees administered the 3-month interview were those who were not enrolled in time for the 1-month interview. The distribution of cases to prenatal sampling versus postnatal sampling was according to the timing of their WIC enrollment (prenatal vs. postnatal).

Table W.4. Expected Sample Sizes and Response Rates

	Mother infant events/interviews	Core sample	Rates	Supplemental sample	Rates	Sub-Total	2nd 24HR	Grand Total
Prenatal sampling	Prenatal WIC Enrollees Sampled	3,097		3,097		6,194		
	Met Screening Criteria	3,035	98%	929	30%	3,964		
	Consented & Enrolled	2,580	85%	836	90%	3,416		
	Live birth	2,245	87%	727	87%	2,972		
Postnatal sampling	Infant WIC Enrollees Sampled	823		823		1,646		
	Met Screening Criteria	659	80%	540	66%	1,199		
	Consented & Enrolled	560	85%	459	85%	1,019		
Prenatal & Postnatal	**Total Sampled**					**7,840**		
	Total Screened					**5,163**		
	Total Consented/Enrolled					**4,435**		
Cohort	Total live infants consented & enrolled pre/post-natal	**2,805**		**1,186**		**3,991**		
Follow-up interviews	Prenatal[a]	2,193	85%			2,193		2,193
	1-Month	2,354	84%	985	83%	3,339		3,339
	3-Month Supplemental[b]			70	15%	70		70
	3-Month Core	2,344	84%			2,344		2,344
	5-Month	2,297	82%			2,297		2,297
	7-Month	2,251	80%	970	82%	3,221		3,221
	9-Month	2,206	79%			2,206		2,206
	11-Month	2,162	77%			2,162		2,162
	13-Month	2,119	76%	883	74%	3,002	212	3,214
	15-Month	2,076	74%			2,076	208	2,284
	18-Month	2,035	73%			2,035	203	2,238
	24-Month	1,955	70%	803	68%	2,758	195	2,953
	Total interviews[c]	***23,992***		***3,711***		***27,703***	***818***	***28,521***

[a] 85% Response rate =2,193/2,580 (Core prenatal sampling consented and enrolled is the only group eligible for prenatal interview)
[b] 15% Response rate = 70/459 (Only those who don’t enroll in time to make the 1-month interview window [a subgroup of supplemental postnatal sampling consented and enrolled] are eligible for 3-month supplemental interview)
[c] Total interviews = Prenatal through 24-months