National Agricultural Workers Survey (NAWS)

Summary of Nonresponse and Design Studies 2020–2021

Submitted: November 12, 2021

Submitted to:
Daniel Carroll
U.S. Department of Labor
Employment and Training Administration
Office of Policy Development and Research
Room N-5637
200 Constitution Ave., N.W.
Washington, D.C. 2021

Submitted by:
Susan Gabbard
JBS International, Inc.
155 Bovet Road Suite 210
San Mateo, CA 94402

This report was prepared for the U.S. Department of Labor, Employment and Training Administration, Office of Policy Development and Research by JBS International, Inc., under contract 47QRAA18D00AE; Task Contract 1630DC-19-F-00019. Since contractors conducting research and evaluation projects under government sponsorship are encouraged to express their own judgment freely, this report does not necessarily represent official opinion or policy of the U.S. Department of Labor.

*Value of **thought.***
*Value of **solution.***

JBS INTERNATIONAL
A CELERIAN GROUP COMPANY

Table of Contents

Introduction

JBS has undertaken a series of studies to examine possible nonresponse bias and to assess the efficiency of the study's design, including three nonresponse studies that assessed item and employer nonresponse rates and the characteristics of agricultural workers from employers who always or sometimes respond to National Agricultural Workers Survey (NAWS) interview requests, a design study that examined how interview allocations would change if they were optimized for statistical efficiency and/or cost reduction, and two design studies that examined how the late start of the fiscal year (FY) 2019 spring cycle and COVID-19 in FY 2020 affected design effects. This document summarizes each of these studies.

The nonresponse studies included:

- Nonresponse Study 1 – NAWS Item Nonresponse Rates
- Nonresponse Study 2 – NAWS Unit (Employer) Nonresponse Rates
- Nonresponse Study 3 – Measuring Differences Between Responding and Non-Responding Grower Populations

The design studies included:

- Design Study 1 – Optimal Interview Allocations for NAWS Sampling Using Underlying National Farmworker Jobs Program (NFJP) Variables
- Design Study 2 – Design Effect Study for Partial Missing Cycles (FY 2019 Spring Cycle)
- Design Study 3 – Design Effect Study for Partial Missing Cycles (FY 2020 COVID-19)

Nonresponse Study 1 – NAWS Item Nonresponse Rates

The first nonresponse study examined nonresponse for questionnaire items (i.e., item nonresponse). Item nonresponse is a measure of data quality representing the percent of survey respondents who did not provide a valid response to a questionnaire item.

Analysis

Item nonresponse was examined for 114 items common to the 2017–2020 NAWS questionnaires. To ensure quality analysis, a consistent set of selection criteria was used when establishing which items to include, beginning with those available within the public access dataset associated with the fiscal years of interest (i.e., discontinued items were excluded), and items that are directly from the questionnaire (i.e., created variables were excluded). These items covered all sections answered by the respondents, except for items in the household grid pertaining to dependents and the work grid. Of the 77 items asked of all respondents, the denominator of the nonresponse rates was the annual number of interviews. For the 37 items asked only if certain criteria were fulfilled (i.e., having a skip pattern), the denominator was the number of respondents who met the criteria for being asked the question. For both kinds of items, the number of valid responses was the numerator. Items with less than 10 eligible respondents were excluded from the analysis.

Results

Overall, the NAWS items showed very low item nonresponse, with most items exceeding 95 percent valid answers and a few items having 85–93 percent valid responses. For items with less

than 70 percent valid responses, OMB requires additional analysis of item nonresponse. No additional analysis was undertaken since all items exceeded the OMB criteria of 70 percent.

For the 77 items asked of all respondents across fiscal years 2017 to 2020 (Table 1), the average nonresponse rate was less than 0.5 percent for each fiscal year, and 0.4 percent across all four fiscal years. Certain items had higher nonresponse rates than others. For example, the Item 18 (Years non-farm work in U.S.) had a nonresponse rate of up to 7.3 percent.

For the 37 items with skip patterns (Table 2), across the years, the average nonresponse rate was less than 2.0 percent for each fiscal year, and 1.4 percent across all four fiscal years. Item 35 (Work transport is mandatory) had the highest annual nonresponse rate of up to 15.5 percent.

Table 1. Item Nonresponse for Items without Skip Patterns, by Fiscal Year

	Fiscal Year			
Item	**2017**	**2018**	**2019**	**2020**
Item 1 - Relation to farmworker in household	0.1%	0.0%	0.0%	0.0%
Item 2 - Gender	0.1%	0.0%	0.0%	0.0%
Item 3 - Married?	0.2%	0.1%	0.1%	0.3%
Item 4 - Country of birth	0.1%	0.0%	0.0%	0.0%
Item 5 - Highest grade completed	0.1%	0.1%	0.1%	0.0%
Item 6 - Country of schooling	0.2%	0.3%	0.0%	0.4%
Item 7 - Schooling in U.S. last 12 months	0.3%	0.7%	0.5%	0.3%
Item 8 - Number of people in household not in Grid	0.0%	0.1%	0.0%	0.0%
Item 9 - Farmworker has Health Insurance?	0.1%	0.1%	0.0%	0.2%
Item 10 - Farmworker spouse has health insurance?	0.4%	0.4%	0.5%	0.3%
Item 11 - Farmworker children have health Insurance?	1.0%	0.5%	0.9%	0.8%
Item 12 - Hispanic	0.3%	0.1%	0.1%	0.0%
Item 13 - Race	0.1%	0.1%	0.2%	0.0%
Item 14 - Speak English	0.2%	0.3%	0.2%	0.0%
Item 15 - Read English	0.2%	0.4%	0.3%	0.0%
Item 16 - First did farm work in U.S.	0.5%	0.1%	0.2%	0.1%
Item 17 - Years in farm work in U.S.	1.2%	0.4%	0.2%	0.1%
Item 18 - Years non-farm work in U.S.	7.3%	6.0%	5.0%	4.0%
Item 19 - Parents did farm work in U.S.?	1.9%	0.1%	0.1%	0.0%
Item 20 - Languages spoken in childhood home	0.0%	0.0%	0.0%	0.0%
Item 21 - Speak English as adult	0.1%	0.4%	0.2%	0.0%
Item 22 - Speak Spanish as adult	0.1%	0.0%	0.0%	0.0%
Item 23 - Speak Creole as adult	0.0%	0.0%	0.0%	0.0%
Item 24 - Speak Mixtec as adult	0.0%	0.0%	0.0%	0.0%
Item 25 - Speak Kanjobal as adult	0.0%	0.0%	0.0%	0.0%
Item 26 - Speak Zapotec as adult	0.0%	0.0%	0.0%	0.0%
Item 27 - Speak 'Other' language as adult	0.0%	0.0%	0.1%	0.0%
Item 28 - English is most dominant	---	0.7%	0.5%	0.0%
Item 29 - Spanish is most dominant	---	0.4%	0.0%	0.6%

Item	Fiscal Year 2017	2018	2019	2020
Item 30 - Creole is most dominant	---	0.0%	0.0%	0.0%
Item 31 - Mixtec is most dominant	---	0.0%	0.0%	0.1%
Item 32 - Kanjobal is most dominant	---	0.0%	0.0%	0.0%
Item 33 - Zapotec is most dominant	---	0.0%	0.0%	0.0%
Item 34 – 'Other' language is most dominant	---	0.0%	0.0%	0.1%
Item 35 - Father born (country)	0.1%	1.5%	0.1%	0.3%
Item 36 - Father born (state or equivalent)	0.6%	1.5%	0.7%	0.2%
Item 37 - Mother born (country)	0.1%	1.6%	0.2%	0.3%
Item 38 - Mother born (state or equivalent)	0.5%	1.3%	0.6%	0.3%
Item 39 - Months of farm work in year before last	0.3%	0.0%	0.9%	0.0%
Item 40 - Hours of farm work last week	0.7%	0.1%	0.2%	0.0%
Item 41 - Pay day wages after taxes	1.7%	1.0%	0.9%	1.2%
Item 42 - Pay day wages before taxes	1.1%	0.8%	1.0%	0.4%
Item 43 - Payment period	0.2%	0.1%	0.1%	0.1%
Item 44 - Hours worked in pay period	1.0%	0.7%	0.4%	0.0%
Item 45 - Method paid	0.0%	0.0%	0.4%	0.0%
Item 46 - Employer provides health insurance if injured/sick from work	0.0%	0.1%	0.0%	0.1%
Item 47 - Workers' comp?	0.1%	0.1%	0.0%	0.0%
Item 48 - If insured or sick outside of work, does employer provide health insurance?	0.1%	0.1%	0.0%	0.1%
Item 49 - Unemployment insurance	0.1%	0.3%	0.0%	0.0%
Item 50 - Years w/ current Employer	0.3%	4.7%	1.6%	1.3%
Item 51 - How farmworker got job	0.1%	0.0%	0.1%	0.1%
Item 52 - Who pays for living quarters?	0.1%	0.1%	0.2%	0.1%
Item 53 - Type of living quarters	---	0.5%	0.1%	0.4%
Item 54 - Location of living quarters	---	3.0%	1.5%	1.3%
Item 55 - Transportation type	0.2%	0.1%	0.1%	0.1%
Item 56 - Job distance from home	0.2%	0.1%	0.1%	0.0%
Item 57 - Household crowding	0.0%	0.0%	0.0%	0.1%
Item 58 - Number of bedrooms	0.0%	0.1%	0.2%	0.0%
Item 59 - Number of bathrooms	0.0%	0.1%	0.2%	0.0%
Item 60 - Number of kitchens	0.0%	0.1%	0.2%	0.0%
Item 61 - Number of other rooms	0.0%	0.2%	0.2%	0.3%
Item 62 - Pay period payment type	0.4%	0.6%	0.4%	0.5%
Item 63 - Payment receipt?	1.5%	1.4%	2.4%	1.4%
Item 64 - Remain in farm work	0.3%	0.5%	0.2%	0.0%
Item 65 - Could you get non-farm work job within a month?	0.5%	1.3%	2.9%	3.3%
Item 66 - Employed	0.0%	0.0%	0.0%	0.0%
Item 67 - Total personal income in last year: farm work	0.6%	0.7%	0.2%	0.1%
Item 68 - Portion of income from farm work	1.4%	0.9%	0.3%	0.3%
Item 69 - Total family income in last year	1.0%	1.0%	0.3%	0.1%

Item	Fiscal Year 2017	2018	2019	2020
Item 70 - Received benefits or services from social programs	0.8%	0.1%	0.1%	0.0%
Item 71 - Own or buying any assets in U.S.?	0.0%	0.9%	0.3%	0.6%
Item 72 - Legal status	0.1%	0.0%	0.0%	0.4%
Item 73 - Handled pesticides in last year	---	0.0%	0.3%	0.3%
Item 74 - Employer provides clean drinking water	---	0.1%	0.3%	0.3%
Item 75 - Employer provides a toilet everyday	---	0.1%	0.5%	0.2%
Item 76 - Employer provides water to wash hands daily	---	0.1%	0.5%	0.2%
Item 77 - Work type	0.0%	0.0%	0.0%	0.0%
Average for Items 1–77	0.4%	0.48%	0.4%	0.3%
Average for Items 1-77, FY 2017-2020	0.4%			

Table 2. Item Nonresponse for Items with Skip Patterns, by Fiscal Year

Item	Fiscal Year 2017	2018	2019	2020
Item 1 - Year first entered U.S.	0.1%	0.3%	0.0%	0.2%
Item 2 - Month first entered U.S.	1.4%	2.5%	0.6%	0.6%
Item 3 - Work type in home country	1.0%	1.4%	1.4%	1.1%
Item 4 - Country lived in before U.S.	0.0%	0.0%	0.0%	0.0%
Item 5 - State or dept before coming to U.S.	0.0%	0.4%	0.1%	0.1%
Item 6 - Number of adults in household not in grid	0.0%	0.3%	1.0%	1.2%
Item 7 - Number of children in household not in grid	0.0%	0.3%	1.0%	1.2%
Item 8 - Number of people in household of unknown age	0.0%	0.3%	1.0%	1.2%
Item 9 - Number of adults in household doing farm work	0.5%	0.3%	1.2%	1.6%
Item 10 - Number of adults related to farmworker in household	0.2%	0.3%	1.0%	1.4%
Item 11 - Number of children in household doing farm work	1.1%	1.1%	3.8%	10.4%
Item 12 - Number of children related to farmworker in household	1.1%	1.1%	3.8%	10.4%
Item 13 - Number of people of unknown age doing farm work [a]				
Item 14 - Number of people of unknown age related to farmworker [a]				
Item 15 - Who pays for farmworker's health insurance?	0.2%	0.3%	0.7%	0.2%
Item 16 - Who pays for spouse's health insurance?	0.5%	0.2%	0.3%	0.3%
Item 17 - Who pays for children's health insurance?	0.7%	0.3%	0.2%	0.6%
Item 18 - How many children under 18 have health insurance?	0.0%	0.2%	0.0%	0.2%
Item 19 - How many children over 18 have health insurance?	0.0%	0.3%	0.2%	0.2%
Item 20 - How well do you speak Spanish?	0.4%	0.0%	0.3%	0.4%
Item 21 - How well do you read Spanish?	0.6%	0.3%	0.6%	0.5%
Item 22 - How well do you speak Creole? [b]				
Item 23 - How well do you read Creole? [b]				
Item 24 - How well do you speak Mixtec?	1.6%	0.0%	3.1%	0.0%
Item 25 - How well do you read Mixtec?	1.6%	2.9%	3.1%	0.0%
Item 26 - How well do you speak Kanjobal? [b]				
Item 27 - How well do you read Kanjobal? [b]				

Item	Fiscal Year 2017	2018	2019	2020
Item 28 - How well do you speak Zapotec? [b]				
Item 29 - How well do you read Zapotec? [b]				
Item 30 - How well do you speak 'Other'?	4.9%	2.3%	0.0%	0.0%
Item 31 - How well do you read 'Other'?	1.6%	4.5%	2.2%	0.0%
Item 32 - Paid as individual or crew?	1.9%	0.8%	3.3%	1.0%
Item 33 - How much do you pay for housing?	0.1%	0.8%	0.5%	0.6%
Item 34 - Pay transportation fee?	5.8%	0.7%	1.5%	2.1%
Item 35 - Work transport is mandatory	15.5%	7.8%	3.4%	7.5%
Item 36 - Which program used to apply for citizenship?	---	1.1%	2.8%	3.1%
Item 37 - Have general work authorization?	2.4%	2.8%	2.8%	3.1%
Average for Items 1–37	1.5%	1.2%	1.4%	1.7%
Average for Items 1–37, FY 2017-2020	1.4%			

[a] Excluded from analysis due to zero farmworkers indicating that the ages of those who live with them are unknown.
[b] Excluded from analysis due to small number of farmworkers (n < 10) eligible to answer these items.

Nonresponse Study 2 – NAWS Unit (Employer) Nonresponse Rates

The NAWS unit nonresponse (employer) bias is calculated by comparing information in the sampling frame on eligible respondents and nonrespondents. While the sampling data is somewhat sparse for nonrespondents, three pieces of information are useful: source used to obtain employer names, North American Industry Classification System (NAICS) code, and geographic location. This analysis used three sources of employer names: a) the Bureau of Labor Statistics' (BLS) Unemployment Insurance (UI) list, b) a commercial list (Data Axle USA), and c) internet searches and contacts with knowledgeable local individuals. Geographic area and source lists are available for all employers, while NAICS codes are available for employers who pay UI taxes and those in the Data Axle USA list.

Analysis

This study examined three characteristics (source of the employer list, NAICS, and geography) and made the following three comparisons:

A. Employers allowing interviews were compared to sampled employers that refused or were unable to be screened (i.e., excluding ineligible employers).
B. Employers allowing interviews compared to eligible employers that refused.
C. Eligible employers compared to employers whose eligibility could not be determined.

Nonresponse bias was calculated using the bias calculation formula from OMB's *Standard and Guidelines for Statistical Surveys* (2006):

$$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_{nr}}{n}\right)(\bar{y}_r - \bar{y}_{nr})$$

where:

$\bar{y}_t$ = the mean based on all sample cases;

$\bar{y}_r$ = the mean based only on respondent cases;
$\bar{y}_{nr}$ = the mean based only on nonrespondent cases;
n = the number of cases in the sample; and
n_{nr} = the number of nonrespondent cases.

Results

The results show that nonresponse rate for the sources was 86–92, 56–59, and 66–81 percent for comparisons A, B, and C, respectively (Table 3). Furthermore, there was low bias (1–4 percent) across the three comparisons and sources. There were more variations in nonresponse rates for NAICS; 75–91, 52–70, and 48–75 percent (excluding NAICS 112) for comparison A, B, and C, respectively. The bias remained low (0–9%). The nonresponse rate for the six regions was 81–89, 50–63, and 61–72 percent for comparison A, B, and C, respectively (Table 4). The nonresponse rates for the 12 regions were 78–94, 44–74, and 57–75 percent for comparison A, B, and C, respectively. The bias for both region six and 12 were low (0–9%). Tables 3 and 4 show nonresponse rates and bias for source, NAICS, and geography by comparisons A, B, and C.

JBS also conducted regression analysis to determine the association between employer characteristics (source, NAICS, and geography) for the three comparisons. The results show that in comparisons B and C, employers selected from the Other source are significantly less likely to participate compared to those selected from the BLS source. In all three comparisons, employers with NAICS 1114 (Greenhouse, Nursery, and Floriculture Production) had the highest likelihood of participating in NAWS, compared to NAICS 1119 (Other Crop Farming). In all three comparisons, employers in four of the six regions (East, Southeast, Midwest, and Northwest) were significantly more likely to participate in NAWS compared to California (with the exception of the Southeast region in comparison C). For the 12 regions analysis, in comparisons A and C, employers in nine of the regions (AP12, CBNP, DLSE, LK, MN12, MN3, NE1, NE2, and PC) had significantly higher odds of participating compared to California. In comparison B, six of the regions (DLSE, MN12, MN3, NE1, NE2, and PC) had significantly higher odds of participating compared to California.

Overall, the results showed that although unit nonresponse rates were high, there was little nonresponse bias between responding and nonresponding employers overall and across NAICS, sampling regions, and list source.

Table 3. Unit Nonresponse Rate and Bias by Source and NAICS.

	A. Nonresponse among all eligible and unscreened employers			B. Nonresponse rate among eligible employers			C. Eligibility Rate		
Variable	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias[1]**	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias**	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias[1]**
Source									
BLS	86%	4%	3%	59%	-1%	-1%	66%	6%	4%
Data Axle	87%	-1%	-1%	56%	1%	1%	71%	-2%	-2%
Other	92%	-3%	-3%	59%	0%	0%	81%	-4%	-3%
NAICS									
111 or 1151 (vs 112)	86%	0%	0%	58%	0%	0%	67%	0%	0%
1111	90%	-3%	-3%	70%	-5%	-3%	67%	0%	0%
1112	84%	2%	2%	59%	-1%	0%	60%	3%	2%
1113	85%	3%	3%	54%	5%	3%	67%	0%	0%
1114	75%	13%	9%	52%	6%	3%	48%	11%	5%
1119	89%	-7%	-6%	58%	0%	0%	75%	-9%	-6%
1151	91%	-8%	-7%	67%	-6%	-4%	73%	-5%	-4%
112	---[2]	---[2]	---[2]	---[3]	---[3]	---[3]	---[2]	---[2]	---[2]

Comparison A = Employers allowing interviews compared to sampled employers that refused or unable to be screen (i.e., excluding the ineligible). [2] Results are not reportable because there were less than four employers with NAICS 112.

Comparison B = Employers allowing interviews compared to eligible employers who refused. [3] No employers with NAICS 112.

Comparison C = Eligible employers compared to employers whose eligibility could not be determined.

NAICS 1111 = Oilseed and Grain Farming. NAICS 1112 = Vegetable and Melon Farming. NAICS 1113 = Fruit and Tree Nut Farming. NAICS 1114 = Greenhouse, Nursery, and Floriculture Production. NAICS 1119 = Other Crop Farming. NAICS 1151 = Support Activities for Crop Production. NAICS 112 = Cattle Ranching and Farming, Hog and Pig Farming, Poultry and Egg Production, Sheep and Goat Farming, Aquaculture, or Other Animal Production.

[1] Bias = $\left(\frac{n_{nr}}{n}\right)\left(\overline{Y}_r - \overline{Y}_{nr}\right)$

Table 4. Unit Nonresponse Rate and Bias by Geography.

	A. Nonresponse among all eligible and unscreened employers			B. Nonresponse rate among eligible employers			C. Eligibility Rate		
Variable	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias[1]**	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias**	**Nonresponse rate**	**Difference between respondents and nonrespondents**	**Bias[1]**
Region 6									
East	83%	4%	3%	50%	4%	2%	66%	1%	1%
Southeast	86%	1%	1%	53%	2%	1%	69%	-1%	0%
Midwest	85%	1%	1%	61%	-1%	-1%	62%	3%	2%
Southwest	89%	-3%	-2%	62%	-2%	-1%	72%	-2%	-1%
Northwest	81%	6%	5%	51%	5%	2%	61%	4%	2%
California	89%	-10%	-9%	63%	-8%	-5%	71%	-5%	-4%
Region 12									
AP12	87%	0%	0%	53%	1%	1%	72%	-1%	-1%
CA	89%	-10%	-9%	63%	-8%	-5%	71%	-5%	-4%
CBNP	87%	0%	0%	62%	-1%	-1%	65%	1%	1%
DLSE	84%	1%	1%	49%	2%	1%	68%	0%	0%
FL	87%	0%	0%	57%	0%	0%	71%	-1%	0%
LK	82%	2%	1%	58%	0%	0%	57%	2%	1%
MN12	80%	2%	2%	45%	2%	1%	64%	1%	0%
MN3	83%	1%	1%	48%	2%	1%	67%	0%	0%
NE1	78%	2%	1%	44%	2%	1%	60%	1%	1%
NE2	80%	2%	1%	52%	1%	1%	58%	1%	1%
PC	81%	4%	3%	53%	2%	1%	59%	3%	2%
SP	94%	-4%	-4%	74%	-3%	-2%	75%	-2%	-2%

Comparison A = Employers allowing interviews compared to sampled employers that refused or unable to be screen (i.e., excluding the ineligible).
Comparison B = Employers allowing interviews compared to eligible employers who refused.
Comparison C = Eligible employers compared to employers whose eligibility could not be determined.
AP12 = KY, NC, TN, VA, WV. CA = CA only. CBNP = IA, IL, IN, KS, MO, ND, NE, OH, SD. DLSE = AL, AR, GA, LA, MS, SC. FL = FL only. LK = MI, MN, WI. MN12 = CO, ID, MT, NV, UT, WY. MN3 = AZ, NM. NE1 = CT, MA, ME, NH, NY, RI, VT. NE2 = DE, DC, MD, NJ, PA. PC = OR, WA. SP = OK, TX. East = AP12, NE1, NE2. Southeast = DLSE, FL. Midwest = CBNP, LK. Southwest = MN3, SP. Northwest = MN12, PC. California = California only.

[1]Bias = $\left(\frac{n_{nr}}{n}\right)\left(\overline{Y}_r - \overline{Y}_{nr}\right)$

Nonresponse Study 3 – Measuring Differences Between Responding and Non-Responding Grower Populations

This study examined the characteristics of agricultural workers from employers who always or sometimes respond to NAWS interview requests. The purpose was to determine whether there was any significant bias introduced by employer nonresponse.

Analysis

The sample consisted of 22,743 agricultural workers from 4,217 employers in NAWS fiscal years 2006–2017. Chi-square analysis with post-hoc tests were conducted on 225 variables with 1,147 categories on agricultural worker characteristics.

Results

The results showed that 68 percent of the variables of interest did not show statistically significant differences between agricultural employers who always respond and those that only sometimes respond at the traditional $p < .05$ level. Due to the increased statistical error rate when conducting multiple hypothesis tests, the *p*-value should be reduced in order to reduce the chance of committing a Type I error (i.e., the error of declaring population parameters to be significantly different based on the sample, when, in reality, they are not). At the more stringent level, $p < .001$, 87 percent of the variables showed no significant differences between the two groups of employers. Table 5 shows the number of significant variables at each significance level for the continuous and categorical variables. Table 6 shows the 25 categorical and 5 continuous variables that were significant at the most stringent level ($p < .001$).

Table 5. Number of Significant variables, total and by type.

Number of significant variables at each level (of 225 total)	***p* < .05** 72	***p* < .01** 46	***p* < .001** 30
Percentage of variables significant	32.0%	20.4%	13.3%
Significant variables by type	***p* < .05**	***p* < .01**	***p* < .001**
Categorical variables	57	40	25
Percentage of categorical variables	25.3%	17.8%	11.1%
Continuous variables	15	6	5
Percentage of continuous variables	6.7%	2.7%	2.2%

Table 6. Variables Significant at the $p < .001$ level, by type.

Categorical Variables	
A23a3: [For the farmworker who has insurance] Who pays, the employer?	**D33A:** While you are working for the grower/contractor, what type of arrangement do you have for living quarters?
A23a6: [For the farmworker who has insurance] Who pays, other?	**D34a:** In what type of living quarters do you live now (housing structure at this location)?
B01: Which of the following describes you? (Ethnicity)	**D35:** Where are your living quarters located?
B20b: When you were a child, in what languages did adults speak to you at home? Spanish	**D37:** How do you usually get to work?
B21b: And now, as an adult, what languages can you speak? Spanish	**D37A:** How far is your current job from your current residence?
B22b: And now, how well do you speak it? Spanish	**D38:** Do you pay a fee to the grower/contractor "raiteros" for rides to work?
B23b: And now, how well do you read it? Spanish	**DMAREGN:** 12 NAWS sampling regions
B24: In which language do you believe you are most dominant (comfortable) conversing?	**E04:** Could you get a U.S.A. non-farm job within a month?
B26a: Where was your father born? In what Country?	**G07B:** Do you own or are you buying the following item in your home country? A house
B27a: Where was your mother born? In what Country?	**Migrant:** Farmworker is a migrant
Crop at time of interview	**NQ05:** The last time you got attention from a healthcare provider who paid the majority of the cost?
D23: If you are injured at work or get sick as a result of your work, do you get any payment while you are recuperating (i.e., worker's compensation)?	**Region6:** Region of interview in 6 codes
D28: Do you work for this employer on a seasonal basis or year-round?	
Continuous Variables	
D52: How many people total sleep in these rooms?	
FWRDAYS: Farm work days	
NFWEEKS: Non-farm work weeks	
C09WEEKS: Non-work weeks	
FWWEEKS: Farm work weeks	

Additional analyses were conducted among the categorical variables to examine the differences between categories. There were 201 categories among the 57 categorical variables that had significant differences between agricultural employers who always respond and those that only sometimes respond. Of the 201 categories, 42 (21%) categories did not show statistically significant differences at the $p < .05$ level. At $p < .001$, 120 (60%) categories were not significant.

In terms of practical significance, 18 (9%) of the 201 categories had a difference of 5 percent or more, 49 (24%) categories had a difference between 3 and 5 percent, and 134 (67%) categories had less than 3 percent difference.

These results can be taken as evidence that there are minimal differences between agricultural workers from responding and nonresponding employers.

Design Study 1 – Optimal Interview Allocations for NAWS Sampling Using Underlying NFJP Variables

The purpose of this study was to see how interview allocations would change if they were optimized for statistical efficiency and/or cost reduction. The current interview allocation is proportional to the distribution of crop workers across geographic areas. The result is that crop worker allocations are concentrated in a small number of sampling regions with large numbers of crop workers, resulting in small allocations and potentially larger variances for estimates in the other regions. The NAWS statisticians calculated optimal interview allocations for each of the three cycles and 12 sampling regions used to stratify the NAWS sample. A parallel analysis was also conducted using the USDA 17 regions. The goal was to gain more information about how to reduce interviewing costs and improve the precision of point estimates.

Analysis

Optimal allocations were calculated for six variables that were used to calculate the three NFJP adjustments:

- Number of farm work days in the past 12 months;
- Number of non-farm work weeks in the past 12 months;
- Total income in the past 12 months;[0]
- Income from agricultural employment in the past 12 months;
- The worker lacked work authorization; and
- Family income was below the poverty level.

Two types of allocations were calculated. The *optimal allocation* achieved both statistical and cost efficiency. The *Neyman allocation* was a special case of optimal allocation that assumed the cost of each stratum was approximately equal and thus calculated statistical efficiency only. For the 17 region analysis, four of the NAWS regions – Appalachian (AP), Corn Belt/Northern Plains (CBNP), Delta Southeast (DLSE), and Mountain I and II (MT12) – were split into Appalachian I and II (AP1 and AP2), Corn Belt I and II (CB1 and CB2), Northern Plains (NP), Delta (DL), Southeast (SE), Mountain I (MT1), and Mountain II (MT2).

[0] The NAWS asks respondents to provide an income category for total income and income from agricultural employment. To analyze income, a midpoint was created for each income range category. For example, category 1=Under $500 and the midpoint is $250; category 2=$500-999 and the midpoint is $750.

Results

The results show both optimal allocation and Neyman allocation would increase interview allocations in the larger crop labor region in all cycles (Table 7). Regions with currently small interview allocations would have even smaller allocations if allocations were optimized for statistical and/or cost efficiency. When the NAWS 12 regions were split into the USDA 17 regions (Table 8), optimal allocations would decrease most interview allocations in the split regions with increases in some regions.

Table 7. Difference between current and optimal or Neyman allocation (12 regions)

Cycle	Region	Optimal allocation	Neyman allocation	Current allocation	Difference if optimal allocation is used	Difference if Neyman allocation is used
Fall	AP	42	52	49	-7	4
	CA	351	320	350	1	-29
	CBNP	30	75	75	-45	0
	DLSE	42	51	47	-5	4
	FL	66	64	53	13	12
	LK	32	42	36	-3	7
	MT12	24	29	27	-2	2
	MT3	26	25	29	-3	-4
	NE1	33	33	29	4	4
	NE2	16	26	24	-8	2
	PC	106	87	88	18	-1
	SP	46	38	42	4	-4
Spring	AP	25	31	29	-3	3
	CA	356	292	317	39	-25
	CBNP	27	46	45	-18	1
	DLSE	41	42	40	1	3
	FL	80	78	64	16	14
	LK	20	23	19	1	4
	MT12	17	20	18	-1	2
	MT3	22	22	25	-3	-3
	NE1	14	17	15	0	2
	NE2	22	23	22	1	2
	PC	68	60	60	8	0
	SP	21	27	29	-7	-1
Summer	AP	48	56	53	-4	4
	CA	375	353	384	-9	-30
	CBNP	63	84	83	-20	1
	DLSE	52	53	49	3	4
	FL	41	53	44	-3	10
	LK	36	44	36	0	8
	MT12	40	34	30	10	4
	MT3	22	18	21	1	-3

	NE1	33	30	27	7	4
	NE2	37	35	32	5	3
	PC	151	132	131	20	1
	SP	22	34	36	-13	-2

Note: Difference between optimal/Neyman allocation and current allocation may not add up due to rounding.

Table 8. Difference between current and optimal or Neyman allocation (17 regions)

Cycle	Region	Optimal allocation	Neyman allocation	Current allocation	Difference if optimal allocation is used	Difference if Neyman allocation is used
Fall	AP1	28	35	32	-4	3
	AP2	10	12	17	-7	-5
	CA	352	326	350	3	-24
	CB1	14	35	32	-18	3
	CB2	4	10	17	-13	-6
	NP	6	16	26	-20	-10
	DL	22	27	22	0	5
	SE	25	31	25	0	6
	FL	54	54	53	2	1
	LK	32	42	36	-4	7
	MT1	10	13	14	-4	-2
	MT2	12	14	12	-1	2
	MT3	33	32	29	4	3
	NE1	23	22	29	-6	-7
	NE2	15	24	24	-9	0
	PC	130	108	88	42	20
	SP	46	38	42	4	-4
Spring	AP1	17	21	19	-2	2
	AP2	6	7	10	-4	-3
	CA	357	297	317	40	-20
	CB1	13	21	19	-7	2
	CB2	4	6	10	-6	-4
	NP	6	10	16	-10	-6
	DL	21	23	19	3	4
	SE	24	25	21	3	4
	FL	66	65	64	2	1
	LK	20	23	19	1	4
	MT1	7	9	10	-2	-1
	MT2	8	10	8	0	2
	MT3	28	28	25	3	3
	NE1	10	12	15	-5	-3
	NE2	20	21	22	-1	0
	PC	82	74	60	23	14
	SP	22	28	29	-7	-1

Cycle	Region	Optimal allocation	Neyman allocation	Current allocation	Difference if optimal allocation is used	Difference if Neyman allocation is used
Summer	AP1	32	38	34	-2	3
	AP2	11	13	18	-7	-5
	CA	376	359	384	-7	-24
	CB1	29	39	36	-7	3
	CB2	9	12	19	-10	-7
	NP	13	18	29	-16	-11
	DL	27	29	23	4	6
	SE	30	32	26	4	6
	FL	34	45	44	-10	1
	LK	35	44	36	-1	8
	MT1	17	15	16	1	-2
	MT2	19	16	14	5	2
	MT3	28	23	21	7	2
	NE1	23	21	27	-4	-5
	NE2	34	33	32	2	1
	PC	184	162	131	53	31
	SP	22	34	36	-13	-1

Note: Difference between optimal/Neyman allocation and current allocation may not add up due to rounding.

Design Study 2 – Design Effect Study for Partial Missing Cycles (FY 2019 Spring Cycle)

The purpose of this study was to examine how the late start of the FY 2019 spring cycle affected the standard errors and design effects of the NAWS. In FY 2019, the start of the spring cycle was delayed from mid-February until mid-May. The effect of the late start can be seen in the number of interviews conducted. In the FY 2019 spring cycle, NAWS conducted 149 interviews, a 49 percent decrease from the FY 2017 spring cycle (294 interviews), and a 60 percent decrease from the FY 2018 spring cycle (379 interviews).

Analysis

We compared the design adjusted standard error and the design effects of the following key variables in FY 2016, 2017, 2018, and the delayed 2019:

- The worker's hourly wage or hourly equivalent wage if a piece rate worker;
- Number of farm employers in the past 12 months;
- Number of farm work days in the past 12 months;
- Number of children in the household.
- The employer was a labor contractor and not an agricultural producer ;
- The worker lacked work authorization;
- The worker had only one farm employer;
- The worker was paid an hourly wage as opposed to a piece rate or salary;

- The number of children in household was three or fewer; and
- The worker was a migrant.

The design effect is calculated as the ratio of the variance of an estimate from a complex sampling plan (the methodology actually used) to the variance under the assumption of simple random sampling. The larger the resulting design effect number, the larger the sample variance is in comparison to what would be expected if the survey were based on the same sample size but selected using simple random sampling. The formula for the design effect can be represented as:

$$\text{Design Effect (DEFF)} = \frac{\textit{Complex Plan Variance}}{\textit{Simple Random Sampling Variance}}$$

Results

It was predicted that smaller sample size in FY 2019 would lead to larger standard errors, and the data confirmed this. There are multiple cases of high design effects, likely as a consequence of higher sampling weights for agricultural workers interviewed in the abbreviated spring 2019 spring cycle. Of the 10 variables of interest, the design effect in FY 2019 was higher in 6 and lower in 4 compared to FY 2018. The average increase in the design effect was 463 percent. That is, the design effects for FY 2019 were approximately 4.6 times larger than those in FY 2018. Table 9 shows the design adjusted standard error and design effects for 2016–2019.

Table 9. Design adjusted standard error and design effects estimates, by FY

Estimated Population Parameter	2019		2018		2017		2016	
	DASE	DEFF	DASE	DEFF	DASE	DEFF	DASE	DEFF
Wage	0.32	7.38	0.24	7.10	0.22	6.06	0.16	6.47
Number of farm employers in the past 12 months	0.04	4.33	0.03	2.03	0.05	3.79	0.02	2.55
Number of farm work days in the past year	10.82	15.47	8.69	13.10	7.31	9.03	7.28	17.13
Number of children in the household	0.07	3.08	0.07	3.38	0.06	2.35	0.05	3.57
The employer was a labor contractor	0.04	30.31	0.03	9.54	0.04	25.44	0.05	55.00
The worker lacked work authorization	0.07	21.92	0.04	9.14	0.04	7.57	0.03	9.27
The worker had one farm employer	0.03	5.93	0.02	3.95	0.03	6.16	0.02	6.02
The worker was paid an hourly wage as opposed to a piece rate or salary	0.03	7.39	0.04	18.98	0.03	10.65	0.02	8.64
The number of children in household was three or fewer	0.15	3.72	0.03	3.79	0.03	3.58	0.03	7.32
The worker was a migrant	0.04	7.43	0.03	8.08	0.02	6.68	0.02	12.35

Note: DASE = Design Adjusted Standard Error. DEFF = Design Effects Estimates.
The worker lacked work authorization was created from CURRSTAT=4 (“Unauthorized”).
The worker had one farm employer was created from NUMFEMPL=1.
The work was paid hourly was created from D11=1 (“Hourly”).
The number of children in household was three or fewer was created from HHDKID <= 3.

Design Study 3 – Design Effect Study for Partial Missing Cycles (FY 2020 COVID-19)

The purpose of this study is to examine how the COVID-19 pandemic in the FY 2020 and the late start of the FY 2019 spring cycle affected the data quality.

The NAWS spring cycle usually runs from mid-February to mid-June. In FY 2019, due to a funding delay, the cycle did not start until Mid-May, 2019. In FY 2020, the Cycle stopped at the end of March due to COVID-19 travel restrictions. The number of completed interviews declined to 149 in FY 2019 and 153 in FY 2020 compared to 379 interviews completed in FY 2018, a more normal data collection year.

Analysis

To understand the impact of the short spring cycles on data quality, we compared the frequencies, design adjusted standard error, and the design effects of the key variables in FY 2016, 2017, 2018, 2019, 2020, and combined 2019/2020 (i.e., two partial cycles FY 2019 and FY 2020 were combined to represent a "full" cycle). The continuous variables included:

- The worker's hourly wage, or hourly equivalent wage if a piece rate worker;
- Number of farm employers in the past 12 months;
- Number of farm work days in the past 12 months; and
- Number of children in the household.

Additionally, the analysis examined the following binary variables:

- The employer was a labor contractor and not an agricultural producer;
- The worker lacked work authorization;
- The worker had only one farm employer;
- The worker was paid an hourly wage as opposed to a piece rate or salary;
- The number of children in household was three or fewer; and
- The worker was a migrant.

The design effect is calculated as the ratio of the variance of an estimate from a complex sampling plan (the methodology actually used) to the variance under the assumption of simple random sampling. The larger the resulting design effect number, the larger the sample variance is in comparison to what would be expected if the survey were based on the same sample size but selected using simple random sampling. The formula for the design effect can be represented as:

$$\text{Design Effect (DEFF)} = \frac{\textit{Complex Plan Variance}}{\textit{Simple Random Sampling Variance}}$$

Results

The frequencies and means of the key variables were consistent with those from prior years (FY 2016–2018) with full spring cycles. While it was hypothesized that smaller sample size in FY 2019 and 2020 would lead to larger standard errors and design effects, the results did not confirm this for most of the variables of interest. There are multiple cases of decreased standard errors and design effects. The most insight comes from the comparison of FY 2020 to FY 2019 and the

hybrid FY 2019/2020 to the most recent full cycle year, FY 2018. Overall, of the 10 variables of interest, the design effect was overall significantly lower in FY 2020 than in FY 2019, with only one exception. When considering the hybrid year 2019/2020, the change was still significantly lower, but to a slightly more tempered degree. Table 10 shows the design adjusted standard error and design effects for 2018–2020.

Table 10. Design adjusted standard error and design effects estimates, by FY

Estimated Population Parameter	2019/2020		2020		2019		2018	
	DASE	DEFF	DASE	DEFF	DASE	DEFF	DASE	DEFF
Wage	0.21	6.89	0.27	6.50	0.32	7.38	0.24	7.10
Number of farm employers in the past 12 months	0.06	2.08	0.06	2.22	0.04	4.33	0.03	2.03
Number of farm work days in the past year	7.19	14.02	9.34	12.16	10.82	15.47	8.69	13.10
Number of children in the household	0.04	1.87	0.05	2.68	0.07	3.08	0.07	3.38
The employer was a labor contractor	0.09	8.67	0.08	6.89	0.04	30.31	0.03	9.54
The worker lacked work authorization	1.31	3.72	1.30	3.65	0.07	21.92	0.04	9.14
The worker had one farm employer	0.10	6.17	0.12	8.52	0.03	5.93	0.02	3.95
The worker was paid an hourly wage as opposed to a piece rate or salary	0.01	1.34	0.01	1.29	0.03	7.39	0.04	18.98
The number of children in household was three or fewer	0.04	1.41	0.03	1.35	0.15	3.72	0.03	3.79
The worker was a migrant	0.04	3.14	0.04	3.37	0.04	7.43	0.03	8.08

Note: DASE = Design Adjusted Standard Error. DEFF = Design Effects Estimates.
The worker lacked work authorization was created from CURRSTAT=4 (“Unauthorized”).
The worker had one farm employer was created from NUMFEMPL=1.
The work was paid hourly was created from D11=1 (“Hourly”).
The number of children in household was three or fewer was created from HHDKID <= 3.