Assessing the Effect of Calibration on Nonresponse Bias in the 2006 ARMS Phase III Sample Using Census 2002 Data

Morgan Earp[1], Jaki McCarthy[2], Nick Schauer3, Phillip Kott[4]
[1]National Agricultural Statistics Service (NASS), 3251 Old Lee Highway Room 305, Fairfax, VA 22030
[2]NASS, 3251 Old Lee Highway Room 305, Fairfax, VA 22030
[3]NASS, 1400 Independence Avenue SW, Washington, DC 20004
[4] NASS, 3251 Old Lee Highway Room 305, Fairfax, VA 22030

Abstract
The United States Department of Agriculture's National Agricultural Statistics Service (NASS) conducts the annual Agricultural Resource Management Survey (ARMS) in three phases. The third phase of the ARMS collects detailed economic data which is highly sensitive. As a consequence, this phase suffers from relatively low response rates for a federal survey. According to the 2006 Office of Management and Budget Standards and Guidelines for Statistical Surveys, response rates lower than 80 percent may not only result in nonresponse bias, but they can jeopardize the future of surveys carried out by federal agencies. NASS has been operating under the assumption that the use of calibrated weights derived from appropriate targets addresses nonresponse bias in the 2006 Phase III ARMS. This assumption was tested using Census 2002 expenditure-sample data.

The results showed that calibrated weights decreased bias levels so that they were no longer significantly different from zero at the 0.05 level for over 90 percent of the variables evaluated.

Key Words: Nonresponse; response rate; bias; calibration weights.

1. Introduction

Survey nonresponse happens; the question is, how do we address it? In 2003, the Federal Government's Federal Committee on Statistical Methodology (FCSM) formed a subcommittee of Interagency Council on Statistical Policy (ICSP) representative nominees to update Federal standards for statistical surveys. The Subcommittee on Standards for Statistical Surveys concluded that in order to ensure the quality, objectivity, utility, and integrity of Federal Government data, nonresponse bias should be assessed when surveys exhibit insufficient response rates. Under the guidance of the FCSM and ICSP, ICSP representatives recommended Federal survey standards and guidelines to the Executive Office of the President's Office of Management and Budget in 2004. After public review, the Executive Office of the President ultimately released the *Office of Management and Budget Standards and Guidelines for Statistical Surveys* on September, 22, 2006.

The United States Department of Agriculture's (USDA) National Agricultural Statistics Service (NASS) helped develop the OMB's new standards and guidelines for statistical surveys. This paper focuses specifically on Standard 3.2. Standard 3.2 addresses response rates and analysis of

nonresponse bias, requiring that "Agencies must appropriately measure, adjust for, report, and analyze unit and item nonresponse to assess their effects on data quality and to inform users" when survey response rates fall below 80 percent. (Office of Management and Budget, 2006, p. 14) Standard 3.2 stipulates that response rate are computed "…using standard formulas to measure the proportion of the eligible sample that is represented by the responding units in each study, as an indicator of potential nonresponse bias." (p. 14)

In 2005, the Agricultural Resource Management Survey (ARMS) Phase III response rate was 70.5 percent (n = 34,937), which fell below the OMB response rate threshold of 80 percent listed in Guideline 3.2.9; therefore, NASS was required by OMB to research the effect of nonresponse bias (Earp, McCarthy, Schauer, & Kott, 2008). In 2006, the Agricultural Resource Management Survey (ARMS) Phase III response rate was 67.6 percent (n = 34,192), which again fell below the 80 percent threshold, and thus NASS was again required by OMB to research the effect of nonresponse bias. Since the Phase II response rate of 81.3 percent exceeded OMB's 80 percent threshold, nonresponse bias assessment was only required for Phase III, the "problem" stage. Specifically, Guideline 3.2.9 states

> Given a survey with an overall unit response rate of less than 80 percent, conduct an analysis of nonresponse bias using unit response rates as defined above, with an assessment of whether the data are missing completely at random. As noted above, the degree of nonresponse bias is a function of not only the response rate but also how much the respondents and nonrespondents differ on the survey variables of interest. For a sample mean, an estimate of the bias of the sample respondent mean is given by:
>
> $$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_{nr}}{n}\right)(\bar{y}_r - \bar{y}_{nr})$$
>
> 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.
>
> For a multistage (or wave) survey, focus the nonresponse bias analysis on each stage, with particular attention to the "problem" stages. A variety of methods can be used to examine nonresponse bias, for example, make comparisons between respondents and nonrespondents across subgroups using available sample frame variables. In the analysis of unit nonresponse, consider a multivariate modeling of response using respondent and nonrespondent frame variables to determine if nonrespondent bias exists. (Office of Management & Budget, 2006, p. 16)

Currently, NASS calculates the unweighted unit response rates (*RRU*) for the ARMS based on the formula provided under Guideline 3.2.2 of the *Office of Management and Budget Standards and Guidelines for Statistical Surveys*:

$$RRU = \frac{C}{C + R + NC + O + e(U)}$$

Where:
C = the number of completed cases or sufficient partials;
R = the number of refused cases;
NC = the number of noncontacted sample units known to be eligible;
O = the number of eligible sample units not responding for reason other than refusal;
U = the number of sample units of unknown eligibility, not completed; and
e = the estimated proportion of sample units of unknown eligibility that are eligible. (p. 14)

Thus, NASS sums the number of positive usables, out-of-business, and non-farms and calculates the percentage this sum represents of the total number of reports to calculate the response rate for ARMS Phase III.

The ARMS is conducted in three phases. Phase I screens for potential samples for Phases II and III. Phase II collects data on cropping practices and agricultural chemical usage, while Phase III collects detailed economic information about the agricultural operation, as well as the operator's household. Phase III is the only phase of the 2006 ARMS with response rates lower than 80 percent.

Due to lower response rates with 2006 ARMS Phase III, the potential for nonresponse bias is greater there. NASS weights the ARMS Phase III respondent sample in such a way that estimated variable totals for a large set of items match "targets" determined from other sources. This is done through a weighting process called "calibration." Calibration is the process of adjusting survey weights so that certain targets are met. NASS uses official estimates of farm numbers, corn, soybean, wheat, cotton, fruit and vegetable acres as well as cattle, milk production, hogs, broilers, eggs and turkeys as calibration targets. For example, after calibration, the calibration-weighted sum of the survey data will equal the NASS estimate for corn acres. In addition to reducing confusion in the user community that might result from NASS releasing alternative estimates for the same totals, calibration weighting produces 2006 ARMS Phase III estimates with generally lower variances and reduces nonresponse biases. This report describes an ongoing research effort aimed at measuring the potential for nonresponse bias in the ARMS Phase III and the success or failure of calibration in removing it.

Nonresponse bias is very difficult to measure directly. Fortunately, an indirect measure of nonresponse bias is available for the 2006 ARMS Phase III, hereafter called simply the "ARMS."

The Census of Agriculture is a mandatory collection of data from all known agricultural operations. NASS has data from the Census on items of interest for many of the ARMS nonrespondents; however, the Census itself is incomplete. An estimated 17.90 percent of all farms were missing from the 2002 Census Mailing List, and 12.26 percent of farms on the List failed to respond to the Census. Moreover, not all ARMS sampled farms could be matched to 2002 Census records. Nevertheless, by comparing the 2002 Census values of ARMS respondents to the full sample of ARMS respondents as a whole, we can measure the difference between the average ARMS respondent and the average of the full sample without any nonresponse adjustment. Additionally, this analysis intends to measure the reduction of that difference from using a calibration-weighting process similar to the one used for the 2006 ARMS.

Although the 2002 Census data do not perfectly match the 2006 ARMS Phase III data, they are highly correlated (see Appendix Tables A-2 & A-3). The present evaluation will effectively compare 2006 ARMS Phase III survey respondents to nonrespondents using their 2002 Census

data. The 2002 Census expenditure data, which were required for all Census reports considered usable in this research, were available for 43 percent of 2006 ARMS Phase III reports.[1]

2. Method

Our analytical data set consists of census values for farms sampled for the ARMS that also provided 2002 expenditure-sample data on the Census. In the 2002 Census only a sample of farms received the long version of the questionnaire which asked the expenditure questions.

The *base sampling weight* for a farm in our analytical data set was its ARMS sample weight before calibration multiplied by its Census sample weight. Each ARMS responding farm was calibrated to produce weighted totals for the calibration variables that were equal to the base-sampling-weighted totals computed from both respondents and nonrespondents. The calibration variables used were inventory/acreage numbers for cattle, corn, cotton, pigs, soybeans, wheat, fruit, vegetables, broilers, and turkeys. Each of these target variables, plus egg and milk production, was used operationally in calibrating the ARMS data.

As in the operational program, the ARMS respondent subset was calibrated independently in 20 regions. These included the 15 leading cash receipts states (Arkansas, California, Florida, Georgia, Illinois, Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Carolina, Texas, Washington, and Wisconsin). The remaining 33 states (Alaska and Hawaii are not sampled for the ARMS) were grouped using the five production regions: 1) Atlantic, 2) South, 3) Midwest, 4) Plains, and 5) West (Figure 1).

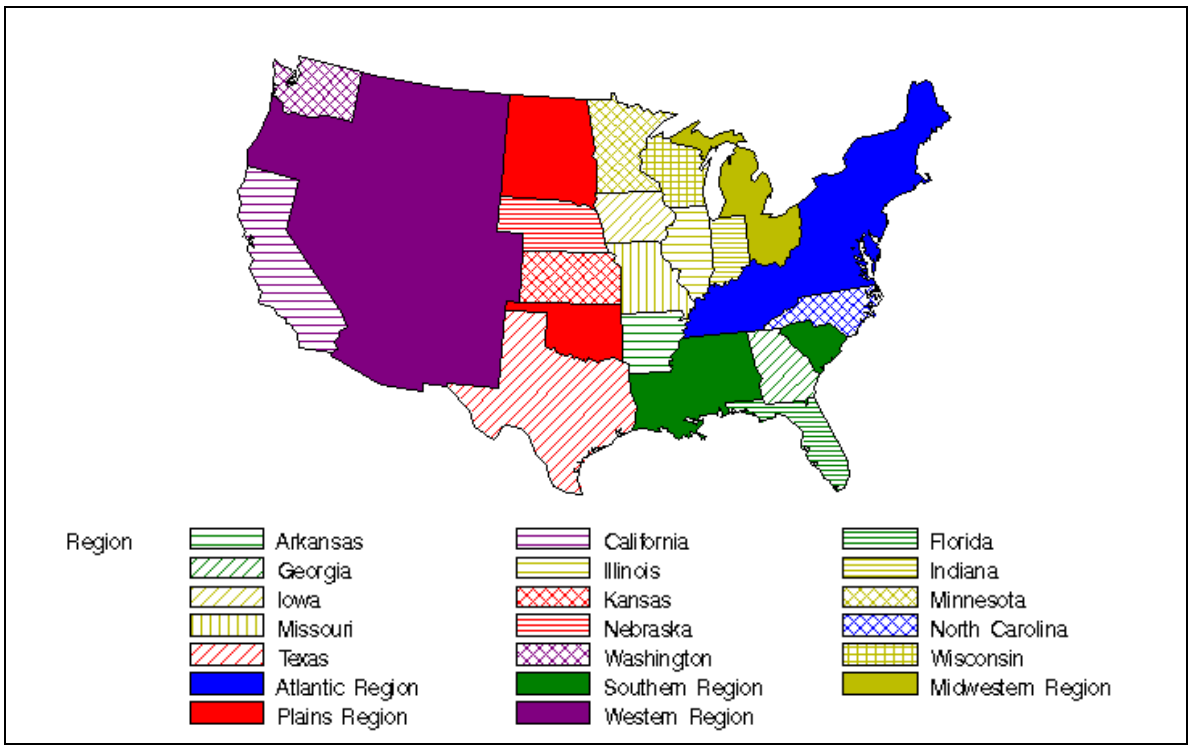


Figure 1. ARMS III Estimation Regions

[1] The match rate for 2006 ARMS Phase III reports with 2002 Census expenditure data was significantly higher for nonrespondents (47.5%) than for respondents (40.5%) ($z = 12.24$, $p < .05$).

Our analysis focuses on 17 specific (non-calibration) variables collected on both the ARMS and the Census:

1. Total Acres
2. Total Sales
3. Acres Rented
4. Cropland Acres
5. Total Production Expenses
6. Crop Expenses
7. Seed Expenses
8. Fertilizer Expenses
9. Chemical Expenses
10. Livestock Purchases
11. Feed Purchases
12. Hired Labor Expenses
13. Machinery and Equipment Value
14. Government Payments
15. Operator's Age
16. Operator's Race
17. Farm Type.

These variables were also included in a similar analysis for ARMS Phase III 2005 (Earp *et al.*, 2008).

Letting $\bar{y}_r$ denote the base-sample or calibrated-sample mean among the ARMS respondent subset for a study variable, and $\bar{y}_t$ denote the corresponding base-sample mean among the entire matched sample, it is a simple matter to compute the relative bias of the former with respect to the latter, relBias = $\frac{\bar{y}_r - \bar{y}_t}{\bar{y}_r}$. The statistical significance of this value is much harder to assess since the samples on which $\bar{y}_r$ and $\bar{y}_t$ are based are complex and overlapping.

Fortunately, we can easily test the persistence or absence of a systematic bias across the 20 regions. To this end, we compute the following measure of bias of an ARMS-respondent mean (before or after calibration) with respect to the Census mean in every region:

$$
\begin{aligned}
M &= \log(\bar{y}_r) - \log(\bar{y}_t) \\
&= \log\left(\bar{y}_r \big/ \bar{y}_t\right) \\
&= \log\left(1 + \frac{\bar{y}_r - \bar{y}_t}{\bar{y}_t}\right) \approx \frac{\bar{y}_r - \bar{y}_t}{\bar{y}_t}
\end{aligned}
$$

This measure is conveniently symmetric, $\log(\bar{y}_t) - \log(\bar{y}_r) = -[\log(\bar{y}_r) - \log(\bar{y}_t)]$, while retaining the scale-invariance property of the relative bias (*i.e.*, multiplying the reported item value on each farm by a fixed factor does not affect the overall relative bias).

The bias measure *M* for a study variable in a region can be treated as an independent random variable. The null hypothesis of no bias (again, either before or after calibration) can be tested against an alternative hypothesis of a persistent bias (*p*%) across all the regions. The conventional

t test based on the 20 observations (one per region) is asymptotically normal under both the null and alternative hypotheses. We follow the standard practice of approximating the distribution of this test statistic with a Student's *t* having 19 degrees of freedom. This may lead to liberal inferences (the inappropriate rejection of the null hypothesis when it is true) because the *M*-values for the study variable may not be normally distributed with a common variance across regions. Nevertheless, by taking logs we create a test statistic that is more nearly normal and homoscedastic than absolute biases would be.

A sign and a signed-rank test of the 20 paired observations for a study variable before and after calibration was conducted. The sign test is not as powerful as the other two tests (*i.e.*, it more often fails to find that *M* is significantly different from 0 when, in fact, there is a persistent bias across the regions), but it assumes neither that *M* is normal nor homoscedastic. The signed-rank test assumes the latter, but not the former. We include all three analyses in our results for completeness.

3. Results

Our results are summarized in Table 1. Chemical expenses, machinery and equipment value, government payments, acres rented, farm type, fuel and oil expenses, operator's age, and cropland acres (Variables 1-6) do not exhibit significant biases using either calibrated or uncalibrated weights. Results slightly varied from those in the previous analysis of ARMS Phase III 2005 data (Earp *et al.*, 2008); total acres operated no longer exhibits significant bias using either calibrated or uncalibrated weights; on the contrary, chemical expenses, machinery and equipment value, and fuel and oil expenses now exhibit significant bias using the uncalibrated weights, but not the calibrated weights (Earp *et al.*, 2008).

In over 90 percent (10/11) of the study variables exhibiting persistent biases using the base sample weights (i.e., variables 7-17), calibration weighting is able to reduce the bias so that it was no longer significantly different from zero using a t-test with $p < .05$. The rate of bias elimination remained consistent from 2005 to 2006, although the rate of bias and the rank of variable bias varied (Earp *et al.*, 2008). All of these variables show a significant reduction in bias levels using a paired t-test. After calibration, only one study variable, total sales, has a significant bias. This result varied from 2005, where fertilizer expense was the only variable exhibiting significant bias. Using the 2006 data, fertilizer expense no longer exhibits significant bias using the calibrated weights, but total sales which did not exhibit significant bias using the calibrated weights in 2005 but does in 2006 (Earp *et al.*, 2008). Bias levels for 2005 fertilizer expenses and 2006 total sales were explored at the regional level to determine if calibration performed better or worse in certain regions. Overall nonresponse bias levels decreased for both 2005 fertilizer expenses and 2006 total sales after calibration adjustment; however, at the regional level nonresponse bias levels did increase after calibration adjustment in Plains, Iowa, Atlantic, Georgia, South, and the West for 2005 fertilizer expenses and in Iowa and Wisconsin for 2006 Total Sales.

As in 2005, the estimated bias of livestock purchases remains the largest among the study variables. Using only the base-sampling weights, this bias was highly significant using all three test statistics. After calibration, although still large in magnitude, the estimated bias was reduced to statistical insignificance using all the tests. For this variable, calibration continues to reduce the bias significantly, if not completely.

Table 1: Mean Comparisons and Indicated Biases for Matching Records Using Base Sampling Weights versus Calibrated Weights

	National Estimates		Regional Estimates (n = 20)				Testing the Effect of Calibration on Regional Bias Means							
Variable	Mean	Bias	Mean[3]	Bias Mean[3]	Bias Mininmum [3]	Bias Maximum[3]	t[3]	p value[3]	Sign[3]	p value[3]	Signed Ranks[3]	p value[3]	t_{paired}[3]	p value[3]
1. Total Acres Operated														
All Matching Records [1]	568.26		563.57											
Matching Respondents [2]	533.99	-6.42%	532.00	-4.24%	-22.97%	18.08%	-1.89	0.07	-1.00	0.82	-48	0.08		
Matching Respondents Calibrated [2]	560.07	-1.46%	563.62	0.07%	-9.24%	14.70%	0.05	0.96	0.00	1.00	-7.00	0.81	-2.28	0.03
2. Farm Type														
All Matching Records [1]	7.77		7.65											
Matching Respondents [2]	7.90	1.65%	7.56	-2.33%	-21.26%	4.57%	-1.61	0.12	-2.50	0.36	-29.00	0.26		
Matching Respondents Calibrated [2]	7.66	-1.44%	7.54	-2.05%	-22.10%	3.30%	-1.62	0.12	-1.00	0.82	-33.00	0.23	-0.35	0.73
3. Acres Rented														
All Matching Records [1]	269.08		269.88											
Matching Respondents [2]	257.28	-4.59%	260.37	-2.56%	-20.77%	22.81%	-0.96	0.35	-2.00	0.50	-27	0.33		
Matching Respondents Calibrated [2]	268.19	-0.33%	273.45	1.96%	-10.07%	21.07%	1.33	0.20	3.00	0.26	34.00	0.22	-1.57	0.13
4. Government Payments (Dollars)														
All Matching Records [1]	4,997.09		5,286.53											
Matching Respondents [2]	4,566.41	-9.43%	5,091.37	-3.44%	-39.85%	21.50%	-0.90	0.38	-1.00	0.82	-16.00	0.57		
Matching Respondents Calibrated [2]	4,870.35	-2.60%	5,262.27	0.89%	-29.48%	19.86%	0.35	0.73	2.00	0.50	16.00	0.57	-1.66	0.11
5. Operator's Age (Years)														
All Matching Records [1]	54.53		54.52											
Matching Respondents [2]	54.76	0.42%	54.54	-6.64E-07	-3.57%	3.89%	0.00	1.00	0.00	1.00	5.00	0.87		
Matching Respondents Calibrated [2]	54.29	-0.44%	54.42	-0.23%	-3.26%	3.21%	-0.50	0.62	-1.00	0.82	-13.00	0.65	1.66	0.11
6. Cropland Acres														
All Matching Records [1]	312.10		318.03											
Matching Respondents [2]	296.93	-5.11%	304.93	-4.61%	-36.09%	23.19%	-1.59	0.13	-1.00	0.82	-42	0.12		
Matching Respondents Calibrated [2]	314.41	0.73%	319.59	0.27%	-20.57%	5.62%	0.21	0.84	2.00	0.50	39.00	0.15	-2.07	0.05
7. Total Production Expenses (Dollars)														
All Matching Records [1]	121,501.64		141,434.15											
Matching Respondents [2]	102,034.92	-19.08%	124,369.26	-13.26%	-72.64%	22.83%	-2.85	0.01	-5.00	0.04	-69.00	0.01		
Matching Respondents Calibrated [2]	118,370.93	-2.64%	138,277.99	-2.36%	-15.38%	11.06%	-1.94	0.07	-4.00	0.12	-56.00	0.04	-2.60	0.02
8. Livestock Purchases (Dollars)														
All Matching Records [1]	12,435.51		11,682.49											
Matching Respondents [2]	7,773.21	-59.98%	8,159.95	-30.17%	-114.05%	25.12%	-3.00	0.01	-5.00	0.04	-65.00	0.01		
Matching Respondents Calibrated [2]	10,925.13	-13.82%	10,489.31	-13.51%	-127.21%	17.52%	-1.82	0.09	-2.00	0.50	-39.00	0.15	-2.21	0.04

[1] Means computed using the base sampling weights for all matching cases with Census 2002 expenditure data (n = 14,633)
[2] Means computed only for ARMS III respondents with Census 2002 expenditure data (n = 9,380)
[3] Regional estimates are based on the 20 ARMS III estimation regions shown in *Figure 1* using only ARMS III respondents

Note: Significant bias and corresponding t scores and p values are identified in **red** font
Significant reduction in bias is identified in **blue** font

Table 1 (Cont.): Mean Comparisons and Indicated Biases for Matching Records Using Base Sampling Weights versus Calibrated Weights

	National Estimates		Regional Estimates (n = 20)				Testing the Effect of Calibration on Regional Bias Means							
Variable	Mean	Bias	Mean[3]	Bias Mean[3]	Bias Mininmum [3]	Bias Maximum[3]	t[3]	p value[3]	Sign[3]	p value[3]	Signed Ranks[3]	p value[3]	t_{paired}[3]	p value[3]
9. Hired Labor Expenses (Dollars)														
All Matching Records [1]	11,301.52		15,523.98											
Matching Respondents [2]	9,019.46	-25.30%	12,831.78	-19.71%	-54.57%	13.46%	-4.69	0.00	-6.00	0.01	-89.00	0.00		
Matching Respondents Calibrated [2]	11,145.84	-1.40%	15,338.80	-5.41%	-47.88%	14.53%	-1.66	0.11	-2.00	0.50	-38.00	0.17	-3.97	0.00
10. Feed Purchases (Dollars)														
All Matching Records [1]	13,443.92		13,664.10											
Matching Respondents [2]	10,836.95	-24.06%	11,523.50	-16.90%	-102.19%	15.11%	-2.78	0.01	-6.00	0.01	-73.00	0.00		
Matching Respondents Calibrated [2]	13,254.52	-1.43%	13,496.59	-3.40%	-20.72%	24.82%	-1.34	0.20	-2.00	0.50	-41.00	0.13	-2.53	0.02
11. Fuel & Oil Expenses (Dollars)														
All Matching Records [1]	4,193.76		4,811.07											
Matching Respondents [2]	3,824.70	-9.65%	4,506.99	-6.83%	-27.90%	23.72%	-2.16	0.04	-1.00	0.82	-48.00	0.08		
Matching Respondents Calibrated [2]	4,258.75	1.53%	4,882.59	1.30%	-6.27%	12.36%	1.09	0.29	-1.00	0.82	14.00	0.62	-3.01	0.01
12. Chemical Expenses (Dollars)														
All Matching Records [1]	5,656.96		7,064.76											
Matching Respondents [2]	4,945.47	-14.39%	6,350.15	-11.17%	-34.99%	21.67%	-3.05	0.01	-4.00	0.12	-65.00	0.01		
Matching Respondents Calibrated [2]	5,611.21	-0.82%	7,087.30	-1.06%	-13.40%	11.35%	-0.80	0.44	-1.00	0.82	-21.00	0.45	-2.62	0.02
13. Machinery & Equipment Value (Dollars)														
All Matching Records [1]	85,927.07		90,622.32											
Matching Respondents [2]	79,194.08	-8.50%	86,031.55	-5.21%	-20.40%	16.37%	-2.21	0.04	-4.00	0.12	-53.00	0.05		
Matching Respondents Calibrated [2]	85,960.07	0.04%	90,746.93	0.58%	-11.68%	12.95%	0.42	0.68	2.00	0.50	15.00	0.60	-2.75	0.01
14. Seed Expenses (Dollars)														
All Matching Records [1]	5,560.46		6,418.56											
Matching Respondents [2]	4,909.03	-13.27%	5,871.17	-10.31%	-45.75%	29.92%	-2.40	0.03	-4.00	0.12	-60.00	0.02		
Matching Respondents Calibrated [2]	5,629.33	1.22%	6,352.64	-1.04%	-23.14%	23.37%	-0.45	0.66	-3.00	0.26	-23.00	0.41	-2.92	0.01
14. Cropland Expenses (Dollars)														
All Matching Records [1]	18,583.65		21,875.20											
Matching Respondents [2]	16,482.46	-12.75%	20,019.46	-9.65%	-37.54%	26.86%	-2.70	0.01	-3.00	0.26	-58.00	0.03		
Matching Respondents Calibrated [2]	18,603.14	0.10%	21,853.61	-0.52%	-11.68%	13.12%	-0.39	0.70	1.00	0.82	-10.00	0.73	-2.77	0.01
16. Fertilizer Expenses (Dollars)														
All Matching Records [1]	7,366.23		8,391.88											
Matching Respondents [2]	6,627.97	-11.14%	7,798.14	-8.26%	-34.14%	27.69%	-2.45	0.02	-4.00	0.12	-60.00	0.02		
Matching Respondents Calibrated [2]	7,362.61	-0.05%	8,413.67	-0.17%	-9.44%	10.55%	-0.12	0.90	0.00	1.00	-5.00	0.87	-2.58	0.02
17. Total Sales (Dollars)														
All Matching Records [1]	144,890.55		171,551.69											
Matching Respondents [2]	121,062.17	-19.68%	149,929.07	-14.46%	-73.49%	11.36%	-3.17	0.01	-5.00	0.04	-71.00	0.01		
Matching Respondents Calibrated [2]	140,104.48	-3.42%	166,388.48	-3.19%	-12.36%	3.30%	-3.18	0.00	-5.00	0.04	-71.00	0.01	-2.62	0.02

[1] Means computed using the base sampling weights for all matching cases with Census 2002 expenditure data (n = 14,633)

[2] Means computed only for ARMS III respondents with Census 2002 expenditure data (n = 9,380)

[3] Regional estimates are based on the 20 ARMS III estimation regions shown in *Figure 1* using only ARMS III respondents

Note: Significant bias and corresponding t scores and p values are identified in **red** font
Significant reduction in bias is identified in **blue** font

4. Discussion

ARMS data are used by farm organizations, commodity groups, agribusiness, Congress, State Departments of Agriculture, and the USDA. The USDA uses ARMS data to evaluate the financial performance of farms and ranches, which influence agricultural policy decisions. The Department also uses Phase III data for objective evaluation of critical issues related to agriculture and the rural economy; therefore, it is essential that measures be taken to minimize the effect of nonresponse bias in ARMS, specifically Phase III.

In assessing the adjustment for nonresponse bias in the 2006 ARMS Phase III, the 2002 Census mean estimates of total production expenses, livestock purchases, hired labor expenses, feed purchases, fuel and oil expenses, chemical expenses, machinery and equipment value, seed expenses, cropland expenses, and fertilizer expenses demonstrated significant bias using just the base sample weights. Although the magnitude of the relative bias of the mean estimate remained high for livestock purchases using the calibrated weights, calibration reduced the magnitude of this bias to statistical insignificance (see Table 1).

For this analysis, the calibration process varied slightly from that of the 2006 ARMS Phase III. Egg and milk production were not included as calibration targets, because these data items were not collected for the 2002 Census. This may help to explain why the magnitude of the estimated relative bias of the mean for livestock in Table 1 remained high even after the data were calibrated. Although it was not possible to use these as calibration targets in this analysis, their use in the ARMS Phase III survey may reduce the bias for livestock purchases in published ARMS data.

According to Guideline 3.2.13 of the *Office of Management and Budget Standards and Guidelines for Statistical Surveys*, NASS should:

> Base decisions regarding whether or not to adjust or impute data for item nonresponse on how the data will be used, the assessment of nonresponse bias that is likely to be encountered in the review of collections, prior experience with this collection, and the nonresponse analysis discussed in this section. When used, imputation and adjustment procedures should be internally consistent, sampled on theoretical and empirical considerations, appropriate for the analysis, and make use of the most relevant data available. If multivariate analysis is anticipated, care should be taken to use imputations that minimize the attenuation of underlying relationships.

Due to the broadness of the ARMS Phase III data user community and the survey's impact on agricultural policy, it is crucial that the calibration process effectively adjusts for nonresponse bias. Assuming that the adjustment process is even more effective than demonstrated here (particularly for livestock purchases and total sales) when all calibration targets (including egg and milk production) are available and used, it appears that NASS is appropriately addressing the issue of nonresponse bias in ARMS Phase III through the calibration process. Furthermore, NASS has expanded the number of calibration targets used since 2006 to include hay acres, peanut acres, rice acres, sugarcane/sugar beet acres, tobacco acres, nursery/floriculture acres, cattle on feed, number of farms by non-estimate states, total number of farms, and number of farms by eight economic classes as opposed to the seven previously used.

Limitations of this analysis include: 1) Inability to replicate the 2006 ARMS Phase III calibration process exactly without egg and milk items; 2) Inability to assess farms not covered or responding to the Census of Agriculture; and 3) Inability to recognize localized biases in the ARMS data (tests were limited to persistent biases across regions).

Knowing that the analyzed data come from the 2002 Census and not from the 2005 ARMS Phase III survey does not limit, but strengthens the analysis. It allows us to focus entirely on the impact of the nonresponse *per se*.

References

Earp, M.S., McCarthy, J.S., Shauer, N.D., & Kott, P.S. (2008), *Assessing the Effect of Calibration on Nonresponse Bias in the 2005 ARMS Phase III Sample Using Census 2002 Data*. Research and Development Division Staff Report RDD-08-01, United States Department of Agriculture, National Agricultural Statistics Service.

Earp, M.S., McCarthy, J.S., Schauer, N.D., & Kott, P.S. (2008). Assessing the Effect of Calibration on Nonresponse Bias in the 2005 Agricultural Resource Management Survey Phase III Sample Using Census 2002 Data. In *JSM Proceedings*, Government Statistics Section . Denver, CO: American Statistical Association.

Earp, M.S., McCarthy, J.S., Schauer, N.D., & Kott, P.S. (2008). *Assessing the Effect of Calibration on Nonresponse Bias in the 2006 ARMS Phase III Sample Using Census 2002 Data*. Research and Development Division Staff Report RDD-08-06, United States Department of Agriculture, National Agricultural Statistics Service.

Kott, P.S. (2005), "Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors" Survey Methodology 32: 133-142.

Hopper, R. (2007). *2006 Agricultural Resource Management Survey (ARMS) Phase III Survey Administration Analysis.* Census and Survey Division Staff Report SAB-07-14, United States Department of Agriculture, National Agricultural Statistics Service.

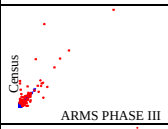
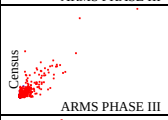
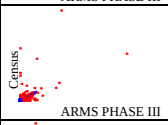
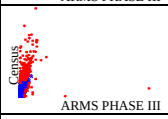
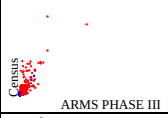
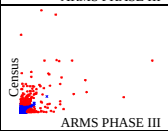
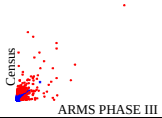
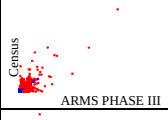
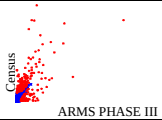
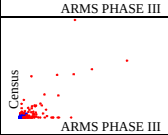
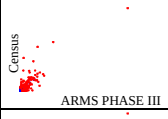
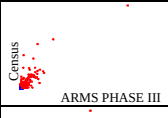
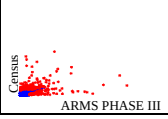
United States. Department of Agriculture. (2002). *2002 Census of Agriculture, Vol. 1, Appendix C.* Washington, DC: U.S. Department of Agriculture.

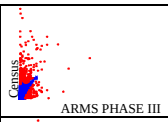
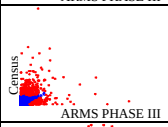
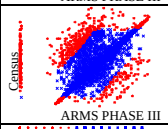
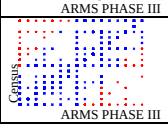
United States. Department of Education. 2003. *National Center for Education Statistics Statistical Standards.* Washington, DC: U.S. Department of Education.

United States. Executive Office of the President. 2006. *Office of Management and Budget Standards and Guidelines for Statistical Surveys.* Washington, DC: U.S. Executive Office of the President.

Appendix

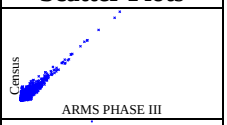
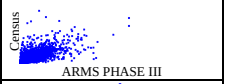
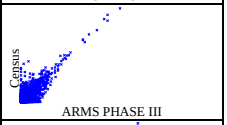
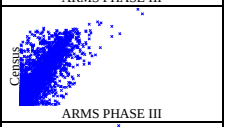
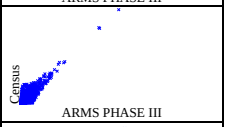
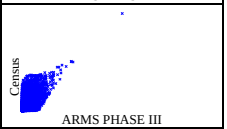
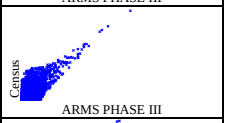
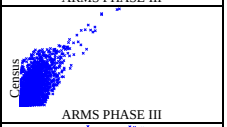
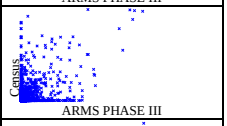
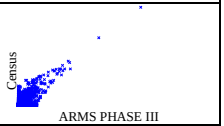
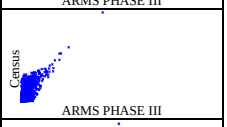
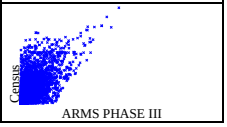
Table A-1: Census 2002 and ARMS Phase III 2006 Variable Correlations with Outliers

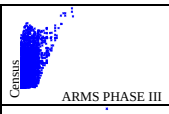
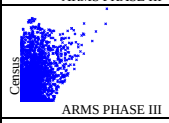
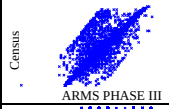
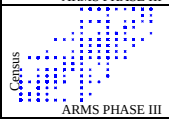
	r	r^2	Scatter Plots
Total Acres Operated	.89399 (*n* = 9,380)	.79922	
Total Sales	.87983 (*n* = 9,380)	.77410	
Acres Rented	.63828 (*n* = 9,380)	.40740	
Cropland Acres	.64025 (*n* = 9,380)	.40992	
Total Production Expenses	.83655 (*n* = 9,380)	.69982	
Seed Expenses	.57250 (*n* = 9,380)	.32776	
Fertilizer Expenses	.71797 (*n* = 9,380)	.51548	
Chemical Expenses	.79302 (*n* = 9,380)	.62888	
Crop Expenses	.75064 (*n* = 9,380)	.56346	
Livestock Purchases	.69380 (*n* = 9,380)	.48136	
Feed Purchases	.83451 (*n* = 9,380)	.69641	
Hired Labor Expenses	.74749 (*n* = 9,380)	.55874	
Fuel & Oil Expenses	.56588 (*n* = 9,380)	.32022	

Machinery & Equipment	.49977 (*n* = 9,380)	.24977	
Government Payments	.47336 (*n* = 9,380)	.22407	
Operator's Age	.63618 (*n* = 9,380)	.40472	
Farm Type	.84887 (*n* = 9,380)	.72058	

1. All correlations were significant at the .05 level.
2. Correlations were only estimated for ARMS respondents.
3. Outliers were flagged using DFFITS, Cook's D, and studentized residuals and are shown in red.

Table A-2: Census 2002 and ARMS Phase III 2006 Variable Correlations with Outliers

	r	r^2	Scatter Plots
Total Acres Operated	.95629 (*n* =9,278)	.91449	
Total Sales	.72669 (*n* = 6,731)	.52808	
Acres Rented	.87279 (*n* = 9,295)	.76176	
Cropland Acres	.86821 (*n* = 8,996)	.88723	
Total Production Expenses	.85281 (*n* = 9,177)	.72728	
Seed Expenses	.68220 (*n* = 9,167)	.46540	
Fertilizer Expenses	.75103 (*n* =9,129)	.56405	
Chemical Expenses	.81290 (*n* = 9,077)	.81290	
Crop Expenses	.81961 (*n* = 9,114)	.67176	
Livestock Purchases	.46853 (*n* = 9,289)	.21952	
Feed Purchases	.63679 (*n* =9,159)	.40550	
Hired Labor Expenses	.83480 (*n* = 9,135)	.69689	
Fuel & Oil Expenses	.68545 (*n* =9,126)	.46984	

Machinery & Equipment	.61670 (*n* = 8,976)	.38032	
Government Payments	.59126 (*n* =8,972)	.34959	
Operator's Age	.89151 (*n* = 8,725)	.79479	
Farm Type	.95712 (*n* =8,842)	.91608	

1. All correlations were significant at the .05 level (*n* = 19,483).
2. Correlations were only estimated for ARMS respondents.