Forest Science, 2023, **XX**, 1–13
https://doi.org/10.1093/forsci/fxad003
Advance access publication 22 February 2023
Research Article

Social Sciences

USDA Forest Service Timber Products Output Survey Item Nonresponse Analysis

Marla Markowski-Lindsay,[1,2,*] Consuelo Brandeis,[3] and Brett J. Butler[2,4]

[1]University of Massachusetts–Amherst, Department of Environmental Conservation, Amherst, MA, USA.
[2]Family Forest Research Center, Amherst, MA, USA.
[3]USDA Forest Service, Southern Research Station, Knoxville, TN, USA.
[4]USDA Forest Service, Northern Research Station, Amherst, MA, USA.
*Corresponding author email: marlal@umass.edu.

Abstract

The Timber Products Output (TPO) survey is used to determine industrial uses of roundwood, reporting on volumes of roundwood received and residues generated by the primary forest industry by tree species and counties of harvest. This knowledge aids stakeholders in making informed decisions about available forest resources and/or harvest intensity. The widespread use of TPO estimates makes it important to understand the scale and scope of missing data in the survey. This first attempt analyzed respondent-level and question-level nonresponse (RLNR and QLNR, respectively) for Northern and Southern TPO regions, comparing response to mill profile questions (those providing general business information) to those related to mill activity (that related to wood processing information). The RLNR differed between regions, by question grouping, survey mode, and mill volume. The QLNR results for selected mill activity questions indicate that the Southern region generally has lower nonresponse than the Northern region. Parametric analysis of RLNR indicated survey mode was significant for both question groups in the Northern region whereas mill type was significant for mill activity questions in the Southern region. The QLNR parametric analysis indicated self-administered surveys in the Northern region were associated with higher nonresponse, and surveys completed by sawmills in the Southern region were associated with lower nonresponse.

Study Implications: Analysis of survey item nonresponse in the national Timber Products Output survey provides analytic background needed to assess the accuracy and completeness of the survey data. Population estimates from the data are used to monitor roundwood production and make informed decisions about forest resources. Responses varied across specific questions, and regional differences appeared to be related to survey mode. Current survey design could be improved to allow for analysis of all survey questions and assessment of the quality of responses. For nonresponse adjustment methods to be comparable across regions, factors causing observed regional difference should first be addressed.

Keywords: wood processing, roundwood production, mill residues, mill activity, timber industry

The USDA Forest Service's Forest Inventory and Analysis (FIA) program, in coordination with state natural resource agencies and university collaborators, assesses past trends, status, and future potential of forests and forest products. The Timber Products Output (TPO) group within the FIA program conducts annual surveys of primary wood processing facilities to track timber removals and their consequential impact on the forests in each state.

Information from the TPO survey is used to monitor roundwood production, reporting on volumes of roundwood products used and mill residues generated by the primary forest industry by tree species and counties of harvest. Reported TPO estimates aid policymakers, forest managers, the forest industry, and others who evaluate trends in forest product removals to evaluate impact on local and regional economies and make informed decisions about available forest resources and/or harvest intensity. Report estimates detailing the variety and volume of roundwood products used influence legislation and regulations affecting the forest industry.

Because of the widespread use of the TPO survey data and the FIA's reliance on estimation from a sample, it is important to fully understand what is behind the TPO population estimates. Of particular concern is understanding the scale and scope of nonresponse, or missing data. Nearly every survey contains instances of nonresponse. Three ways to understand the scale and scope of nonresponse include assessing individuals who fail to return a survey (unit-level nonresponse); respondents who only complete a subset of the questions asked of them (respondent-level nonresponse); and the percentage of times that a valid response is not provided for a specific question but should have been (question-level nonresponse) (Butler et al. 2021). Respondent and question-level nonresponse are collectively referred to as item nonresponse.

This study was a first attempt to analyze TPO survey item nonresponse. We evaluated respondent-level nonresponse as well as question-level nonresponse for both the Northern and Southern TPO regions. The Western TPO region was

Received: August 19, 2022. Accepted: January 12, 2023.

not included, as unedited response data were not available at the time of the study (See figure 1 for states included in regions).

Such an analysis enables us to better understand item nonresponse. Item nonresponse analysis provides information to help identify survey questions that might have higher propensity of nonresponse and, therefore, present a higher risk of information loss during TPO data reporting and evaluation. The analysis also allows us to explore the pattern of item nonresponse in the TPO survey, to determine whether nonresponse can be considered random or nonignorable (i.e., when nonresponse correlates with a variable of interest). If the data are not missing at random, imputation methods could produce biased results (van Buuren 2018). In this sense, the analysis allows identifying categories of questions that are consistently incomplete, informing recommendations for the design of future surveys to improve item response and therefore TPO estimates.

Background

Survey nonresponse, be that unit or item nonresponse, is a type of nonsampling error that will bias population estimates when such nonresponse follows a nonrandom path (i.e., when respondents and nonrespondents differ on variables of interest and the missingness cannot be explained by observed characteristics). Assessing whether missingness can be assumed to occur at random is therefore critical to support the validity of population estimates (Fulton 2018; Stocké 2006). Similarly, when question-level nonresponse is high, analysts must consider the implications of this missingness when interpreting results. The FIA TPO's recent move from a periodic census to an annual sampling of mills coupled with widespread use of data imputation (see, for example, van Buuren [2018]), necessitates careful evaluation of the scale of missing data (at the unit and item-level) prior to estimation.

Evaluating the nonresponse pattern requires information from both respondents and nonrespondents, which is usually accomplished by using known characteristics as well as exogenous data sources available for both respondents and nonrespondents (Tomaskovic-Devey et al. 1995). As summarized by De Leeuw et al. (2003), variables that can influence nonresponse include those characterizing the respondent, the interviewer, aspects of the survey form (i.e., organization, clarity, length, etc.), and mode of data collection. Research examining respondents' characteristics for surveys of businesses or organizations, although limited, indicates business size as a factor in item and unit nonresponse (Thompson and Washington 2013). Perceived interest level to survey topic (Rogelberg and Stanton 2007; Tomaskovic-Devey et al. 1995), the rank/role of the respondent within the organization, and level of staff available to respond to the survey are also mentioned as influential factors to survey participation (Fulton 2018; Wagner and Kemmerling 2010). Survey length and difficulty of questions asked (detail and sensitivity) have also been found to influence response (Rogelberg and Stanton 2007), whereas work by De Leeuw et al. (2003) suggests in-person (face to face or phone) interviews could reduce item nonresponse over self-administered mail surveys (i.e., a survey that is completed by respondents without interviewer's assistance).

TPO Survey

The TPO survey has been a part of the USDA Forest Service (Forest Service) FIA program since 1948, with information on primary wood processing mills used to complement the FIA's removals information. Information from mill surveys is used to determine type and amount of industrial roundwood received, county of origin, tree species used, and the uses for bark and other wood residues. Although national in scope, the program is administered regionally, with each of the three TPO regions (Northern, Southern, and Western regions) gathering information from their constituent states (figure 1). Although the program has a core set of questions, survey forms vary by region to accommodate regional needs (example 2019 survey forms are included in Supplement 1 and Supplement 2). Starting in 2019 (for the 2018 survey), the TPO program switched from periodic survey frequencies of all active mills that ranged from 2 to 7 years to annualized surveys of a mill sample. The Northern and Southern

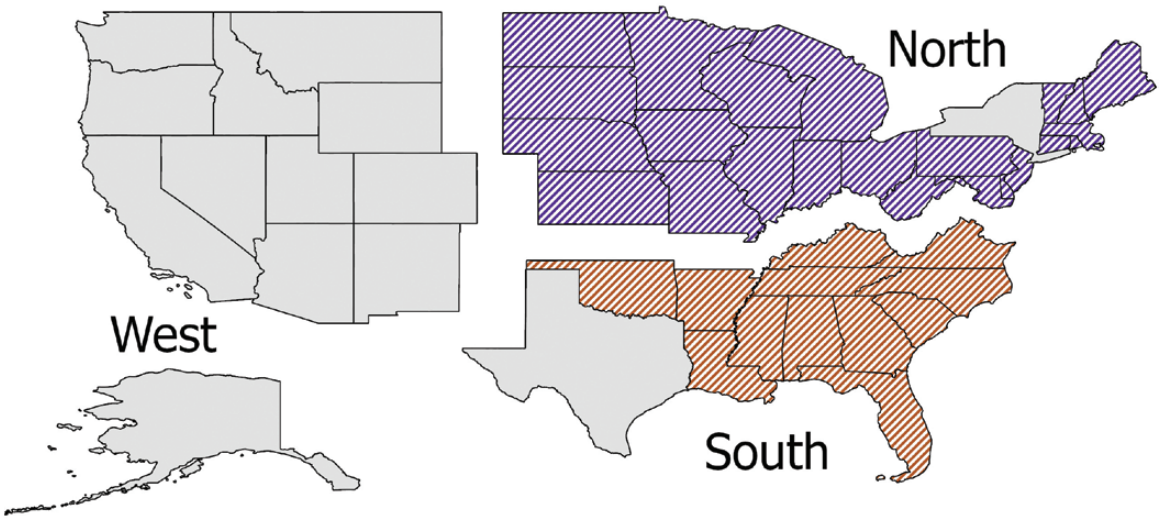

Figure 1. USDA Forest Service's Timber Products Output regional administrative divisions; shaded states with cross-hatching show participants in the 2019 survey for the Northern and Southern regions.

regions adopted the new methodology in 2019 across all participating states except Maine, and the Western region fully implemented the annual survey the following year. Annual surveys are conducted at the beginning of each year, collecting data on the preceding year. The 2019 survey data analyzed in this article is for Northern and Southern region information collected during 2020.

Samples were drawn using a methodology adapted from Coulston et al. (2018). First, the mill population was stratified by state, mill type, and measure of size (MOS). Then, an MOS threshold was determined for each mill type based on the mill type's state population. Mills at or above the established threshold were sampled with certainty (i.e., all these mills are included in the sample). Mills in states with small mill populations (i.e., 20 or fewer mills) were sampled with certainty. Within a state, mill types with a population of fewer than five mills are also sampled with certainty. All other mills were stratified further by MOS, and two mills were drawn at random from each noncertainty MOS stratum. The approach aims for a 40% sample size for each subpopulation and to capture at least 80% of a state's roundwood production. A primary wood processing facility can be classified under more than one mill type, as classification is based on the type of roundwood processed at that facility. For instance, a mill facility receiving and processing both sawlogs and pole logs will be classified as a sawmill and as a pole mill and would be included in the sawmill stratum and the pole mill stratum. Either one or both of those mill types in this mill facility could be chosen for the final sample based on the methodology outlined above.

Survey administration varies by region and depends on TPO partners' involvement in the data collection process. Although both regions use self-administered mail surveys (via post and electronic mail) as their primary survey delivery mode, nonresponse follow-up is handled differently.

In the Northern region, nonresponding mills were contacted via interviews; that is, by telephone, and survey information is asked verbally. Most state surveys were conducted by the University of Massachusetts Family Forest Research Center, administered using a modified Dillman method (Dillman et al. 2014), whereas state partners used self-administered or interview modes to survey mills in Vermont, Missouri, New York, and a subset of Minnesota mills.

In the Southern region, nonresponding mills were contacted by electronic mail and via interviews, conducted either through in-person mill visits or by telephone. Surveys were conducted by Forest Service personnel and by state agency partners.

The survey collected various information about each mill. For this study, we categorized questions into either mill profile (Profile) or mill activity (Activity) questions, to facilitate analysis and discussion. Profile questions are those providing general business information and include, for example, mill name, address, year established. Activity questions are those related to wood processing information and include, for example, roundwood amount, species type and origin. The full list of the types of questions included in each category is shown in Table 1. See Supplement 1 for the Northern region and Supplement 2 for the Southern region for exact wording of questions.

Data for Analysis

Data used for the item nonresponse analysis reported in this article come from the 2019 TPO survey administered in 2020.

Table 1. Types of survey questions assigned to mill profile and mill activity categories.

Mill profile	Mill activity
Mill name	Roundwood amount & unit of measure
Mill address	Roundwood length
Mill phone/fax	Roundwood diameter
Company name	Annual mill capacity volume & unit of measure[a]
Company address	Procurement radius[a]
Company website/email	Salvage use and volume[a]
Physical mill address	Urban wood use and volume[a]
Mill county	Species origin
Contact name/title	Species amount
Contact phone/fax/email	Number of species by origin
Mill type	Amount of product produced & unit of measure
Mill last year processed[a]	Type of product produced
Mill status	Percent dressed of sawlog product
Portable mill	Exports and volume[a]
Number of employees	Residue use: bark, coarse, shavings, sawdust
Year mill established	Residue amount: bark, coarse, shavings, sawdust & unit of measure[b]
Omit directory	Equipment
Receive report	Boiler/hog fuel use[a]
	Boiler/hog fuel volume, origin & unit of measure[b]
	In-woods chips[b]

[a]Northern region only
[b]Southern region only

This article includes item nonresponse information for the entire Northern region sample and, due to availability, for a subset of mills sampled in the Southern region, corresponding to mill responses provided using a fillable PDF form. Western data were not available in the needed format at the time of this analysis. The combined sample size is 647 mills, made up of 464 Northern mills and 183 Southern mills. This section briefly discusses unit nonresponse; specifically reporting the cooperation rate, which reflects the number of responses from eligible contacts (American Association for Public Opinion Research (AAPOR) 2016).

The Northern region survey sample had a starting sample size of 1,100 mills. One hundred and thirty mills were removed from the sample due to undeliverable addresses and closed or idle status (no roundwood consumed in 2019), for an effective sample size of 970 mills. Of these, 464 mills returned a completed or partially completed survey for a 48% cooperation rate reflecting mills across twenty-one of twenty-three of the Northern states (figure 1). New York did not participate in the 2019 TPO program, and Vermont data were unavailable in the needed format for analysis. Of these 464 mills, 254 responded by mail and 210 responded by phone.

The Southern region survey sample had a starting size of 756 mills, 39 of which were ineligible (closed or idle). Of the eligible sample, 532 mills provided a response for an overall 74% mill response, encompassing twelve of the thirteen southern states (figure 1). Texas was a nonparticipant in the 2019 program. The 183 questionnaires available for this item nonresponse analysis (covering eight of twelve participating states) were those provided using the survey's electronic PDF form either self-administered (including postal-delivered surveys transcribed to the PDF form by state partners) or via interview. Table 2 provides a summary of the number of mills included in the analysis by mill type and region.

Methods and Analysis

Item Nonresponse Coding

To assess whether a question was considered a nonresponse, we recoded the data associated with the 647 records in our analysis set based on whether the question was a "qualifying" or "nonqualifying" question for a given mill. A qualifying question is one that should have an answer; if the mill responded to the question, it is coded as 1 and if not, -1. A nonqualifying question is one not required to be answered by the mill because of the survey skip pattern (these questions are coded as -2) or because the question was never asked (these questions are coded as -3). This last case was found in the interview mode of the Northern survey, which was an abbreviated form of that region's self-administered survey. In our analysis, we only include those questions positively identified as having a qualifying response or nonresponse. For example, the Northern survey asks the mill to report the origin and volume of residues if any were used for hog fuel or industrial fuelwood; a blank in this section could mean nonresponse or that the question did not apply because no residues were used for hog fuel or industrial fuelwood. Questions such as these are left out of the analysis.

Although most questions on the TPO survey required mills to check a single box or provide a percentage, amount, or location, certain questions were "check banks" where respondents were asked to check all that apply. For purposes of this item nonresponse analysis, each check bank question was collapsed into a single variable indicating whether they checked at least one option. Nonresponses to check bank questions were coded in the manner previously explained. Qualified respondents to the mill residue question were determined based on mill type (see Supplement 3).

Table 2. Number of mills used in analysis, by mill type and region.

Mill type	Northern	Southern	Total
Sawmill	403	128	531
Veneer/plywood	11	10	21
Composite panel/ engineered wood product	0	8	8
Biomass/energy plant	8	9	17
Log home	7	0	7
Pole	2	12	14
Post	7	3	10
Miscellaneous mills[a]	26	13	39
Total	464	183	647

[a]Miscellaneous mills include bark or mulch mills, concentration/export yards, and other mills.

Methods

Respondent-level Nonresponse

We calculated the respondent-level nonresponse (RLNR) for each mill in our analysis set. To do this, we determined the number of valid questions (VQ) for each mill (m) (i.e., the number of questions the mill should have answered) and the number of unanswered valid questions (i.e., nonresponses) (NR) by that mill, NR_m. The ratio (NR_m/ VQ_m) is that mill's RLNR (Eq. 1).

$$RLNR_m = \frac{NR_m}{VQ_m} \tag{1}$$

We calculated and averaged RLNR (AV_RLNR) across mills within each category i = {Profile, Activity}, per Table 1. To calculate AV_RLNR by category i, we summed the RLNR for mills in category i and divided by the number of mills in category i (N_i) (Eq. 2).

$$AV\_RLNR_i = \frac{\sum_{m_i=1}^{N_i} RLNR_{m_i}}{N_i} \tag{2}$$

Question-level Nonresponse

The question-level nonresponse (QLNR) was determined for each question (q). We first determined the number of qualifying respondents for a specific question (N_q); that is, the sample of respondents for whom the question should have been answered. We then determined the number of times that a valid response was not provided for that specific question (QNR_q). The ratio (QNR_q/ N_q) was that question's QLNR (Eq. 3).

$$QLNR_q = \frac{QNR_q}{N_q} \tag{3}$$

Our focus for calculating QLNR was on essential elements of the survey important for understanding timber product output. As such, we report on a selected group of Activity questions that address unique regional timber product output characteristics (Table 3).

Preliminary Analysis

We conducted a preliminary analysis to determine whether Northern and Southern survey data were sufficiently different to require parametric analysis to be performed for each data set individually. Although there are similar questions asked of both regions, each region's survey was substantially different due to additional questions, question wording, or answer format. We performed two nonparametric tests to compare RLNR from the Northern and Southern regions: equality of distributions ("ranksum" in Stata17) and equality of medians ("median" in Stata17) (Mann and Whitney 1947; Wilcoxon 1945). Nonparametric tests were used given the nonnormal distribution of our nonresponse variable. The nonparametric test results support the hypothesis of variation by region, with both the Wilcoxon and equality of medians tests providing strong evidence (Prob < 0.0001) to reject the null hypothesis of equal distribution and equality of means, respectively. These results supported analyzing each region separately.

Parametric Analyses

We next looked more deeply at the respondent-level and question-level nonresponse results (RLNR and QLNR, respectively) to see whether there were systematic patterns of variability by region and to assess the random nonresponse assumption.

For the RLNR parametric analyses, we modeled mill-level RLNR, which ranges from zero (indicating full response from the mill) to 1 (indicating full nonresponse from the mill) for two subgroups of data. As described previously, each mill's RLNR is a ratio of the number of unanswered valid questions for the mill and the number of questions the mill should have answered (Eq. 1). We grouped the data such that we analyzed RLNR for each region (i.e., Northern, Southern) and question type (i.e., Profile, Activity) combination. Hence, we analyzed four RLNR models. The question types included in the RLNR are those that were common to both regions (See Table 1).

Based on De Leeuw et al.'s (2003) identified factors influencing item nonresponse, we hypothesized that for each of these four models, RLNR may vary by survey delivery mode (i.e., self-administered vs interview) and respondent characteristics—namely, mill type (i.e., whether the mill was a sawmill or not) and mill volume. For the

Table 3. Mean question-level nonresponse (QLNR) by mill activity question and region.

Question category	Question	Mean QLNR (Qualifying respondent sample size)	
		Northern region	Southern region
Roundwood	Roundwood amount	0.11 (464)	0.04 (182)
	Roundwood length	0.36 (464)	0.12 (182)
	Roundwood diameter	0.38 (464)	0.12 (182)
Species	Species origin	0.35 (464)	0.07 (182)
	Species amount	0.18 (464)	0.11 (182)
	Species by origin	0.42 (464)	0.15 (182)
Mill products	Amount of product produced	0.25 (254)	0.27 (182)
	Type of product produced	0.04 (463)	0.21 (182)
	Percent dressed of sawlog product	0.63 (213)	0.44 (54)
Mill residues	Residue use: bark	0.20 (438)	0.24 (169)
	Residue use: coarse	0.17 (423)	0.15 (150)
	Residue use: shavings	0.45 (119)	0.15 (59)
	Residue use: sawdust	0.26 (223)	0.18 (146)
Equipment	Equipment	0.03 (464)	0.06 (183)
Northern region-specific activity questions	Procurement radius	0.29 (210)	
	Annual mill capacity volume	0.13 (464)	
	Urban wood volume	0.31 (193)	Not applicable
	Salvage volume	0.31 (359)	
	Log export amount	0.30 (23)	
	Finished product export amount	0.35 (254)	
Southern region-specific activity questions	In-woods chips		0.04 (183)
	Residue amount: bark		0.31 (169)
	Residue amount: coarse		0.17 (150)
	Residue amount: shavings	Not applicable	0.17 (58)
	Residue amount: sawdust		0.24 (146)
	Boiler/hog fuel volume		0.14 (7)
	Boiler/hog fuel origin		0.29 (7)

Southern region models, we also hypothesized potential effects from data collector characteristics, given that data collection was carried out in part by partnering states. As such, we included a state indicator as a proxy for data collector characteristics.

For the dependent variable in each of the four RLNR models, we assumed that zero values and (0,1) proportions were generated by the same data process. We analyzed these data using a generalized linear model with a logistic link function and a binomial family (McCullagh and Neldor 1989). Further, for the Southern analysis, we applied a likelihood ratio test to select between a full model including all identified covariates and a restricted model, excluding state indictors. With a *P*-value of 0.998, we could not reject the test's null hypothesis that the smaller model fit our data better. Therefore, our final specification for all four models included only survey delivery mode, mill type, and mill volume (labeled self-administered, sawmill, and mill volume, respectively). See Table 4 for sample and subsample statistics for these variables.

For the QLNR models, we focus on the fourteen shared Activity questions in Table 3 and examine whether the explanatory factors described above significantly affected the likelihood that qualified respondents provided an answer to each modelled question. Based on the preliminary analysis, we grouped the data based on region and ran models for each of the fourteen Activity questions for a total of twenty-eight separate models. Here, for each mill that was qualified to answer the question, we had a dependent variable that reflected whether the question was answered or not (1/0). (For this analysis, we conformed to the standard that providing a response is "1" and nonresponse is a "0". Although this approach appears to examine question-level response, we are simultaneously examining the inverse, QLNR.) We explored whether the probability of qualifying mills' responses to each Activity question varied by respondent mill information: survey delivery mode, mill type, and mill volume. The dependent variable for these models is binary, corresponding to whether a mill responded to the question or not; as such, we analyzed the models using a logistic regression

$$Prob_q\left(R_{qm}=0,1 \mid X_m\right)=\pi_q=\frac{\exp(\beta_0+\beta_1 X_m)}{1+\exp(\beta_0+\beta_1 X_m)} \quad (4)$$

where q = each Activity question, R_{qm} is the response to the q-th Activity question (1 if answered and 0 otherwise) by respondent m, and X_m, a vector of observed variables for each mill respondent m that includes an indicator for sawmills (=1 if a sawmill and 0 otherwise), an indicator for survey delivery method (=1 if the survey was self-administered and 0 if interviewed), and the reported volume of mill receipts.

Results

RLNR

The AV_RLNR across Profile and Activity category questions shows opposite results between the North and South. The Northern AV_RLNR results indicate much lower nonresponse rates for Profile questions than for Activity questions, whereas Southern AV_RLNR results indicate much lower nonresponse rates for Activity questions than Profile questions in the survey (figure 2).

The AV_RLNR by mode in the North shows Activity question nonresponse to be higher than Profile question nonresponse for both survey modes. For the self-administered mode, the AV_RLNR for Activity questions is 29% and AV_RLNR for Profile questions is 18%. For the interview mode, AV_RLNR for Activity questions is 13% versus an AV_RLNR of 6% for Profile questions (figure 3A). In the South, the AV_RLNR by mode also shows differences between question categories, but higher nonresponse for interview versus self-administered surveys. For the self-administered mode, AV_RLNR for Activity questions is 15% and AV_RLNR for Profile questions is 32%. For the interview mode, AV_RNLR for Activity questions is 22% versus an AV_RLNR of 36% for the Profile questions (figure 3B).

Figure 4 shows the distribution of RLNR for Profile questions overlaid with RLNR for Activity questions by region. The overlap of results between Profile questions and Activity questions for the North (figure 4A) confirms the result found in the average RLNR's for roughly the entire frequency distribution. The distribution consistently shows greater response (lower nonresponse) for Profile questions over Activity questions. The frequency distribution of the RLNR for Profile questions compared with that of the Activity questions for the South clearly shows differences between question categories (figure 4B). Although the distribution of nonresponse leans further towards the left (lower nonresponse) for the Southern subset of Activity questions than Profile questions, we observed the highest nonresponse (over three quarters of questions unanswered) corresponding to Activity questions.

The relationship between mill volume and RLNR for the Activity questions by survey mode shows differences across regions. In the North, all mills larger than 10,000 thousand cubic feet (MCF) that responded to the self-administered survey completed between roughly 80% to 100% of the Activity questions. In general, in-person survey responses for the Activity questions from mills smaller than 10,000 MCF were more complete than their self-administered survey counterparts. With only five exceptions, the mills with nonresponses 50% or greater were self-administered survey respondents (figure 5A). Approximately 9% of all Northern

Table 4. Summary statistics of independent variables used in parametric analyses.

Variable label	Both regions		Northern region		Southern region	
	n	Mean (standard deviation [SD])	*n*	Mean (SD)	*n*	Mean (SD)
Self-administered (yes=1, no=0)	634	0.5 (0.5)	464	0.5 (0.5)	170	0.5 (0.5)
Sawmill (sawmill=1, else=0)	647	0.8 (0.4)	464	0.9 (0.3)	183	0.7 (0.5)
Mill volume (continuous, green tons)	647	2570.3 (6243.2)	464	978.9 (3041.4)	183	6605.4 (9591.0)

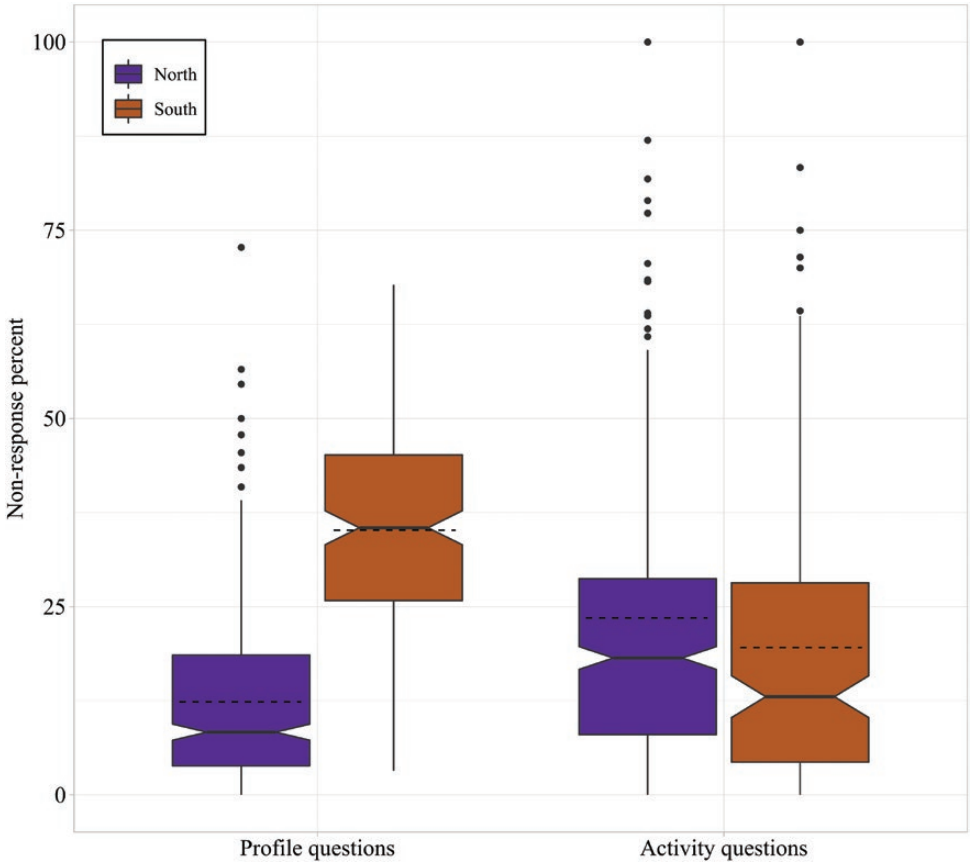

Figure 2. Respondent-level nonresponse for Profile and Activity questions by region. Boxplot shows median, hinges (25th and 75th percentiles), and whiskers (1.5 multiple of interquartile range). Dashed line indicates AV_RLNR (i.e., mean).

mills responding to the survey had RLNR for Activity questions 50% or greater (42/464).

In the South, RLNR for Activity questions was not concentrated by mill volume or survey mode. Graphical representation of the data indicate that Southern region mills are dispersed across a range of mill volume sizes; although many mills were concentrated below the 5,000 MCF level like with the Northern region, many more mills were distributed between 5,000 MCF and 50,000 MCF. Results indicate that the South had fewer mills than the North with RLNR 50% or greater (figure 5B). Note that survey mode was unreported by data collectors for thirteen mills in the Southern region.

Despite the overall lower RLNR for Activity questions (Northern region self-administered survey excluded), some mills provided very limited data. A closer look at the data itself (not provided here for confidentiality reasons), shows nine Northern mills and six Southern mills had Activity question RLNRs of 75% or greater. In the Northern region, all but 1 were sawmills; 3 had mill volumes between 3 and 9 MCF and 6 had mill volumes between 230 and 5,540 MCF. In the Southern region, all six were sawmills with mill volumes between 670 and 45,000 MCF.

QLNR

Results for a selected group of shared Activity questions by region correspond to the overall result that the South had lower nonresponse to these questions than the North. Across these shared Activity questions, the Northern QLNR ranged from 3% to 63%, whereas the Southern QLNR ranged from 4% to 44% (figure 6). Table 3 shows the QLNR for each Activity question in the analysis.

For the North, QLNR was 25% or greater for more than half of the shared Activity questions evaluated, whereas for the South, QLNR exceeded 25% only for two questions, both related to mill products produced (amount produced and percent dressed). For both North and South, the greatest nonresponse came from "Percent dressed of sawlog product," unanswered 63% of the time for the North and 44% of the time for the South. In the North, the lowest QLNR (3%) was to the "Equipment" question; in the South, the lowest QLNR (4%) was to the "Roundwood amount" question (figure 6).

For selected regional-specific Activity questions, the QLNR exceeded 25% for all Northern region questions except "Annual mill capacity volume." Mills in the North most often did not provide the finished product export amount, with a QLNR of 35%. For the Southern region questions, QLNR exceeded 25% for "Residue amount: bark" (31%) and "Boiler/hog fuel origin" (29%). Figure 7 and Table 3 provide details.

Parametric Analysis

The effect of observed mill and survey characteristics on RLNR varied by region and, for the South, it also varied by the set of questions analyzed. For the North (figure 8A), survey mode (self-administered indicator) was statistically significant when considering Profile questions as well as Activity questions. Respondents using a self-administered survey had

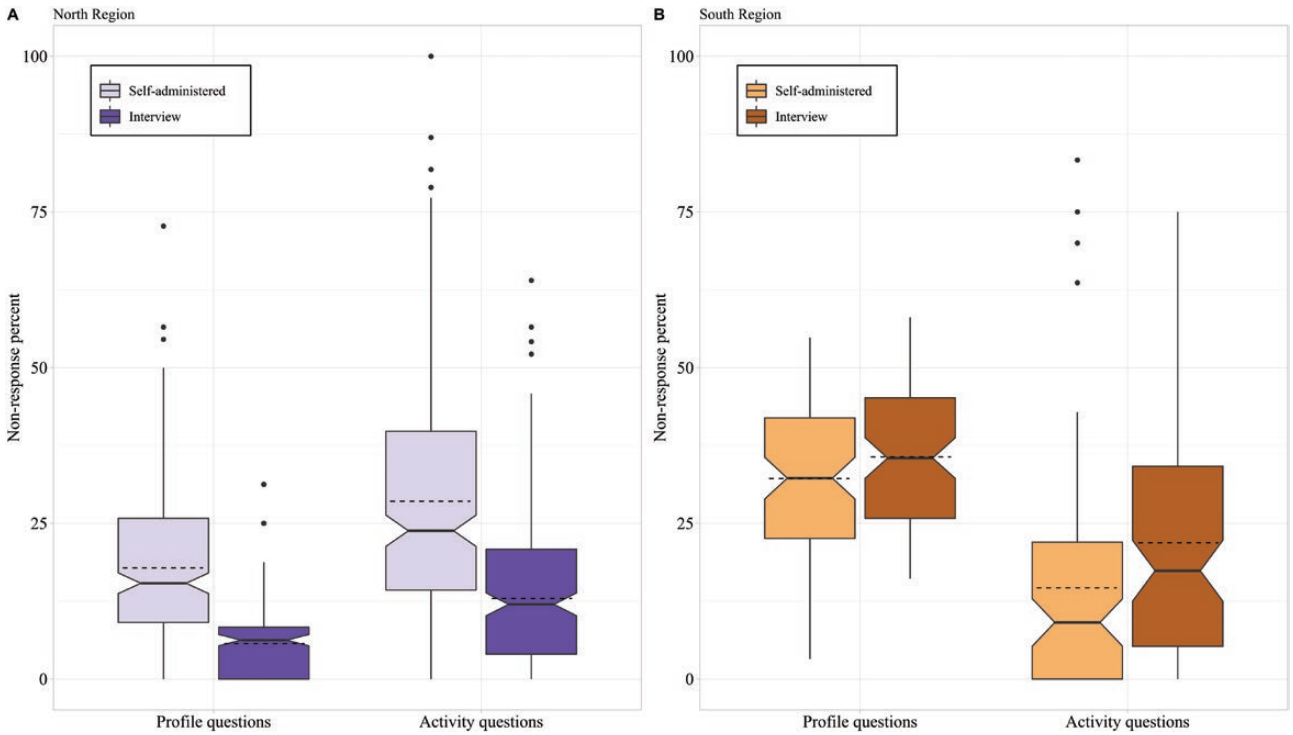

Figure 3. Respondent-level nonresponse for Profile and Activity questions by mode and region: Northern region (A) and Southern region (B). Dashed line indicates AV_RLNR (i.e., mean).

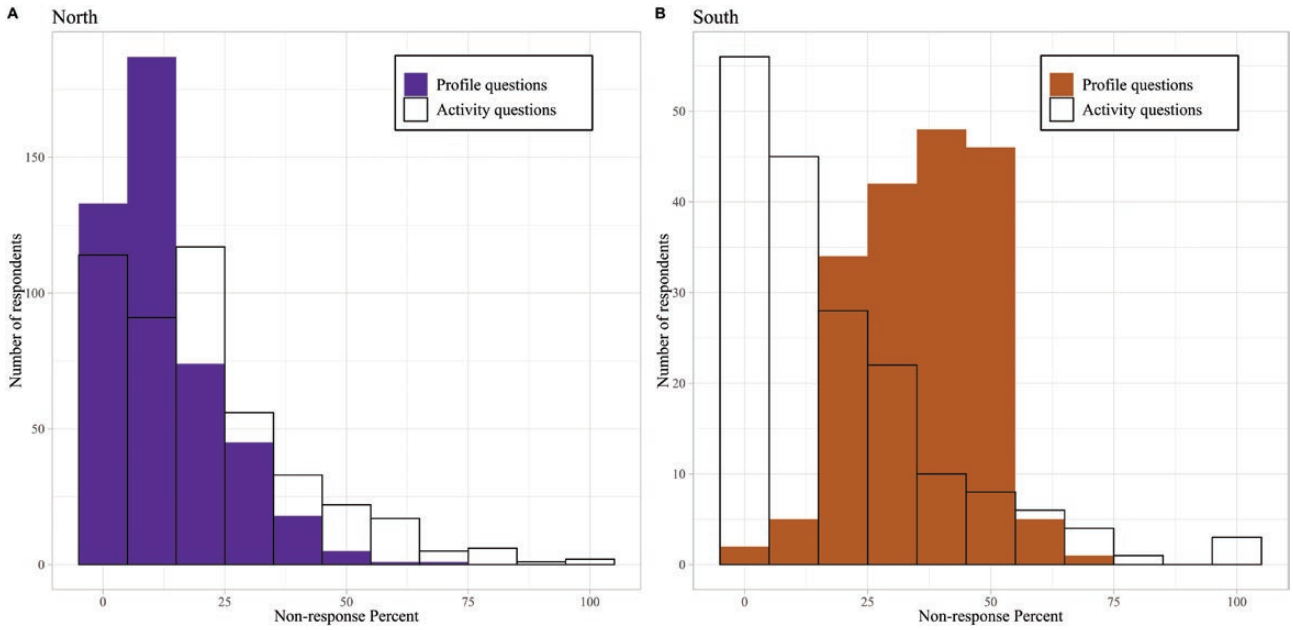

Figure 4. Respondent-level nonresponse percentages by question category for Northern region (A) and Southern region (B).

odds of nonresponse roughly 3.6 and 2.7 times higher than interviewed respondents, for Profile and Activity questions, respectively. None of the variables were statistically significant when looking at Profile questions for the South, whereas a sawmill respondent decreased the odds of nonresponse for the Activity questions. In fact, the odds of nonresponse by other mill types were roughly three times higher than that of sawmills (1/0.325 = 3.1; see Table 5, figure 8B). Marginal results for the RLNR parametric models are provided in Supplement 4.

The QLNR models analyzing the likelihood of response to individual Activity questions, given observed mill and survey characteristics, resulted in statistically significant effects from the survey delivery mode indicator for the North. With self-administered surveys, all but one question showed decreasing odds of response (Table 6). We observe a 114% increase in odds of response to the county (species origin) question, yet species percent (a related question) shows decreasing odds of response (or higher nonresponse) with a self-administered survey. With sawmills, the odds of response to the type of products

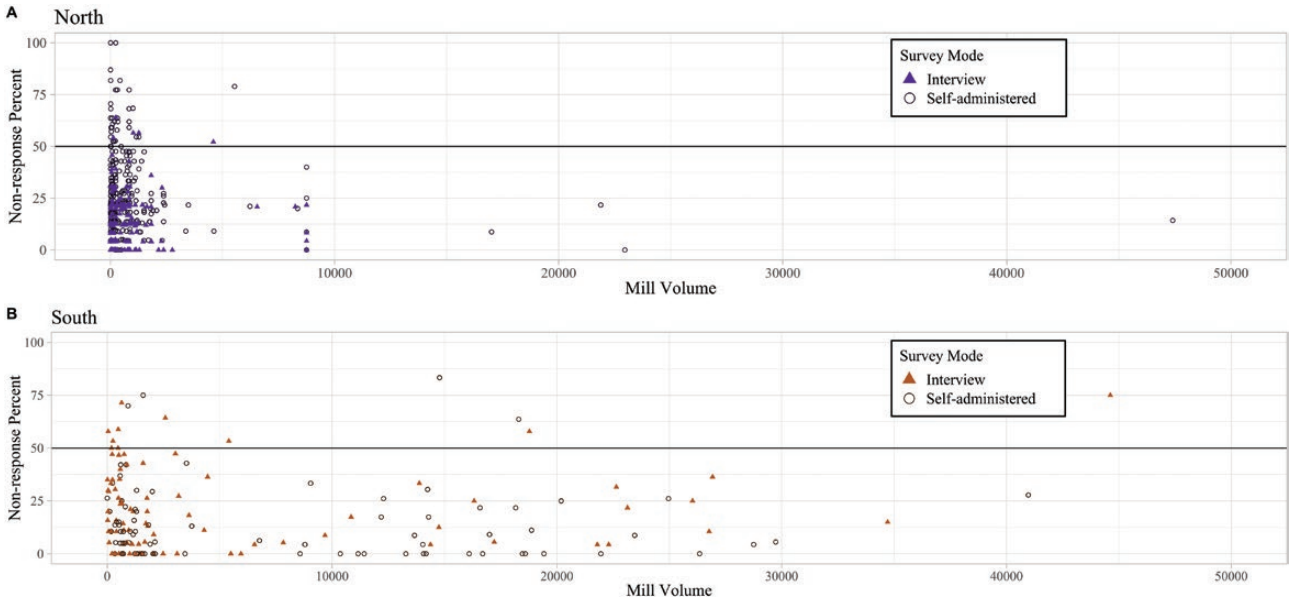

Figure 5. Respondent-level nonresponse percentages for Activity questions by mill volume (MCF) for Northern region (A) and Southern region (B).

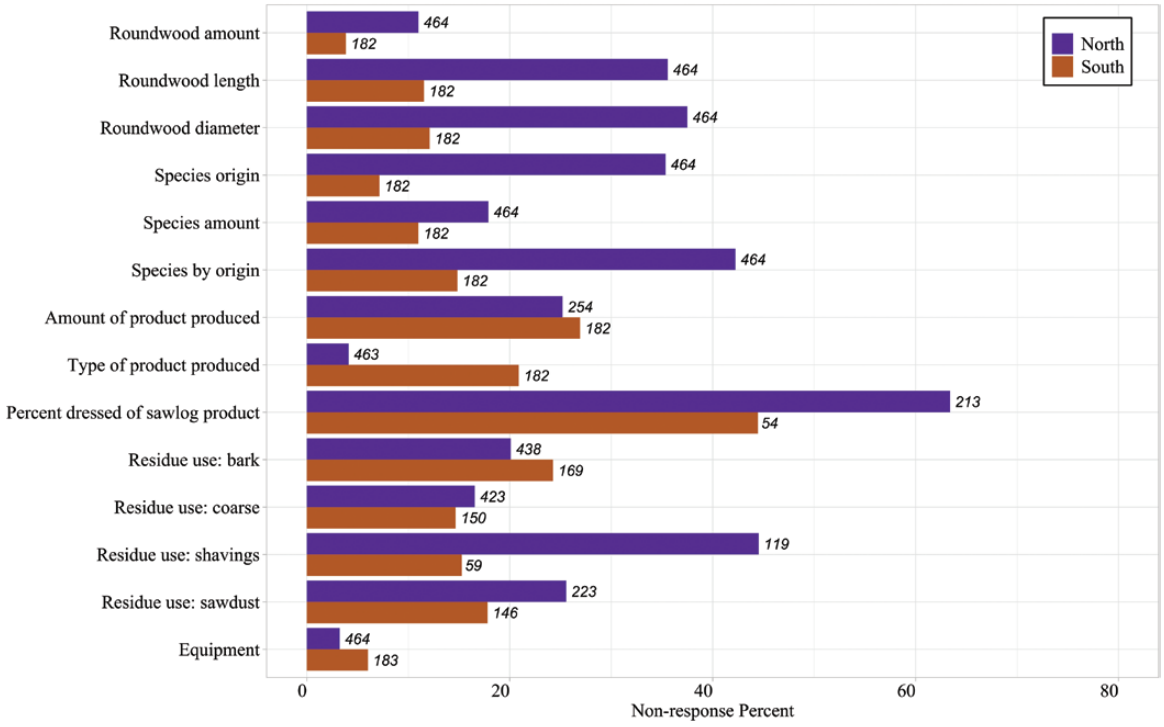

Figure 6. Question-level nonresponse percentages for shared Activity questions for the Northern and Southern regions. Qualifying respondent sample size indicated in italics next to the corresponding bar.

produced question was over four times higher than other mill types, but there was a decrease in the odds of a response to the bark residue use question (odds of other mill types to respond to the bark residue question were over three times higher than that of sawmills). Although mill volume was statistically significant for three of the fourteen Activity questions analyzed, the magnitude close to one indicates no effect.

Southern region data displays a different pattern (Table 6). Self-administered surveys increased the odds of response for the percent of species by origin question over three times but did not show a statistically significant effect for any other Activity question examined. With sawmills, odds of response increased significantly for over half of the examined questions. Mill volume was statistically significant for only two questions (percent of species by origin and use of shavings residue). However, as with Northern survey questions, odds ratios for mill volume were almost 1, indicating no effect.

Comparing results across datasets shows sawmills increasing the odds of response to the product type question in both surveys, with sawmill respondents 4.4 and 3.8 times more likely to respond in the North and South, respectively.

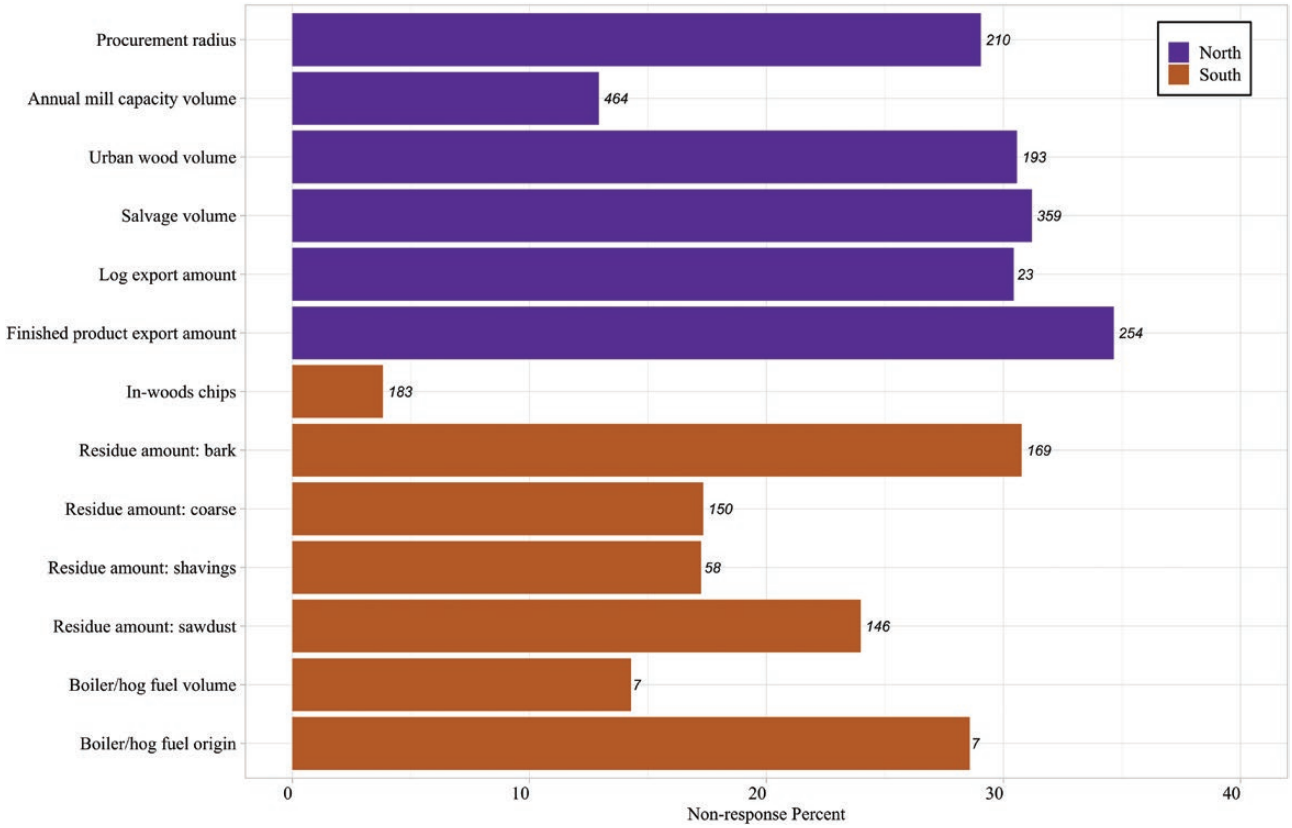

Figure 7. Question-level nonresponse percentages for selected region-specific Activity questions for Northern and Southern regions. Qualifying respondent sample size indicated in italics next to the corresponding bar.

Marginal results for the QLNR parametric models are provided in Supplement 4.

Discussion

Our approach to collating TPO survey data and analyzing item nonresponse provides a means to better understand item nonresponse in general for the North and South TPO regions and allows us to make comparisons across regions. The survey mode effect in the Northern region was statistically significant (at a 95% confidence level) for both Profile and Activity questions, indicating higher odds for nonresponse when surveys were self-administered rather than delivered through interviews. Although the finding appears to support the conclusions of De Leeuw et al. (2003), we draw readers' attention to the potential confounding effects due to differences in the instruments used for data collection (see Supplement 1 which shows that the interview survey is an abbreviated version of the self-administered questionnaire.). That notwithstanding, the interview effects on item nonresponse suggest that the presence of the interviewer may be eliciting some social desirability bias in the survey-taking process, or it might be a result of the interviewer providing clarification to the respondent, thereby enabling them to better answer the questions. In the Southern region, survey mode differences were not significant (95% confidence level) for either category of questions, with RLNR from interviewed respondents not much different from the self-administered RLNR. The Southern region used the same set and number of questions for both modes (see Supplement 2); however, the self-administered survey form is a fillable PDF that provides some guidance to survey respondents. This fillable PDF guidance might influence the lack of significant difference between the results associated with the two survey delivery modes.

When graphically assessing RLNR by mill volume, an interesting survey pattern emerges. In the North, mills with the largest volume (over 10,000 MCF) responded to the self-administered survey. That is, they responded to the TPO survey without follow-up prompting. In addition, these larger mills also tended to provide a more complete form (with 22% or less of the Activity questions left unanswered). Although the Northern pattern could be indicative of unobserved characteristics that favor larger mills' response compared to their smaller counterparts (e.g., more infrastructure, resources, and/or interest to respond to the survey questions), a similar pattern was not observed in the South, where response rate from larger mills was similar for both survey modes. Results from the empirical analysis show no statistical evidence of mill volume effects on the odds of nonresponse for either region or question type category.

Responses to Activity questions are critical to the TPO program. For TPO estimates, several Activity questions are imputed if they are left blank, so the more complete these are, the more accurate the TPO estimates. Key questions, including "Roundwood amount," "Species amount," "Species origin" (in the North only), and "Species by origin," as well as all "Residue use" for both North and South were left unanswered more than 10% of the time and likely led to reduced accuracy of the TPO reports by at least that much, or in some cases potentially much more.

However, we found that item nonresponse was random, conditional on our observed covariates. In the Southern region, nonresponse to some survey questions depended on

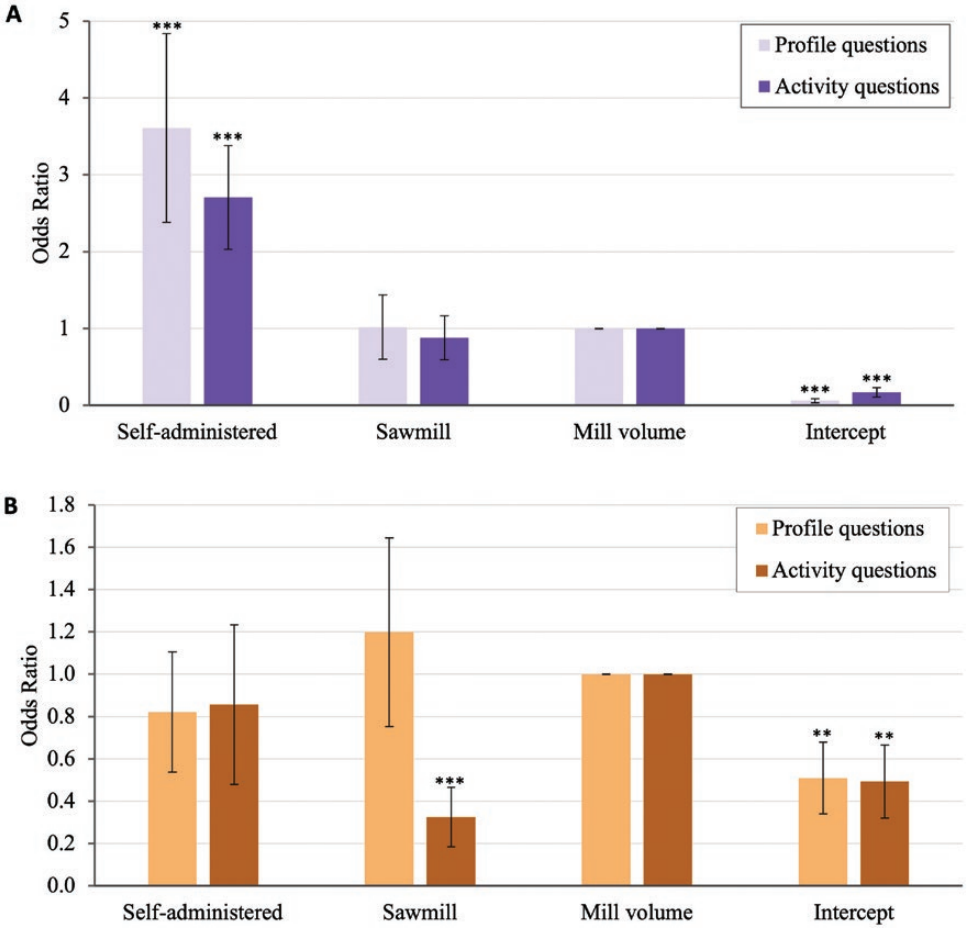

Figure 8. Odds ratio of respondent-level nonresponse covariates with standard error bars, for Profile and Activity questions by Northern region (A) and Southern region (B). Asterisks denote significance level (*$P \leq 0.1$, **$P \leq 0.05$, and ***$P \leq 0.01$).

Table 5. Odds ratios from respondent-level nonresponse generalized linear model regressions for Northern and Southern regions by mill profile and mill activity group questions.

Variable	North		South	
	Profile Questions	Activity Questions	Profile Questions	Activity Questions
Self-administered	3.610*** (0.000)	2.706*** (0.000)	0.822 (0.569)	0.857 (0.725)
Sawmill	1.017 (0.967)	0.879 (0.692)	1.199 (0.627)	0.325*** (0.009)
Mill volume	1 (0.746)	1 (0.449)	1 (0.874)	1 (0.784)
Intercept observations	0.060*** (0.000) 464	0.171*** (0.000) 464	0.509** (0.042) 170	0.494** (0.043) 170
Akaike information criterion (AIC)	0.5444148	0.7569537	0.91691	0.7281487
Bayesian information criterion (BIC)	-2778.751	-2744.67	-841.9774	-819.4834

P-values in parentheses; *$P \leq 0.1$, **$P \leq 0.05$, ***$P \leq 0.01$.

the mill type (sawmill indicator) whereas in the Northern data, survey mode emerged as a significant explanatory variable. Mill volume, the primary variable of interest for TPO, was found either not significant (95% confidence) or with neutral effect on nonresponse (odds ratio=1). Therefore, item missingness can be considered missing at random, where survey-item missingness can be explained by observed covariates but is not dependent on the variable of interest.

Although the Equipment category appeared to be a silver lining in the data collection of both regions, it is important to note that this question, like type of product produced in the Northern survey and mill residue use, is a check bank question. If the mill checked at least 1 item in the list, then the question was considered answered. Currently, we are unable to assess the quality of the answers associated with these types of questions. If we were, it is likely that the QLNR for these questions would be higher. We suggest that check bank questions could be revised to have a forced response that enables analysts to understand if any of the categories in the check bank are relevant. For example, ask whether the mill

Table 6. Odds ratios from logistic regressions of question-level response by region[a].

Dependent variable	North				South			
	Variable	Odds [SE] [b]	*n*	R^2 [c]	Variable	Odds [SE][b]	*n*	R^2 [c]
Roundwood amount	Self-administered	0.44*[0.1]			Self-administered	0.88[1.2]		
	Sawmill	0.80[0.4]	464	0.047	Sawmill	4.90[6.4]	169	0.122
	Mill volume	1.00*[0.0]			Mill volume	1.00[0.0]		
Roundwood length	Self-administered	0.04***[0.0]			Self-administered	0.60[0.3]		
	Sawmill	0.66[0.2]	464	0.291	Sawmill	4.96**[2.9]	169	0.078
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Roundwood diameter	Self-administered	0.06***[0.0]			Self-administered	0.47[0.3]		
	Sawmill	0.73[0.2]	464	0.253	Sawmill	6.22**[3.7]	169	0.097
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Species origin	Self-administered	2.16***[0.4]			Self-administered	0.77[0.5]		
	Sawmill	1.45[0.4]	464	0.028	Sawmill	0.76[0.6]	169	0.043
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Species amount	Self-administered	0.08***[0.0]			Self-administered	0.74[0.4]		
	Sawmill	0.95[0.3]	464	0.150	Sawmill	2.08[1.2]	169	0.039
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Species by origin	Self-administered	1.29[0.2]			Self-administered	3.04*[1.6]		
	Sawmill	1.22[0.3]	464	0.005	Sawmill	0.75[0.4]	169	0.095
	Mill volume	0.98[0.3]			Mill volume	1.00**[0.0]		
Product amount	Self-administered	1.00[.][d]			Self-administered	1.53[0.6]		
	Sawmill	2.61*[1.0]	254	0.055	Sawmill	7.00***[2.9]	169	0.164
	Mill volume	1.00*[0.0]			Mill volume	1.00[0.0]		
Product type	Self-administered	0.23*[0.2]			Self-administered	1.94[0.9]		
	Sawmill	4.97***[2.5]	463	0.109	Sawmill	3.85**[1.6]	169	0.101
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Percent dressed	Self-administered	1[.][d]			Self-administered	0.88[0.6]		
	Sawmill	1[.][d]	213	0.002	Sawmill	1[.][d]	52	0.060
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Residue use: bark	Self-administered	0.03***[0.0]			Self-administered	2.01[0.9]		
	Sawmill	0.25*[0.2]	438	0.241	Sawmill	0.76[0.4]	156	0.052
	Mill volume	1.0[0.0]			Mill volume	1.00[0.0]		
Residue use: coarse	Self-administered	0.07***[0.0]			Self-administered	0.61[0.4]		
	Sawmill	1.15[0.7]	423	0.161	Sawmill	14.13***[8.9]	138	0.189
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		
Residue use: shavings	Self-administered	1[.][d]			Self-administered	0.24[0.2]		
	Sawmill	1.59[1.3]	119	0.019	Sawmill	1[.][d]	57	0.307
	Mill volume	1.00[0.0]			Mill volume	1.00**[0.0]		
Residue use: sawdust	Self-administered	1[.][d]			Self-administered	0.58[0.3]		
	Sawmill	14.77**[14.7]	223	0.068	Sawmill	3.89*[2.7]	134	0.102
	Mill volume	1.00*[0.0]			Mill volume	1.00[0.0]		
Equipment	Self-administered	0.09*[0.1]			Self-administered	0.96[0.8]		
	Sawmill	5.74***[3.1]	464	0.155	Sawmill	6.96*[6.2]	170	0.105
	Mill volume	1.00[0.0]			Mill volume	1.00[0.0]		

[a]Dependent variable coding: 1=answered question, 0=nonresponse.
[b]SE=Standard Error; Significance: * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.
[c]McFadden R^2.
[d]Variable omitted due to collinearity.

has a piece of equipment or not (yes/no boxes) rather than allowing it to check "all that apply." The same could be done for questions related to product produced and mill residue use.

Although item nonresponse was high for some mills (as noted in the results section), analyses of available survey data showed an overall low RLNR, with item nonresponse centered mostly around questions that help identify mills (Profile questions) rather than questions addressing mill operations (Activity questions). Activity questions are used to generate TPO estimates; therefore, when not available, those data were

imputed using previously known data, exogenous information, or a combination of both.

Conclusion

Analysis of item nonresponse provides information to assess patterns and potential bias in our estimates. Users of TPO data for other applications (e.g., carbon studies) might focus on specific questions and how accurate or complete the answers to these questions are (e.g., residue use). We were also able to identify areas where modifications could lead to improved response; for example, revising check bank questions. Although our analysis provides valuable information to the TPO program, current survey design does not allow identification of eligibility for all survey questions nor does it allow us to identify the quality of responses. Results from our analysis could differ significantly if all questions had been available for analysis or if we had been able to measure and incorporate validity of responses for all questions. We also analyzed a specific subset of the surveyed mills for the Southern region, which might not represent the full sampled set. Future research is needed to evaluate the significance of those factors on our item nonresponse rates. Further, future research should aim to include Western region item nonresponse to facilitate a national comparison.

The effect of data collection mode is unclear, with interview mode a significant factor for increasing odds of response for the Northern region but no statistical significance (95% confidence level) for Southern region response. This disparity between regions could be an effect of the survey delivery mode, survey format (in the case of the Southern region), or other unexplained differences. Further analysis of follow-up modes and the impact they might have on item nonresponse is needed.

Data imputation generates biased estimates when nonresponse depends on variables of interest. In the TPO case, the volume a mill consumes (mill receipts or mill volume, as labeled in our analyses) constitutes key information. If item nonresponse depends on mill volume, we can conclude nonresponse to be nonrandom and nonignorable in nature. Results from our parametric analyses indicate mill volume as not statistically significant across respondents. However, when examining each question response pattern, we found mill volume statistically significant but with odds of 1.0, indicating essentially the same effect for both responses and nonresponses. Although evidence of random item nonresponse allows for methods to estimate those missing data points to provide TPO estimates, likely unbiased by item nonresponse, the strength of question-level imputation methods currently used by TPO is a topic for future analysis.

Analysis of item nonresponse can help researchers identify patterns that could lead to survey instrument improvements. Further, analyzing item nonresponse can help identify the best method to calculate population estimates from the sample and interpret results. However, research assessing methodology to address item nonresponse in the TPO surveys needs to consider the observed regional differences in item nonresponse. For TPO nonresponse adjustment methods to be consistent and comparable across regions, factors causing the observed regional difference should be addressed. In addition, these methods would need to account for unit nonresponse to the survey, which is a topic currently under study.

Acknowledgements

The authors gratefully acknowledge comments on the manuscript provided by James Bentley, Erik Berg, Chelsea McIver, Todd Morgan, Ron Piva, and Dillon Thale.

Funding

Funding support for this project came from the USDA Forest Service, Northern Research Station (Agreement No. 20-JV-11242305-078).

Conflict of Interest

None declared.

Supplementary Materials

Supplementary data are available at *Forest Science* online.

Literature Cited

American Association for Public Opinion Research (AAPOR). 2016. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys*. Oakbrook Terrace, IL: AAPOR. https://www-archive.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf.. Date accessed January 25, 2023.

Butler, B.J., S.M. Butler, J. Caputo, J. Dias, A. Robillard, and E.M. Sass. 2021. *Family forest ownerships of the United States, 2018: Results from the USDA Forest Service, National Woodland Owner Survey*. Madison, WI: USDA Forest Service, Northern Research Station. https://doi.org/10.2737/NRS-GTR-199.

van Buuren, S. 2018. *Flexible Imputation of Missing Data*. Boca Raton, FL: CRC Press. 415 p. https://www.routledge.com/Flexible-Imputation-of-Missing-Data-Second-Edition/Buuren/p/book/9781138588318.

Coulston, J.W., J.A. Westfall, D.N. Wear, C.B. Edgar, S.P. Prisley, T.B. Treiman, R.C. Abt, *et al*. 2018. "Annual Monitoring of US Timber Production: Rationale and Design." *Forest Science* 64 (5): 533–43.

De Leeuw, E.D., J. Hox, and M. Huisman. 2003. "Prevention and Treatment of Item Nonresponse." *Journal of Official Statistics* 19 (2): 153–76.

Dillman, D.A., J.D. Smyth, and L.M. Christian. 2014. *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 4th ed. Hoboken, NJ: Wiley & Sons. 528 p.

Fulton, B.R. 2018. "Organizations and Survey Research: Implementing Response Enhancing Strategies and Conducting Nonresponse Analyses." *Sociological Methods & Research* 47 (2): 240–76.

Mann, H.B., and D.R. Whitney. 1947. "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other." *Annals of Mathematical Statistics* 18 (1): 50–60.

McCullagh, P., and J. Neldor. 1989. *Generalized Linear Models*. London, UK: Chapman and Hall.

Rogelberg, S.G., and J.M. Stanton. 2007. "Introduction: Understanding and Dealing with Organizational Survey Nonresponse." *Organizational Research Methods* 10 (2): 195–209.

Stocké, V. 2006. "Attitudes Towards Surveys, Attitude Accessibility and the Effect on Respondent's Susceptibility to Nonresponse." *Quality & Quantity* 40 (2): 259–88.

Thompson, K.J., and K.T. Washington. 2013. "Challenges in the Treatment of Unit Nonresponse for Selected Business Surveys: A Case Study." *Survey Methods: Insights from the Field* 1–16.

Tomaskovic-Devey, D., J. Leiter, and S. Thompson. 1995. "Item Nonresponse in Organizational Surveys." *Sociological Methodology* 25: 77–110.

Wagner, S.M., and R. Kemmerling. 2010. "Handling Nonresponse in Logistics Research." *Journal of Business Logistics* 31 (2): 357–81.

Wilcoxon, F. 1945. "Individual Comparisons by Ranking Methods." *The Biological Bulletin* 1 (6): 80–3.