

TESTING A NEW RESPONSE MODE FOR AN ESTABLISHMENT SURVEY: REPORTING RESPONSE RATES IN A PILOT STUDY OF TOUCHTONE RESPONSE

Karen L. Goldenberg*

U.S. Bureau of Labor Statistics, 2 Massachusetts Ave. N.E., Room 4985, Washington, DC 20212

Keywords: Touchtone data entry; AAPOR standard definitions; response rate computation

1.0 Introduction

The U.S. Bureau of Labor Statistics (BLS) conducts the Annual Refiling Survey (ARS) each year. During the ARS, state agencies mail questionnaires to approximately one third of the establishments registered with their state Unemployment Insurance programs. These questionnaires are sent to sampled establishments for verification or updating of information in state records.

In an effort to reduce the data collection workload, BLS conducted a five-state pilot test in 2002 to see if preselected respondents would self-respond using a touchtone telephone system. Successful touchtone responses eliminate the need for mail returns, with the potential for substantial savings in postage and labor. Therefore, another component of the pilot test was an embedded experiment to assess whether various contact approaches would increase TRS participation. The intent was to measure the effectiveness of each communication by looking at the "response rate," defined as the percentage of units using TRS.

During analysis of the data, the meaning of the response rate definition became less clear. Reported response rates usually reflect a single response mode, or the total response to a survey when one or more modes are used in nonresponse followup. That is not the case when the variable of interest is response to one component of a multi-modal survey, but not all respondents are eligible to use this mode of response.

The purpose of this paper is to explore response rates for a multi-mode establishment survey, and to use AAPOR's *Standard Definitions* (2004) as a framework for discussion. I begin by raising issues that apply to response rates for establishment surveys and that are different from surveys of households or individuals. I describe the Annual Refiling Survey, which illustrates many of the above issues, and the Touchtone Response System Pilot Test, conducted as part of the ARS in 2002. Since the focus of this paper is response rates, I turn to the AAPOR presentation for mail surveys of named individuals (AAPOR 2004). I attempt to extend the AAPOR model to the ARS, which is a mail survey of named businesses. Finally, I use both BLS and AAPOR response rate formulas to examine the results of the TRS Pilot Test and the embedded experiment.

*The opinions expressed in this paper are those of the author and do not reflect official policy of the Bureau of Labor Statistics. The Touchtone Response System Pilot Test represents the combined efforts of Michael Searson, Branch Chief, and team members Sudarshan Jakhu, Mark Sauer, Linda Unger, James Walker, and the author, with assistance from Vinod Kapani. Gordon Mikkelson and Amishi Dedhia provided access to additional data while this paper was being written. Tom W. Smith and William Mockovak provided helpful comments. All errors are the responsibility of the author.

2.0 Computing response rates for establishment surveys

Establishment surveys have a number of attributes that complicate response rate calculations. They include:

- *Definition of the sample unit.* Surveys of households have generally accepted definitions, with rules to determine who is a member of that household. Establishments tend to be more complex. At the most basic level, an establishment consists of a single unit performing a single type of economic activity at a single location. When a business or organization grows, it may expand to other locations. When other locations are considered a part of the same administrative unit, we consider the business a multi-establishment firm or "multi." Large organizations, especially corporations that cross political boundaries such as states, may be organized as "enterprises." Establishment survey sample units may be defined at any of these levels.
- *Unit count versus units weighted by employment.* Sample units in a household survey generally contribute equally to the population of interest, within sampling strata. This is not true for the establishment population, which is highly skewed. BLS data on private industry show that 60 percent of establishments in 2002 had fewer than 5 employees, and those establishments employed 6.6 percent of U.S. workers. In contrast, establishments with 500 or more workers accounted for 0.2 percent of all establishments, but 18.0 percent of workers, for the same year. (BLS 2003a, Table 3). Survey organizations may define response rates in terms of the number of units, units weighted by employment, or both.
- *Mandatory versus voluntary surveys.* Responses to some government agency surveys are required by law, although penalties for nonparticipation are rare. Some establishments only participate in mandatory surveys. In the U.S., reported response rates are higher for mandatory Federal surveys (Osmint et al., 1994).
- *One respondent, multiple establishments.* Many enterprises and large, multi-establishment businesses maintain the data requested for a survey at a centrally located headquarters or administrative office. In large surveys, one respondent could easily receive requests for data about several—or several hundred—sample units. While larger firms maintain data in automated systems, the burden on any one respondent to extract those data becomes far greater than the burden associated with a single establishment. In addition, these respondents may be at a distance from the sample unit and can speak only to the information available in business records. Nonresponse at large "multis" has a greater impact on estimation.
- *Third party respondents.* A variation on the 'one respondent, multiple establishment' theme arises for data held and reported by third parties, such as accountants and payroll services. Surveys associated with the collection of administrative data are particularly vulnerable. Like respondents

from large multis, third parties are removed from the sample unit and may have limited knowledge about it apart from records. In addition, they may charge a fee to prepare the information requested on the survey. However, the third parties may be the only source of information available for some establishments. Survey organizations are not always aware that data have been prepared or provided by third parties.

- *One establishment, multiple respondents.* Some establishment surveys require input from several departments or groups within an organization, especially for questionnaires concerning complex financial data (e.g., Willimack et al., 2002).
- *Mergers, acquisitions, and divestitures.* Businesses often combine units or sell components through mergers, acquisitions, and divestitures. These processes can affect sampling plans, data collection, and response rate computations. Two or more sample establishments can merge and combine reporting; one sample establishment can absorb a non-sample unit and report for both; and so forth. Personnel from the reconfigured establishment sometimes opt out of a survey after a merger or acquisition because they have more work to do and fewer people with whom to do it (Fisher et al., 2003).
- *Multiple response modes.* It has become common practice in establishment surveys to offer respondents a choice of response modes. Some modes are more conducive than others to maintaining high levels of response and to identifying establishments whose response status has changed, both of which will affect response rates.
- *Standard data systems.* Many large establishment surveys create or support data systems that include various ways of recording or reporting the response status of individual units. For example, BLS has standard data collection/accountability status codes (Ferguson et al., 2003). Individual surveys may build upon these basic codes in much more detail (e.g., BLS, 2003b, Appendix Q). The level of detail in the standard data system determines the amount and type of data available for measuring response by characteristics such as mode.

3.0 The Annual Refiling Survey

The ARS illustrates a number of these response rate issues for establishment surveys. The survey itself is an integral component of the BLS Quarterly Census of Employment and Wages (QCEW) program, also known as ES-202 or the Covered Employment and Wages Program. QCEW is a Federal-State Cooperative Program managed by BLS in conjunction with state Unemployment Insurance (UI) programs. Each of the fifty states, plus the District of Columbia, Puerto Rico, and the Virgin Islands, operates a UI program within Federal guidelines but under its own UI laws. Employers have a legal obligation to participate in the program by registering their establishments with their state agencies, receiving a UI Account Number in that state, and by paying quarterly state UI taxes. Under QCEW, states follow BLS-developed statistical methodology and use BLS-provided standard data systems to maintain files of administrative records associated with the UI tax payments (Searson, 2002). These

files are submitted quarterly to BLS, where they are compiled into the Longitudinal Database (LDB). The LDB is a file of approximately 8.2 million employer establishments, coded geographically, by industry, and by size. The LDB serves as the sampling frame for all BLS establishment surveys, and is now used to publish quarterly detailed employment data by industry and data on business employment dynamics.

Through the state agencies, the ARS is the mechanism by which states and BLS update and maintain the LDB. During the ARS, each state agency contacts roughly one-third of the employers registered in that state (covering approximately 2 million U.S. business establishments annually)¹. These contacts emphasize data *verification* rather than data *collection*. Employers receive a short questionnaire that has been pre-printed with the business name, mailing address, physical location of the sample unit(s), economic activity (industry), and other items. The questionnaire for single-unit employers consists of one sheet of paper, printed front and back. For multi-establishment employers, the number of pages depends on the number of establishments included under the UI account in that state. In either case, respondents are asked to verify or correct the preprinted information, to supply any missing location or industry data, and in the case of "multis," to add new locations and delete any that are no longer operating. A business is asked to indicate if it has closed or moved out of the state shown on the form. A single-establishment employer can complete the form in about 5 minutes.

An important feature of the ARS is that it is conducted separately by each state, using procedures developed by BLS and standardized software systems designed to support that those procedures. Standard software controls sample selection, forms printing, receipt control, data entry, and generating summary reports for management purposes. States operate within their own computer environments, and print their questionnaires on their own mainframe printers or at a central Service Center. Questionnaire content is carefully controlled, but the varied hardware and printing platforms result in some differences in the appearance of printed forms across states. Some states also print bar codes on their questionnaires to facilitate response processing.

Sample. The sample unit for the ARS is the UI account, which may represent a single business with one employee or all of the locations in a state for a large multi-establishment firm. BLS prepares detailed sample specifications annually for the ARS so that the refiled establishments are representative of all U.S. employer establishments that year. Sample units are selected within each state based on specified positions of the Federal Employer Identification Number (EINs), with a back-up procedure based on the UI Account Number if the EIN is missing (BLS, 2001a). The product of the sample selection procedure is a *control file* of the selected establishments for the year's ARS.

Mailing. States prepare their own questionnaire mailing packages, which consist of employer-specific forms, postage-paid reply envelopes, and state-specific cover letters. Questionnaires for single-unit employers and small multi employers

¹ The terms "employer", "establishment," and "business" are used interchangeably.

are mailed in window envelopes that display the preprinted mailing address through the window. All cover letters are printed on state letterhead, describe state and BLS uses of the data and include the authorizing legislation if the ARS is mandatory in the state.

The field period begins around October 1. Most states complete all data collection and processing by June or July of the following year. The standard systems allow states to send the second or third mailings to alternative addresses. Since the mailing address could be for a third party, a different address might reach an establishment that was missed during the first mailing.

3.1 The Touchtone Response System Pilot Test

The ARS is a paper-intensive activity for the states. While it involves a significant amount of work, the number of businesses with changes to economic activity or geographic information is relatively small, roughly 10 percent of units per year. The Touchtone Response System (TRS) was proposed as a means of reducing the states' volume of paper handling.

The TRS allows respondents whose ARS information is correct and up-to-date to notify their states of this fact by means of a short automated telephone call. In addition to being easy for respondents, touchtone eliminates the cost of return postage and data entry, and cuts down on the number of returned envelopes and forms the states must process. Since large states mail tens of thousands of forms a year, even a small response by touchtone can have a noticeable effect on workload. BLS has a long history of using Touchtone Data Entry (TDE) for its ongoing Current Employment Statistics Survey. However, the experience is more limited for one-time surveys such as the ARS (Searson, 2002). The 2002 TRS pilot test was implemented in five states as a demonstration of the TRS system and to test the viability of touchtone response for the once-every-three-year ARS.²

It is important to recognize that the pilot test did not involve a split sample and was *not* an experiment to see if offering touchtone would increase survey response rates. The sample units that were invited to respond by touchtone were systematically different from units not selected for the pilot test in terms of the completeness of the data on their records.

Procedures. Five states participated in the test. Arizona, Colorado, Ohio, Tennessee, and Texas all work in the same ARS processing and printing environment, ensuring that forms in test state mailings had a common appearance. The ARS is mandatory in Colorado and Ohio and voluntary in Arizona, Tennessee, and Texas.

TRS was incorporated into the test states' regular ARS activities. The touchtone option was offered to all single-unit employers in those states if they met the selection criteria for the FY 2002 ARS, and could be reasonably expected to respond with no changes. That is, the businesses were thought to be active single units, had complete address information, and satisfied several other criteria. The data system was programmed to identify TRS-selected and non-selected units

and to assign a series of Collection Mode Indicator (CHI) codes for the mailing experiment.

At a minimum, employers who were selected for the TRS were informed of that fact directly on the ARS questionnaire. An item on the back of the form contained a message encouraging respondents to use TRS if there were no changes to their preprinted information. The message provided the toll-free number, the state code, and the UI account number for that respondent. Non-TRS-selected employers received the same ARS form, but saw a message in the same item asking them to "return the form within 14 days using the postage paid envelope provided." One concern going into the pilot test was that most respondents would have completed the questionnaire before they learned about the telephone option, and so might not bother to call. However, we did not have the option of printing anything about TRS elsewhere on the questionnaire.

3.2 ARS Response Monitoring

The ARS has a complex and detailed set of procedures for monitoring response, documented in the *ES-202 State Operations Manual* (BLS, 2003b). The standard data systems store survey response information for each establishment on the state control file. Response codes, developed for processing purposes, focus on changes that affect the economic data published by QCEW, especially industry and geographic information (state, county, and in selected states, township). Response codes differentiate between forms that were mailed with no response to date, forms waiting to be processed (to prevent remail), forms with no changes to key data, and forms with changes to economically-relevant data. All returns except refusals are considered complete, under the assumption that data are being verified and no change means correct information. Response codes identify forms returned by the U.S. Postal Service (USPS) for any reason, businesses that have closed, and refusals. Using the standard data systems, states prepare monthly reports showing response codes and response rates for singles, subunits of multis, and total units.

To accommodate the TRS Pilot Test, standard QCEW systems were programmed with an additional response code to identify successful TRS responses, as well as codes to identify businesses that responded by both phone and mail.

4.0 Response Rates: Extending AAPOR's Standard Definitions

During the 1990s, the American Association for Public Opinion Research (AAPOR) developed a tool for describing survey outcomes. The goal of AAPOR's *Standard Definitions* (2004) is to encourage full disclosure of survey methods, to standardize the use of outcome measures, and to provide researchers with a basis for making comparisons across surveys. *Standard Definitions* considers three types of surveys: in-person household surveys, random digit dialing telephone surveys (of households), and mail surveys of specifically named individuals. While the ARS does not fall into any of these categories, it is a mail survey of named businesses. As such, it seemed to be a useful framework for computing response rates.

According to *Standard Definitions*, survey cases can be divided into four groups: responses, eligible cases that do not

² See Jakhu and Sauer (2002) for a more detailed description of touchtone data collection.

respond, cases of unknown eligibility, and cases that are not eligible. These categories are delineated in considerable detail in *Standard Definitions*, Table 3, and should be applicable to the ARS as a survey of named businesses, where the researcher seeks a response from each specific sampled unit. Table 1 presents an abbreviated version of AAPOR's Table 3.

On closer examination, the AAPOR model does not fit the ARS as well as had been anticipated. The model described in the AAPOR document assumes that "the named respondent is at the target address or otherwise still eligible for inclusion" (2004:22). While eligibility can be defined in numerous ways (e.g., Osmint, 1994), linking eligibility to an address seems overly restrictive, unless the sample is address-based—which the ARS is not. Rather, an eligible sample unit is an establishment that had employees or paid wages at some point during the four calendar quarters prior to the time the sample was drawn. Businesses that have closed since sample selection are eligible and encouraged to respond, so that they can provide information that will allow a state to declare them inactive. In addition, the AAPOR definition requires that proxy respondents be identified, while the ARS has no way of knowing whether the actual respondent is a third party (proxy) representing the business.

The ARS starts out with considerable *a priori* knowledge about each sample unit. The categories in AAPOR's Table 3 for "Unknown eligibility, 'non-interview'" are more suited to situations where all the information a researcher has is a name and address. Consequently, AAPOR Category 3.10, "Nothing known about respondent or address," does not fit. AAPOR Category 3.20, "Unknown if eligible respondent in unit," again suggests an address-based sample. Given a mail survey of named individuals, Category 3.20 does not seem to capture a fundamental aspect of the mailing, which is whether or not it reached the individual. At the same time, breakouts under category 3.2 include [mail] "Refused by addressee"—which could also be a subcategory of refusal if eligibility is known. If the mail is refused by the named individual, at least the researcher knows that the mail was deliverable. Most of the remaining breakouts under "Unknown eligibility, 'non-interview'" are in fact noncontacts where delivery did *not* occur (AAPOR Categories 3.24 through 3.36) and no additional information has been provided by the USPS. AAPOR Group 3.4 cases ("Returned with forwarding information") may or may not remain noncontacts, depending on whether there is enough time to forward the form to a new address. Regardless, the ARS has one code for all "Post Office returns." State staffs attempt to locate a new mailing address for returned questionnaires, and to mail again if possible.

Another important difference between the AAPOR model and the ARS is the assumption that the specifically named person is one sample unit. In the ARS context, that is not the case. Many of the ARS survey contacts are for multi-establishment firms, where the sample unit is the named business, but the business consists of many locations or worksites. ARS contacts are associated with "master records" for the business. However, the survey does not monitor the number of master record contacts, only the number of associated subunits. Response rate calculations are based on the number of single

units and multi subunits rather than the number of individual contacts. As described above, there are many ramifications of multi-unit businesses for response rates, and these (justifiably) fall outside of the scope of the AAPOR model.

Table 2 compares the AAPOR final disposition codes and ARS Response Codes. Since the sample is eligible for ARS by definition, Categories 3.0 and 4.0 on the ARS side are empty. The third column of Table 2 contains a symbol for use in response rate formulas. The AAPOR terminology has been maintained to the extent possible.

The QCEW program requires states to achieve a minimum usable response rate of 75 percent for the ARS each year. The formulas used to compute response rates are printed on the management reports and documented in the program manual (BLS, 2003b). There are two response rates. *Usable response* is defined by the ARS as the number of units—single establishments or subunits of multis—that are actively in business with response codes that reflect "no change" or economically-significant changes, divided by the sum of [(change plus no change responses) plus nonresponses, pending cases, and refusals]. "Out of business" units and postal returns are excluded from both the numerator and denominator of the usable response rate. *Total response* includes usable responses plus pending cases, refusals, forms returned by the USPS, and "out of business" establishments in the numerator, and all of the above plus nonresponse in the denominator.³ Using the symbols shown in the third column of Table 2, the ARS formulas used in FY 2002 are:

$$\text{Usable Response Rate: } \frac{I}{I + O + PN + R}$$

$$\text{Total Response Rate: } \frac{I + PN + OOB + NC + R}{I + PN + O + NC + OOB + R}$$

Both usable and total response are computed for units and weighted by employment. States may meet the response rate requirement by the percentage of units or the percentage of employment represented by those units.⁴

The ARS Total Response Rate appears to be equivalent to AAPOR Contact Rate 3, at least when computed for single-unit establishments. The ARS Usable Response Rate is not equivalent to an AAPOR definition. It is closest to AAPOR Response Rate 5, since all units are eligible, but the AAPOR formulation keeps all noncontact (USPS returns) and other nonresponse in the denominator while the ARS formula excludes USPS returns. In any event, it is not possible to compute AAPOR equivalent rates for the entire 2002 ARS. The data are no longer accessible.

Noncontacts warrant an additional mention here. Establishments have a reason for existing, whether to make a profit

³ The actual formulas include a small number of additional special-purpose items in both the numerator and the denominator. The effect of these items on computed response rates is negligible.

⁴ There is an apparent inconsistency between collecting data from units that are out of business and not considering those units as usable response. However, policies regarding response rates have been in place for many years. It wasn't until 1999 that ARS questionnaires systematically offered respondents a place to indicate that they were out of business or no longer operating in the state. "Out of business" could also refer to establishments that were sold, merged, or acquired by another firm.

or to provide a service. In order to be viable, they have to be visible to their customers or clients. Therefore, if the USPS returns a mailing as undeliverable, and the state staff is unable to locate another address for that establishment, it may be reasonable to consider the establishment "closed." ARS policy is to include nonresponding establishments from one ARS cycle with the first mailing the next year. If these "carryover" establishments still do not respond, and they have no reported employment or wages, they are presumed to be out of business and their status is changed to inactive. Carryover records are tabulated as responses if forms are returned, but are otherwise excluded from response rate calculations.

5.0 Response Rates for the TRS Pilot Study

The TRS Pilot Test should more closely approximate the AAPOR model of mail surveys of named individuals than the larger ARS. First, the touchtone option was only offered to single-unit establishments, satisfying the expectation that the named business is one sample unit. Second, there are eligibility criteria for TRS. Although establishments were preselected for participation, only units that met those criteria could respond by telephone. This eligibility could not be known in advance, since it is confounded with the dependent variable (and is essentially the reason for conducting the ARS). Finally, noncontacts and other nonresponses are cases of unknown eligibility for the TRS. On the other hand, establishments that are not TRS-eligible are still expected to respond, but by mail.

Goldenberg (2003) separated the pilot test sample units into four groups based on the ARS response rate schema: usable responses through TRS; usable responses returned by mail; non-usable responses per the ARS definition (out of business, undeliverable, etc.); and nonrespondents. Since one objective of TRS is to reduce the amount of paper handling, we look first at touchtone in the context of all TRS-selected forms mailed for the 2002 ARS.

ARS Response Summaries for TRS-Selected Units.

Table 3 shows the distribution of ARS sample cases for TRS-selected units participating in the pilot test. We display the data at two time periods: the end of the first mailing period, in February, and at the end of the ARS cycle for the year in August. The right-hand columns show TRS response as a percentage of mailed forms. For these reference points, usable response under the ARS definition for the TRS-selected units averaged 66 percent (ranging from 52 to about 74 percent across the five states) after the first mailing, and increased to about 84 percent (ranging from 78 to 86 percent) at the end of the cycle.

Because the ARS collects data from units that are out of business, these data can be recomputed to include "out of business" cases as part of usable response. This makes the response consistent with AAPOR Response Rate Number 5, which defines all sample units as eligible for the survey. Under this definition, it seems reasonable to argue that any unit providing usable data should be counted as a response. Following the AAPOR definition, usable response rates are about 2 percent higher (68.2 percent) after the first mailing and about 6 percent higher (88.4 percent) at the end of the

cycle. These rates appear in Table 3 under Percent of Forms Mailed.

Readers familiar with household surveys may question the fact that refusals have been combined with other nonresponses. In fact, overt refusals among the TRS-selected cases were negligible—170 cases by the end of the ARS cycle. It is worth noting, however, that 82 percent of the overt refusals came from larger single-unit establishments, with 50 or more employees.

TRS Response Relative to TRS Eligibility. The impetus for introducing touchtone response to the ARS was to reduce workload. Therefore, what we really wanted to measure was the proportion of TRS responses relative to the number of units that were eligible to respond using this method. We call this the "workload savings" TRS response rate, since it shows how many records are updated automatically and excludes respondents whose records needed updating. To compute this rate, we must distinguish between units that are eligible and not eligible for TRS, along with those whose eligibility is unknown.

One limitation of the ARS data system is that it does not capture all of the information needed to identify "ineligible, data changed" responses. More specifically, the system cannot readily identify forms returned by mail with address changes. As a result, we don't know how many mail respondents used mail because they were not eligible for touchtone. However, we found that about 3.3 percent of respondents unsuccessfully attempted to respond by phone, and subsequently mailed in forms with address or other changes. Therefore, we reduced the number of mail responses by 3.3 percent, so that the response rate reflects true TRS eligibility.

Both the ARS and AAPOR response formulations exclude TRS-ineligible responses from the denominator. The main difference is the presence or absence of USPS returns. The ARS formulation is:

$$\frac{T}{(T + eM) + (U-R + U-O)}$$

where

- T = Touchtone response
- eM = Estimate of mailed responses that could have used touchtone
- U-R = Unknown eligibility, refusal
- U-O = Unknown eligibility, other nonresponse

Conceivably units whose status was still pending at the end of the cycle could also fit into the unknown eligibility group, but with only 22 records in this category, the impact on final rates is minimal. "Out of business" units are not eligible for touchtone, so they do not factor into the numerator or denominator for the workload rate.

Using AAPOR's response rate number 1, the AAPOR formula would be:

$$\frac{T}{(T + eM) + (U-R + U-NR + U-O + U-NC)}$$

where the additional element is

- NC = Unknown eligibility, noncontact (USPS return)

Using these formulas, the workload savings response rate under the ARS definition was 21.8 percent at the end of the first mailing, and 28.1 percent at the end of the FY 2002 refiling cycle. The slightly more restrictive AAPOR definition has no effect on the rate in February, but lowers it slightly to 27.8 by the end of the cycle. The change is solely attributed to the presence or absence of USPS-returned forms in the denominator. Regardless of the number, the pilot test states collectively had 28,000 fewer pieces of paper to handle in FY 2002 than they would have otherwise.

6.0 The Response-Enhancement Experiment

The pilot test contained an embedded experiment to see whether the use of different contact materials would affect the level of participation in TRS, compared to the state's standard cover letter. A cover letter that highlights the touchtone system, a TRS flyer, or a combination of TRS-specific letter and flyer should call additional attention to the TRS, so respondents receiving them—especially respondents who actually read the enclosures!—might be more likely to call than those receiving only the standard letter. The experiment is based on a null hypothesis that there is no difference in the proportion of TRS-selected employers responding to TRS between those who received the standard ARS cover letter (control group) and those who received one of the test treatments. Ironically, the ideal result would be no difference among the four groups, because there are labor and material costs associated with using different letters and adding flyers. Regardless of the outcome, the design provides a measure of the response gains, if any, from each mail condition.

Experimental Procedures. TRS-selected sample units were randomly assigned to one of the four conditions: a control group (standard cover letter only) or one of the three treatments. Initial specifications called for each of the three test groups to have 3800 cases, with the remaining cases assigned to the control group. This practice was followed for all states except Tennessee, where an alternate randomization procedure was used because the number of TRS-selected cases was too low. All states used the same cover letter and flyer, modified per state legal requirements and to show local assistance telephone numbers (BLS 2001b; 2001c; 2001d).

TRS mailing lagged behind the regular mailing because of startup delays in the system. Once cleared to begin, states varied in actual mailing dates. However, 80 percent of the forms were mailed in December, 2001, 14 percent in late November, and 6 percent in early January, 2002. Four of the five states mailed their control group forms first and the experimental treatment groups later. While they did not all mail the forms in the same sequence (e.g., Group 2 was second in four states and last in one), most went out within 10 days of the first mailing.

Analysis. Because of the labor-intensive preparations for the experiment, participating states were only required to use the experimental treatments for the first mailing. However, all TRS-selected establishments continued to receive the control group treatment (standard letter with TRS invitation) in subsequent mailings. To look specifically at the effect of the experimental treatment, analysis of the data is based on forms returned prior to the second mailing. TRS was an option

throughout the ARS cycle, however, and the analysis shows the final tally at the end of the cycle.

This analysis looks at TRS responses as a proportion of responses that were eligible for touchtone. It excludes any forms with changes to industry or county, sample units that were out of business, and cases with other changes that disqualified them from using TRS. It also excludes nonrespondents and noncontacts. However, it includes all mail responses that indicated there were no changes to the pre-printed data on the form. Based on the estimation of 3.3 percent of mail forms used in Section 5.0, the denominator is slightly overstated and the TRS response rates slightly understated.

Results. The question addressed by the experiment was whether we could increase TRS response by sending a flyer, a TRS-specific cover letter, or both, as compared to the standard ARS cover letter. Figure 1 shows the usable TRS response for all five pilot test states, at the end of the first mailing in February and again at the end of the ARS cycle in August. The chart shows that there is a clear pattern across groups. TRS response rates range from about 20 percent to 23 percent at the end of the first mailing, and from 24 through 28 percent by the end of the cycle. TRS response is lowest for the control group, demonstrating that there is a definite response benefit to calling attention to TRS through a flyer or letter. For both time periods, the difference in response rates between sending only the standard letter and sending *any* TRS-specific communication is statistically significant with probability $p < 0.001$.⁵ In addition, the combination of TRS letter and flyer is always higher than either the TRS letter or the flyer, and all are higher than the control group. However, the difference between the TRS letter and the flyer is quite small.

How important were the treatments to response for the first mailing? Comparisons of the control group with any of the treatments are statistically significant at $p < 0.001$. Using the letter does not produce results that are significantly different from using the flyer. Contrasting the flyer alone with the combination of TRS letter and flyer, the difference is statistically significant at the $p < 0.01$ level, while the difference between the TRS letter alone and the combination of letter and flyer is statistically different at the $p < 0.05$ level.

Most of these differences persist to the end of the ARS cycle. In fact, the effect of the treatments increases throughout the cycle in four of the five states, even though the states were not required to maintain the different treatments after the first mailing. Most of the statistical relationships continue as well, especially those between the control group and any of the treatments. Again, there is no statistical difference between the flyer alone or TRS letter alone in comparison to each other, and by the end of the cycle the response rate difference between the TRS letter and the combination TRS letter and flyer was no longer statistically significant. However, the combination performed better than the flyer, a difference that retained statistical significance at the $p < 0.05$ level.

⁵ Comparisons based on t-tests between response rates for control group and flyer, control group and TRS letter, control group and combination, flyer and TRS letter, flyer and combination, and TRS letter and combination.

What does this mean for states preparing to send out TRS mailings? The differences are not overwhelming, but there are gains over using just the standard letter. There are additional printing costs but no additional mailing costs for a TRS specific letter, which averaged a 2 to 2.5 percentage point gain over the standard letter in terms of overall response. The flyer by itself improved results but not as much as the letter, and represents an additional printing and possibly mailing cost. The biggest gain comes from the combination of the TRS letter and flyer, averaging 3.7 percentage points across all states. In terms of the units that actually respond to the ARS, the benefits are more pronounced: 4 to 5 percentage points more for a letter or flyer, 7 percentage points more for the combination. On the other hand, roughly one-fifth of sample units that received only the standard cover letter responded using touchtone by the end of the first mailing, a figure that rose to one-fourth by the end of the cycle—so a sizable minority of respondents called the TRS system even without the additional mailing materials to bring TRS to their attention. In short, it's a toss-up. Some states may find that a likely increase in response is large enough to justify the additional printing and mailing costs, while others may decide that touchtone will save resources without extra printing and handling.

7.0 Discussion

The purpose of this paper was to explore response rates for a multi-mode establishment survey, so as to assess the results of a new mode component added to the BLS Annual Refiling Survey. AAPOR's *Standard Definitions* initially seemed to offer a framework for exploring response to a survey of named businesses, but many of the AAPOR model's assumptions were incompatible with the operations of the ARS and the then-experimental Touchtone Response System. In addition, the standard data systems do not allow for many of the detailed assessments that the AAPOR model recommends, while others that seem appropriate are not part of the AAPOR model. However, we used the ARS response codes and the collection mode indicators to determine the outcome of every TRS-selected record. By decomposing the data in this way, we compared standard ARS response rates with rates based on the formulas specified by AAPOR.

This paper does not delve into the technology or file-handling procedures needed to make the Touchtone Response System practical, but the pilot test served as a demonstration of their viability. Almost a fourth of the mailed forms resulted in a response using touchtone, with a slightly-higher rate of 28 percent based on sampled units that were eligible to use it. The experiment revealed a distinct pattern, where using a TRS-specific letter or a TRS flyer resulted in a higher level of response than printing a message on the questionnaire alone, and the combination of letter and flyer yielded the highest rate of all. However, the statistically significant differences in response were not large, averaging under 4 percent across the pilot test states. The additional printing and handling costs

could offset the potential gains from the use of the additional materials.

As for the Touchtone Response System, QCEW program managers were pleased with the pilot test results. They made TRS available to all states for FY 2003 and forty states participated. It is now a required component of the ARS, except in Puerto Rico and the Virgin Islands. In addition to the workload efficiencies, the savings in postal costs have allowed resources to be shifted to other program initiatives.

8.0 References

- American Association for Public Opinion Research. 2004. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 3rd Edition*. Lenexa, KS: AAPOR.
- Bureau of Labor Statistics. 2001a. "FY 2002 ARS Selection Criteria." Internal document, updated March.
- _____. 2001b. "State Instructions for the Touchtone Response System," November.
- _____. 2001c. "TDE Specifications," November.
- _____. 2001d. "Procedure for Assigning Cases." August, revised November.
- _____. 2003a. *Employment and Wages. Annual Averages 2002*. Washington, DC.
- _____. 2003b. *ES-202 Operating Manual*. Washington, DC.
- Ferguson, G.R., S.H. Cohen, and R.J. Rosen. 2003. "Survey Response Measurement at the Bureau of Labor Statistics." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association
- Fisher, S., J. Bosley, K. Goldenberg, W. Mockovak, and C. Tucker. 2003. "A Qualitative Study of Nonresponse Factors Affecting BLS Establishment Surveys: Results." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
- Goldenberg, K.L. (with S. Jakhu, M. Sauer, M. Searson, L. Unger, and J. Walker). 2003. "Results of the FY 2002 Touchtone Response System Pilot Test." Internal report, Bureau of Labor Statistics. January.
- Jakhu, S. and E.M. Sauer. 2002. "New Data Collections Using Touchtone Data Entry." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
- Osmint, J.B., P.B. McMahon, and A. Ware Martin. 1994. "Response in Federally-Sponsored Establishment Surveys." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, pp. 997-982.
- Searson, M.A. 2002. "Automated Data Collection Strategies for the Covered Employment and Wages (CES) Program." *Proceedings of the Section on Survey Research Methods*. Alexandria, VA: American Statistical Association.
- Willimack, D., E. Nichols, and S. Sudman. 2002. "Understanding Unit and Item Nonresponse in Business Surveys." In Groves, R.M., D.A. Dillman, J.L. Eltinge, and R.J.A. Little (eds.), *Survey Nonresponse*. New York: Wiley, 213-228.

Table 1 AAPOR's Standard Definitions, Table 3:
Final Disposition Codes for Mail Surveys of Specifically Named Persons (Modified)

<p>1.0 Returned questionnaire</p> <p>1.1 Complete</p> <p>1.2 Partial</p>	<p>3.0 Unknown eligibility, "non-interview"</p> <p>3.10 Nothing known about respondent or address</p> <p>3.20 Unknown if eligible respondent in unit</p> <p>3.21 No screener completed</p> <p>3.23 USPS category: refused by addressee</p> <p>3.24 USPS category: returned to sender due to various USPS violations by addressee</p> <p>3.25 USPS category: cannot be delivered</p> <p>3.30 Unknown Whereabouts, Mailing Returned Undelivered</p> <p>3.40 Returned with Forwarding Information</p> <p>3.90 Other</p>
<p>2.0 Eligible, "Non-Interview"</p> <p>2.10 Refusal & Break-off</p> <p>2.20 Non-Contact</p> <p>2.30 Other</p>	<p>4.0 Not Eligible, Returned</p>

Source: AAPOR, 2004, p. 41

Table 2. Comparing AAPOR's Final Disposition Codes and ARS Response Codes

AAPOR Category Code	Final ARS Response Code	Symbol
1.0 Returned questionnaire, usable data - Complete - Partial	RC 41: No economically relevant changes RC 42: Respondent thinks industry is wrong but code is correct RC 43: Successful TRS response RC 46: Changes to industry or geographic area RC 64: Out of business	No distinction between 'complete' and 'partial' response I I I OOB
2.0 Eligible, "non-interview" 2.1 Refusal and break-off 2.2 Non-contact 2.3 Other	RC 65: Refusal RC 63: Post office return (would encompass all of AAPOR categories 3.24-3.4) RC 01, 02, 03, 04: Mailed, 1, 2, 3, or 4 times, no response RC 31: Pending (received, not yet processed)	R NC O PN
3.0 Unknown eligibility, "non-Interview" 3.1 Nothing known about respondent or address 3.2 No screener completed 3.23 Refused by addressee 3.24 – 3.36 USPS categories for mail not delivered 3.4 USPS return with forwarding information	Does not apply; whole sample is eligible	
4.0 Not eligible, returned	Does not apply; whole sample is eligible	

Table 3. Summary of ARS Response for TRS-Selected Units
All TRS Pilot Test States, End of First Mail Period and End of ARS Cycle

Mail Receipts	End of First Mailing (February)	End of ARS Cycle (August)	Response Measures	End of First Mailing (February)	End of ARS Cycle (August)
	Number of Forms			Percent of Forms Mailed	
Usable response			TRS as percent of total mailed	19.7	23.6
TRS (RC 43)	23,746	28,438	ARS definition		
Mail response (RC 41, 42, 46)	55,905	72,573	Usable response excluding OOB	66.1	83.8
Total usable response	79,651	101,011	Non-usable response	2.4	5.7
Nonusable response			AAPOR definition		
Out of business (OOB)	275	1,374	Usable response including OOB	68.2	88.4
USPS return	2,575	5,603	Non-usable response	0.0	1.2
Other nonusable	10	22	Nonresponse	31.6	10.6
Total nonusable	2,860	6,999	Total forms mailed	100.0	100.0
Nonresponse (includes refusals)	38,056	12,557			
Total forms mailed	120,567	120,567			

Figure 1. TRS Response by Collection Mode Indicator, February and August, 2002

