

# Considerations for Sampling from a Skewed Population: Establishment Surveys

Marcus E. Berzofsky and Stephanie Zimmer<sup>1</sup>

## Abstract

Establishment surveys often have the challenge of highly-skewed target populations (e.g., very few large establishments). Therefore, when designing a series of surveys for the same population, reducing burden on establishments that will be selected with certainty across the surveys is important. In this paper, we examine the idea of developing a panel of law enforcement agencies in the United States whereby we can remove overlapping questions while leveraging the information obtained in all surveys when analyzing each individual survey. Challenges to this type of design that will be explored include differing response rates across surveys; different analytic subpopulations of interest; and differing precision requirements. In assessing the advantages and disadvantages of this type of design we look at a pair of surveys conducted by the U.S. Bureau of Justice Statistics. The surveys cover the use of body worn cameras by officers and an omnibus survey to better understand the officer population across law enforcement agencies. The United States has approximately 18,000 law enforcement agencies, but only two percent have 250 or more officers. Since these agencies are all in the major metropolitan areas and employ the majority of law enforcement officers they are essential to include in any representative survey. For smaller agencies, a random sample will be used to identify agencies, but participation rates are often relatively low. Therefore, recruiting a panel of agencies that can be used across surveys and reduce the burden for completing each survey may be advantageous for minimizing survey costs and burden for participating agencies. Our presentation will present the different sample design options considered and discuss how the final allocation balances the goals of both surveys.

**Key Words:** Sample design, minimizing respondent burden, panel designs, law enforcement agencies

## 1. Introduction

One key goal of a survey is to obtain a representative sample from the target population. To ensure that rare populations are included survey designers will alter the probabilities of selection for the rare populations to increase their inclusion probabilities. When the same population is being sampled for multiple surveys the sampling units in the rare population may be selected multiple times. This may lead to respondent fatigue in the rare population. In this paper we discuss methods that can be used to increase the participation of rare populations from a series of surveys.

### 1.1 Skewed Populations

---

<sup>1</sup> Marcus E. Berzofsky, RTI International, 3040 Cornwallis Rd Research Triangle Park, NC 27709, email: [berzofsky@rti.org](mailto:berzofsky@rti.org). Stephanie Zimmer, RTI International, 3040 Cornwallis Rd Research Triangle Park, NC 27709, email: [sazimmer@rti.org](mailto:sazimmer@rti.org).



|         |      |       |      |       |      |       |      |       |
|---------|------|-------|------|-------|------|-------|------|-------|
| 1-1.5   | 6.5  | 100.0 | 0.2  | 0.2   | 0.8  | 100.0 | 0.0  | 0     |
| 2-4.5   | 18.9 | 93.5  | 1.6  | 1.8   | 7.1  | 99.2  | 0.4  | 0.4   |
| 5-9.5   | 24.1 | 74.6  | 4.4  | 6.2   | 17.5 | 92.0  | 2.0  | 2.4   |
| 10-24.5 | 26.0 | 50.5  | 10.8 | 17.0  | 30.6 | 74.6  | 7.8  | 10.2  |
| 25-49.5 | 13.0 | 24.5  | 12.0 | 29.0  | 19.1 | 44.0  | 10.7 | 20.9  |
| 50-99.5 | 6.3  | 11.5  | 11.4 | 40.4  | 11.8 | 24.9  | 13.0 | 33.9  |
| 100+    | 5.2  | 5.2   | 59.6 | 100.0 | 13.0 | 13.0  | 66.2 | 100.0 |

## 1.2 Statement of the Problem

From a survey methodology standpoint the problem a skewed distribution causes is one of representation. Large establishments often function very differently than small establishments. Therefore, survey analysts not only want to make overall estimates (where a proportional allocation would suffice) but subdomain estimates for the larger establishments. This often requires the oversampling of larger establishments.

Furthermore, because of the extreme skewness found in many establishment populations, there is a very finite number of large establishments. Therefore, when an oversample is conducted a census of the larger establishments will need to be selected.

Moreover, each type of establishment is sampled across a multitude of surveys. Each with the same goal of ensuring adequate estimates from large establishments. The result is that larger establishments are overburdened and may have lower participation rates.

## 2. Proposed Solution

In this paper we propose the use of a panel to help mitigate some of the challenges associated with the repeated selection of some sampling units in the population.

### 2.1 Advantages

A panel design offers several advantages over a set of disjointed surveys. These advantages include

- *Minimal impact on larger establishments.* Large establishments will be selected with certainty regardless of a panel design or not. Therefore, the panel aspect does not increase the burden on them.
- *Reduction in burden for any given survey.* By connecting surveys through a panel, basic survey items about the establishment do not need to be repeated in each survey. Rather basic information about the establishment can be updated on a periodic basis – perhaps annually – thereby reducing the burden on all establishments.
- *Improved analysis.* By having a panel of establishments, the set of respondents for each survey is the same. Therefore, outcome information from one prior surveys can be used in the analysis of future surveys. This can help increase the analytic use of each survey.

### 2.2 Challenges

A panel design, while posing many advantages over a set of independent surveys, has challenges as well. In this section we discuss each challenge and some mediation strategies.

### *2.2.1 Participation and Retention*

The key to a successful panel is panel retention. If establishments do not initially participate and then continue to participate throughout the panel's life cycle the benefits of reduction in burden and improved analysis cannot be achieved.

One mitigation strategy to increase participation and retention is incentives. For establishments the incentive often cannot be monetary – who at the establishment would receive the incentive? However, there are non-monetary incentives that may be equally as effective. For example, often establishments have a great interest in the survey topic and want to know how their establishment compares to other, like establishments. Therefore, one incentive type is a report that is provided to the establishment with tailored information. This tailored information could include the establishment's responses next to the average responses from other like establishments.

Another mitigation strategy is a rotating panel design. Smaller establishments would not normally be selected multiple times under independent survey samples. Therefore, to reduce the burden on them that a panel design may impose, a rotating panel design could be used whereby after a certain number of surveys the establishment is rotated out of the panel.

### *2.2.2 Single Sample Size*

Under a panel design a single set of establishments is selected. However, for any given survey, the number of responding establishments needed for adequate power may vary greatly. Therefore, the panel must be able to accommodate the largest necessary sample. This may cause unnecessary burden for other surveys that do not require as much sample. Furthermore, the optimal allocation of the sample may not be the same for each survey.

One mitigation strategy for this is to draw subsamples for surveys that require less sample. Under the subsampling strategy the benefits of the panel are retained but not all panel participants are utilized minimizing burden on the panel.

To mitigate the issue of the sample allocation, a design that minimizes the design effect across a multitude of outcomes can be employed. If the topics for the surveys are known in advance, an allocation that optimizes across all key outcomes can be utilized.

### *2.2.3 Panel nonresponse*

Panel nonresponse is when a selected establishment either initially refuses to participate in the panel or drops out before all surveys have been administered. If panel nonresponse is large this will minimize the benefit of the improved analysis advantage. Furthermore, for a given survey iteration, if not enough panel members participate then the precision for that particular survey may be compromised.

One mitigation strategy for panel nonresponse is the use of replacement establishments. These establishments will only be used for the single wave to help ensure adequate precision for that wave and will not be incorporated into future surveys. There are two possible methods to identify a replacement establishments: (1) a reserve (or replicate) establishment or (2) a substitute establishment. A reserve establishment design needs to be considered at the initial sample selection stage. Under this method, a random subset of establishments is selected within each stratum. When a panel member does not participate a random replacement from the reserve sample can be selected. Alternatively, a substitute

approach can be used. Under a substitute approach a nonresponding establishment is replaced with another establishment that has the smallest “distance” to the nonrespondent establishment and is not currently in the panel. The “distance” is defined based on key survey objectives such as establishment size and location. The substitute establishment assumes the survey weight of the nonresponding agency for that panel wave.

Because of panel nonresponse, two types of weights will need to be developed: (1) panel weights and (2) survey specific weights. Panel weights will sum to the population but only include the establishments that are a part of the longitudinal cohort. Replacement establishments will not be included in this weight. Survey specific weights will include all participating establishments for that survey wave. This weight will include the replacement establishments.

#### *2.2.4 Births, Deaths, Marriages, and Divorces*

One systemic problem with frames for establishment surveys is that they quickly are out of date. In the U.S., approximately 10% of business are new (births) and 10% close in any given year. In some establishment types – such as law enforcement agencies – the establishments merge (marriage) or separate (divorce) over time. A sample of establishments for a panel may quickly have a lot of undercoverage due to exclusion of new establishments and a lot of ineligible businesses due to closures.

To mitigate this issue frame maintenance needs to be an ongoing and continuous process. Furthermore, supplemental samples of new establishments (that had no probability of selection at the time of the initial sample) can be drawn to help supplement the initial panel. These establishments may need to have an initial survey to provide the basic establishment information that the other panel members already provided.

### **3. Application**

We apply the proposed methods to law enforcement agencies in the United States.

#### **3.1 The Problem**

The U.S. Bureau of Justice Statistics (BJS) conducts multiple surveys of law enforcement agencies each year. In all surveys, agencies with 100 or more officers (5.2% of the agencies) are selected with certainty. These agencies employ 59.6% of local police officers and 66.2% of sheriff’s officers. BJS is interested in a design that can maximize participation from the self-representing agencies while ensuring adequate representation from smaller agencies.

#### **3.2 Design**

The proposed design will be a stratified simple random sample of establishments. Agencies will be stratified by their size (6 strata) and the agency type (2 strata). As typically done by BJS, agencies with 100 more officers will be selected with certainty.

For purposes of optimizing the allocation of the sample, two law enforcement surveys are considered: (1) The Law Enforcement Management Administration Survey (LEMAS) and (2) Survey of Body Worn Camera Use (BWC). LEMAS is the flagship law enforcement agency survey conducted by BJS. It is administered every 4-5 years. This is an omnibus survey with multiple key outcomes of interest. The BWC is a short web survey on the

prevalence and utilization of body worn cameras. For the allocation, several key outcomes from LEMS were considered as well as the prevalence of body worn cameras.

To determine the optimal allocation, four allocation schemes were considered:

1. Proportional to number of agencies in stratum
2. Proportional to number of officers in stratum
3. Proportional to square root of the number of officers in stratum
4. Bethel allocation which optimizes the allocation based on cost and precision constraints

**Table 4** presents the results of the allocation comparison. Expected relative standard errors (RSEs) were used as the basis for comparison. The methods proportional to number of agencies, proportional to square root of number of officers and the Bethel all had similar RSEs for all outcomes. Among those three methods, the proportional to number of agencies was chosen as the allocation method because it was the easiest to implement.

**Table 4.** Relative Standard Errors for Key Outcomes by Allocation Method

| Estimate                 | Prop. to Agencies | Prop. to FTEO | Prop. to SQRT FTEO | Bethel |
|--------------------------|-------------------|---------------|--------------------|--------|
| Operating Budget         | 2.0%              | 1.9%          | 1.9%               | 1.9%   |
| Entry Salary             | 0.5%              | 0.8%          | 0.6%               | 0.5%   |
| Body Worn Camera Usage   | 2.9%              | 5.6%          | 3.9%               | 3.0%   |
| % Community Policing     | 1.3%              | 2.8%          | 1.8%               | 1.4%   |
| Full-time Civilian       | 2.2%              | 2.1%          | 2.1%               | 2.3%   |
| % Female Full-time Sworn | 6.5%              | 11.2%         | 8.1%               | 6.6%   |

We are still in the process of developing methods for panel maintenance and retention. Law enforcement agencies cannot receive monetary incentives, but can receive donations to charity in their name and are in need of data comparing themselves to other agencies.

#### 4. Summary

A panel design can be one solution to mitigate the problems of sampling from a skewed population such as establishment populations. While this method offers many advantages over independent sets of surveys from the same population, there are many challenges that need to be addressed during the survey design. This paper describes how one can mitigate each of these challenges. We present an application of the proposed methods to law enforcement agencies in the United States.

#### Acknowledgements

The authors would like to thank the Bureau of Justice Statistics (BJS) for their valuable contributions to this research. However, we would like to note that the views expressed in

this poster are those of the authors only and do not reflect the views or position of BJS or the Department of Justice.