Arkansas Act 1220 Evaluation: Multi-stage Stratified Surveys with Probability Proportional to Size (PPS) Sampling

Zoran Bursac, Martha Phillips, C. Heath Gauss, LeaVonne Pulley, Delia Smith West, James Raczynski
University of Arkansas for Medical Sciences, Fay W. Boozman College of Public Health

Abstract

Arkansas Act 1220 mandated a process for nutrition, physical activity standards, and related policy changes among schools statewide. A baseline probability survey of families was conducted to characterize behavioral intentions to change diet and exercise, assess concerns about obesity, identify possible adverse outcomes, and collect self-reported BMI. With a stratified three-stage survey that accounted for geographical-school type-school size differences, we randomly sampled schools-families-students with probability proportional to the size (PPS) of school enrollment using Sampford's method. Responses were weighted to include stratified design features, probability of selection at each stage, and adjustments for non-response. The final sample included 1,551 families from 113 schools. Our BMI distribution closely matched the one collected using census-like methods. We found a majority of the parents of at risk or overweight children do not perceive them as such. Based on the lessons learned, we recommend increasing the sample of schools per strata in the first stage of the design, to more easily obtain a larger sample of families and gain more power at the sub-group analysis level.

Keywords: Stratified multi-stage sampling, Arkansas Act 1220, Sampford's method, Policy evaluation.

1. Introduction

In April 2003, Act 1220 of 2003 was passed by the Arkansas General Assembly and signed into law by the governor, creating a comprehensive program to combat childhood obesity in the state. The major provisions of the Act required annual body mass index (BMI) screening for all public school students with the results reported to parents, restricted access to vending machines in public elementary schools, public disclosure of schools' contracts with food and beverage companies, creation of district advisory committees made up of parents, teachers, and local community leaders, and creation of a statewide Child Health Advisory Committee to recommend additional physical activity and nutrition standards for public schools [3]. As a part of Act 1220's evaluation, baseline surveys of superintendents and principals have been conducted so that process and impact data can be collected to characterize the implementation of Act 1220 policies and variation of implementation, and examine attitudinal and behavioral impact on school measures [3]. The focus of this paper, however, is on the design of a baseline survey of random families, including adult and youth respondents, to obtain the data on their general knowledge about weight control and health-related behavior patterns of Arkansas youth.

2. Methods

2.1 Act 1220 Sampling Design

This survey followed a multi-stage stratified design. In order to ensure a representative sample of schools from Arkansas, all eligible schools were stratified based on geographic region (north, northwest, southwest, central, and east) (Figure 1), type of school (elementary, middle, and high school), and size of school (small, medium, and large school).

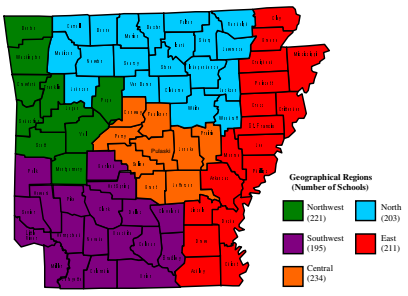

Figure 1. Geographical regions of Arkansas used for stratification and sampling

The goal of the survey was to sample 110 schools in the first stage of sampling, 20 families per sampled school in the second stage, and 1

child or adolescent per family in the third stage. The Common Core Data (CCD) from the National Center for Education Statistics (NCES) for school year 2002-2003 was used as a sampling frame. The type of school was already defined in the CCD data set, while the size of school was created based on tertiles of enrollment of eligible schools.

Geographic regions were determined by dividing the state into five most meaningful sections based on the geographic makeup of the state, such that the distribution of enrollment of schools was approximately uniform. The sampling frame consisted of 1,064 eligible schools after elimination of 4 alternative schools and 94 schools with enrollment less than 100. Schools with enrollment less than 100 were excluded to reduce the number of phone calls needed to locate eligible families. Households with listed phone numbers within a 5-15 mile radius of a school were sampled. Telephone interviews were conducted with families whose children attend the sampled public schools. The interviewer asked for the adult in the household most suitable to answer questions about the health, diet, and physical activity of the family. If there was more than one eligible child in the household, one child was randomly selected. If the eligible child was 14 or older, the adolescent completed a portion of the interview in addition to the parent; otherwise, the parent served as a proxy.

2.2 Sampling Theory

If all M_{h_i}, the size measure for school i in stratum h, are known, we can select schools with probabilities proportional to their size M_{h_i}. Brewer's method [1,2] selects two units from each stratum, with probability proportional to size and without replacement. The selection probability for unit i in stratum h equals

$$\pi_{h_i} = \frac{2M_{h_i}}{M_{h.}} = 2Z_{h_i} \quad .$$

Brewer's algorithm draws the first unit with probability

$$\frac{Z_{h_i}(1-Z_{h_i})}{D_h(1-2Z_{h_i})}$$

where

$$D_h = \sum_{i=1}^{N_h} \frac{Z_{h_i}(1-Z_{h_i})}{(1-2Z_{h_i})}$$

is the divisor needed for conversion into probabilities. Then, a second unit is selected with probability

$$\frac{Z_{h_j}}{1-Z_{h_i}}$$

where unit i is the unit drawn first. Since our goal was to sample 110 schools and we had 45 strata, we needed more than 2 schools per strata. Sampford [2,4] extended Brewer's method to samples of size $n_h > 2$. For this method, it can be shown that the selection probability for unit i in stratum h equals

$$\pi_{h_i} = n_h \frac{M_{h_i}}{M_{h.}} = n_h Z_{h_i} \quad .$$

Sampford's method first selects a unit from stratum h with probability Z_{h_i} and all subsequent units j with probabilities proportional to

$$\frac{Z_{h_j}}{1-n_h Z_{h_j}}$$

with replacement.

2.3 SAS Application

Sampling, statistical calculations, and all analyses were performed using SASv9 [5]. SAS® PROC SURVEYSELECT was used to draw a random sample of schools with METHOD = PPS_SAMPFORD option. The number of units drawn per strata was altered between 2 and 3 using the N = (<list>) option. Enrollment was used as a measure of SIZE and three stratification variables described above were used with the STRATA statement. Responses were weighted to include stratified design features, probability of selection at each stage, and adjustment for adolescent non-response.

3. Results

3.1 General Summary Measures from the Survey

Sampling resulted in 113 schools being selected from which we obtained 1,551 family responses. From those 1,551 family responses, we further received 1,110 parent proxy responses and 209 adolescent responses. The cooperation rate of eligible families was 76%. The sum of weights for schools added to 1,039, for families to

387,568, and for all children to 453,180. Responders for the family were predominantly females (81%), and 17% were minority.

3.2 Specific Results from the Survey

Each child's weight category was determined based on child's gender and age, parent's report of the child's height and weight and appropriate CDC BMI percentiles. The distribution of the weight categories derived from the survey data closely resembles the distribution reported by the Arkansas Center for Health Improvement (ACHI), which collected the data on all public school children in Arkansas (Figure 2) [6]. ACHI used the measured height and weight of each child for calculating BMI and categorization of weight. Thus, ACHI data were direct observations, and Act 1220 Survey data were parental reports.

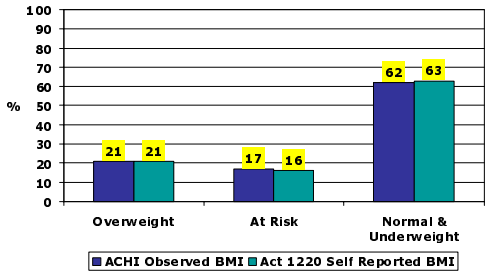

Figure 2. Weight category distribution from two sources (ACHI and Act 1220 Survey)

We also found that 60% of the parents of children whose BMI percentile puts them in the at risk or overweight categories do not perceive their child to be at risk or overweight [3].

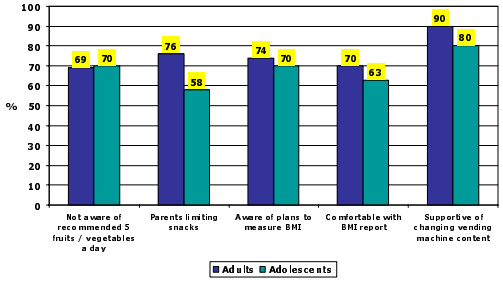

Figure 3. Selected baseline outcomes for adults and adolescents

Several selected results are presented in Figure 3 to show the comparison between parent and adolescent responses to a few questions that were asked of both [3].

4. Discussion

This design allowed us to capture and handle complexities associated with the goals of the Act 1220 evaluation. It became apparent that it was more feasible to interview a number of families proportional to schools enrollment (1-2%) than it was to locate 20 families from schools that had an enrollment of 100-200. Based on the experiences from this project, we recommend increasing the number of sampled schools per strata for similar surveys, in order to increase the target for the interviews and to gain more power at the sub-group analysis level.

Acknowledgements

Support for this research was provided by the Robert Wood Johnson Foundation grant 050262.

References

1. Brewer, K. R. W. 1963. A model of systematic sampling with unequal probabilities. *The Australian J of Statistics,* 5;5-13.

2. Cochran, W. G. (1973). Sampling techniques. 3rd ed. New York: John Wiley & Sons, Inc.

3. Raczynski J.M., Phillips M. M., Bursac Z., Pulley L., West D., Birdsong M. C., Evans V. L., Gauss C. H., Louvring M. S., Walker J. F. (2005). Establishing a baseline to evaluate act 1220 of 2003: An act of the Arkansas general assembly to combat childhood obesity. *University of Arkansas for Medical Sciences, College of Public Health*.

4. Sampford, M.R. 1967. On sampling without replacement with unequal probabilities of selection. *Biometrika,* 54;499-513.

5. SAS Institute Inc., SAS 9.1 Help and Documentation, Cary, NC: SAS Institute Inc., 2004.

6. Thompson, J. et. al. 2004. The Arkansas assessment of childhood and adolescent obesity. *Arkansas Center for Health Improvement*.