

## Discussion

*Donsig Jang*<sup>1</sup>

I would like to thank the committee for giving me the opportunity to participate in the Hansen lecture series as a discussant.

Sharon Lohr chose an interesting title for her Morris Hansen lecture – “The Care, Feeding, and Training of Survey Statisticians.” In particular, the word “care” caught my attention. Sharon’s lecture calls for our commitment to “caring” for statisticians – promoting not only the growth of the next generation of statisticians but also the professional development of our current colleagues in this rapidly changing world.

My discussion of the “care and feeding” of survey statisticians is based on my experiences in survey sampling, both as a practitioner and as an instructor on sampling courses. Some of my points may sound redundant because Sharon already discussed them in her lecture, but I strongly agree with many of her arguments.

### 1. Survival of the Species

In his Senate confirmation hearing, Dr. Robert Groves (director of the U.S. Census Bureau) responded to Senator Akaka’s comments about the expected exodus of Census employees from the bureau as they reach retirement age. Dr. Groves expressed his concern about the lack of academic programs to train new employees for the U.S. Census Bureau, which is the largest employer in the nation for survey statisticians. Others have raised similar concerns – especially in the past two decades – about the lack of survey statisticians “in the pipeline.” For example, Kalton (2002):

There is a notable lack of educational and training programs in survey research around the world. Few universities provide, or are equipped to provide, these programs. They seldom have the full range of faculty needed, and most lack instructors with the wealth of experience required to teach the applied aspects of survey research effectively.

In response to these shortages, new university programs have been created to train survey researchers. For example, the Federal Interagency Consortium on Statistical Policy helped establish a special program in 1993: the Joint Program for Survey Methodology. There are also a few universities with strong survey research programs. These programs have helped fill the “pipeline” with survey statisticians, and we can now see many graduates of the programs working in the field. But do we require more academic programs to turn out even more survey statisticians, as Dr. Groves testified?

<sup>1</sup> Mathematica Policy Research, 600 Maryland Avenue, SW, Suite 550, Washington, DC 20024, U.S.A.  
Email: djang@mathematica-mpr.com

To help answer this question, we need a better understanding of the “supply and demand” for survey statisticians.

## 2. Supply of Potential Survey Statisticians

A student can become a survey statistician without majoring in statistics, but for the purpose of this discussion I am only considering graduates with master’s or doctoral degrees in statistics as “potential survey statisticians.” Figure 1 shows the number of master’s and doctoral degrees awarded to statistics majors in the U.S. from academic year (AY) 2004 to 2008. While the number of overall awards conferred by U.S. post-secondary institutions steadily increases from year to year, growth in the number of statistics degrees awarded has been very slow, if not completely flat.

In 2008, master’s or doctoral degrees in statistics were awarded to about 1,700 graduates – only three-tenths of a percent of all graduates in the nation. If we narrow the pool to only U.S. citizens and permanent residents, the number is even smaller: in 2008, U.S. citizens and permanent residents made up about 40 percent of all graduates who majored in statistics, or 735 graduates. Although students with nonstatistics majors can become survey statisticians, we can see here that the maximum pool of graduates each year for potential survey statisticians would be no more than 700, if we limit them to the U.S. residents graduating with statistics major.

The number would shrink again if we look at the specific educational backgrounds of these graduates. For example, since survey sampling is a foundation of survey methodology, sampling expertise is often viewed as necessary for survey statisticians. But how many graduates are exposed to survey sampling? According to Sharon’s survey of 80 statistics/biostatistics programs, only about 60 percent of schools offer sampling courses, and even in those schools, students may not take sampling unless the school makes it a requirement. So, if we further restrict the pool of potential survey statisticians to graduates who have taken sampling courses, we are left with no more than 50 percent of graduates majoring in statistics, or about 300 to 400 graduates.

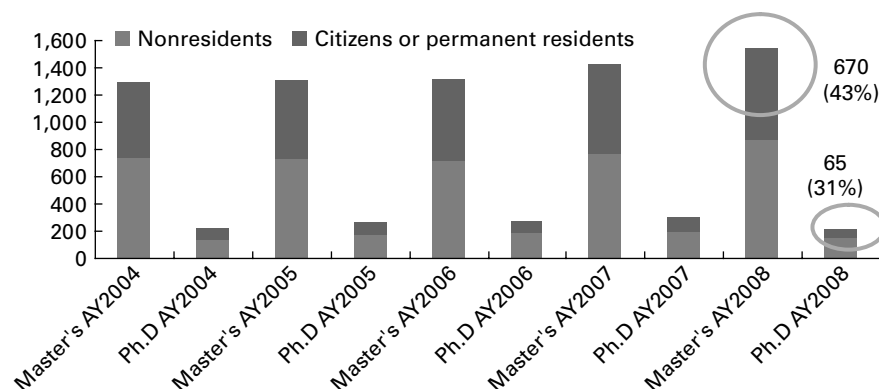


Fig. 1. Number of Statistics Degrees Awarded in U.S. from AY 2004 to AY 2008. Source: Integrated Postsecondary Education Data System (IPEDS) Completions Data.

Is this enough to fill all available positions for survey statisticians? Also, keep in mind that not all potential survey statisticians may choose to enter the field – for example, some may take sampling classes but then decide that sampling is “boring,” or they may simply find other career paths more appealing.

### 3. Low Demand?

Another reason students may not pursue a career in survey research is the perceived lack of demand for survey statisticians – a perception reinforced by the small number of ads posted on online job search sites and in other listings. For example, in the September 2009 edition of *Amstat News* (the monthly membership magazine of the American Statistical Association), only 3 out of 28 job postings were for survey statisticians. Furthermore, the *Amstat* online job search database returns only one job opening for survey statisticians as of early October, 2009. So where is the demand for survey statisticians that many of us have been talking about? What kind of message does this send to graduate students? Students may assume that survey sampling is not an area with potential growth – a misperception that must be changed.

### 4. Awareness

As Sharon pointed out, sampling has been an isolated subject, even in statistics programs in many universities. But perhaps we in the survey research community can do more to enhance awareness of survey sampling among students. For example, we could place more job postings in *Amstat News*, which I believe is most visible to graduate students majoring in statistics.

As Sharon said, high salaries would definitely attract more people to the field. We should also provide graduate students with more internship opportunities in government and private survey research agencies. This will allow students to actually see how exciting real survey research can be. We need to have more schools offer sampling courses with experienced sampling instructors. Finally, we should make sampling classes more interesting to change the image of survey sampling from boring to exciting.

### 5. Designing Effective Sampling Courses

Sampling courses should be informative, interesting, and well-taught by experienced instructors. The curricula must include the fundamental design-based sampling framework. This is even more important nowadays because, as Sharon pointed out, nonprobability samples are emerging in response to the ever-evolving information age. But we should not stop there; instructors must also expose students to the applied aspects of sampling. Indeed, one reason why students may shy away from survey sampling, even after taking sampling classes, is that they often do not get a chance to see survey sampling “in action.”

#### 5.1. Getting Students Excited About Sampling

Use of real survey data in the classroom might also motivate students to tackle challenging issues, making sampling more exciting and meaningful. A great quantity of public data for government surveys is now available through websites or upon request and can be used in

the classroom. Analysis software developed for complex survey data can also help make classroom examples more interesting. For example, SAS (which is available in computer labs in almost all universities) contains newly added procedures for the analysis of complex survey data. R, a freeware program, also has a special module to deal with complex sample data.

## 5.2. Experienced Instructors and High-Quality Sampling Texts

Instructors should have a solid grasp of sampling theory as well as real survey experience. However, from Sharon's map showing survey research method section members in the country, we can see that there is a nationwide shortage of survey statisticians to serve as instructors, except in a few areas such as Washington, DC. We need more experienced survey statisticians throughout the country, especially in academic institutions.

High-quality sampling books are also essential, both to help instructors expand their knowledge of sampling and to provide students with a broad view of survey sampling. Here are a few of the sampling books on my bookshelf:

- *Sample Survey Methods and Theory I and II* by Hansen, Hurwitz, and Madow 1953
- *Survey Sampling* by Kish 1965
- *Sampling Techniques* by Cochran 1978
- *Practical Statistical Sampling for Auditors* by Wilburn 1984
- *Sampling* by Thompson 1992
- *Model-Assisted Survey Sampling* by Särndal, Swensson, and Wretman 1992
- *Practical Sampling Techniques* by Som 1996
- *Theory of Sample Surveys* by Thompson 1997
- *Sampling: Design and Analysis* by Lohr 1999
- *Finite Population Sampling and Inference: A Prediction Approach* by Valliant, Dorfman, and Royall 2000
- *Sampling Methods: Exercises and Solutions* by Ardilly and Tillé 2006
- *Sampling Statistics* by Fuller 2009

All these books cover basic sampling theory. Two books address both sampling theory and real survey application in a well-balanced manner: *Sample Survey Methods and Theory I and II* (by Hansen et al.) and *Sampling: Design and Analysis* (by Lohr). Until recently, I used the former book mainly for reference. But after I was asked to be a discussant of Sharon's article, I read a few chapters of the book thoroughly and found it to be very insightful. The authors recognize the importance of real survey application as well as fundamental sampling theory. In their book, they introduce three real surveys in which readers can actually see basic sampling theory and methods applied.

Sharon's book is very comprehensive, covering both fundamental sampling theory and real survey examples, and she did an exceptional job of making the information clear and understandable, even for nontechnical readers. Sharon's book also covers many of the current survey sampling topics that are not usually covered in standard sampling books, such as sampling weights, nonresponse, and complex data analysis. Additional topics will likely be covered in the second edition of her book. I am currently using her book to teach two-semester graduate-level sampling courses at George Mason University.

## 6. Change

Besides enhancing awareness of survey sampling among graduate students through more exciting sampling courses, we “adult statisticians” (to use Sharon’s term) must undergo a “paradigm shift” in our own sampling work. Some of us, especially practitioners like myself, probably use the design-based approach too much in tackling design and estimation of survey samples. For example, we tend to use a design-based approach not only for variance estimation but also for nonresponse weighting adjustments (using a weighting cell method) and for imputation (using a hot-deck method); the latter two are based on a quasi-randomization approach.

However, we can no longer rely on the design-based approach in this era of information overload. For example, survey statisticians used to use the final response status and limited frame information to perform nonresponse weighting adjustments. But nowadays, data collection has become multimode with more complex data collection protocols, including incentive provision and other customized treatments. So using the final response status for nonresponse adjustments may not be enough to identify the cause of nonresponse and to estimate characteristics of nonrespondents. Survey statisticians have now started to look at the paradata – the information about the survey process – for a source of statistical adjustments to reduce nonresponse bias.

Change is coming. The complexities and information overload stemming from the survey process require survey statisticians to be more receptive to other approaches besides design-based, especially models and mainstream statistics. In addition, design-based and model-based approaches are merging, and we must be open to this paradigm shift. It is not a choice – it is a “must” for survey statisticians as we enter into a new era of survey research. Change is challenging, of course, but it will help us greatly expand the horizons of survey research application so that we can draw more people into the field in coming years.

## 7. Reference

- Kalton, G. (2002). Training Needs for Survey Statisticians in Developed and Developing Countries. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1746–1750.

Received March 2010