CHAPTER 14

Selected Topics in Sample Design and Estimation Methodology

In this chapter we will discuss some sampling and estimation techniques that have been developed for special purposes. More specifically, we will first discuss some sampling methods that have been developed for use in health surveys conducted in developing countries, and then examine techniques that have been used in developed countries in order to meet special needs.

14.1 WORLD HEALTH ORGANIZATION EPI SURVEYS: A MODIFICATION OF PPS SAMPLING FOR USE IN DEVELOPING COUNTRIES

The World Health Organization (WHO), through its Expanded Program on Immunization (EPI), aims to ensure the availability of immunization to all children of the world by the year 1990. As part of this program, it has developed methodology for estimating, by use of relatively quick and inexpensive sample surveys, immunization levels of children in areas that would be targeted for special immunization programs if the immunization levels were low. The methodology makes use of a modification of probability proportional to size (PPS) sampling developed originally in the United States [1] and later modified for use in the Smallpox Eradication Program in West Africa [2].

Although it can be adopted to meet other objectives, the major objectives of an EPI survey is to estimate the immunization coverage (i.e., the proportion of children having all their required immunizations) in a given target area (this could be a village, town, city, etc.). Over the years, the practice evolved in these surveys of selecting 30 sample clusters and 7 children within each cluster, which would yield a sample size of 210 children. In fact, these surveys became referred to as "30×7" surveys.

Although hundreds of these "30×7" surveys have been conducted, the rationale for the choice of the sample size is rather unusual. Assuming that 50%

of the target population is covered by immunization, a simple random sample of size 96 would be required to be 95% confident of estimating this proportion to within 10 percentage points of the true value. Simple random sampling in the environment in which EPI surveys are conducted, however, is not cost effective, and so some cluster sampling technique would normally be employed. If one assumed a design effect of 2, the required sample size would be doubled, and one would need 192 (2 times 96) individuals in the sample. The developers of the EPI methodology were determined to use 30 clusters. Therefore, a sample of 30 clusters with 7 individuals per cluster would yield slightly more than the desired sample size. This is described in more detail by Lemeshow and Stroh [3].

Clusters in the EPI surveys depend on the particular population being surveyed and are generally villages, cities, towns, health service districts, etc. for which population information is available. At the first stage, a PPS sample of 30 clusters is taken (as described in Chapter 11). At the second stage, however, for budgetary and logistic reasons, the process of selecting the seven individuals differs markedly from traditional cluster sampling. Instead of taking the seven subjects at random from all available subjects, the EPI methodology calls for randomly selecting a starting household (HH), collecting all relevant information on eligible subjects in that HH, and then proceeding to the HH whose front door is physically closest to the HH just visited. This process of visiting next closest HHs and collecting information on *all* eligible individuals in those HHs continues until the required seven subjects are studied.

The particular procedures used for selecting the first HH at random at the second stage of sampling depends on the nature of the cluster. For example, in a rural area, the interviewer might be instructed to go to a centrally located landmark in the cluster (such as a church, market, etc.), randomly select a direction in which to walk (i.e., north, south, east, or west) and count the number of households (K) found in that direction from the starting point to the town boundary. A random number between 1 and K would then be selected which would designate the first HH chosen. As an operating rule, data are collected on all eligible children in the household contributing the seventh child, even if that results in including more than 7 children in a cluster.

The method as described above of selecting children within sample clusters in EPI surveys has been found to be relatively easy to implement in the field. The costs of taking a true random selection at the second stage would be prohibitive for the resources typically allocated to these EPI surveys. In an effort to evaluate the effect on estimates of immunization coverage of this nonrandom selection of individuals within clusters, a computer simulation model was developed [4]. This model compared the results of the EPI selection method to the results obtained by more traditional methods using artificially created populations having specified characteristics. It was found that although the EPI method performed poorly within particular clusters when there was pocketing of nonimmunized individuals, the resulting estimates, over all clusters, tended to be accurate to within 10 percentage points of the true population levels. This was within the goals specified by the EPI Program and may be attributed to the fact

that a large number of clusters is selected and biases occurring within clusters tend to average out over the set of 30. Thus, the method appears quite useful when used for the target areas as a whole, but could provide highly unacceptable estimates if used for particular clusters or subgroups.