

# 학교보건연구에 있어 국민학생을 대상으로 한 표본설계법

최재원\*·김정근\*\*

## I. 서론

본 연구는 우리나라 전국의 국민학교 학생을 대상으로 할 때의 표본추출을 위한 설계법에 관하여 검토하는 것을 목적으로 시도된 것으로 직접 계산한 수치를 이용하여 예를 들었다.

여기에서 제시하고자 하는 설계법은 3단계 표본추출법(three stage design)이다. 첫번째 단계는 도(province)등을 선정하는 것이고 두번째 단계는 선정된 도로부터 학교를 뽑는 것이다. 세번째 단계는 추출된 학교로부터 학급을 추출하는 것이다.

이때 표본으로 추출된 아동들을 대상으로 우리는 건강이나 영양상태 등에 대한 측정치(measurement)나 반응치(response)를 수집하게 된다.

그러나 필요한 경우 네번째의 단계(four stage)를 도입하여 학급내에서 일부분의 학생을 추출하기 위한 방법을 연장하여 생각하여 볼 수도 있다.

## II. 전제사항

실제로 표본을 추출하기에 앞서 표본추출설계(sampling design)에서의 몇가지 기본적인 부분

을 다음과 같이 살펴보기로 하자.

### 2. 1. 관련자료의 사용 가능성

관련자료가 있으며 또한 사용할 수 있다면 새로운 조사계획을 짜는데 활용할 수 있다. 수정된 또는 이와 유사한 조사를 하여야 할 때, 표본조사계획(sample survey design)을 짜기 위하여 과거의 경험과 자료를 활용할 수 있다.

### 2. 2. 세 가지의 자료원

국민학교 아동으로부터 얻을 수 있는 자료의 내용은 3가지로 구분할 수 있다. 첫째 형태는 신장, 체중, 비만도, IQ, 또는 시력과 같이 측정하거나 검사하여 얻는 자료이다. 둘째는 연구대상기간 중에 특정 상태나 사고에 관한 질문의 응답내용들이다. 셋째는 검사결과, 점수, 그리고 기타 가족의 배경 등에 관하여 학교기록으로부터 얻는 자료들이다.

### 2. 3. 일반적인 접근방법

모집단으로부터 표본을 택하는 몇 가지 일반적인 접근방법을 경비, 모집단형태, 기타 모집단의 특성에 관하여 살펴 보아야 한다. 예를 들면 지역 또는 명단(area or list)을 통한 표본추출, 패널 또는 그렇지 않은 표본추출, 동일하거나 동일하지 않은 표본추출을, 그리고 한단계 또는 여러

\* 미국 국립보건통계센터

\*\* 서울대학교 보건대학원

단계 표본추출법(multistage sampling)등을 언급할 수 있다.

위에서 언급한 일반적인 원칙에 대하여 검토를 한 후 단순확률표본(simple random sampling), 층화확률표본(stratified random sampling), 집락추출(cluster sampling)등 특정한 표본추출방법에 관하여 토의하게 된다.

#### 2. 4 일시적 또는 연속적 표본조사

표본조사는 일회에 그칠 수 있고 또는 연속적인 것일 수도 있다. 연속적인 경우 조사 기간에 걸쳐 모든 기간동안 조사를 하게 되며, 그 결과 보고는 월별 또는 분기별로 하게 된다. 대부분의 보건에 관한 표본조사에서는 특정한 자료항목의 계절별 편의(seasonal bias)를 최소한으로 하도록 하기 위하여 일년동안에 걸쳐서 하게 된다. 이때 매분기별 자료를 보고하든지 또는 축적하든지 한다.

표본조사를 몇 년 동안에 걸쳐 하게 되면 시계열(time trend)에 의한 추세나 미래의 결과에 대한 예측(prediction)을 알아 볼 수도 있다.

#### 2. 5 모집단(universe)

표본을 추출하게 되는 모집단(population or universe)은 명확하게 정의되어야 한다. 각 가지 학교와 학급들이 있다. 학교와 학급의 특성은 매우 다양하다. 이때 우리의 관심은 다음과 같은 것 중의 하나에 국한된다. 즉 특수학교는 물론 유치원, 국민학교, 중학교 등이다. 한편 우리는 이 학교의 특정한 학년에 연구의 초점을 맞출 수 있다. 예를 들면 국민학교의 6학년만을 대상으로 연구를 시도하려면 우리나라의 모든 6학년 학생을 대상으로 표본을 추출하여야 한다. 여기에서는 한국의 국민학교에 관심을 두는 것으로 한다.

#### 2. 6 층화경계(stratification boundaries)

관리목적으로 우리는 각 지역의 특성에 따라 지역이나 인구를 층화한다. 예를 들면, 우리나라는 첫단계에서 특별시 및 직할시, 도 등으로 층화하고 두번째 단계에서는 구, 시, 군으로 층화할 수 있다.

#### 2. 7 표본설계를 위한 기본변수(key items)

표본조사를 설계할 때 구하고자 하는 것 중에서 중요한 몇 가지의 통계량에 초점을 맞추는 것이 좋다. 대부분의 경우 전형적이고 중요한 통계량 몇 가지를 선정하여 설계를 위한 계산 등에 필요한 기본변수로 사용한다. 아동들의 건강평가를 위하여 신장 및 체중, 결석일수, 만성질환진단내용, 그리고 영양상태 등을 기본변수로 고려할 수 있다.

#### 2. 8 표본수

분산, 고정된 예산, 기타 표본수에 대한 제약점 등을 기초로 함수를 설정한다. 모수( $n_1, n_2, p$  등)에 관한 이러한 함수의 적정화(optimization)를 통하여 적절한 학교의 수, 각 학교로부터의 학급의 수를 얻을 수 있다(Hansen, Hurwitz, and Madow, 1953, Vol. 2, p. 199).

#### 2. 9 통계량의 표본분산

표본이 작으면 큰 표본의 경우보다 표본분산은 훨씬 커지게 된다. 큰 표본은 분산을 작게 하기 주니 경비가 훨씬 더들기 때문에 표본추출에 따른 예산내로 수용할 수 있도록 표본분산은 탄력성이 있어야 한다. 대규모의 표본조사에서는 우리가 분석하고자 하는 변수의 유형에 따라 1% 또는 10%의 표본을 택한다. 표본분산은 모집단분산의 불편적(unbiased)이고 일관성(consistent) 있는 추정치이다.

#### 2. 10 표본추출에 따른 경비

각 개체의 비용추정을 통하여 표본의 수집에 대한 다음과 같은 내용에 대한 전반적인 비용을 추정할 수 있다. 즉,

1. 계획과 예비조사(pretest),
2. 자료수집 : 사무실 사용에 따른 관리비(overhead cost), 기구, 컴퓨터 그리고 기타 관련되는 제 비용과 표본추출, 여비, 아동들의 건강상태를 면접 또는 측정할 때 관련되는 비용(이 경우는 학교당 단위개체에 대한 경비와 각 아동들이 속한 학급당 비용을 포함한다),
3. 연속적인 표본조사일 경우 매 5년마다 또는

필요한 햇수마다 표본설계를 개정할 때의 비용,

4. 자료처리 및 코딩,
5. 자료분석,
6. 보고서 인쇄비용 등이다.

사용할 수 있는 전체 예산이 확실하지 않으면 총비용과 개체를 표본추출할 때의 경비는 가능한 일찍 결정하여야 한다. 이 경비의 규모는 조사할 표본수를 결정하는데 결정적인 영향을 미친다. 그러나 실제예산이 다소 차이가 나는 경우 그 이론은 여전히 적용된다.

### 2. 11. 일정표

표본조사의 구체적인 일정을 밝혀 주면 표본조사를 잘 할 수 있게 된다.

## Ⅲ. 표본조사설계

우리는 먼저 모집단을 정의하고 그 모집단의 층화 (stratification), 그리고 도, 학교, 그리고 학급의 세 단계표본추출의 골격(three stage master sample), 그리고 관심의 대상이 되는 변수를 포함한 자료수집의 일반적인 구성형태(geral scheme) 등을 정의한다.

### 3. 1 모집단

모집단은 우리나라의 모든 국민학교라고 정의한다. 그리하여 이 모집단으로부터 표본을 추출하고 이 표본으로부터 반응치(response)를 측정한다.

### 3. 2 모집단의 층화

모든 국민학교를 8개의 도, 3개의 특별시와 직할시(광주와 대전은 인구수의 규모로 하여 편의상 각 도에 편입) 등으로 하여 분류한 후, 각 단위내에서 학교의 크기에 따라 5개의 층으로 구분한다고 하자. 여기에서는 한 지역에 대해서만 고려하기로 하고 필요하다면 11개의 행정구역의 각각에 대하여 동일한 원칙을 적용할 수 있다.

한 학급에 60명의 학생이 있다고 가정하고 이 대상지역의 층아동수가 속한 지역에 대하여 층을 정의한다.

<표 1> First Stage Strata of 11 Areas

Province or Big City	No of Classes (artificial number)	Probability of Selection
Seoul	2000	200/920
Pusan	1000	100/920
Daegu	700	70/920
Kyonggi	500	50/920
Kangwon	300	30/920
Kyongpuk	1200	120/920
Kyongnam	1100	110/920
Chonbuk	700	70/920
Chonnam	600	60/920
Chungbuk	400	40/920
Chungnam	700	70/920
	9200	1

한 학급에 60명의 학생이 있다고 하면 <표 1>의 11개의 층에서 하나로서 첫번째 층을 잡아서 두번째 단계의 표본추출 방법을 가장적인 학급수를 이용하여 예시하여 보자.

<표 2> Sub-stratification of School

Sub-Strat	School size	No of school	No of class
1	<12 class	10	120
2	12-18-23	20	360
3	24-30-47	15	450
4	48-54-99	10	540
5	100 class<	2	200
			1,670

위의 가상적인 예는 표본추출방법을 예시하기 위한 것이다. 100개의 학급이 표본으로 필요하고 선정된 도 등에 있는 모든 국민학교 중에 모두 1,670 학급이 있다고 하자.

이러한 표본아동들의 선정은 아래에 언급하는 학교추출과 학급추출의 두 단계 확률법칙에 따른다.

그러한 선정은 두 단계들의 표본추출 목적이 된다. 두번째 단계에서는 크기에 비례한 확률(probability proportional to size)로 표본을 추출하고 세번째 단계에서는 복원 추출의 동일확률(equal probability with replacement)로 학급을 뽑는다. 이는 <표 3>에 나타나 있다.

<표 3> Selection of the 100 sample classes

St	Pro. of Sch	No. of Sch	Sample school no. $n_1$	Class per Sch. $n_2$	$n_1 \times n_2$	Prob of class	100 sample classes
1	$\frac{120}{1670}$	10	0.72	12	8.64	$\frac{1}{120}$	7.2
2	$\frac{360}{1670}$	20	4.31	18	77.70	$\frac{1}{360}$	21.6
3	$\frac{450}{1670}$	15	4.04	30	121.26	$\frac{1}{450}$	26.95
4	$\frac{540}{1670}$	10	3.23	54	174.61	$\frac{1}{540}$	32.3
5	$\frac{200}{1670}$	2	0.24	100	23.95	$\frac{1}{200}$	11.98

여기에서 우리나라의 모든 국민학교의 수와 규모에 관한 자료를 문교부등지를 통하여 입수할 수 있고 이를 토대로 모든 학교의 전체적인 목록을 작성할 수 있다고 하자.

그 선정된 표본의 구조에 관한 형태(frame-work)는 세단계 층화 표본추출법이 되는데 이때 도와 학교는 각각 1차 및 2차 층, 학급은 3차 집락(cluster)이 된다.

첫번째 및 두번째 단계에서 도와 학교를 뽑을 때 크기가 큰 대상은 표본에 뽑힐 확률이 커지고

작은 것은 상대적으로 낮아진다.

세번째 단계에서는 학급을 동일한 확률로 뽑히 이는 복원추출(with replacement)로 뽑힌 것으로 가정한다.

학급을 단순확률표본(simple random sampling)으로 뽑는데, 이때는 학급의 크기는 60명으로 동일하여야 한다. 이 학급은 최종 표본추출단위가 된다.

이 표본으로 선정된 학급의 모든 어린이는 면접조사나 측정의 대상이 된다. 그 표본으로부터

의 모집단의 평균과 분산의 추정방법은 부표 A에 제시하였다.

### 3.3 학교나 학급의 신설 또는 폐쇄

이미 표본으로 선정된 학교나 학급이 신설되거나 폐쇄되는 경우에는 주의깊게 취급하여야 한다. 학교가 신설되었다면 표본추출과정에서 별도로 관리하여야 하는데, 특히 연속적으로 표본조사할 하는 경우에는 특히 이 점을 유의하여야 한다.

## IV. 패널(panel)

학교표본을 4개의 패널을 구성하여 목록(master sample)을 작성할 수 있고 각 패널은 우리나라 학교 아동들의 확률표본이 된다. 학교의 규모가 아주 다양하다면 각 패널에 4가지의 크기의 학교를 거의 동수로 포함하게 한다.

1) 또 패널은 다른 목적으로도 사용할 수 있다. 하나의 특정한 단위기간으로 또는 예비조사로 사용할 수 있다. 단위기간이 분기라면 4개의 패널은 일년에 해당된다. 표본을 추출하는 연구자가 충분한 예산을 갖고 있는지를 알 수 없다면 하나 이상의 패널을 구성하거나 예산의 정도에 따라 표본수를 조정하기 위하여 몇 개의 패널을 탈락시킬 수 있다.

## V. 하나의 패널을 이용한 예비조사

정확한 표본수와 그 구성내용을 최종적으로 결정하기 전에 적은 수의 학교를 대상으로 예비조사를 실시하여야 한다. 이것에 대해서는 몇 가지 타당한 이유가 있다. 실제 자료를 관찰하면 어떠한 적정화도 정확하지 못하다.

예비조사는 이미 구체적으로 제시된 표본추출설계에 대한 전반적인 윤곽을 잡기 위하여 필히 실시하여야 한다. 이렇게 하여 주요한 오류를 사전에 방지할 수 있는데 그렇지 않으면 본조사후에 이를 바로 잡을 도리가 없다.

필요한 현장 경험을 얻고 신중하게 고려한 추출설계법을 평가할 수 있는 더 정확한 자료를 수집하기 위하여 최종적으로 학교와 학급수를 결정

하기 이전에 작은 수의 학교를 (이를 주요 패널이라고 한다) 일정한 기간동안 조사할 필요가 있다. 그 주요 패널은 이를 예비표본으로 사용한 후 최종 설계법의 핵심부가 된다.

## VI. 추정 및 표본분산

합계, 평균, 비율 등의 모수는 세 가지 기본적인 구성요소에 의하여 추정된다. 세 가지 기본요소라 함은 1) 추출확률의 역수(reciprocal)에 의한 확대(inflation), 2) 비용담 보정(nonresponse adjustment)의 두 가지 수준, 3) 알려진 합계에 따른 비(ratio) 보정의 두 가지 수준등을 말한다. 비용담에 대한 보정은 학교내에서 이루어질 수 있다.

표본을 구하게 되면 항상 표본분산(sample variance)을 갖게 된다. 가중치(weighted)를 택하게 되면 표본분산은 줄어들게 되지만 가중 자료(weighted data)를 이용하게 되면, 그 추정치의 분산을 구하기가 훨씬 어려운 경우가 많다.

## VII. 자료의 질관리

가끔 유의한 자원을 할당하여 조사되는 학교로부터 얻는 자료의 질과 이에 따른 코딩 그리고 자료처리 등의 질에 대한 감시를 하도록 한다.

각 학교로부터 조사한 자료의 5% 계통표본(systematic sample)을 이용하여 코딩의 타당성을 조사할 수 있다.

오차율(error rate)이 100개의 코딩마다 15개를 초과하는 경우 그 자료를 사용하는 데는 문제가 있다. 유난히 빠르게 코딩하는 사람을 관리하는 방법은 그의 오차율이 50명의 코딩 결과의 평균 오차율로 부터 2배의 표준편차보다 클 경우 그는 재훈련시켜야 한다.

## VIII. 보고체계

표본으로 추출된 학교로부터 얻는 자료는 연구를 담당하는 책임자에게 주기적으로 보고하여야 한다. 이 담당자는 관찰(surveillance)하고 체제를 관리(control)하기 위하여 자주 현지를 방문

하게 된다.

### 참 고 문 헌

Hansen, M. H., Hurwitz, W. N., and Madow. W. G. (1953) : Sampling Survey Methods and Theory, Vol. II. Theory, John Wiley and Sons, Inc.

Kendall, M. G., and Stuart, A. (1968) : The Advanced Theory of Statistics. Vol. 3, Hafner Publishing Company, New York.

### 부 록 A

수리적 모형 : 학교의 수를  $N_i$ ,  $i$ 번째 학교의

학급수를  $N_{2i}$ ,  $i$ 번째 학교의  $j$ 번째 학급의 아동 수를  $N_{3ij}$ 라 나타낸다고 하자.

이에 따른 표본수는 동일하지 않은 표본 밑수와 함께 소문자  $n$ 으로 나타낸다.

$ij$ 번째 학급의 아동들은 모두 조사대상으로 택하므로 표본추출과정을 거치지 않고 표본으로 취해졌다고 가정한다. 필요하다면 그 학급에서 다시 표본을 뽑을 수도 있다.

$$\bar{y} = \frac{N_1}{n_1} \sum_{i=1}^{m_1} \frac{N_{2i}}{n_{2i}} \sum_{j=1}^{m_{2i}} \frac{N_{3ij}}{n_{3ij}} \sum_{k=1}^{N_{3ij}} y_{ijk}$$

모평균의 추정치는 위의 공식에 의하여 추정되며 분산  $\text{var}(y)$ 를 갖는다 (Kendall and Stuart, Vol. 3 참조).

# How To Take a Sample From Korean Primary School Children

Jai Won Choi\* • Jong Kun Kim\*\*

## I. INTRODUCTION

In this paper we present a sampling design to take a sample from the entire Korean school children, and illustrated the sampling with a numerical example.

It is a three stage design. The first stage is the selection of province, and secondly of schools from the selected province, and the third stage is the selection of classes from selected schools.

We may take the measurements or responses on the health or nutrition status from all the children in the sampled classes.

If necessary we may extend the three stages to the four stages including one more stage of the selection of school children from selected classes.

## II. CONSIDERATIONS

Before starting a sampling design, we have to consider some principal components aside from actual sampling design.

### 2. 1 Available Relevant Data

When any relevant data is available, we may use them in developing a new survey. When an

updated or a similar survey is needed, we might use previous experiences and data in order to plan a sample survey design.

### 2. 2 Three Sources of Data

There may be the three types of information from school children. The first types are the measurements or examinations such as height and weight, fatness, IQ level, and eye sights. The second types are the responses to questions on certain conditions or accidents during a reference period. The third type is the information from the school records regarding test results, grades, and other family backgrounds.

### 2. 3 General Approaches

Some general approaches in taking a sample from a population should be discussed in relation to costs, population structures, and other features of population. For example, it could be an area or list sampling, panel or nonpanel sampling, equal or unequal probability sampling, and one stage or multistage sampling.

After discussion of general principles mentioned above, we may further discuss specific sampling methods such as simple random, stratified random, or cluster in selection of a sample.

---

\* National Center for Health Statistics

\*\* School of Public Health, Seoul National University

#### 2. 4 One Time or Continuous

A survey could be a one time or continuing one over the years. If one time, it should cover the time within a specified survey period. Most health surveys are done during one year period in order to minimize seasonal bias for certain data items. Every quarterly data could be published or accumulated.

Continuing survey are often done through the years, and provide valuable data on time trends and predictions.

#### 2. 5 Universe

The population or universe from which a sample is taken should be defined clearly. There are various schools and classes. Our interest may be restricted to one of the following types : kindergarten, primary school, middle schools as well as special schools. We may also restrict our study to one of the grades in those schools. For instance, if we only want to study the 6-th grades in primary schools, we take a sample from the entire 6-th graders in Korea. We assume that we are interested only in the Korean primary schools.

#### 2. 6 Stratification Boundaries

For the purpose of administration, we may stratify the areas or population according to their characteristics. For instance, the Korea can be stratified by DO or Special City at the first stage, and GOON or GOOCHUNG for the second stage.

#### 2. 7 Key Items for Designing Purposes :

In designing a survey, it is necessary to focus on a limited number of statistics among the objectives. In some manner, a few typical and important statistics will be chosen as key items for design computations. For the evaluation of health of school children, we may consider certain variables such as the height and weight,

school loss days, diagnosed chronic conditions, and nutrition status.

For optimization of sample size, some types of statistics like aggregates, mean, ratio, percent, and trends should be decided. The variance of the key statistics is needed in optimization model discussed below.

#### 2. 8 Sample Size

We may use a function with variance, fixed cost, and other sample constraints. Optimization of this function with respect to parameters ( $n_1$ ,  $n_2$ , and  $p$ , say) would give optimum number of schools and classes in the school (Hansen, Hurwitz, and Madow, 1953, Vol. 2, p. 199).

#### 2. 9 Sample Variance of Statistics

A smaller sample gives a much greater sample variance than a large sample. The sample variance must be flexible in order to accommodate sampling budgets because a larger sample gives a smaller variance, but costs more than a small sample does. In a large sample survey, we often take a 1 or 10 percent sample depending on the type of variables that we want to analyze. Sample variance is often the unbiased and consistent estimator of population variance.

#### 2.10 Budget

We could estimate overall costs of manpower, equipments, and other resources for collection of a sample.

A. Plan and protect

B. Data Collection ... Overhead costs of office, tool, computer, and other common costs. Cost related to the sampling, travel, interview or examination of children's health status ; this cost may be divided into Unit Cost per School and Unit Cost per Class derived from the individual children.

C. Redesign cost for every five or so years if



it is a continuing surveys.

D. Data processing and coding

E. Analysis of data

F. Publication

While total available budget is often uncertain, a firm's figure of overhead and fixed costs and individual sampling costs should be decided as early as possible. This figure had a significant impact on sample size, but the theory would remain the same even if the actual budget had shown a little different figure.

### 2.11 Time Table

Detailed schedule of a survey is essential for the conclusion of a successful sampling.

## III. SURVEY DESIGN

We define the population, its stratification, three stage master sample of provinces, schools, and classes, and along with it, outlined a general scheme of data collection including the variables of interest.

### 3. 1 Population

The population is defined to be the entire primary school children in South Korea. We take a sample from this population, and take responses from this sample.

### 3. 2 Stratification of Population

All primary schools in Korea are stratified into 5 primary strata according to the size of school within each of 11 administrative areas—8 provinces and 3 special cities. Now we consider only area, and if necessary, we can apply the same principle to each of the 11 administrative areas.

Assuming 60 students in a class, we use one first stratum selected from the 11 strata in Table 1 in order to illustrate the second stage selection, using artificial numbers of classes.

Above artificial example will be used to illustrate the sampling. If a sample of a 100 classes is required for the sample, and if there are 1,670 classes in the entire primary schools in the selected province.

The selection of these sample children follows a two stage probability rules described below, selection of schools and of classes.

The selection are the subject of sampling in two stages. we sample with probability proportional to size for the second time in sampling of schools, and equal probability with replacement in the third stage sampling of classes as illustrated in Table 3.

We suppose that the information on the number and size of all the primary schools in Korea is available for public use at the ministry of education, and that we make the listing of Master School Inventory.

The framework of the chosen sample structure is three stage stratified design in which provinces and schools are the primary and secondary strata, and the classes are the tertiary clusters.

In the first and second stage drawing of provinces and schools, large ones should have a higher probability of being in the sample and small ones, a relatively low probability.

Third stage sampling is done with equal probability in selecting classes, and we assume that this is done with replacement.

A simple random sample of classes may be taken if class sizes are about the same of 60 children. The classes are the ultimate sampling units.

All the children in the sample classes are the subject of interview or measurement. The estimate of population mean from the sample and its variance are given in Appendix A.

### 3. 3 Births or deaths of schools or classes

Creation or dissolution of a school or class in-

cluded in the original sample frame should be carefully handled. If schools are newly created, they should be considered in the sampling process separately, especially in any continuing sample survey.

#### IV. PANELS

We may divide a master sample of schools into four panels, and each panel would be a probability sample of the Korean school children. If the size of schools varies greatly, each panel could include approximately the same number of four different sizes of schools.

A panel could be used differently. One panel may be used in one time period or pretest. If one time period is a quarter, then the four panels would cover the entire year. If it is not known whether or not a sampler has enough fund, he could make more than one panels, and drop some panels to adjust sampling size according to the available fund.

#### V. PRETEST WITH ONE PANEL

Prior to making a final determination as to the exact sample size and its composition, the decision was made to survey a small number of schools as pretest. There are a few good reasons for doing this. No optimization would be accurate with looking at the actual data.

A pilot study must be done to get an overall picture on the sampling design specified already. It often prevents major mistakes which we can not correct later.

In order to gain necessary field experiences, and to collect more precise data to evaluate the contemplated design, we need a pretest with a small number of school, or one panel, and use this results in deciding on the number of schools and classes in the final sample. This panel is to become an integral part of the final design after

its use as a pilot survey.

#### VI. ESTIMATION AND SAMPLING VARIANCE

The population parameters—total, mean, or proportion—are estimated by three principal components : Inflation by the reciprocals of the probabilities of selection ; two levels of non-response adjustments ; and two levels of ratio adjustment to the known totals. Nonresponse adjustment may be carried out within school.

A sample always induce the sampling variance. When weighted, sample variance is often reduced but it is much more difficult to find the variance of an estimate based on the weighted data.

#### VII. QUALITY CONTROL OF THE DATA

Significant resources are often allocated to monitor the quality of the abstracts received from the reporting schools and subsequent coding and processing of the data.

A five percent systematic sample of the reports from each school can be used to check the validity of the coding.

If the reports are rejected if the error rate exceeds 15 errors per 100 codes. For individual scooter control, a scooter is retrained if his coding error rate exceeds by two standard deviation the average error rate obtained from a set of 50 batches.

#### VIII. REPORTING SYSTEM

The responses from sample schools should be periodically reported to a central office. Personnel from the central office need to visit the field operation for surveillance and control of the system.