a selective review of evaluative studies

SCHOOL HEALTH SERVICES

MCH Collection
Document Number 290

Provided by the Maternal and Child Health Library, Georgetown University
SCHOOL HEALTH SERVICES

a selective review of evaluative studies

BRONSON PRICE

Analytical Statistician
Division of Research

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
Social Security Administration • Children's Bureau • 1957

Provided by the Maternal and Child Health Library, Georgetown University
A large number of studies that bear on the evaluation of health services in schools have been made over the past century. Scattered through the scientific literature, they have seldom been assembled and compared. In order to summarize their findings, determine the adequacy of the statistical methods used, and consider their implications for future studies, the Children's Bureau undertook a critical review of this material.

"School Health Services—A Selective Review of Evaluative Studies" is the result. This monograph is one of a series on problems of evaluation that the Division of Research of the Bureau is conducting. It is the work of Dr. Bronson Price, analytical statistician, who examined over 1,000 references before selecting the material for this review.

At present, there is a great deal of interest in evaluative studies of public health services, including school health services. We hope that this monograph, by analyzing a large number of such studies, will be of help to research workers in designing future investigations in this important field.

KATHERINE B. OETTINGER
Chief, Children's Bureau.
This page is blank in the original document.
# CONTENTS

## INTRODUCTION

### 1. STATISTICAL RATES AS CRITERIA
- Mortality rates
- School illness data
- Studies of illness trends
- Accident rates
- Selective Service findings
- Studies using Selective Service data
- Correction rates
- Summary

### 2. SURVEY FINDINGS
- Official surveys
- The Academy study
- Nursing services
- Dental services
- 'Profile' surveys
- Cost data
- 'Background' factors

### 3. THE USE OF EXPERT JUDGMENT
- Evaluations of State programs
- Evaluations of city programs
- Comment

### 4. SAMPLING AND RE-EXAMINING THE CHILDREN
- Major studies
- Adaptations for routine use
INTRODUCTION

THERE IS BROAD AGREEMENT that the aim of school health services is the protection and improvement of children's health. There is also agreement, at least in principle, that evaluation should show how well this purpose is being achieved.

There are, however, many different views about best ways of improving children's health. These differences of opinion have inevitably affected the content of school health work, and indeed they account for most of the existing variations in the programs.

The variations in school health services are most marked in respect to: (a) the distribution of responsibilities between local education and health authorities; (b) the amount of responsibility assumed for children under private physicians' care; (c) the extent of treatment provided, especially for indigent cases; and (d) the quality and frequency of "periodic" or regular examinations arranged for the children.

These variations have led to a great many kinds of programs and procedures. Their diversity is a likely source of uncertainty about evaluative methods, if only because the evaluator must wonder whether any methods could be generally applicable to such a wide range of activities.

Another probable source of uncertainty about evaluative methods is the fact that, from the start of an evaluative study, we ordinarily hope to do more than learn how well a program's purpose is being fulfilled. We also hope to be able, and we are usually expected to be able, to specify what changes in procedures might be desirable. This requires that we attempt to evaluate not only the program's achievement but also the effectiveness of each procedure or component activity used in the program.
This may be attempting too much, at least by evaluative methods now available. That is, we may be trying to do more than should be expected from an ordinary evaluative study, considering the paucity of baseline facts so far established.

Perhaps we should expect objective methods to give a firm answer regarding only one aspect of evaluation—namely, how well programs are actually protecting and improving the children's health. Re-examination of a sample of the children should yield a good answer to that question for almost any type of program. The information obtained by this evaluative method would always help a little, and sometimes it might help a lot, with the problem of advising on procedural changes.

Then, for the rest of an ordinary evaluative study, we might leave answers to questions about the program's component procedures to the opinions of panels of the best experts available. They would be asked to make group judgments which, though outrightly subjective, would be developed in an organized way.

At the same time, we would of course hope and expect to be discovering more satisfying answers to procedural questions through specially designed experimental studies. Indeed, we could well stress to the public and to administrative authorities that no type of evaluative effort should have higher priority than controlled comparisons of alternative procedures for conducting the case finding and follow-up work.

The foregoing statements are offered as "candidate" viewpoints, or hypotheses that seem worth weighing as we review the evaluative studies done so far.

**Scope of review**

The studies to be covered are grouped in five main sections, depending on whether the work chiefly concerns: (1) statistical rates as criteria of effectiveness; (2) survey findings; (3) the use of expert judgment; (4) sampling and re-examining of the children concerned; or (5) experimental approaches.

Except where noted otherwise, the discussion will concern health services in elementary schools, both because the secondary grades receive a relatively small part of the total effort invested in school health, and because practically all of the evaluative studies reported to date have dealt with the problems of elementary schools.

Likewise, except where specifically noted to the contrary, the term "health services" will be used to mean medical, nursing, and dental services provided to individual children in or through the schools, and will not include health instruction, physical education, or the inspection of school premises.

This review attempts to be selective rather than exhaustive.
For comprehensive reviews of the earlier literature on school health, the reader should see the works of Kerr (1926) and Wood and Rowell (1927). No equally comprehensive review of more recent literature is available, but the lack of such a work is not necessarily serious. For, despite the great volume of material that has been published in the past 30 years, the changes during that period in school health theory and practice have been small compared to the changes that occurred in the preceding decades. There would seem to be a need for reviews that are selective and constructively critical, and the writer has attempted to supply one such review. He has included suggestions and critical comments where they seemed appropriate, hoping that they will have stimulus value to other writers and that, after being criticized and weighed in their turn, some of the comments may be useful in connection with future studies.

There is reason to believe that the field of school health has been handicapped, and that much needless confusion has occurred, through investigators' frequent failure to use simple association tables and correlation coefficients. The reviewer has been told that this situation has arisen mainly because physicians object to the use of $2 \times 2$ tables and the ordinary ways of generalizing from them. If so, it is strange that such tables have long been used in medical literature, and that the necessity for continuing their use is stressed in the physicians' own journals (see, for example, the lead editorial in *J. Am. Med. Assn.* 143:1260–61, August 5, 1950). In any event the reviewer has not hesitated to re-cast the findings of school health studies into association tables, and has also computed the correlations whenever they could help to clarify the findings.
STATISTICAL RATES AS CRITERIA

NO ONE SHOULD EXPECT such statistical indexes as mortality and illness rates to give more than a partial picture of children's health status. Nevertheless, such data have the merit of furnishing evidence which, as far as it goes, is objective. Accordingly, these types of data are often studied for their possible relevance to school health evaluation. Rapeer (1913) and Keene (1929) were among those who stimulated these efforts by reviewing certain early data and by calling for more intensive studies. Have the hopes held for statistical criteria of effectiveness been justified?

Let us consider in turn the data on mortality, illnesses, accidents, Selective Service findings, and "correction" rates, bringing out the more pertinent information actually available from each type of statistics before examining limitations.

Mortality rates

In 1953, among each 100 deaths of children aged 5-14 years, accidents caused 40, cancer 12, influenza and pneumonia 6, rheumatic fever and heart diseases 5, congenital malformations 4, nephritis 3, and poliomyelitis 3.

The remaining 27 deaths were due to a large number of conditions which, taken singly, had only small effects on the mortality rate as a whole. A striking example is the fact that, on the average, only 1 of the 27 deaths was due to some disease in the group formerly called the "main contagious diseases of childhood." Those diseases were diphtheria, smallpox, measles, scarlet fever, and whooping cough. It was largely through the successful attacks on these and other infectious diseases that the
overall mortality rate of school-age children was cut by 88 percent from 1900 to 1953 (or from 3.9 to 0.5 per 1,000 children aged 5–14 years).

It is reasonable to credit some part of this reduction to school health programs, if only because the school, through its contacts with parents, has served as a vantage point for aiding immunization programs and other health work with infants and preschool children. However, the amount of credit that should go to school health programs is quite indeterminate. Aside from immunization programs, preschool children have been exposed to health programs that were considerably less comprehensive than the programs for school-age children. Yet the mortality rate of the preschool group has dropped even more than the rate for the school-age group, namely 93 percent from 1900 to 1953 (or from 19.8 to 1.3 per 1,000 children aged 1–4 years).

In this light it would seem difficult to make direct use of percentage changes in school-age mortality rates for purposes of evaluating school health programs. It appears safer to examine the rates to see where school health work may have failed. Evaluative effort of this kind is probably the more important part of the discussions of reductions in childhood mortality rates published by Wheatley (1941 and 1947), Smith (1948), Maxwell and Brown (1948), and Maxwell (1950).

The last two of those reports provided convenient presentations of the more pertinent and recent school-age mortality rates, as well as discussions of possible relations between those rates and certain defects actually found and treated in the school health program of New York State. The relationships discussed were not very marked or convincing. It is nevertheless noteworthy that the 1948 report by Maxwell and Brown was almost unique in that it related the findings of a school health program in a particular area to the childhood death rates for the same area.

Indeed, evaluations of particular school health programs have rarely included the mortality rates for the children concerned. Even if there were no reason, in the past, to question the value of using childhood death rates, their value for the present and future is doubtful because too few childhood deaths occur in an ordinary school system to give its mortality rate much statistical meaning. During recent years the number of deaths among children aged 5–14 have totaled less than 16,000 annually in the country as a whole. This means that only the very largest school systems have had more than 10 fatalities per year, and by no means all childhood deaths are from causes which, at least as yet, can be considered preventable.

Finally, although it may be valid to compare death rates
for a particular State or large city at different periods of time, it is less sound to compare the death rates for different areas. At least as regards childhood rates, the best evidence available on this question was obtained by Clark and Burdick (1952). They compared the death rates of children under age 15 with numerous other indices of children's health status in 14 northeastern States. They found, for example, that Pennsylvania's childhood mortality rate was one of the highest (worst) in the whole group of States, while Wisconsin's rate was among the lowest (best) rates. Yet Pennsylvania was above average and Wisconsin was below average in several of the other indices studied, including children per pediatrician, children under care in clinics, and hospital beds for children.

All such variables correlated in the expected directions with the childhood death rates, but the coefficients were not high enough to suggest that the death rates were valid indications of the relative amounts of care received by different child populations. Although not restricted to the childhood age range, correlations consistent with this generalization were reported for the 48 States by Hirschfeld and Strow (1946).

**School illness data**

School absences are our main source of information on morbidity in the school-age range. The most useful index obtainable from school illness data is the "absence rate," representing the proportion of school days lost due to illness. Another index, termed the "case rate," represents the frequency of separate instances of illness. The case rate is less important than the absence rate but is essential if one wishes to compute the average duration of absences caused by a given illness.

To obtain these rates it is first necessary to distinguish between the absences which are, and those which are not, due to illness. Collins (1925) has shown that teachers can do this with substantial validity. The absences which are not due to illness need no discussion here except for remarking that they usually account for a relatively small amount of absence, compared with illness, in schools that are well administered and well supported by the community.

For the absences attributable to illness, the number of days lost and the number of cases (instances) of absence are usually counted throughout a school year. When each of these numbers is divided by the number of children enrolled, one obtains an absence rate and a case rate representing the average child's experience per school year. However, since the school year is not 180 days for all school systems or for all years, it is desirable to
reduce the rates to a basis of 100 days. This has the advantage of stating the absence rate as a simple percentage of school days, and it puts the case rate on a uniform basis for all school years.

Probably the most representative illness rates available for schools in this country are the figures obtained in a study which the Metropolitan Life Insurance Co. directed and reported (1950). The study yielded the rates shown below for 7,700 elementary school children in seven cities of California. The records of the children's absences were kept from January to June, 1947.

SCHOOL ILLNESS DATA FOR 7 CITIES IN CALIFORNIA,
Metropolitan Life Insurance Co., 1947

<table>
<thead>
<tr>
<th></th>
<th>Absence rate (days lost per 100 school days)</th>
<th>Case rate (instances of absence per 100 school days)</th>
<th>Average duration (days lost per instance of illness)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory diseases</td>
<td>3.52</td>
<td>1.39</td>
<td>2.5</td>
</tr>
<tr>
<td>Communicable diseases</td>
<td>.98</td>
<td>.11</td>
<td>8.9</td>
</tr>
<tr>
<td>Digestive disorders</td>
<td>.46</td>
<td>.32</td>
<td>1.4</td>
</tr>
<tr>
<td>Skin disorders</td>
<td>.38</td>
<td>.08</td>
<td>4.8</td>
</tr>
<tr>
<td>Accidents (including injuries)</td>
<td>.22</td>
<td>.08</td>
<td>2.8</td>
</tr>
<tr>
<td>Other conditions</td>
<td>1.06</td>
<td>.46</td>
<td>2.3</td>
</tr>
<tr>
<td>All illnesses</td>
<td>6.62% of school days</td>
<td>2.44 cases per 100 school days</td>
<td>2.7 days per case</td>
</tr>
</tbody>
</table>

Mason (1953) has obtained a more recent set of data on illnesses of elementary school children in California. The classification of school illnesses which he used is probably the most detailed and widely applicable scheme so far developed. However, Mason's findings are of limited value because they are based solely on one area, Redwood City. It therefore seems appropriate here to cite only the absence rate which he obtained for all illnesses, together with the main components of that rate.

The total absence rate in Mason's study was 5.12 percent of the school days. Using the same categories and sequence as in the table above, the component parts of that total rate were, respectively, 3.32, .58, .38, .07, .27, and .50 for the respiratory, communicable, digestive, skin, accidental, and other condi-
tions. In general, these rates correspond about as well with those in the table above as one could expect, considering the different classifications of illnesses used in the two studies.

Discussion

Both studies show that respiratory diseases are, by wide margins, the leading cause of school absence for illness. Since the duration of cases of respiratory diseases is not especially long, the very high absence rate for them is mainly due to the great frequency of such illnesses.

Although communicable diseases comprise the next ranking category, they are, fortunately, a poor second. They occasion long absences when they do occur. But except for that fact, communicable diseases would cause a relatively insignificant amount of absence. To a large extent the durations of these illnesses reflect protective regulations rather than acute phases of the diseases concerned.

The third ranking category comprises the digestive disorders. In contrast with communicable diseases, the digestive disorders would occasion a high absence rate except for the fact that their durations tend to be short.

The greatest disparity in the two sets of rates concerns absences for skin disorders, for which .38 and .07 percent were obtained in the separate studies. At least in part, the low rate of .07 percent might reflect some special success of Redwood City's efforts against the "nuisance diseases" of school-age children. Partly, too, the difference could be due to the fact that the higher rate was obtained in 1947 and the lower rate in 1952. For, during that interval, new treatments for children's skin troubles were being used with increasing success. Finally, the rates could have been affected by the fact that the studies used different classifications of illnesses. For example, a significant cause of school absence in most areas is ivy poisoning. We know that Mason always grouped such cases with accidents rather than with skin disorders, but we do not know how regularly this may have been done in the larger study.

Regarding absence rates for accidents, which generally include injuries, the two studies obtained the values .22 and .27 percent. These rates are probably as similar as could be expected considering possible classification differences; e.g., we can be sure only in Mason's study that sunburn, like ivy poisoning, was always classed with accidents.

It is worth stressing that the relatively low ranking of accidents in both these studies in no way belies the seriousness of accidents as a cause of death in school children. Rather, the contrast between the low absence rate and the relatively high
mortality rate for accidents reflects the fact that, on the average, situations in which children are exposed to diseases are much less dangerous than situations in which children are exposed to accidents.

In view of the seriousness of accidents and the increasing importance of school safety programs, we may well ask whether the routine collection of data on absences due to accidents is useful for evaluating those aspects of school health programs having to do with safety. This problem will be examined after consideration of studies regarding trends in the larger causes of school absence.

Studies of illness trends

There have been two outstanding reports on trends in school illness rates. Each report was a "retake" in a school system where conditions had been carefully studied nearly two decades previously.

The study by Linde and others (1950) was a re-survey of the illnesses found among New Haven school children, employing the same methods as those used 19 years earlier in the same area by Wilson and his associates (1931). Since both surveys were limited to absences of three or more days duration, the figures obtained for the absence rates are not comparable with the rates found in other studies and need not be cited here. So far as changes in rates are concerned, however, the trend found for a rate based on three or more days of absence ought to give a fair indication of the trend in the ordinary absence rate for the same cause, provided that the given cause is a major component of the absence rate.

In any event, over the 19-year interval the two surveys of New Haven children showed a 35 percent increase in the absence rate for respiratory diseases, while there was a 25 percent decrease in the absence rate for communicable diseases. The absence rate for all illnesses remained practically unchanged (1 percent increase). Although the overall case rate had increased substantially, its effect on the absence rate was cancelled by a decrease in the average duration of absences.

The other important trend study was the report by Ciocco, Cameron, and Bell (1941) summarizing a series of school illness surveys in Hagerstown, Md. The absence rate was based on all days of illness (rather than on three or more days), but it was for both elementary and high school children and was therefore somewhat lower than the ordinary rate based on elementary school children. Again, however, we are interested in the trends rather than absolute magnitudes of the rates.

Comparison of the absence rate for both elementary and
high school children in 1939-40 with the same rate in 1922-23 (reported by Collins, 1924, page 2418) showed a 14 percent increase in the rate during the 17-year interval. There had been a moderate decrease in the average duration per case, but this was more than counterbalanced by a rise of 35 percent in the case rate.

The authors went on to show that the increase in the case rate—and presumably the increase in the absence rate also—was due almost entirely to increased absences for colds and digestive disorders. The investigators nevertheless found reason to doubt that the true prevalence of these conditions had changed. The increases in their rates seemed attributable, instead, to "greater care or precautions taken now by parents (due to) the health propaganda by private and public agencies regarding the need for early treatment of minor disorders." If this were true it would of course raise some question about the value of the propaganda mentioned, but this possibility was not noted.

Discussion

It is clear from these reports that childhood illness rates as a whole have not changed to anything like the extent that children's mortality rates have changed with advances in public health. For such outstanding diseases as diphtheria and smallpox, the trend of the illness rates parallels the trend of the mortality rates, and both trends reflect the success of immunization programs. But these successes are largely offset, in the illness data, by higher absence rates for other causes. The extent to which this may be due to real increases in frequency of the other diseases, or perhaps to changes of the kind mentioned by Ciocco's group, is not known.

Aside from questions regarding trends or changes in the practices of parents, data on illnesses have long been considered ambiguous for purposes of school health evaluation, simply because a poor program might fail to send children with infectious conditions home as often, or to keep them there as long, as an optimum program should. For example, it was found in the 1945 survey by the New York State Education Department that the rate for all causes of absence was about the same in schools with better and poorer programs, but the rates for illness alone were higher where programs were better.

In evaluative studies, then, illness statistics might "work in reverse" to some extent, either for purposes of comparing the results of one program at different times, or for comparing different programs at a given time.

It should be added that although the possibility of this reverse effect is a serious deficiency of illness data so far as
evaluative work is concerned, the use of illness data for predicting certain care needs is not thereby invalidated. That is to say, even though the illness rate for a whole group of children may involve some bias, this fact has only a slight bearing on the value of an individual child's past illness record as an index of his need for future check-ups.

The relation between absence rates and findings in medical examinations is evidently quite low (see data of Collins, 1922; and Downes, 1930), but that does not necessarily cast doubt on the value of watching the illness records of individual children. Downes (1945) found that the children who would most need health supervision in the coming year could be predicted fairly well by identifying the children whose records for the past year showed they were absent at least twice from conditions other than communicable diseases, skin infections, and tonsillectomies.

**Accident rates**

Since data on accidents are sometimes used for evaluating the aspects of school health programs that have to do with safety, it is important to inquire whether accident rates are subject to less difficulty than other school illness rates. According to the previously cited findings of Mason and the Metropolitan Life Insurance Co., the absence rate for accidents is approximately .25 percent. Although this rate seems low, it is high enough so that, for example, more than 400 school days would be lost due to accidents during one school year in a school of about 1,000 pupils. From a purely statistical viewpoint, therefore, accident rates would seem to be practicable for evaluative purposes, at least in sizeable school districts.

However, the possibility of serious trouble in schools' reporting of accidents is evident if we compare the figures in certain series of regularly published rates with the rate of about .25 percent obtained in special surveys.

One series of figures dates from 1928, when a group of schools interested in reducing childhood accidents began reporting to the National Safety Council on student injuries “requiring a doctor's attention or causing absence from school of one-half day or more.” In recent years these reports (see National Safety Council, 1955) have covered about five percent of the country's school children. For the reporting elementary schools, the data would indicate that the proportion of school days lost due to accidents is .04 percent. This is so much lower than the rate of .25 percent found in special surveys that the difference cannot reasonably be attributed to the Council's somewhat restricted definition of accidents.

8
The other important series of accident rates concerns school children in Kansas, where for over a decade the State Board of Health has asked all schools to make monthly reports on accidents. Currently the participating schools make up about 40 percent of the State's total enrollment. For reporting purposes, accidents are defined as "injuries requiring medical attention or resulting in one-half or more days absence from school." This is practically identical with the definition used by the National Safety Council. However, the data reported from the Kansas elementary schools indicate an absence rate of only .02 percent, or half the figure reported to the Council. (See Hood, 1956.)

It is possible that the voluntary and unofficial nature of the Council's reporting program tends to attract a relatively high proportion of schools with special interest in safety work, and this factor may have enough effect on the completeness of reporting to account for the Council's higher rate.

Be that as it may, an even more serious consideration is the fact that much higher figures are reported where accident rates are obtained in special surveys than where they are collected routinely. Rates that are useful for evaluative purposes ought to be rates that are obtainable routinely. Until someone develops a procedure for classifying and reporting accidents that works almost as well under routine conditions as in special surveys, it would not appear sound to attempt to use accident rates for evaluating school safety programs.

Selective Service findings

Despite their imperfections, the Selective Service findings provide the best available inventory of the health status of our youth, and they deserve the relatively large amount of attention given them in connection with school health evaluation.

Unfortunately, the statistical limitations of the Selective Service data have led to confusion, and at times to overstatement, regarding the seriousness of the findings. Before reviewing the studies in which the Selective Service data have been utilized for school health evaluation, let us see what the findings actually were, particularly in the younger registrants.

For the very reason that the data on younger men are of more interest in connection with child health programs than the data on registrants of all ages, Selective Service officials issued, in 1943, a special tabulation of the findings for a sample of 45,000 men aged 18-19 years.¹ The results need to be given here in only

¹ In the comprehensive report which Selective Service issued in 1947, data are given for an additional sample of 170,000 men aged 18-20. The data from
approximate terms, since the detailed breakdowns are readily available in the source article by Rowntree and others (J. Am. Med. Assn., Sept. 25, 1943; see also the convenient summary by Goldstein, 1951, for data on registrants of all ages).

Among each 100 registrants aged 18-19 years, 75 were classified as I-A and inducted for full military service. Approximately 50 of these men had no defect, while 25 had some defect not serious enough to affect their I-A status. Another 4 men out of the 100 were classified as I-B and accepted for limited service only. The remaining 21 men were IV-F, or disqualified for any military service.

As is customary, the Selective Service report grouped the I-B and IV-F men together to make up the "rejection rate," which was thus 4 plus 21, or 25 out of each 100 registrants aged 18-19. What was the distribution of these 25 men with respect to their principal defects or causes of rejection? The tabulation showed:

- 4 had eye conditions
- 3 had musculoskeletal defects (including flat feet)
- 3 had mental disorders
- 3 were illiterate or dull
- 2 had cardiovascular defects
- 2 had hernias
- 2 had ear conditions
- 1 had a neurologic defect
- 1 had a venereal disease
- 4 had miscellaneous defects

25 percent rejected for full military service

In the above figures only one defect, i.e., the "primary" one causing rejection, was counted for each man. However, the Selective Service report also gave the frequencies of all the defects found, regardless of their severity. This information deserves more detailed attention than the data on rejections because it is relatively comprehensive in nature and is less oriented toward specific military needs.

that sample were used here to estimate the proportion of I-A men with defects, and also the proportion of rejectees in the I-B and IV-F categories. Otherwise, the data of the 1947 report have not been used because its breakdown of findings on the younger men was not as detailed as the breakdown given in the 1943 report, and there is no reason to suppose that the sample reported in 1947 was more representative than the sample reported in 1943, so far as the younger men were concerned. However, the two sets of findings have been compared, and the data of the 1943 report were not used here until it was verified that the differences between the two samples were small and attributable largely to the inclusion of men aged 20 in the 1947 report.
In tabulating the frequencies of defects, those I-A men who had defects were grouped together with the I-B and IV-F cases, and each man was counted as many times as he had a defect. When the defects were tabulated in this way and related to the total number of registrants, an average of approximately .7 defects per man was found. More precisely, there were 69.4 defects per 100 men, although it should be recalled (see above) that the defects making up this rate were concentrated in approximately half the men, while the other half were free of defects.

Shown below are the prevalence rates for the main groups of defects, or those which occurred with a rate of at least 1.0 or more per 100 men. Each figure at the right gives the rate for one of the specific defects that is included in the figure for the broader group. Both sets of figures represent the number of defects of a given kind found per 100 men, regardless of the presence or absence of defects other than those named.

Teeth, mouth and gums........11.0 (9.1 due to caries or its results)
Musculoskeletal (incl. feet).....10.9 (4.5 flat feet)
Eye conditions..................10.2 (7.1 vision defects)
Under- and overweight..........5.5 (2.6 underweight)
Illiteracy and dulness.........3.7 (2.8 illiteracy)
Genitalia.........................3.4 (1.8 varicocele)
Mental disorders................3.3 (2.0 psychoneurotic)
Cardiovascular..................3.2 (0.5 functional murmurs)
Hernia and relaxed rings.......2.5 (2.0 hernias)
Ear conditions....................2.5 (0.4 hearing defects)
Nose defects......................2.1 (1.3 nasal obstructions)
Neurologic conditions..........1.8 (0.3 epilepsy)
Skin disorders....................1.6 (0.9 acne)
Venerial diseases.................1.5 (1.2 syphilis)
Total of prevalence rates for all other defects ..................6.2

Total ................................69.4 defects per 100 men

Regarding the findings as a whole, it is important to note that certain evaluative comments were offered by the Selective Service authorities themselves. They pointed out that, from a military viewpoint, the instructions given to the examining physicians had set “fairly high” standards. The result was that a substantial proportion of the men classified as IV-F, as well as most of those classified as I-B, “would be acceptable for military duty in any army in continental Europe” (Rowntree and others, 1942).

Noteworthy also was the viewpoint of the civilian physicians who conducted most of the examinations. It was probably well expressed by an official of the American Medical Association...
(Bauer, 1942), who stated that Selective Service accepted “not merely healthy young men, but the very best and healthiest.”

Studies using Selective Service data

Two approaches have been made to using the Selective Service findings in connection with school health evaluation. In one approach, which we might term “analytical,” the findings have been studied to distinguish defects which, presumably, could have been prevented or corrected in school health programs. The other approach has been that of “linking” the records on the health status of boys while in school to the Selective Service findings on the same individuals as young men.

Analytical studies

Typical of the analytical approach were the studies of Perrott (1941-47) and Britten and Perrott (1941a and b). These reports included a general comparison of the findings on World War I draftees with the findings on Selective Service registrants. Although the definitions and standards used in the two wars were so different that exact comparisons of the data for the two periods could not be made, the authors found marked similarity in the two sets of findings so far as “the important causes of rejection” were concerned.

The same authors pointed out that, among those more important causes, the most preventable or remediable conditions seemed to be defective vision, defective teeth, underweight, hernias, tuberculosis, and venereal diseases. Similar interpretations are found in the other analytic studies that have been made of the problem, and the ensuing discussion will therefore give chief attention to differences among the studies.

For each main group of defects in the Selective Service findings, a brief discussion of what was known or believed regarding preventive or curative measures was included in the analysis by Davis and Arena (1948).

The analysis by Mace (1944) was perhaps the most discriminating of those offered so far, since he drew the clearest distinctions between the conditions which a school health program might hope to prevent or correct, and the conditions for which there was no such prospect in the immediate future. For conditions of the latter type, Mace made the point that prevalence rates, whether based on Selective Service findings or other data, are not suitable criteria for judging the success of a school program. For, in respect to those conditions the program’s aim is not to reduce prevalence but to provide optimal adjustments in the school life and other activities of the children concerned.
It is worth remarking that this point, though not new, is increasingly receiving the recognition it deserves (see, for example, Fowler, 1954, on hearing defects and Lanciano, 1955, on eye defects).

The analysis by Schmidt (1945) was of special interest in two respects. His main conclusion was that the Selective Service findings reflected “medical needs rather than lack of physical training.” He arrived at this view without benefit of record-linking data, yet this interpretation of the Selective Service findings was practically the same as the conclusion drawn in an important record-linking study (Lyon’s) to be reviewed later.

The other point of interest in Schmidt’s discussion was his suggestion that, where the funds for school health work are limited, the program should not be spread thin over all pupils, nor should adequate services be limited to selected children. Instead, Schmidt said, the program should provide adequate services for all pupils needing them in a few schools, with extension of the program to other schools as funds permit. This suggestion will be recalled in the discussion at the end of this section regarding ways in which evaluative studies might utilize Selective Service data to better advantage.

It would be difficult to say that any of these analyses have established much about school health programs which was not already known. And it appears equally hard to say that better results have been obtained with the record-linking approach, at least in the ways it has been applied so far.

Record-linking studies

One of the record-linking studies was made by Greer (1948). He supplied a brief report on a study of graduates of 5 North Carolina orphanages, where all of the children had received regular pediatric care. The data are said to have “shown that 1,138 men and women who grew up in these institutions were accepted by the armed services, and only 16, or 1.4 percent were rejected.” Figures were not given separately for the boys, but they presumably comprised about half of the graduates, and even if we assume that all 16 rejections were among them, the rejection rate would still have the remarkably low value of 3 percent.

Greer noted that, to some extent, the orphanage graduates were a physically pre-selected group, since the orphans with severe handicaps had been sent to special hospitals. He felt, nevertheless, that the results “definitely” showed the favorable effects of the medical care given the children.

Since there was some pre-selection of the children, and since it is not clear that Greer started with full lists of the orphan-
age graduates of specific years and followed up all or nearly all of these graduates, one cannot well accept the published data as very convincing evidence of the effects of health care.

The important, if small-scale, study by Lyon (1945) made use of full lists of high school graduates. Lyon was the school superintendent in Norwich, N. Y., and he was also a member of the local Selective Service board. This enabled him to study the Selective Service records on all 353 boys who had graduated from the Norwich high school over a 6-year period. Among them, 42 had not been medically examined by Selective Service because of deferment or because the examining process had not yet reached them.

Of the 311 graduates who were examined, 27 had been rejected. Comparison of the rejectees' school health records with their Selective Service findings showed that "most" of the defects causing rejection had been discovered by the school, and that the school health staff had followed up the individual cases in ways that seemed satisfactory. However, the majority of the defects involved were "either hereditary or they resulted from such diseases as asthma, rheumatic fever, and infantile paralysis."

Lyon said the data showed that: (1) a more intensive physical training program would have had no appreciable effect on the Selective Service findings; but that (2) a few of the rejections might have been prevented by better school health services, especially in elementary school. This relatively well documented interpretation is practically the same as the one drawn in the analytic study by Schmidt (see above).

Of additional interest in Lyon's study was the disparity between the rejection rate of 9 percent (or 27/311) among the high school graduates, and the comparatively high rate of 30 percent among all the local registrants having the same ages as the high school graduates.

To explain this disparity, Lyon said that the health status of high school graduates was, in general, relatively superior, and that most rejections occur in that majority of the registrants who do not complete high school. He urged that Selective Service verify this generalization by tabulating rejections separately for the registrants who were high school graduates and those who were not.

Information on this question does not seem to be available in Selective Service reports or the literature on school health. However, Lyon's general point appears reasonable because, as he indicated, severe mental or physical handicaps almost certainly reduce an individual's likelihood of completing high school, and, in addition, boys who leave school early go without health supervision for a relatively long period, compared to high school gradu-
ates, before being examined by Selective Service.

The record-linking report by Ciocco, Klein and Palmer (1941) is well known for showing, as the authors stated, "that appreciable indications of Selective Service findings already exist in childhood." It is less well known that the data are instructive as to those Selective Service findings which are most and least predictable from school health records.

Starting with the names of men examined by Selective Service in Hagerstown, Md., the authors located the school health records, made about 15 years earlier, for 411 individuals out of some 1,400 registrants on the original list. During the period when the school records were made the Hagerstown schools had no health service program of consequence, but the children happened to have been subjects of studies in which they were examined by physicians with special training and experience in school medical surveys.

The 411 cases comprised 186 individuals who were later accepted by Selective Service, and 225 who were later rejected (including I-B men). Although the information available from the school records was frequently incomplete, it was lacking about equally often for the groups who were accepted and rejected later by Selective Service.

Eight kinds of defect were studied. For 4 of them (dental, visual, cardiovascular, and ear conditions) the presence or absence of the given defect as recorded in school could be related to the same defect's presence or absence in Selective Service findings. For the other 4 conditions (defects of posture, tonsils, nutrition, and weight), it was necessary to relate the defect's presence or absence, during the school period, to Selective Service acceptance or rejection regardless of cause.

Without using correlation coefficients, the authors brought out the fact that, for each type of defect, there was at least some association between the school findings and the Selective Service findings. From the presentation given in the report, however, one cannot easily judge the ranking of the defects in respect to the amount of association they showed with the Selective Service findings. Since the data readily permit the computation of 2x2 or "point" correlations, they are used here to indicate degrees of association, which mean, in this context, the extent to which the Selective Service findings are predictable from school health records. The data indicate that this predictability was substantial for heart conditions (correlation of .57); it was quite marked also for ear conditions (.48) and for visual defects (.36); but it was low (.22 or less) for dental, tonsillar, postural, nutritional and weight conditions.

The correlations have practically no bearing on the frequen-
cies of the defects, and, by themselves, they of course indicate nothing about the relative importance of the various defects for either school health programs or Selective Service.

Nor does the low association found as regards teeth raise doubt about the seriousness of caries according to the standards of Selective Service or any other criterion. In the case of dental defects the low correlation probably reflects the fact that caries’ effects are largely remediable, and indicates that, during the 15-year interval, some of the boys with high caries attack rates had obtained enough fillings, and had thus saved enough teeth, to pass Selective Service standards.

However, one might ask whether similar reasoning applies to the data on tonsillar conditions, i.e., whether the chief method used for treating such defects is effective in relation to Selective Service acceptance.

Unlike the study’s information on teeth, the information on tonsils did include data regarding treatment, i.e., tonsillectomies. Of all 306 boys for whom the schools had recorded the condition of the tonsils, 40 were reported as having normal tonsils; 59 as having their tonsils “removed,” and in the remaining 207 the tonsils were “diseased.” We may leave aside the 40 cases recorded as normal, and ask whether, in the others, removal of the tonsils was associated with Selective Service acceptance. It was, but only to the extent of a correlation of .08. This result is consistent with the generally negative findings by Kaiser (1932) as to the value of tonsillectomies.

As regards deficiencies in posture, nutritional status and weight, such experimental evidence as is available (Clement and others, 1950; Kaiser and others, 1926; and Schwartz and others, 1928) indicates that these conditions are not substantially changed by the treatment usually used for them. Thus the low associations with Selective Service acceptance or rejection would not seem readily attributable to treatment. Instead, the low associations are probably due in some part to the low relationships between bodily habitus in childhood and bodily habitus in adulthood, and partly also to qualitative differences between the physical features considered in school health examinations and those considered by Selective Service examiners.

We may conclude that the available studies using the record-linking approach have produced results of general interest in connection with school health, but have not helped much in evaluating effectiveness.

Possible future studies

In all the evaluative work done with the Selective Service findings, the main merit of these findings is simply the fact that
they provide an independent check on the health status of the groups studied. Recognizing that this merit is genuine, one must wonder whether there are not other checks that have the same virtue of independence, while being no less valid for the purpose—and perhaps more readily applied in most situations—than the Selective Service findings. The answer appears to be that another check exists, and that even though it is not as yet commonly employed, its frequent use should be feasible in the future. This topic is considered in Section 4 herein.

What is missing from all the studies that have used the Selective Service findings is a clear-cut comparison of: (1) a group exposed, during school age, to one or more health service procedures believed to improve health status; and (2) an initially similar group who were exposed to different procedures or to no procedures of a systematic kind.

It is possible that this type of study could still be made. Two such differently exposed groups might be found in some area where there was substantial evidence that the groups were initially similar. The criterion of effectiveness would be the results of either Selective Service examinations or other specially conducted examinations. It would not be essential to link individual findings from the criterion examinations to the school findings, except to the extent of making reasonably sure that each individual appearing in the criterion examinations was also a member of the original school group.

It would be much better, of course, to plan a long-term study in which the similarity of the groups could be insured in advance. This could be done best along the lines of Schmidt's suggestion (see above). That is, with the adoption of each main part of the health service, it would be instituted in selected schools, while one or more comparable schools would not receive it, at least until there was expansion of the program as a whole. Possibilities of this nature are discussed in Section 5.

Correction rates

The purpose of a "correction rate" is to show, for a given school health program, what proportion of the children's need for medical attention the program is meeting. The denominator of the rate usually represents defects known to exist at a particular time, and the numerator represents the defects that were corrected or placed under care during a subsequent period.

Sometimes the denominator is the number of children having defects, while the numerator is the number of children whose defects were corrected. This is probably the best form of the rate, but is seldom used for lack of a generally accepted way of reckon-
ing with the children who have more than one defect. That ques-
tion is commonly avoided by expressing the rate in terms of
defects, regardless of the fact that two or more defects may
occur in the same child.

Example of contrasting rates

Compared with certain other problems, the question of
whether the rate is in terms of defects or children with defects
is a small matter. Further difficulties in correction rates, as well
as the range of values typically found, may be illustrated by com-
paring the findings in two recent studies, both of which happen
to have been conducted in Pennsylvania. One of the two sets of
findings is from the report of Mather and others (1955). Their
study was an important experimental test of certain follow-up
procedures, and will therefore be discussed in detail in Section 5.
Here we need only note that the study began with a sample of
children found to have medical defects in Pennsylvania's regular
school health examinations; that certain routine and special pro-
cedures were used to get the defects corrected; and that over a
period of scarcely 3 months the routine procedures yielded a cor-
rection rate of 46 percent. In this rate the denominator was the
number of children with defects, while the numerator was the
number of children whose parents said, when interviewed, that
they had seen a physician or had at least got in touch with one
regarding the child's defects. Thus the measure of corrective
action was essentially the proportion of cases in which parent-
physician contacts had been made.

The other set of data was specially assembled by Philadel-
phia's school health staff for publication in the report of the Penn-
The figures are probably the most complete and accurate -data on
correction rates so far collected. Yet it is noteworthy that, in
publishing the Philadelphia data, the Commission indicated that
the figures were of considerable interest and value, but did not
suggest that these or any other correction rates should be regarded
as models for general use.

The data were for 75,000 remediable medical defects found
at the start of the 1951-52 school year. (The number of children
having these defects was not mentioned, but was presumably
about 60,000.) Of the 75,000 defects, 13 percent were found to
have been treated by the close of the 1951-52 school year, while
17 percent were treated during the next year, and 12 percent
were treated during the third year. Thus, over 3 years a total
of 42 percent of the defects were known to have received correc-
tive action, as measured by private physicians' reports or school
physicians' re-inspections of the children concerned.
For 14 percent of the defects, nothing could be learned because the families of the children had left the area and could not readily be traced. For 4 percent, the family physicians disagreed with the school's diagnoses or did not think medical care was needed, and another 4 percent were neither treated nor found to be present in subsequent examinations by the school physicians. Finally, it was found that 36 percent of the defects still existed and had not received treatment in the 3-year period.

If the first of these two studies had run for a school year, rather than for 3 months, before the interviewers had asked the parents about contacts with physicians, it is obvious that a correction rate considerably higher than 46 percent would have been reported—perhaps 70 percent would be a fair estimate.

If the second study had run for one school year and then stopped, a correction rate of about 13 percent would have been reported, according to the findings shown above.

Since the two studies differed in more than one way we cannot say that the disparity between 70 percent and 13 percent is due entirely to the use of different criteria of corrective action, but we may suspect that the largest part of the difference was due to that factor.

This inference does not, by itself, raise doubt about the appropriateness of the rate used in either study. The rate in terms of parent-physician contacts used in the first study had the virtue of being comparatively simple, and this rate may be considered a reasonably valid measure of effectiveness for school health programs which aim to see that children needing care are brought to physicians' attention. The rate used in the second study is relatively difficult to obtain because it usually requires some re-inspecting of the children, and yet it is clearly the requisite rate for programs aiming to see that care is received by as many as possible of the children who need it.

As Rapeer (1913) and Buck (1923) brought out in their early discussions of the general subject, the difficulties in developing correction rates which are comparable from place to place or time to time are mainly practical. Yet those practical difficulties are quite as serious as if they were theoretical in nature. Aside from the problems already discussed, there is the fact that families are changing their residence at an increasing rate. It is true that families with children of elementary school age move from one home to another in the same metropolitan area more often than they move from one city to another. But even among that group of families, most of the moving involves changes of schools, and this markedly complicates the record-keeping that is required for compiling accurate correction rates.

In many schools the staff engaged in health services is
continually changing, and changes in supervisory staff are likely to involve changes in definitions of both defects and corrections. Most correction rates are in terms of "remediable" defects, and the distinction between remediable and non-remediable defects is often a matter of the supervisor's judgment. And if, for example, epilepsy is counted as a remediable defect, the supervisor must decide what type or stage of an epileptic child's treatment is to be counted as a correction. Sometimes the non-remediable defects are included in the rate's denominator. This is reasonable if the numerator includes the cases for whom the program is providing suitable adjustments in the school work and other activities of the individual children concerned, but few published correction rates are clear about such matters.

Finally, the comparability of correction rates is affected by the accuracy of case finding. Unusually good case finding was reflected, very probably, in Philadelphia's finding that family physicians disagreed with school physicians only 4 percent of the time, and that the school physicians missed only 4 percent of the defects at later examinations. More typical figures were those which Walker and Randolph (1941) found, showing that "from 50 to 85 percent of the children reported on a particular examination as having a defect of heart, lungs, or nutrition were reported as normal at the second examination" in 6 Tennessee counties where the case finding was done under competent local health department auspices.

Evaluative studies

School systems in both Pennsylvania and New York have collected data from correction rates in the expectation that the figures might be useful for evaluating results of the laws requiring biennial (Pennsylvania) or annual (New York) school health examinations. The accumulated data have been used or at least examined in a number of the evaluative studies conducted by authorities in both States. A general review of their findings is probably more worthwhile than further examples of the correction rates and problems involved in them.

In publishing recent data for Pennsylvania, school health director German (1954) noted that about 1,900,000 children, comprising nearly all of the State's enrolled pupils, were examined during the two school years 1951–53. Some 36 percent of the examined children were reported as having remediable defects which were neither corrected nor, presumably, under treatment. The trend of this percentage, which had decreased to 36 percent from approximately 50 percent in 1946–47, was cited as an indication of the effectiveness of the State's 1945 school health law. We may remark that, at least in principle, the use of this simple
and direct statistical index would seem to have as much merit as any procedure so far used or proposed. German's report went on to give both the number of children (183,000) said to be "under treatment" and the number of "corrections completed" (165,000) in 1951-53, but no attempt was made to use these figures in any type of correction rate.

Support for German's caution was evidently found in the Pennsylvania Joint State Government Commission's evaluative study (Davis, 1955) of the school health work throughout the State as a whole. This was the Commission whose report, as noted above, included special data from Philadelphia. The Commission also gave attention to the problem of interrelating the data on remediable defects and corrections, as reported from other areas of the State. From this effort it was concluded, without elaboration, that "no significant relationship can be established for the State as a whole between discoveries and corrections of given types of remediable defects."

The schools in New York State, except for the larger cities, are required to report each year on the number of "defects found" and the number of defects "brought under treatment." The latter figure is divided by the former for purposes of a correction rate, which is published annually for at least 10 separate groups of medical defects as well as for total defects. As to how the total defects reported for a given year relate to the defects reported as brought under treatment in the same year, the most specific information which the reviewer has found was a statement by Maxwell and Brown (1948). They noted that an analysis of the records had shown that "two-thirds of the defects are new ones each year, while only one-third are untreated defects recorded for more than one year."

The same authors reported that the correction rate for all medical defects, as defined above, rose from 39 to 61 percent over the 20-year period 1925-26 to 1945-46. The prevalence rate of reported defects, per 100 children examined, decreased from 56 to 38. The authors recognized that the trend of these rates was considerably affected by changes in both examining and treatment practices. Such changes were critically examined for several kinds of defect, without reaching a general conclusion as to whether the children's health status had improved over the two decades concerned. An analysis was made of the 1945-46 prevalence rates and correction rates for each type of defect in each grade from kindergarten through high school, and from this analysis it was concluded that "few significant defects are not brought under care by the time the child completes twelfth grade."

Attention was also given to the data of New York State schools in two earlier evaluative studies. Winslow (1938) exami-
ined the correction rates of schools in 18 communities throughout the State. He could find no relationship between the statistics on medical services rendered and the percent of children who had defects, and he concluded that whatever value the routine medical inspections might have had could not be measured from the data which the schools had collected. Finally, when the New York State Education Department (1945) sought to use the correction rates to evaluate certain local programs, it was found that the recording of corrections had not been adequate for that purpose.

In evaluative studies that have critically considered correction rates as they are ordinarily obtained, the similarity of the conclusions is at once striking and rather discouraging.

It seems fair to conclude that, although there is no great problem about defining and obtaining a correction rate appropriate to a special purpose or project, it is probably idle to hope that any practical method can be devised for routinely obtaining corrections rates useful for evaluative purposes.

The time and trouble involved in compiling the rates would seem better spent in other ways, including careful and frequent estimation, in line with German's general procedure, of the proportion of children who need medical attention and are not receiving it.

**Summary**

As criteria of school health effectiveness, mortality rates are not very helpful because, for evaluating large programs, they are apparently poor indexes of the amounts of care that different groups of children receive; and, for evaluating smaller programs, childhood fatalities are too infrequent to make the use of death rates statistically sound.

School illness data, including data on accidents and injuries, are evidently liable to serious biases of reporting, with the result that poor programs may appear superior to good programs in terms of illness or accident rates as ordinarily reported. Although comparison of different schools is unsafe, a given school should be able, theoretically, to use illness rates as one check on the effectiveness of its program. This will be sound, however, only if uniformity of illness reporting can be maintained regardless of any interim changes in the program that might be found necessary or desirable.

Studies making use of Selective Service findings have not, as yet, proved much about school health effectiveness. It might still be possible to organize studies that will make suitable use of Selective Service findings, but it seems doubtful whether those findings are any better criteria of effectiveness than, for example, examina-
tions by specially qualified physicians. It is quite possible also
that studies using expert physicians' examinations as criteria are
easier to conduct than studies using Selective Service findings.

It is feasible to define and use correction rates appropriate
to special projects, but there continue to be severe practical difficul-
ties, perhaps on an increasing scale, in the way of compiling
correction rates that are adequate for routine evaluative purposes.
It is possible that solutions of the problem can be found in the
future, but if so, they will probably be along lines different from
the attempts made so far.
ALTHOUGH FACT-FINDING rather than evaluation per se is the purpose of most statistical surveys, the findings are often used in evaluative studies. It therefore seems worth while to review briefly the more important surveys of school health services. While this is a convenient point to do so, it should be noted that survey data are usually gathered also in evaluative studies of the kind discussed in the next Section, namely, studies utilizing expert judgment, and that there is no sharp line between those studies and the surveys reviewed here. An attempt is made, not to cover school health surveys comprehensively, but to select and discuss surveys that have either background value or methodological interest in connection with future evaluative work.

**Official surveys**

The pattern of several government surveys was set by the school health survey conducted in 1910 by the Russell Sage Foundation (see summary by Gulick and Ayres, 1913). Questionnaires were sent to some 1,300 school superintendents of that time, asking, for example, whether their schools had medical inspection programs, and if so, whether the work was administered by health or education authorities and how many physicians and nurses were employed.

Similar but increasingly detailed surveys were conducted as of 1923, 1930, and 1940 by Rogers and his associates in the Office of Education. The report of the 1940 survey (see Rogers, 1942) was notable for its inclusion of a review of previous governmental and non-governmental surveys in relation to other developments in the history of school health services.
The latest survey in this series was conducted in 1950. It was designed by C. H. Maxwell and reported by Kilander (1952, 1953a, 1953b, and 1955). Whereas most previous surveys had been restricted to places of at least 10,000 inhabitants, coverage was greatly increased in the 1950 survey by including places with 2,500 to 10,000 inhabitants.

As the first step, the school superintendents were sent a form whose chief purpose was to identify the schools which had some health service. The form indicated that, for purposes of the survey, a school had health service if medical or dental examining was "available" to the school. This was, in effect, a broader definition of school health service than the definition indicated on the survey form used in 1940, which asked for the numbers, kinds, and salaries of physicians actually employed for school health work.

In both 1940 and 1950, the survey forms were sent to all places having at least 10,000 inhabitants, so for that group of cities a comparison of the main findings is valid. The response rate of those cities was 74 percent in 1940, and 96 percent in 1950. Of the respondents, the proportion saying they had health services was 98 percent in both surveys.

It is likely that the definition used in 1950 tended to increase both the response rate and the proportion of respondents who considered that they had health service, compared to what would have been found if the 1940 definition had been used.

In consequence, the available evidence is not clear regarding either the direction or the amount of change, if there was any change at all during the decade concerned, in respect to the proportion of schools having health services.

The form used as the first step in the 1950 survey also asked whether at least one physician was available to the school. Of the respondents, 63 percent said yes. Similar questions regarding nurses, dentists, and dental hygienists yielded affirmative answers regarding them, respectively, in 85, 40 and 16 percent of the schools.

Another question on the same form concerned administrative control of the school health work. The respondents who had said they had health service reported that it was run by school boards in 60 percent of the places, by local health departments in 11 percent, by joint education-and-health authorities in 23 percent, and by other agencies in 6 percent. These percentages resembled those found in earlier decades, except that the 23 percent administered by joint education-and-health authorities represented a moderate increase over the 1930 and 1940 figures for that category.

To complete the survey, a detailed questionnaire was sent
to an appropriate sample of the places reporting that they had health service. A response rate of 79 percent was obtained to this questionnaire. Much information was obtained regarding the frequency of examinations, the roles of nurses and teachers, the school’s work with parents, and the school dental program.

In general, it may be said that the survey provided substantial information on school health services in about three-fourths of the places having over 2,500 inhabitants. At the same time, this means that the survey did not cover roughly half the children of school age, and that the greater part of the children not covered in the survey were located in rural areas.

The American Medical Association cooperated in planning the survey, and sent a separate questionnaire to the local medical societies over the country. Since the medical societies are organized on a county basis, the scope of this survey was not limited to urban places. The response rate was 53 percent.

One of the questions unique to this survey (see Hein and Dukelow, 1950 and 1951) was whether the community served by each medical society had “some method of assuring needed medical care for children whose families cannot afford to pay for services.” About 82 percent of the responding societies answered yes. Of all such service provided for underprivileged children, the average estimate of the proportion paid for by public funds was 40 percent.

In reply to a question as to whether the local schools had the services of a part-time or full-time physician, 54 percent of the responding societies said yes. Although this finding is reasonably consistent with the figure of 63 percent which Kilander’s report showed for the proportion of urban places with at least one physician “available,” neither figure provides a clear indication of the number of physicians participating in school health work or the volume of their services.

Data on the number of physicians employed at least part-time in school health work, and on the ratios of such physicians to the pupils served, are available in the surveys reported by F. W. Hubbard (1950), Smith (1951), Weaver (1954), and Schloss and Hobson (1956). These studies show that well over 90 percent, and perhaps over 95 percent, of the school physicians are employed on a part-time basis, but the reports do not indicate the amount of service that the part-time physicians provide.

**The Academy study**

Although no breakdown of full-time and part-time physicians was attempted in the report, the study of child health services conducted by the American Academy of Pediatrics (1949a and b) yielded the best information available on school physicians.
The study also obtained important information on several other aspects of school health, which were set forth in detail by J. P. Hubbard and others (1949).

**Numbers of physicians and nurses**

The data on school physicians were obtained through State and local health departments. Their staffs canvassed all of the public elementary schools in the country, thus insuring a response rate of practically 100 percent.

In this survey, which for brevity is usually termed the Academy study, school physicians were defined as those who conducted, in the schools, regular or special examinations of the children for purposes other than athletic participation. Such physicians totaled nearly 8,000, including approximately 1,600 health officers, 6,000 general practitioners, 250 pediatricians, and 150 other specialists.

| Physicians and Nurses per 100,000 Children in Public Elementary Schools, 1946 |
|-----------------------------|-----------------|-----------------|
|                             | United States 1 | Urban 1         | Rural 1         |
| Total physicians            | 44              | 52              | 31              |
| Health officers             | 9               | 6               | 13              |
| General practitioners       | 33              | 43              | 17              |
| Pediatricians and other specialists | 2         | 3               | 1               |
| Total nurses                | 65              | 77              | 44              |
| Full-time nurses            | 25              | 34              | 9               |
| Part-time nurses            | 40              | 43              | 35              |

1 To obtain the rates in the column headed "United States," the number (see text) of personnel in each of the categories at the left was multiplied by 100,000 and divided by 18,000,000, which was the average of the enrollments in public elementary schools in the school years 1945-46 and 1946-47. The denominators used for the columns headed "Urban" and "Rural" were 11,420,000 and 6,580,000, respectively. These urban and rural groupings are the areas which are designated respectively as "metropolitan-adjacent" and "isolated" in the report of the Academy study.

These figures were obtained in 1946. Today the absolute number of participating physicians is probably greater in all four groups. Yet, with the possible exception of the "other" specialists, it is unlikely that the relative frequencies of the groups have changed much, either in relation to each other or in relation to the school population served.

For the school physicians identified by the study, the rela-
tive frequencies per 100,000 children in public elementary schools during 1946 are shown in the accompanying table, which includes the rates for urban and rural areas as well as for the country as a whole.

The table also shows rates of the same type for the 4,440 full-time and 7,280 part-time nurses whom the study, through the same procedure, identified as serving public elementary schools in 1946.

The differences between the rates for the urban and rural areas are marked, but are not too surprising in view of the urban-rural differences known to exist for other public health activities.

The rate of 44 physicians per 100,000 children means that there was one participating physician, usually part-time, for each 2,200 enrolled children. The nurses' rate of 65 per 100,000 children means that one full-time or part-time nurse was engaged in school health work for each 1,500 enrolled children.

**Hours of medical service**

However, these ratios, like the rates per 100,000 pupils on which they are based, are of doubtful value because of the uncertain meaning of the term “part-time.” The important question is not how many different individuals provide service, but how much service they provide. A limited amount of information on the latter question is available from the Academy study. Although it is the only information of its kind available, it is worth examining less for that reason than for its relevance to the problem of obtaining better information in future surveys.

The previously noted breakdown of the 8,000 physicians serving in public schools indicated that nearly four-fifths were general practitioners or pediatricians, and we may believe that about the same proportion holds today. Moreover, the proportion is probably similar in non-public and public schools. It is thus apparent that the bulk of the medical service in schools as a whole is provided by general practitioners and pediatricians. With the exception of a few individuals who are full-time supervisors, such physicians serve schools on a part-time basis. This means that practically all of them are engaged concurrently in private practice.

In another part of the Academy study, a schedule was sent to each of the 75,000 general practitioners and 8,500 pediatricians engaged in private practice at the time of the survey. One section of the schedule included the question: “During the past month how many hours did you spend in school health services?” The study thus sought to obtain the total hours of service provided by all general practitioners and pediatricians serving part-time
in both public and non-public schools. As already noted, this service did not represent the total of physicians' services provided in the schools, but it was much the largest part.

Only half of the general practitioners and two-thirds of the pediatricians completed the section of the schedule that included the question on school health. But if we arbitrarily assume that the average non-responding physician spent as much time on school health as his responding colleague, and if we allow for certain seasonal factors, the data indicated that 2,450,000 hours of service were provided. Since the survey mainly concerned the calendar year 1946, the hours of service are best related to the average of the enrollments (both public and non-public) in the school years 1945–46 and 1946–47, which was 20,300,000 children. The ratio 2,450,000/20,300,000 yields .12 hours per child as the average amount of medical service provided in the schools by non-supervisory general practitioners and pediatricians.

Relevance for a future survey

Considering the fact that only a minority of children need attention in any one year, and the fact that the physician's role in the school is mainly to conduct, verify or improve the case finding and not to provide treatment, an average of .12 physician-hours per child is a very substantial amount of service. If this figure could be confirmed by sound survey methods, and if it could be shown that the participating physicians have, on the whole, training and experience appropriate to their work, we could feel sure that the keystone of school health services was firmly in place. But the figure .12 hours per child should not be relied upon at all until it has been checked by adequate survey methods.

An adequate survey would be designed to insure a high response rate and to reduce dependence on memory. The survey would cover not only the services of general practitioners and pediatricians, but also the work done for the schools by health officers and other specialists.

For these purposes it would be desirable to set up four equivalent samples of school districts, each one covering both urban and rural areas, and to use a different sample for each quarter of the school year. At the end of each quarter the schools in the particular sample would be asked, preferably by interviewers traveling directly to the schools, for the number of hours of physicians' services that were provided to the children through the school health service. So far as possible this information should be ascertained from records.

It would be very desirable to obtain breakdowns of the hours of services according to the kinds of training the physicians have received, and also according to the amounts of experience
they have had in school health work. Such questions are at least as important as the commonly used questions regarding administration and financing of the programs by education or health agencies, although those questions warrant inclusion also.

**Nursing services**

The same survey should, if possible, obtain analogous information on the services of school nurses. A variety of information relating to that subject is already available, but it badly needs to be supplemented with data from special surveys.

The most recent of the annual reports on nurses by the Public Health Service (see its “Census of Nurses”) shows that, in Continental United States during 1955, approximately 7,730 nurses were employed by school boards, while about 12,270 nurses were employed by local health agencies, some of which operated in combination with voluntary agencies.

We may assume that the great majority of the 7,730 nurses employed by school boards are full-time employees, and that they are roughly comparable with the 4,400 full-time nurses identified in 1946 by the Academy study (see above). However, we do not know how many children were served by the full-time nurses employed by schools in 1946, 1955, or any other recent year.

Concerning the 12,270 nurses employed by health agencies in 1955, no estimates seem to be available for the country as a whole regarding either the time they spent on school health or the number of children they served.

For both the nurses employed by school boards and those employed by health agencies, the 1955 report gives valuable information about the academic and public health training which the nurses have received. However, it would be desirable to find out in a future survey how much training or experience each group has had in work with school-age children. Finally, it would be important to learn how long the nurses have been with the schools where they are found at the time of the survey, since it is well recognized that continuity of service contributes to the effectiveness of a nurse's follow-up work.

Using data from the same series of censuses, Tibbitts and Levine (1953) showed that during the 15-year period 1937-52, of all public health nurses, the proportion employed by local health agencies showed a moderate rise from 44 to 51 percent, while the proportion employed by school boards rose comparatively rapidly from 20 to 29 percent. The decrease for the non-official agencies was from 34 to 18 percent. These figures are of background interest in connection with the problem of providing school nursing service, but they leave us without a clear picture regard-
ing trends in the amount of that service, per child, which has been available through either school boards or health agencies.

**Dental services**

A recent survey sponsored by the American Dental Association (see Moen, 1955) has provided important new data on school dental programs. Questionnaires were sent to the school superintendents in the 3,530 cities having over 2,500 inhabitants in 1955. The response rate was 63 percent. Of the respondents, 60 percent said they had programs which were "using in any way the services of dentists, dental hygienists, or dental assistants." The findings noted below are based on the replies of the 1,340 superintendents (60 percent of 63 percent) who reported programs.

The survey indicated that in 74 percent of the programs dentists were performing, at intervals of one, two or more years, inspections of the teeth of all children in the schools concerned. Since this practice is often thought to be a poor use of dentists' time, it is perhaps encouraging that only a third of the time, or in only 25 percent of the programs, dentists were performing such inspections every year.

Dental hygienists were performing such inspections at various intervals in 29 percent of the programs, and in nearly half of those programs the hygienists' inspections were given annually. The published figures do not show how often the inspections by dentists and those by hygienists were done in the same schools. It is nevertheless clear from the two sets of figures that inspecting is a very common activity in school dental programs.

To the extent that the findings from inspections are used to keep the school, the parents, and the children informed about the program's results, the inspections are of course an important evaluative activity, as will be discussed later (Section 5). Unfortunately, the superintendents were not asked whether the results of dental inspections were used in evaluating the programs. The survey did, however, bring out the rather surprising fact that, following the inspections, the children are referred to private dentists in less than half of the programs where inspections are carried on. Thus, except in relation to evaluation, the usefulness of the inspections is uncertain in many programs.

In 8 percent of the programs, fluoride treatments were given by dentists. In another 12 percent, the same treatments were given by dental hygienists, who devoted about 40 percent of their time to this work.

Additional forms of dental treatment, including fillings and extractions, were given to underprivileged children in 44
percent of the programs. In another 11 percent of them, such treatment was available for children whose parents "requested the service."

In reply to a question regarding the locations in which the inspecting and other dental services were performed, school buildings were named by 77 percent of the superintendents, while 20 percent specified private dentists' offices, and 16 percent mentioned a health agency, a community clinic, or some other place. The figures total over 100 percent because some programs utilize more than one location.

Asked whether children were excused from school to go to private dentists, 92 percent of the superintendents said "yes." In part this may reflect favorable effects of recent efforts to get all schools to adopt a "released time" policy for dental care (see, for example, Menzner, 1954a and b).

Although the report of the survey does not mention the numbers of full-time or part-time dentists, hygienists, and assistants who were participating in the programs, those numbers could be estimated, if desired, from the percentage distributions of personnel that are included in the published results. However, it is perhaps just as well that the numbers of personnel were not cited, since most of staff members were part-time, and the amounts of service they provided in the various programs were not ascertained.

**Hours of dental service**

For information on the amount of dental treatment we must turn again to the study of the American Academy of Pediatrics (1949b and c). Except that it did not cover the work of dental hygienists, the study yielded comprehensive information on the volume of dental care that children receive, both in public clinics and in the offices of private dentists. Since this part of the Academy study was not directly concerned with schools, the data were related to child population data rather than to school enrollments.

Each community dental clinic, whether tax-supported or voluntary, and whether in a school or elsewhere, was asked to estimate the "dentist-hours of service" given to children under age 15 throughout a 12-month period. Some 970,000 hours were reported, with 28 percent given by voluntary organizations and 72 percent by schools and other tax-supported agencies. These figures represented practically complete coverage of the community dental clinics, which were canvassed by staff members of local health agencies in the same way as the schools were canvassed for the data on numbers of physicians and nurses.

By far the greatest part of the 970,000 hours of "clinic"
care—probably about 850,000 hours—went to an estimated 630,000 children aged 5–14, almost all of whom were underprivileged. The average number of hours per school-age child receiving the clinic care was therefore about 1.3 hours (850,000/630,000) per year.

The 630,000 children who received the clinic care were only a small fraction (1/35) of the 22,000,000 children who were of ages 5–14 years. What of the dental care received by the remaining 21,370,000 children, who were not served by the tax-supported or voluntary agencies?

To obtain data on that question each of the 66,000 dentists in private practice was asked to record, for one day, the number of hours of service given to patients of specified age groups. The dentists were assigned different days on which to report, and the days were randomized to include weekdays, Sundays, and holidays in due proportion.

The response rate was only 43 percent, and it is unfortunate that steps were not taken to raise it. If, however, we assume that the non-respondents gave as much service as the respondents, the data indicated that 17,780,000 hours per year were spent on patients aged 5–14. Thus the average amount of time spent per school-age child in the non-clinic group may be estimated as .8 hours (17,780,000/21,370,000). If a sample of the non-respondents had been queried, a sounder estimate would have been obtained, and the result might have been somewhat less than .8 hours. We can say, in any event, that the obtained figure was strong evidence that the average child in the non-clinic group received considerably less dental care than the 1.3 hours which the average child in the clinic group received.

Especially with respect to non-clinic children, it is known from the study of Klein and Palmer (1940) that some of the children in that group receive much attention while others receive little or none, and that the differences in amount of care are markedly associated with the parents’ income levels. Therefore, if the data for the non-clinic group had included separate figures for children in families above and below median income, the results might well have shown, say, that scarcely .4 hours of service was provided for the average child in lower income families, while 1.2 hours or more was provided for the average child in higher income groups. The data obtained in the dental part of the Academy study are thus consistent with the generalization made in respect to other aspects of health care—that substantial care is received by members of families whose income is either very low or above average, while inadequate care is received by the group whose incomes are not quite low enough to make them eligible for clinic care.
"Profile" surveys

A series of surveys of a different type began with the American Child Health Association's study of health conditions in 86 cities (Palmer, 1925, and Palmer and others, 1925). The Association sent teams of surveyors to all cities having 40,000 to 70,000 inhabitants. Using a schedule of several hundred questions arranged under 84 headings, the staff gave attention chiefly, though not exclusively, to child health services. The surveyors' report is of interest now mainly for the remarkably "modern" notes, so to speak, which were struck in the descriptions of individual cities. In the same vein as today's discussions, the 1925 report stressed the need for better cooperation between schools and parents, the need for better cooperation between school physicians and private physicians, and the lack of a rational basis for many of the city-to-city variations in child health services.

In the light of experience with the original schedule, the surveyors were able to develop a new form covering the more significant questions under 11 topics. For each of the 86 cities, it was possible to re-cast the survey data in terms of the new form. This made it possible to chart each city's "profile" on the 11 topics. When these profiles were sent to the participating cities, some of them were able to use the material effectively in budget justifications.

The principles evolved in this work were applied later (1941) in an extensive revision of the "Appraisal form" which the American Public Health Association had developed for studying community health programs. That form developed into a whole set of forms called the "Evaluation Schedule." It was designed for use with what were termed "Health Practice Indexes," which were essentially an extension of the "profile" idea developed in the study of 86 cities.

Beginning in 1941, interested local groups used the Evaluation Schedule to obtain and report to the Association on many different kinds of statistical rates in their communities. In the Indexes, all of the rates of a given kind from the various communities were presented on a single page, with a horizontal line representing each community rate, and with the lines ranked in order of size. When the completed charts were sent to the participating communities, each one could readily see where it stood, compared to other reporting communities, with respect to whatever rates it had sent in.

About 10 percent of the Evaluation Schedule was devoted to "School Health," and this section was one of the most popular parts of the Schedule. Some 16 items in that section called for rates that were fairly easy to obtain as percentages, and 20 other
items could be used in the manner of a check-list, with scoring simply in terms of the presence or absence of particular conditions or practices in the school health program (see Palmer, 1951, for a full description of the scoring of this and other sections of the schedule). The selection and phrasing of the material indicated that the authors subscribed to most features of the Astoria plan of school health services (Nyswander, 1942).

When the last of the Indexes, or chart reports, was published in 1950, over 200 communities encompassing about 21 million people were reporting to the Association on at least some parts of the Schedule. However, for reasons that are not clear, the process of collecting the rates was terminated after publication of the 1950 Indexes. In 1954 the Association completed a drastic revision of the forms in the old Schedule, which was renamed “Guide to a Community Health Study.” In the new section on school health services, emphasis on the principles of the Astoria plan was continued, but with much less attention to statistical matters. The main elements of the revised section were statements regarding desirable school health practices. The Schedule went on to ask whether those practices were being followed, and provided space for writing essay-type statements on how local conditions might be improved.

Cost data

The available information on the cost of school health services is unsatisfactory, but is well worth reviewing to bring out problems needing attention in future studies. Sketched below, in increasing order of importance, are four such problems.

Public versus non-public schools

1 Most of the available data on school health expenditures are for public schools, and relatively little is known about expenditures for the 15 percent of school children who are in non-public schools.

However, there does not appear to be substantial reason to assume that a marked difference exists in per-pupil expenditures of the two groups of schools. The question would therefore seem to be a low-priority problem. Nevertheless, as opportunity permits, the question ought to be studied in a small-scale special survey or in connection with some large survey.

Elementary versus secondary schools

2 Another not-too-serious problem concerns the fact that, in reporting on expenditures for health services, school systems usually provide the data for the elementary and secondary
grades combined, and, on a national basis, no data are available regarding expenditures for the elementary grades alone.

It is known that, on a per-pupil basis, expenditures are much larger for the elementary than for the secondary grades in almost all school districts. This consideration, together with the fact that the elementary children comprise a majority of the total group, means that the average per-pupil expenditure in the elementary grades alone must be only a little higher—perhaps by some 10 percent—than the average per-pupil expenditure for the elementary and secondary grades combined.

**Health department expenditures**

A more serious problem to be reckoned with in future surveys concerns the fact that most of the available figures on school health costs represent expenditures made by the schools alone. It is known that, in many if not all of the States, considerable amounts are spent for school health services by health agencies, and yet no estimate of the overall extent of their contribution is available.

The school is clearly the most convenient place for assembling information about funds expended from both sources, and a recommendation in accordance with that principle was made in a handbook on school accounting procedures issued in 1948 by the Office of Education and the Association of School Business Officials (see Foster and Akerly, 1948). It was recommended that schools should not only keep a record of their own expenditures for health services, but, wherever a health agency also expended funds for that purpose, the school served by the health agency should ascertain, and keep a separate record of, the contribution made by the health agency. It was pointed out that this procedure would permit the collection of data on school health expenditures that would be comparable from one school system to another.

A coordinate proposal was included in the handbook issued a few years later by the Office of Education and the National Council of Chief State School Officers (see Reason and others, 1953). This handbook contained recommendations as to the kinds of information which State education departments should collect from local school systems. One of the recommendations was that the State education departments should collect, in separate categories, the expenditures which the schools and the health agencies made for school health services.

At some time in the future, data in line with these recommendations may become available for a large number of States. At present, however, the available data on the amounts which health agencies are contributing to school health services are too sparse to give useful indications of the national picture. The latest
school year for which data have been collected is 1953–54. As part of its regular biennial survey covering that year, the Office of Education asked the State education departments to furnish tabulations in accordance with the recommendations described above. The results showed that only Oklahoma, Vermont, and the District of Columbia had been able to collect data on the amounts spent for school health by health agencies. The Office of Education therefore did not publish figures in that category.

The report of the survey (see Schloss and Hobson, 1956) nevertheless gave the expenditures made by the schools alone. These figures totaled $58,269,000 for the 40 states that responded to the question concerning expenditures made by schools.

To see what an expenditure of this size means on a per-pupil basis, we may divide $58,269,000 by the figure 21,630,000, which was the total average daily attendance in the schools of the same 40 States during the same year (1953–54). The quotient is $2.70. This represents the per-pupil expenditure for health services which was made by the schools alone in the 40 States. If the same type of data were available from all of the States, it is likely that the per-pupil figure would still be close to $2.70.

As noted in our discussion of problem 2 above, all such data are for the elementary and secondary grades combined. To obtain a rough estimate of the per-pupil expenditure in the elementary grades alone, we may increase the figure $2.70 by some 10 percent. We thus arrive at approximately $3 as an estimate of the per-pupil expenditure, made by the schools alone, for health services in the elementary grades.

**Physical education expenditures**

Lest our discussion be burdened with too many problems at once, we have spoken of expenditures for school “health services” as though we could assume that practically all of the funds reported in this category were spent for health services alone. However, the correctness of that assumption is open to serious doubt. This circumstance seems to be the most difficult of all the problems involved in judging school health costs.

The obstacles in the way of obtaining precise information on this question are illustrated by certain findings in the survey.

---

2 In using average daily attendance rather than enrollment as our divisor, we have followed the usual practice of the Office of Education. It should be noted that enrollment figures are always larger than figures for average daily attendance. If, therefore, we had used enrollment rather than average daily attendance as our divisor, we would have obtained less than $2.70 as the per-pupil expenditure, and the smaller quotient would not represent the full amount expended per child actually present in the schools’ day-to-day operations.
of urban schools which Hubbard (1950) conducted for the National Education Association. One part of Hubbard's survey concerned personnel engaged in health services, as distinct from personnel engaged in physical education or recreation. On that part of the survey only a fourth of the superintendents who returned the questionnaire were able to furnish information which Hubbard considered acceptable. The survey also asked for information on expenditures for health services, as distinct from expenditures for the other two functions. The answers to this question were even less satisfactory than the answers to the questions regarding personnel, and Hubbard could only report that the survey "failed to produce the type of information desired" regarding expenditures for health services alone.

The trouble arises because many schools administer health service, physical education, and recreation as coordinate parts of one program. This, of course, is what the schools have long been encouraged to do by authorities in the field of school health services. It was not the intention of those authorities that the tying together of health services, physical education, and recreational activities should extend to school accounting practices. But apparently no official recommendation for the separate accounting of health services was made until 1948, and by that time many school systems had become accustomed to grouping together the expenditures made for health services and various other activities relating to health.

A contribution toward correcting this situation was included in the 1948 handbook by Foster and Akerly (see above). They specifically recommended that "all costs for physical education or health instruction, including physical examinations, tests, and weighing that are considered part of the instructional program, should be charged to instruction," rather than to health services. The later handbook by Reason and others (1953) included essentially the same recommendation, and added the practical suggestion that, when a staff member is engaged in providing both health service and physical education, the school should either prorate his salary, or "include the salary under instruction if more than half the work load consists of teaching."

If and when most school systems follow these recommendations for the separate accounting of health services, it is likely that this problem, like problem 3, will be resolved. However, all experience with State reporting systems indicates that they cannot be changed rapidly, so it is unlikely that satisfactory data will become available in the immediate future. A thorough check on the present situation has not been attempted by the reviewer, but he has looked for the pertinent information in a number of the recent reports of State education departments. He has found that,
in at least some instances, the figure given in the State report as expenditure for "Health promotion," "Health, physical education and recreation," or "Health services and other coordinate activities" was the same or nearly the same as the figure which, in the Federal biennial survey, was reported as expenditures for "School health services."

If a substantial number of States are likewise continuing to report the total spent for health services and one or more other activities as the amount spent for health services alone, it is obvious that the figures available from the biennial survey are biased in the direction of overstating the expenditures actually made for school health services.

This direction of bias is, of course, the opposite of the direction concerned in our discussion of problem 3, which dealt with the lack of national data on the expenditures made by health agencies for school health services.

Although there is no sound way of judging which of these two biasing conditions is the larger, there is reason to suspect that both of them are sizeable, and that they may, to a considerable extent, cancel each other's effects. Indeed, it seems very likely that, if accurate allowances could be made for both biases, the figure $3 would not be lowered to less than $2.75 nor raised to over $3.25.

Thus, pending an adequate sample survey (see page 29) which not only covers funds contributed by both schools and health agencies, but which distinguishes clearly between funds spent for health services and funds spent for health-related activities, it seems fairly sound to use $3 as a working figure for the per-pupil expenditure for health services in elementary schools of the country as a whole.

"Background" factors

Since school health services are frequently regarded as an expensive type of public health activity, it is worthwhile to relate the approximate per-pupil cost, which we have estimated as something like $3, to the per-capita cost of public health work as a whole.

Expenditures for public health

The annual report on selected civilian health programs compiled by the Social Security Administration (1955) shows that, for the United States including its outlying parts, a total of $984,000,000 was expended in the year 1953-54 on "community health services."

This category includes the programs of Federal, State and
local health agencies, crippled children's programs, maternal and child health programs, and the services furnished to public and non-public schools by health departments. It does not include the health services which public and non-public schools finance from their own budgets. To remedy that omission, it seems fair to increase the published figure of $984,000,000 by about $70,000,000. (The latter figure is a rough estimate made by the writer, with consideration of the fact that, in 40 States, the public schools spent a total of $58,269,000.) It should be noted that even if the estimate of $70,000,000 is in error by as much as $10,000,000, the effect of that error amounts to less than 1 percent in the total $1,054,000,000 (i.e., $984,000,000 plus $70,000,000).

Since a sum of approximately that magnitude was spent in 1953–54 for civilian health programs in continental United States, Alaska, Hawaii, Puerto Rico and the Virgin Islands, the proper base for obtaining the per-capita expenditure is the total civilian population of the same areas on January 1, 1954, which was 165,529,000. Using this as the divisor yields $6.32 as the per-person cost of public health work.

In this light the estimated per-pupil cost of $3 for health services in elementary schools does not seem excessive, even if we grant that school children are recipients, not only of those particular services, but also of considerable health service from the many other public programs included in the per-capita expenditure of $6.32.

However, opinions as to whether school health costs are large or small are likely to be affected by the background figure against which those costs are judged, and the size of the background figure depends on the definition of public health used for the purpose. The definition which we employed in computing the figure $6.32 not only excludes health programs for veterans and other military-related health programs, but it also excludes public programs of hospital care, hospital construction, medical care, medical rehabilitation and "health expenditures made in connection with public welfare." Clearly, if some of these programs had been included in our definition of public health, the cost of school health services would have appeared a good deal smaller, relatively, than in the comparison made above.

The effect of background figures is especially striking at the local level, where they sometimes give rise to serious misunderstanding. If a school's health service is provided by the local health department, the budget of that department becomes, naturally enough, the background figure against which the school health costs are judged. But, as Mountin and Haldeman (1953) have stressed, the health department's budget usually covers only a part of what the community is actually spending on public health.
That is, for the administration of certain activities which should be considered as part of the public health program, the community gives control to other units of government, and the expenditures made for those activities do not appear in the health department's budget.

The result is that, for purposes of judging school health costs, health department budgets are misleading. Community leaders concerned with child health may or may not have fully understood why this was so, but the fact that school health costs appear high in relation to health department budgets must have been obvious for a long time, and, almost certainly, that situation is one of the reasons for the slowness with which the jurisdiction of school health work has been changed from school boards to health departments.

We may add, in anticipation of the discussion of a survey by Moss (1945) in the next Section, that she found the jurisdictional problem was well solved where the school's budget included the funds for health services, but with the money earmarked for purchasing all or most of the services from the local health department. This may not be quite as good as Britain's solution of the jurisdictional problem through making one man responsible for two jobs (Henderson, 1955), but it is clear that, for American conditions, the solution noted by Moss merits wide consideration.

**Expenditures for schools**

The amount spent, per pupil, for all phases of schooling provides another important type of background figure. Schooling can be regarded as the investment of certain kinds of effort in the making of future citizens, and school health services can be considered as one of those kinds of effort. How large is the effort which is being invested, through the schools, in health status, compared with the total effort being invested in the schools?

The previously mentioned report of Schloss and Hobson showed that, in 1953–54, current expenditures (exclusive of capital outlay and interest) made by the public elementary and secondary schools in the country as a whole amounted to $265 per pupil. As noted earlier, data are not available on a national basis for elementary schools alone. However, certain information is available regarding the per-pupil expenditures made, separately, by the elementary and secondary schools in several hundred cities (see Herlihy, 1955). The figures available from the cities are for instruction only, but that category accounts for the greater part of all current expenditures by schools. The information from the cities indicates that in 1953–54, the per-pupil *instructional* expenditure for elementary schools alone was about 22 percent less.
than the corresponding expenditure for elementary and secondary schools combined.

Even if we assume that the national figure $265 should be reduced by somewhat more than 22 percent to make it reflect the per-pupil expenditure in the elementary schools alone, the reduced figure would still amount to about $200.

We may thus conclude that the per-pupil expenditure of about $3 for school health services accounts for less than 2 percent of the total current expenditure made per child in elementary school.

Information on expenditures is of consequence, and more detailed figures should be obtained if possible in a future survey. But whether the amount spent on school health services is considered high or low in relation to other expenditures, the more important question is whether school health services are effective for their purpose. The remainder of this review is addressed to that problem.
THE USE OF EXPERT JUDGMENT

STUDIES employing "expert judgment," as used here, include studies which rely on statistical or other information that is available from first-hand inspection, interviews, or existing records, as distinct from information obtainable only by re-examining the children concerned or through controlled experiments.

Although the difference between the reports covered in the preceding section and some of those to be reviewed here is a matter of degree, we have tried to include in this section only those survey-type studies where the chief intention was evaluation, and not simply fact-finding with a view to its possible use in evaluation.

Evaluative reports relying mainly on expert judgment fall naturally into studies of State programs and studies of one or more city programs. It is convenient, and possibly of some special interest, to review in chronological order the reports under each heading. The studies to be considered are heterogeneous in nature. In part this arises from the fact that we have selected studies representing a wide range of approaches and findings. It also appears, however, that heterogeneity is characteristic of the studies relying on expert judgment, especially where only a few experts are used.

Evaluations of State programs

Massachusetts

The first study of school health in this country was initiated by Horace Mann and was the prototype of several later studies utilizing expert judgment. About 1838, soon after he was made secretary of the Massachusetts board of education, Mann wrote to school boards throughout the State, asking for reports on physi-
cal conditions in the schools that might be affecting the children adversely. Probably at Mann's suggestion, the State board of education asked William Alcott to evaluate the materials sent in from the local districts. Alcott was qualified for the task, not only because he was a schoolmaster, but because he had a special interest in medicine, as evidenced by the fact that he obtained a medical degree a few years after publishing his evaluation of the material collected by Mann.

Most of Alcott's report (1840) was a recital of the schools' deficiencies in respect to space, heating, ventilation, toilets, and seating. The presentation was interspersed with claims, chiefly on a priori grounds, that the physical inadequacies of the schools were having quite serious effects on the children's health or behavior.

From a historical viewpoint, the most important part of Alcott's report was his plea that a large city should conduct the experiment of having physicians continuously "watch over the physical education and management" of school children. For, he said, "wherever large masses of people, whether children or adults, are accustomed to assemble and remain together for hours in succession, there, in the present ignorance of mankind in regard to the natural laws of the Creator, will be oft repeated transgressions of those laws... and wherever the natural laws are thus continuously disobeyed, preachers of those laws are required." Alcott went on to suggest that the duties of the school physicians should include instruction of teachers in selecting indi-

---

3 Shortly before the publication of Alcott's report, Lorinser (1886) visited 50 schools in Germany and wrote a famous essay on "protecting" children's health by having them study less and exercise more. While saying he could not go into detail on the medical problems involved, he declared he had found evidence that school work had alarming effects on the blood, on digestion, and even on reproduction. Lorinser's report stimulated several other authors to claim that school work or bad school conditions caused a variety of maladies, including myopia, goitre, scoliosis, epilepsy, and chorea.

These claims were reviewed by the illustrious public health pioneer Virchow (1870). He pointed out that the evidence was weak, and said that not until "comparative statistics" had been collected by competent physicians could one tell "how far certain diseases are connected with school conditions." This relieved the fears of parents and school authorities for over a decade. However, it did not deter Cohn, who had already published an extensive statistical study (1867), from continuing to assert that school work caused myopia. That alleged fact was used by Cohn (1886), in the first text on school health, as his chief justification for calling on all schools to employ school physicians.

Virchow's suggestion regarding "comparative statistics" was taken up all too literally by a number of authors. An example of the masses of nearly meaningless statistics that were collected was the study of Warner (1893). After conducting individual visual inspections of 50,000 British school children, he reported that 11 percent were defective in anatomical development, while 10 percent showed deficiencies in what he called "nerve signs," and 4 percent had "nutritional" deficiencies. In a critical review, Kerr (1897) was able to moderate the effects of such reports by pointing out that most of them were conducted by "unscientific observers," and by calling for "exact studies" in schools and psychological laboratories.
vidual children to be "presented to the medical man at his semi-weekly, weekly, or monthly visits" to the school.

Thus Alcott stated the philosophy of school health work, including emphasis on teacher observation, at least half a century before major cities began to have programs (in the late 1890's), and a full century before teacher observation was widely adopted.

New York State

Although significant evaluations of city programs began to be made soon after 1900, Alcott's report seems to have been the only important evaluation of a State program until Winslow (1938) published his report on the effects of New York State's law requiring annual examinations.

Accompanied by two other experts, Winslow visited 18 communities believed to be representative in respect to the areas of the State affected by the law. Records of the school health program were examined and extensive schedules were filled out during or after interviews with administrators, physicians, and nurses serving the schools.

The main conclusion drawn was that the requirement of annual examinations was wasteful, and that the law should be repealed. In place of annual examinations, said Winslow, each child should be given "a comprehensive examination three times during school life," and selective examining should be conducted the rest of the time in accordance with the recommendations of earlier reports, among which Franzen's 1933 study of city programs (discussed below) was perhaps the most outstanding.

Winslow's report probably affected laws in other States, but it did not accomplish the repeal of New York State's law. Moreover, Winslow's study and the Tennessee study to be reviewed next failed to prevent the enactment of Pennsylvania's 1945 law requiring biennial medical and dental examinations.

Tennessee

Walker and Randolph (1941) evaluated Tennessee's program, which had been guided for over a decade by recommendations of the State health department. Those recommendations included provisions for examining all children every two years.

As their evaluative method, the authors analyzed the records which had accumulated during the period 1930-36 in six of the State's counties. Since the purpose of the study was evaluation of the procedures recommended by the State health department, the basis used for selecting the six counties was the fact that, in the selected counties, the local health authorities had been able to carry out the State's recommendations under fairly constant conditions.
In all, the records of 56,000 children were studied. The main finding was that a substantially greater frequency of uncorrected defects had been recorded for the sixth-graders of 1930, whose exposure to the program had been relatively brief, than for the sixth-graders of 1936. This was taken to indicate that the program had been moderately effective, although the authors admitted that trends in examiner “fashions” could have accounted for at least some of the difference between the findings of 1930 and 1936.

Information in the records was also used to compute certain correction rates, and these rates were cross tabulated with such variables as whether the children had received preschool supervision, whether the parents had attended the examinations at school, whether school nurses had made home visits to stimulate corrections, and whether certain defects had been found in earlier examinations and reported to the parents.

The tabulations showed statistical associations in expected directions, ranging in degree from moderate to slight. The associations were interpreted as confirming the desirability of preschool supervision and of parents' attendance at examinations, while casting some doubt on the value of home visits, and raising serious questions about the desirability of biennial examinations.

In general, the study findings did not seem unreasonable. But, for the degree of conclusiveness which the data could carry, it was unnecessary, and it was rather wasteful of professional time, to tabulate data from such a large number of records.

California

California’s program was surveyed and evaluated by Moss (1945). She sent an extensive questionnaire to the State’s 40 full-time local health departments and obtained complete information from 37 of them. We may take space to note only two of her many findings and recommendations, all of which are of interest in connection with the administration of State programs.

Jurisdictional problems appeared to be solved best where the funds for school health services were made a part of the school’s budget, but the money was tagged for purchasing the services from the health department. This arrangement was encouraged, though not required, by State law. Substantial conformity with the intent of this law was reported in 10 of the 37 health jurisdictions, and Moss found that interest in the plan was increasing. Considering the problems noted in the discussion of costs in the preceding section, the plan encouraged by California’s law would seem advantageous to all concerned.

Moss reported that most of the State’s local health depart-
ments had not only issued standing orders permitting nurses to recommend treatment for, and sometimes to treat, the “nuisance” diseases, but had also directed the nurses to inspect the children for “gross evidence of health disorder.” As will be brought out in the discussion of screening methods (Section 5), this type of pre-selection of the children to be examined by school physicians is worth testing against the use of teacher observation, which is now employed relatively often for the same purpose.

**Washington State**

A different method of pre-selecting the children to be examined by physicians was urged by Williams (1946) as the central recommendation in his evaluation of Washington State's school health program. The evaluative procedure used by Williams was somewhat unusual in that, except for citing certain data which were readily at hand, he relied directly on his judgment and experience as a physician long associated with school health work.

He said Washington's new medical school should select and train “a new type of health worker for the schools,” to be called a “health examiner.” He stressed that, up to the present, most of the 6 or 8 years of training given to medical students has been devoted to specialties that are of little or no use to school physicians. For, he said, whenever a physician serves schools he is “told not to assume in any way the functions of the family physician or other medical agencies in the community.” He called the practice of using physicians in this way unimaginative and wasteful of professional resources.

As an alternative, Williams urged that the new medical school select individuals already familiar with basic sciences and train them, as “health examiners,” to detect abnormal conditions in children. They would not be trained in treatment or diagnostic methods but would be able to make accurate referrals to physicians, who would be responsible for all differential diagnosis. The examiners would be part of the school's full-time staff, and would therefore be “more effective (than visiting physicians or nurses) in controlling the spread of communicable disease, by sending home children showing early signs of contagion.”

Other recommendations arising from Williams' evaluation were relatively routine in nature, except for his proposal that “significant facts” from the children's health records, especially regarding any uncorrected defects, should be entered on the children's report cards along with their school marks.

**Oklahoma**

Hiscock (1951a) has stated the rationale of his evaluative methods, and has provided a good example of their use in his
study (1951b) of child health services in eastern Oklahoma. He believes that improved motivation toward action is the most important goal of evaluative studies. He feels experts too often publish excellent reports only to have them forgotten because the evaluative process did not sufficiently involve those who must act on the findings.

In eastern Oklahoma, Hiscock and his associates set up a large study group headed by a prominent business executive and several leading citizens, who acted as area chairmen. Funds were obtained from local health departments, business groups, and voluntary organizations. The area chairmen and their committees were guided by Hiscock's group in the kind of statistics and other information they should seek, following procedures like those outlined in the APHA's Evaluation Schedule (discussed in Section 2). Then Hiscock and his staff aided the members of the study group with writing a report, in such a way that they not only saw what changes were needed but "learned about the agencies concerned with the health of the children, and the methods that are used to coordinate the health efforts of parents, school personnel, health departments, voluntary health agencies and professional associations." On occasion, too, this general procedure has the merit of bringing out the fact that "there is a lack of coordination," Hiscock reports.

Believing that one important part of school health evaluation is study of the health status of the children concerned, Hiscock (1951a) considered the value, for that purpose, of statistical rates of mortality, illness, immunization, and physical defects. He found they were not as suitable as one might hope, and yet he believed that they, together with information on the qualifications of the school's health personnel, should be given consideration as part of the evaluative process. He saw special need for research looking to better "correction rates" than those which are ordinarily available.

**Pennsylvania**

In an evaluation of Pennsylvania's program, the Joint State Government Commission (see Davis, 1955) asked outstanding experts in the State to act as an advisory panel. The panel "developed all statements pertaining to medical practice and medical opinion" for the study. The chief question at issue was the effectiveness of the State law, enacted in 1945, requiring that school children be given biennial medical and dental examinations.

The children's individual records contained information on defects found and corrected, and this material was the first type of statistical data considered by the panel. As we reported earlier in the discussion of correction rates, this approach was not found
fruitful, since the panel could establish "no significant relationship" between the data on defects found and the data on defects corrected.

The other type of statistical information considered was cost data. The total of all State and local expenditures made for school health was estimated as $7,500,000 in the school year 1952-53. The panel stressed that this figure was large in comparison with the $2,100,000 spent annually for the State's six other health and health-related programs for children. Granting the relevance of these figures, it was unfortunate that the report did not attempt to reduce them to a per-child basis, and did not give some emphasis to the fact that the school health program was the only program designed to serve all of the State's children once they reached the school-age range.

From their analysis of the available statistical and administrative information about the program, the panel members judged that the biennial examinations were not warranted, and that "if a portion of the funds now devoted to biennial examinations were spent in other ways, the health level of the school population would be materially improved."

If provision were made for interim examinations of children referred on the basis of screening tests and observations by teacher or nurse, the experts felt that three regularly scheduled examinations in the child's school career should suffice for ordinary medical supervision. (We may note that the panel might have strengthened its hand a little by pointing out that in 1941 Walker and Randolph had recommended the same substitute for Tennessee's biennial examinations, but this was not mentioned.) Only one routine dental examination was needed, "primarily to alleviate a child's fear of exposure to dental treatment," the panel said.

In place of the existing fees of $1.50 for each medical examination and $.75 for each dental examination, the panel urged that physicians and dentists be paid for the time they actually spent on school health work. The physicians should be required to note degrees of severity with respect to any defects that are not entirely matters of presence or absence. It was suggested that some form of the Cornell Medical Index (see Brodman and others, 1951) be used to enable parents to inform the school about any adverse conditions in their children.

These and other recommendations in the report were, basically, a selection of practices regarded as desirable elsewhere, including practices based largely on preceding judgmental evaluations. Except for the fact that some of the recommendations were tailored to fit conditions in Pennsylvania, the panel's findings could have been made without extensive study. This was not, of
course, a reflection on the recommendations. But in relation to future evaluations, the Pennsylvania study is an outstanding example of studies that raise a question as to the amount of time a panel can profitably spend on assembling and analyzing detailed information about a program, if main reliance is to be placed, anyway, on expert judgment.

Evaluations of city programs

Rapeer's study

Except for the 1908 study of New York City's program that will be discussed in Section 4, the first outstanding evaluation of urban school health programs was by Rapeer (1918). He was a school official who, for his doctoral thesis, visited 25 cities having at least a few school physicians and nurses.

Through first-hand study and observation of each city, Rapeer was able to obtain systematic information on a large number of variables thought to be important at that time. He tabulated the data in a variety of ways, bringing out the wide variations that existed in school health practices, and examining possible relationships between the practices and apparent results of the programs.

On the whole, the tabulations showed a marked lack of association among most of the variables considered. One of the two possible exceptions concerned the jurisdictional problem, and the other concerned the value of what Rapeer called the "nurse-alone" plan.

Regarding the jurisdictional problem, Rapeer found indications that in most of the 9 cities where health departments ran the school health work, the programs were less efficient than in the 16 cities where the schools operated the services. This finding, however, was based largely on Rapeer's subjective judgment, and in any case it is worth recalling that health departments were relatively new and inexperienced organizations compared with city school boards.

The information regarding the nurse-alone plan was too sketchy to establish a consistent trend, but it led Rapeer to think that "most of the children needing care and treatment" could be found by nurses, provided they were given some training by, and worked under the supervision of, experienced school physicians.

Rapeer's study of the "defects" found in the various school health programs convinced him that, along with the cases of real consequence, many unimportant conditions were being reported in the schools' case finding work. He declared that it was "better to concentrate all energies on the worst cases than to disgust
parents and family physicians with notices of trivial ailments.” To aid concentration on “worst cases” he worked out what he thought were reasonable figures for frequencies of defects which a school should expect to find in large groups of children. He believed, for example, that no more than 7 percent of the children should be referred for poor visual acuity, while about 2 percent might be expected to have “malnutrition including anemia,” and not over 1 percent should be found with any other specific kind of defect. Other authors before and after Rapeer have offered similar lists of expected frequencies of defects, but there is little indication that the lists have influenced case finding practices perceptibly.

Studies by Ayres and Clark

The study of Cleveland’s program reported by Ayres (1917) and the study of Minneapolis’ program reported by Clark (1921) were important examples of early evaluations of programs in individual cities. Each study utilized a staff of experts who, today, would probably be called administrative analysts. They conducted the evaluations by preparing full descriptions of the existing programs, and adding detailed recommendations based on their judgment and on “standards” of physician-pupil and nurse-pupil ratios current at the time of study.

The reports included considerable praise for the programs under study. Indeed, the more or less outright purpose of the studies was less to change the practices in Cleveland or Minneapolis than to stimulate interest in improving school health programs elsewhere. However, it may have been just as well if the studies did not greatly affect practices in other cities. The recommendations regarding the Minneapolis program, for example, included an elaborate system of forms for record keeping. All experience since publication of the report indicates that the best that could be said today about the forms recommended in the 1921 study is that they are object lessons in the need for continuously reviewing such forms and eliminating parts of them whose use is not in accord with good practice or cannot be supported with evidence.

Phair’s study

In an evaluation of the programs of 12 Ontario cities, Phair (1933) sent staff members of his health department to study the children’s records and to work up special correction rates. The rates covered a 2-year period and were computed in a uniform manner for all of the cities. Phair then compared the rates where the programs did and did not include free treatment, and where the nurses did and did not make home visits routinely. He also com-
pared the rates for programs where the physicians did most of the case finding, with the rates where all of it was done by nurses. Finally, he compared the rates for programs which did and did not encourage parents to attend the examinations.

The working assumption implicit in the study was that historical accidents had led cities which were basically similar to develop different kinds of programs, thus resulting in "natural experiments" that were reasonably well controlled. However, Phair admitted that, especially in respect to the problem of effects of free treatment, there was uncertainty about the initial comparability of the cities.

The correction rates of the various cities indicated: (1) that free treatment was not effective "to the extent formerly presumed"; (2) that nurses' routine visits to the homes were not generally effective, although the visits seemed to facilitate care given to indigent children by charitable agencies; (3) that the "nurse-alone" plan of case finding and follow-up was satisfactory, insofar as the correction rate for that plan was 36 percent as against 29 percent in programs where physicians were responsible for the case finding; and (4) that not much was accomplished by having parents attend the examinations, since the correction rates for cities using and not using this procedure were 30 and 27 percent, respectively. Of all corrections achieved in the cities which did not encourage parents to attend the examinations, "60 percent had been effected without any other effort on the part of the staff than the original notification."

To see whether children's progress in school was improved by treating their defects, Phair sorted the children whose records showed they had had defects into those who did and those who did not receive suitable treatment. He then compared the subsequent school marks and attendance records of the two groups. The results, he said, were "not convincing enough to justify their inclusion" in his report. Lastly, with respect to the control of contagious conditions, he found that unless an epidemic was already under way, so few incipient cases of communicable disease were identified that he was "not able to demonstrate that the school health staff materially aided in the control" of such conditions.

It seems to the reviewer that Phair's study, although apparently not well known, was among the best of the school health evaluations that have relied on judgment and the use of already existing information.

Franzen's 1933 Report

Franzen (1933) summarized, chiefly through an analysis of correlational statistics, the series of evaluative studies which he
conducted for the American Child Health Association between 1929 and 1933. The main instruments of the studies were batteries of pencil-and-paper tests and questionnaires. With them Franzen sought to get at the habits and knowledge of pupils, as well as policies, procedures, personnel, records, and other matters relating to school health work.

Three squads of field workers were trained in using the battery of tests and questionnaires, which the squads then administered in selected schools of 70 different cities. The findings were put through an elaborate statistical mill, the object of which, said Franzen, was “grouping of selected items and interpretation yielding the elements of school health procedures which bring school health results.” The methodology was exploratory, and, at some points, admittedly rather arbitrary or “circular” in nature.

The conclusions stressed the importance of “cooperative determination of procedures” by nurses and teachers; the value of training and supervising teachers in the selection of children for referral to physicians; the importance of having the nurses brief themselves thoroughly from the school records before starting out on home visits; and the need for auxiliary staff to round up the case-history material and screening data before a physician began examining a child.

It is true that Franzen’s correlational analysis lent support to these findings, but it seems equally true that most of the conclusions could have been arrived at through administrative analysis alone.

Woodruff’s Report

The report of the committee chairmained by Woodruff (1941) was an outstanding example of judgmental evaluation of one part of a school health program. The problem was the value of New York City’s special or “open-air” health classes. Woodruff had been among the supervisory physicians who, in 1910–15, were responsible for placing children in the open-air classes, and it is therefore of considerable interest from a historical viewpoint that his report recommended discontinuance of those classes.

Woodruff and the other members of his committee conducted their evaluation along four lines: (1) a review of the literature on the subject; (2) a review of certain studies made by local authorities other than Woodruff’s group; (3) study of data worked up specially for this evaluation; and (4) the narrative reports of selected physicians and educators who were asked to make on-the-spot visits to the classes.

Part (1) indicated that the trend of findings from previous studies showed the special classes were of little or no value, and led Woodruff’s group to believe that some causes of below-par
conditions in children “may lie beyond any assistance that the school can give.”

Part (2) included consideration of three sets of data: (a) the findings of one school physician's re-examinations of 130 children in the open-air classes, indicating that such children could be helped more by other community facilities than by segregating them in the special classes; (b) 45 case histories from Nyswander's concurrent study (reviewed below), which found that “the diagnosis of below par . . . obscures exceedingly complicated conditions—medical, social, and economic”; and (c) findings of a health department survey that had urged a different and greatly restricted basis of selecting the children for special classes.

For part (3), Woodruff's group obtained special information on a sample of 340 of the 5,000 children who were in the city's open-air classes at that time. This information was largely confirmatory of the various materials studied in part (2). Finally, for part (4), consideration was given to the observations submitted by 28 pediatricians and 16 university educators who, independently of each other, had made on-the-spot visits to a large proportion of the classes.

The net result of this study was a strong recommendation that the city dispense with the open-air classes, and that, within the regular classes, a lightened school program be provided for the below-par children.

Woodruff's report went on to suggest specific ways of improving the selection of below-par children and of dealing with them in the regular classes. Otherwise, however, the central recommendation of the report was practically the same as, and was even built upon, findings to the same effect in earlier studies. This raises some doubt as to whether it was necessary to bother with the relatively expensive parts (3) and (4) of the study, when those parts could scarcely prove more than (1) and (2). Perhaps it was hoped that the inclusion of parts (3) and (4) would make the evaluation more convincing. If so, that hope was apparently not well founded, for New York City and several other cities continued to place below-par children in special classes for at least a decade after publication of Woodruff's extensive report.

Nyswander's Astoria study

The study of New York City's program by Nyswander (1942) was the most extensive judgmental evaluation yet conducted. It is usually called the Astoria study, after the Astoria district of Queens where the work was done. The study's more important recommendations, which chiefly concerned teacher observation, teacher-nurse conferences, and parents' attendance at examinations, are termed the Astoria plan.
The study was characterized by common-sense approaches. Some of the results were of outstanding value, especially as regards methods of training school health staff, ways of gaining the cooperation of parents and family physicians, types of record forms, and procedures for transferring the forms from school to school as families moved from one part of the city to another (which about one-third of them did annually).

The study included much trial-and-error development and testing of new techniques. However, no controlled comparisons of alternative procedures were attempted. This being so, the interpretation of Nyswander's findings had to depend on judgment to a large extent, and for that reason we have grouped her work with evaluative studies which have relied mainly upon expert judgment.

With respect to two important questions, Nyswander's uncontrolled tests produced findings of quite uncertain value. Because the study is widely thought to have provided satisfactory evidence concerning those questions, the nature of Nyswander's data on them deserves attention.

One question was the correctness of the very old and respected view (first stated by Alcott, 1840) that teachers can be trained to select or "pre-select" the children for school physicians to examine. To develop evidence on this question, Nyswander's staff trained teachers to keep notes on each child's illnesses, appearance and behavior, while at the same time training nurses to aid the teachers in the task of observing and recording. Also included in the procedure were semi-annual conferences of teacher and nurse at which the teacher's notes and other information available on each child were discussed, and a selection was made of children to be referred to school physicians. The nurse was encouraged to inspect at least some of the children whom the teacher thought needed attention, but the nurse did not inspect, even on a sampling basis, any of the other children.

When the teachers and nurses of one school had been trained in this procedure, they applied it to all of the children in grades 1-8, thus selecting 241 children. The total number of children from whom the 241 cases were selected was not reported. The selected children were then examined by school physicians. They found that 194, or 80 percent of the 241 children, had some type of defect.

The same physicians examined all children in the school's entering class, which comprised 426 children. Of them, 188 children, or 44 percent, were found to have some type of defect.

The difference between the figure 44 percent found for the entering group and the figure 80 percent found for pre-selected children from grades 1-8 was said to "reflect favorably on the
value of the teacher-nurse conference.” It is doubtful, however, that this inference should have been drawn. As Nyswander noted, it was uncertain that the entering children and the children of grades 1–8 were comparable with respect to the proportions of children having defects. Moreover, the figure 80 percent was of questionable meaning for the purpose, because there is good statistical reason to believe that a figure higher than 80 percent would have been found if the teachers and nurses had simply selected fewer children. Conversely, if the teachers and nurses had selected a considerably larger number of children out of the total group in grades 1–8, the physicians might have found that only 50 percent, instead of 80 percent, were correctly selected.

The school in which the work was done had made a special effort to persuade parents to attend the medical examinations. The other important question tackled in the study was whether this effort was instrumental in getting parents to seek appropriate treatment of the defects found.

To test that question Nyswander analyzed the data on the 382 children who, in the work just described, were found to have defects. This group comprised the 188 entering children who had been identified by the physicians alone, and the 194 children of grades 1–8 whom both teachers and physicians had identified.

It was found that treatment of the children’s defects was sought oftener by parents who had attended the examinations than by parents who had not attended them. Nyswander showed this was true by computing certain percentages. The percentages were valid as far as they went, but they did not yield a clear picture of the degree of association between attendance at the examination and whether treatment was sought. To obtain such a picture we need to consider the data in terms of the basic “scatter” or 2 x 2 table of the findings, which was as follows:

<table>
<thead>
<tr>
<th>Parent sought treatment</th>
<th>Parent did not</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent attended examination</td>
<td>329</td>
</tr>
<tr>
<td>Parent did not</td>
<td>62</td>
</tr>
</tbody>
</table>

Despite the fact that neither attendance at examinations nor the seeking of treatment concerns a “graded” variable, it is admissible to generalize these findings with the point correlation coefficient, and in fact there is no better way to see how much
relationship or association exists between the variables concerned. The correlation computes as .07, showing that the association is quite low.

Another adequate, if less general, way of getting at the statistical association is to see how the 180 cases in the "both" category, or the category of parents who both attended the examinations and sought treatment, compares with the number to be expected in that category by chance alone. This is done by taking \( \frac{320 \times 209}{382} \), which yields 175. This figure means that, simply through chance, 175 cases should be expected in the "both" category. Since the 175 expected cases are nearly as many as the 180 cases actually found in that category, it is evident from this approach, too, that the association is low.

The hypothesis under consideration is that more treatment will be sought for the children if special efforts are made to have parents attend the examinations. With Nyswander's approach, her findings would not have been conclusive even if they had shown a high correlation. But if she had found, say, a coefficient of at least .30, the results would have provided some presumption that the hypothesis was correct. The correlation of .07, however, is so low that it could easily arise from a tendency for the examinations to be attended relatively often by the more conscientious parents, i.e., by the parents who would be equally moved to seek treatment if the school merely advised them by phone about their children's needs.

Thus the Astoria study did not prove that case finding was relatively efficient with the teacher-nurse method of selecting children for medical examinations, or that follow-up was improved by special efforts to have parents attend the examinations.

It is equally true that the study did not disprove either of those propositions. Regarding the first of them, Jacobziner (1951) found certain evidence that the selection of children by teacher observation needed to be supplemented by examinations of all the children every few years, but his data were not conclusive. As regards efforts to have parents attend the examinations, the practice might be made more effective than it was found to be by Phair and Nyswander, if attention were concentrated on the parents of entering children and parents who fail to seek treatment for severe defects. It is possible, too, that the case for encouraging attendance of all parents should be made on grounds quite different from those commonly assumed. As Jacobziner and Culbert (1953) have pointed out, the school can make very good use of the parents' visits "to inculcate the need for home safety and accident prevention."

However, the more important questions left open by the
Astoria study and its predecessor studies can hardly be solved by further work using judgmental methods alone. There is need for controlled comparisons that can show what procedures or combinations of procedures may be more effective than others. Even if such comparisons should reveal that the more commonly used procedures are about equally effective, that information would represent a very substantial improvement over our present knowledge regarding case finding and follow-up practices.

Study by Wheatley and others

The last judgmental evaluation that we will consider is the study of New Orleans' program by Wheatley, Harper, Hagan and Swanson (1950). It might be considered a model study of its kind. The first two of the authors were pediatricians with wide experience in school health work and public health programs. The third author was a dentist and Federal official informed on dental health programs, while the fourth was a nurse and State supervisor of school nursing programs. Certain administrative details are of interest because they were virtually a part of the evaluative method. For expenses of the study, the local council of social agencies was able to obtain $2,000 from maternal and child health funds. A part of the money went for the services of Elizabeth McFetridge, who was skilled in preparing reports of studies of this nature. The four experts gave her draft material as rapidly as it was developed, and she was present at the frequent meetings held by the experts to work out their recommendations. This procedure enabled McFetridge to be writing parts of the report while the study was going on, and a nearly final draft was ready at the end of the experts' 10-day stay in the city.

The experts spent a substantial part of their time observing the routine procedures actually used in the school health program. They also interviewed the staff members, supervisors, officials of both the public and the parochial schools, and representatives of the local medical, dental, and hospital groups. A relatively small amount of time was spent in studying records and accumulated information, except as regards data on costs.

A new budget was outlined which, through a joint education-and-health authority, would provide for a moderate increase in the amount of physician-hours of service received by the children, and a more substantial increase in the dentist-hours of service. The dental program in the public schools was in need of special attention, the experts said, since there was evidence that the public school children were receiving less dental care than the children in parochial schools.

The report recommended that the school physicians specify
"which defects should have priority for correction," and that they ignore some defects whose correction would be "desirable but unessential." It was urged that more effort be made to have family physicians provide periodic examinations, in order to save "public funds which could be better spent in health services for children who cannot afford them otherwise." Otherwise, the Astoria plan of selecting children for special examinations was recommended. It may be wondered whether that plan would have been recommended unqualifiedly if the panel of experts had included a school physician from a program which had used both the Astoria plan and a plan like the one employed in Philadelphia.

Comment

From the studies reviewed in this Section it is obvious that judgmental evaluation is liable to the charge that "its results depend on who the experts are." As judgmental studies are ordinarily conducted, the findings could usually be predicted in advance by a neutral observer. He would not necessarily need to know what programs were being evaluated. If given the names of the experts chosen for the work, he could often forecast the recommendations by simply finding out what experience the experts have had and what views they held before beginning the studies.

This does not mean that judgmental studies have not made contributions, or that the general method should be discarded. It does mean that, wherever feasible, the method should be supplemented with other procedures, such as re-examining the children and conducting controlled tests.

Even though judgmental evaluation will always be liable in some degree to criticism like that noted above, studies utilizing expert opinion can and should be made much less vulnerable in the future than they have been in the past.

Whether the judgmental method is used alone or in combination with other procedures, there should be several experts, and, if possible, they should represent the various professional groups involved in school health services. However, the sheer number of experts and the professions they represent are not as important as arranging for the inclusion of experts with experience in programs of different kinds, especially as regards case finding and follow-up practices.

No other step can do as much to moderate the difficulties inherent in the judgmental method as insuring that there is representation, not only of important professions, but of important practices as well. Some of the groups that can help with the problem of arranging suitable representation of practices are the American Academy of Pediatrics, the American Association for Health, Physical Education and Recreation, the American Dental
Association, the American Medical Association, the American Nurses Association, the American Public Health Association, and the American School Health Association.

It is likely that experts with widely different kinds of experience will have to come greater distances, on the average, than would a group of experts with similar backgrounds. The more broadly based panels are, however, well worth the additional travel funds which they may require. If some other feature of the study has to be restricted to make up the difference in cost, it might be desirable to limit the stay of the experts, or the amount of money that is made available for working up special data about the program.

A full report is desirable, but the study plans might call for a processed rather than a printed report, if the difference in cost will permit the selection of experts representing a wider range of programs than would be feasible otherwise. Later, a relatively brief form of the report, containing practically everything that will be of interest to persons not directly concerned with the given program, can be published in one or more of the professional journals. Some of the most important school health evaluations have been reported only in the form of expensive "separates" and are not generally accessible except through the time-consuming process of inter-library loans. The editors of professional journals have often helped by seeing that these reports received appropriate reviews, but many editors would prefer that the original authors re-wrote and submitted their studies in forms suitable for periodical publication.

Whatever type of panel is set up, its members should be discouraged from reporting only the recommendations on which general agreement is reached. It goes without saying that disagreements among the experts should not be dramatized in the report, but the factual reporting of significant disagreements is a very different matter. Few facts or principles are as yet firmly established in the field of school health, and it would be idle for experts to give the appearance of thinking otherwise. Both the interests of the programs being evaluated and the interests of other programs will be served best if legitimate differences of opinion are reported as such, perhaps with mention of factors in the experiences of the experts that may help others to understand and interpret the views presented in the report.

Although there should be no limit to questions upon which the members of the panel should pass judgment if they wish to do so, it may be in order to note some kinds of problems for which judgmental methods seem particularly appropriate.

One is the jurisdictional problem, involving questions of how the administrative and financial responsibilities for the pro-
gram under consideration should be distributed between health and education authorities. It is conceivable that, a rather long time hence, optimum solutions of this problem might be aided by controlled comparisons. For the present and immediate future, however, communities could well seek advice on this problem from outside experts, whose judgments should be guided largely by the history, present resources, and future expectations of each community concerned.

Similar factors should guide a panel of experts in judging how much of a program's budget should be expended for part-time personnel as against full-time or supervisory personnel, and in recommending the types of diagnostic aid or treatment by specialists or clinics that are most suitable in particular programs.

In these matters as in others, the panel members could well be invited by the local authorities to distinguish between changes which should be made within a year or two, and changes which, although planned immediately, would not become effective for several years, or perhaps only when certain related community developments occurred.

Finally, the panel might be able to make certain recommendations regarding the program's case finding and follow-up procedures. Since, however, these procedures lend themselves well to experimental comparisons, and since there is cause for hope that such comparisons will soon improve our knowledge of what is effective, the panel's recommendations should be qualified to allow for the adoption of whatever new procedures or combinations of procedures may be found optimum in future tests.
SAMPLING AND RE-EXAMINING THE CHILDREN

THE METHOD DISCUSSED in this Section was first used in a study conducted nearly half a century ago by the New York City Department of Health (1908). The results were extensively reviewed a few years later in the text on school health by Cornell (1912). After relating the findings of the New York City study to his own experience as director of Philadelphia's program, Cornell epitomized the re-examination method by saying that if school health authorities wish to find out what is going on in their programs, they should "look at the children, not at the records."

Major studies

The 1908 study

About 1905 New York City had begun to employ part-time physicians to inspect all school children for physical defects. Parents of children found to need care were sent reply postcards to take to their family physicians, who were urged to use the cards for advising the school about their findings and any treatment given the children.

The investigators saw that there was need for evaluating not only the postcard follow-up but also the case finding on which the follow-up efforts were based. In this respect the investigators were well ahead of their time, and their approach deserves consideration in some detail.

The investigators first compared the reports of different
school physicians working in similar schools and found surprisingly large differences in the percentages of children reported as having defects. It was evident that the physicians had different ideas about how severe the children's adverse conditions should be in order to warrant reporting as "defects." However, the investigators realized that this fact was not the only kind of information needed to evaluate the case finding. For, even if two physicians working in similar schools had reported practically the same percentages of children with defects, this would not necessarily mean that the physicians were selecting the same kinds of cases, and it would not indicate whether the regular physicians were selecting the same cases as better qualified physicians would select if they examined the children in the same schools.

An answer to the latter question was sought by having a special group of physicians re-examine a sample of children who, only a short time before, had been examined by the regular school physician. The sample consisted of 20 children in each of 15 schools, making a total of 300 children.

With respect to each defect, the investigators reported the number of cases found by the regular examiners, the number found by the special examiners, and the number that were in common (i.e., the cases found by both the regular and the special examiners). It is worth stressing that, in using these numbers to report the findings, the investigators were employing the most economical way of presenting all of the data needed to set up the "2 x 2 scatter" for each defect.

With respect to "defective vision," for example, the three numbers were 72, 101, and 51, in the order stated above. By subtracting these figures from 300, the following scatter is obtained:

<table>
<thead>
<tr>
<th>Re-examinations by special physicians</th>
<th>Children with defect</th>
<th>Children without defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examinations by regular physicians</td>
<td>Children with defect</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td>Children without defect</td>
<td>228</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>51</td>
<td>199</td>
</tr>
</tbody>
</table>

The correlation coefficient computes as .44, indicating that so far as visual problems were concerned, the examining was
either unreliable or the work of the regular physicians did not always cover the same visual conditions as were covered in the examinations by the special physicians.

Moreover, comparison of the numbers 72 and 101 indicated that, wholly aside from the question of how many cases were in common, the regular and special examiners were using somewhat different levels of severity to distinguish between "defective" and "normal" vision. The special examiners evidently thought that more cases of moderate visual deficiency should have been reported as "defects" by the regular examiners.

For most of the other defects the difference happened to be in the other direction. The data showed, for example, that the special examiners thought too many cases of "malnutrition" and "pulmonary disease" were being reported by the regular physicians.

The data indicated, however, that there was no general relationship between the problem of whether the level of severity was too high or too low, and the more important problem of how well the results of the original and special examiners corresponded, in the sense of the overall association or correlation between the two sets of findings.

The investigators did not attempt to prove, statistically, that the two problems were independent of each other. But the whole approach taken in the study shows that the investigators realized the necessity of taking into account both the level of severity and the extent of association with a criterion. Probably no better statement has been made about the role of case finding in school health services than the concluding statement of the New York City investigators. They said the examinations should neither "alarm parents unnecessarily" by reporting defects of no great importance, nor "fail to find defects that are actually present." The latter phrase reflected the investigators' awareness that some form of re-examination procedure and an association table appropriate to the findings were essential for evaluative purposes.

**Franzen's "pathway" study**

New York City was also the locale of several other studies using re-examination methods, one of the best known being the "pathway" study by Franzen (1934). He was not specially concerned with evaluating New York City's program as such, but wished to make use of the program to learn the reasons why children's defects "often went unattended" despite extensive efforts by the school to secure care. The study dealt with a limited number of defects for which special examining methods had already been
developed in connection with the work summarized in Franzen's 1933 report (see Section 3).

The study's general approach is well illustrated by citing the methods and results of the work done on visual defect. The acuity of 5,132 children in the fifth and sixth grades was tested, using a special Snellen chart. To distinguish the "defect" cases from the "normal" children, use was made of the scoring procedure and "cut-off" score which had been found suitable in the earlier studies.

The school health staff had already tested the same children with the ordinary Snellen chart and had applied a conventional scoring scheme and cut-off score to distinguish defect cases. Below is the scatter for the two sets of findings:

<table>
<thead>
<tr>
<th>Special Snellen (Franzen)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Children with defect</td>
<td>5,132</td>
</tr>
<tr>
<td>Children without defect</td>
<td>712</td>
</tr>
<tr>
<td>Conventional Snellen</td>
<td></td>
</tr>
<tr>
<td>Children with defect</td>
<td>637</td>
</tr>
<tr>
<td>Children without defect</td>
<td>4,495</td>
</tr>
</tbody>
</table>

It was clear that the conventional Snellen's cut-off had been set to distinguish nearly as many of the children as the cut-off used with the special Snellen, so the two procedures were roughly comparable so far as questions regarding level of severity were concerned. Moreover, there was good agreement between the two procedures as regards their overall association, since the correlation for the scatter was .79. (For most of the other defects covered in the study, there was less agreement in respect to severity level, correlation, or both.)

The study gave no further consideration to the 4,420 children passing the special Snellen test even though some of those children were wearing glasses. Instead, attention was concentrated on the 712 children failing the special Snellen. In line with the procedure to be suggested later we may note incidentally that, as his basic study group, Franzen could well have used the 549 cases identified by both tests, perhaps giving attention also to certain cases in both "odd" categories inside the scatter (i.e., the 163 cases identified by the conventional Snellen alone and the 88 cases identified by the special Snellen alone).
Among the 712 children who had failed the special Snellen tests, it was found that the majority, comprising 370 children, already had glasses. Although the study gave most attention to the remaining group who lacked glasses, the majority group who were wearing glasses were not ignored. Their parents were asked whether the school's follow-up work or the independent recommendation of a private physician had stimulated them to seek eye care for their children. Two-thirds of the parents said the school's efforts were responsible. Moreover, a sample of 28 of the 370 children with glasses were examined by an eye specialist at the city health department. He found that even though most of the children were receiving substantial benefit from the glasses they were wearing, glasses of a more or less different kind were needed by all but 2 of the 28 children. Franzen pointed out that, in some part, this finding was due to real changes which had occurred in the children's eyes after their glasses were prescribed. He therefore recommended that schools routinely test all children who were wearing glasses as well as the children who were not.

As the first step in studying the 342 children who had no glasses but needed them according to the special Snellen test, a sample of 100 were tested by the health department eye specialist. He reported that 98 of the 100 children definitely needed glasses, and this was taken as indicating that the validity of the special Snellen test was high. We may remark that, although this check indicated the test's validity was at least substantial, the evidence would have been more complete if the eye specialist had been asked to examine equal numbers of the children failing and passing the special Snellen test and had not been told which children were which.

It was assumed, for study purposes, that glasses were actually needed by all 342 of the children. To find out what had happened to prevent appropriate care for this group, Franzen examined the children's records and interviewed the children's teachers and parents.

For 67 of the children not enough information could be obtained to pass judgment. For the remaining 275 cases it was found that the cause of the trouble could be attributed to one or another of 9 rather complex categories or "steps" in an arbitrarily conceived "pathway to correction."

In what follows, Franzen's 9 categories have been reduced to 5 categories for purposes of summarizing the results, which were reported as percentages of the 275 children who lacked care.

The lack of care seemed attributable to some inadequacy of the school's follow-up work in 56 percent of the cases, while an additional 23 percent of the cases were due to inadequate case finding. Thus the school appeared to be at fault over three-quarters
of the time. In another 11 percent of the cases, family physicians had doubted the need for care. In 8 percent the parents could not pay for glasses, and for the remaining 2 percent of the cases, glasses were in process of being obtained.

Analogous sets of percentages were reported for various other defects covered in the study. The percentages varied markedly from one type of defect to another. The data are of much interest provided it is remembered that only untreated cases were used in computing each set of percentages, and that the cases with which the school had been most successful were left out of account.

This was admissible inasmuch as the study's main purpose was not to evaluate the given program but to learn why defects needing attention had not received it. Nevertheless, the results would have appeared almost as striking, and they would certainly have been less confusing, if the treated as well as the untreated cases had been taken into account in any percentages or rates computed to generalize the findings.

**Study of Molner and Blanchard**

In their report on Detroit's program, Molner and Blanchard (1941) gave a brief but significant account of the use of re-examinations to evaluate the work of private physicians. The latter group had been encouraged to take over the bulk of the examining required in the school health program, and critics were asking about the quality of the examinations.

A sample of the children examined by the private physicians were re-examined by physicians employed by the schools, and the findings were compared with the reports of the private physicians. The details of the re-examinations and their results are not given in the report, which states simply that "careful study revealed that certain inadequacies did exist."

The findings were presented to the leaders and appropriate committees of the local medical and pediatric societies. Articles on the problem were published in the bulletin of the medical society, and discussions were held with groups of physicians.

As a result the private physicians' reports to the school on the findings of their examinations became more complete, and the examining itself was "much better." Presumably this was ascertained by further application of the re-examination procedure, although the report of the study does not so state.

**Wilzbach's study**

In one of the most direct applications of the re-examination procedure yet made, Wilzbach (1944) evaluated the health status of the junior and senior high school students in Cincinnati. The study was regarded, in part, as a test of the effectiveness of
Cincinnati's school health program. That program had included medical examinations for students entering the first, third, fifth and ninth grades, and the examining program had been supplemented with substantial efforts by the schools to see that "the needed corrections were secured."

The 5,600 students comprising the junior and senior classes were re-examined by 18 staff physicians and 43 staff nurses of the Cincinnati health department. Since that department had long operated the city's school health program, it would seem likely that some of the physicians who conducted the re-examinations had participated in the earlier examinations of the children, although the report of the study did not mention the extent to which that was true. The report indicated that private physicians had performed some of the original examinations, but without noting how often this was the case.

The re-examinations were conducted "much as the examinations were made in the large induction centers for selectees" at that time.

One object of the study was improvement of the students' fitness, and the re-examining was accompanied by a campaign to get the students to obtain whatever treatment they needed.

Among the 5,600 students, the re-examinations showed approximately 2,300 defects which were uncorrected, or, less often, inadequately corrected. (The proportion of students having at least one such defect was not mentioned, but, considering the large number of defects found, we may estimate that this proportion was over one-fourth of the students.) Some of the larger categories of uncorrected defects found, and their approximate prevalence rates among the students, were as follows: visual defects 6 percent, diseased tonsils 2⅓ percent, impaired hearing 2 percent, and heart diseases 1⅓ percent.

Wilzbach did not consider that these findings were very serious, and, on the relatively favorable side, he stressed that among all 5,600 students, only 12 showed positive reactions to serologic syphilis tests and only 2 had active tuberculosis. He did not attempt to relate the study's findings to particular aspects of the school health program, but his report as a whole indicated that he felt the program had accomplished a good deal for the students' health status.

Comments on Wilzbach's study

On a more or less intuitive basis Wilzbach probably had good reason to make this judgment, but there are two considerations which should be noted here for their relevance to future evaluative work of this kind.

As Lyon stressed in his 1945 study of the Selective Service
findings (see Section 1), many young people do not go to high school, and the health status of those who do is probably better, age for age, than the health status of those who do not. In consequence it is possible that the findings from re-examinations of high school students tend to make the health services provided in elementary school seem more effective than they actually were.

Independently of bias which might arise in that way, there is a second factor which is likely to bias matters in the opposite direction if it is not considered, but which, if recognized, might be used to some advantage in studies of high school students. We refer to the fact that urban families move frequently, with the result that many children attend elementary school in one city and high school in another. This in turn means that the high school students in any given city are likely to include, in roughly equal proportions, young people who attended elementary schools elsewhere and those who graduated from the local elementary schools. Consequently, even if high school students were not a selected group in respect to general health status, a given city's high school population is not, as a whole, very representative of the individuals who received health service in that city's elementary schools.

We should add that this problem differs only in degree from the problem which arises when re-examination is used in evaluative studies of elementary school children. By the time the re-examining is organized, some of the children exposed to the program under consideration have gone elsewhere, and their places have been taken by children who may or may not have received some health service but who, in any case, have received little or no service under the given program.

If, at the start of the re-examining, there are grounds for thinking that the health status of the children who went elsewhere was not typical of the children who remained, effort should be made to see how true that might be, perhaps through correspondence with the schools to which the children went. Ordinarily, it will not be unreasonable to disregard the children who went elsewhere, and thus to assume that the remaining students are typical of the whole elementary group exposed to the program. It is, however, important to differentiate between the children who recently entered the school and the children who have been attending for some time. In elementary schools the newly entering children are usually a small group, and it will ordinarily be best to exclude them from consideration in the study. But if administrative or other special circumstances require their inclusion, the findings on them should be reported separately from the findings on the other children.

When, as in Wilzbach's study, junior and senior high school
students are studied, the youth who are relative newcomers are
a large enough group to be worth some special attention, and it
may be feasible to classify them into at least two groups. Among
the juniors and seniors as a whole, the study could then distin-
guish, for example, the students who had been exposed to the
local school health program 6 or more years, those exposed 3 to
5 years, and those exposed 2 years or less.

It would not be sound to assume that the three groups of
students were fully comparable in “original” health status, so
to speak, but one could reasonably assume they were comparable
enough so that, if the local program were distinctly superior to
other programs, that fact should be evident in the findings of
re-examinations.

If possible, of course, the re-examining should be done
by physicians who have not been associated with the program
under consideration, and indications of the elementary schools
attended by the students should be removed from the students’
medical histories furnished to the examiners.

Studies by Wallace’s group

In a continuing series of studies, Wallace and others
(1954a, 1954b, and 1955) have been using re-examination pro-
cedure to evaluate New York City’s special classes of children
with visual, cardiac, orthopedic and other handicaps.

A sample of the children in each type of class is usually
designated. The children’s histories are assembled, and, as deemed
necessary, the children are re-examined by experts. Some of the
experts are, and some are not, associated with the original place-
ments of the children in the classes.

For example, in the work which Wallace’s group (1954a)
reported on visually handicapped children, a consulting ophthal-
mologist examined a sample of 182 children comprising approxi-
mately 15 percent of the enrollment in the sight-conservation
and Braille classes. He and the chief of the bureau of handicapped
children then judged the appropriateness of each child’s placement,
using the criteria for placement which the bureau had set down
officially a few years earlier. The placements of about one-fourth
of the children in the sample were judged as inappropriate. Recom-
mendations were made both as regards individual children and
as regards future procedures for making and reviewing place-
ments.

The chief aim of Wallace and her associates has been
the securing of information that will be of immediate practical
value for the programs concerned, and from that viewpoint refine-
ments of the re-examining process are not of great consequence.
It is nevertheless worth noting that some of the samples of
children could well be judged, independently, by two different but professionally comparable experts—or, even better, by two similar pairs of experts. This would tend to improve the final judgments arrived at, and the scatter of the two sets of findings would give some idea as to how often inappropriate placements arise solely from the unreliability of the judging process.

**Yankauer’s studies**

The Astoria plan of school health services was evaluated by Yankauer in two studies (1947 and 1951) utilizing re-examination procedures.

In the first study, which Yankauer conducted after the Astoria plan had operated in New York City for 6 years, he examined the sixth-grade children in 2 schools. The district in which the schools were located was a low-income area, and “adequate free or low-cost health, medical, and dental services are easily accessible outside of the school. In addition, considerable mass health education has been carried on in this district, and it has served as a citywide training center for school physicians.”

Yankauer’s purpose was not to compare the schools’ case finding with what he could find in examinations conducted independently of the school records. He aimed, first, to identify as many of the children’s defects as possible, and for this purpose he combined what he could learn by examining the children himself with information from the school records and from the parents, school physicians, and family physicians concerned. Having ascertained the children’s defects in this way, his chief object was to see how many of the defects had, and how many of them had not, already been found by the schools.

From his own examinations and the other sources, it appeared that there were 77 medical defects of some consequence among the 114 children whose cases could be studied thoroughly. Administrative and other difficulties having little or no relationship to the problem at issue did not permit full study of the other 35 sixth-graders in the two schools.

Of the 77 defects, the schools had already discovered 53, while the other 24 were unknown to the schools. Yankauer noted, in effect, that somewhat fewer than 53 defects might have been discovered by these schools if they had not been involved in the training program for school physicians, but he believed the study as a whole indicated that the Astoria case finding procedures had “functioned satisfactorily in these two schools.”

In view of the inexpensive and readily accessible medical services in the area concerned, Yankauer recognized that the study could not yield conclusive data regarding the effectiveness of the Astoria plan’s follow-up procedures. He nevertheless went
on to cross-tabulate the defects which had and had not been found by the schools, against the program's success or failure in bringing the defects under care. Below is the resulting 2 x 2 table:

<table>
<thead>
<tr>
<th></th>
<th>Defects found by schools</th>
<th>Defects not found by schools</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defects brought under care</td>
<td>77</td>
<td>52</td>
</tr>
<tr>
<td>Defects not brought under care</td>
<td>53</td>
<td>24</td>
</tr>
</tbody>
</table>

The table shows that the schools were successful in securing care for most of the defects which the program's case finding had revealed. Conversely, and perhaps naturally enough, most of the defects which the school had not found were not under care. The overall extent to which these generalizations hold is best indicated by the point correlation coefficient, which is .67.

In his second (1951) study, which was concerned only with case finding effectiveness, Yankauer reported upon 334 children of grades 2–8 in nine schools. These schools, too, were in low income areas where inexpensive care facilities were available, but the schools had not been used as training centers for school physicians. The study also differed from the previous one in that, although Yankauer directed the re-examining, most of it was done by staff physicians of two voluntary agencies. Again the findings were combined with the information available in the school records. It is of incidental interest that 15 defects which the school had discovered were missed by the re-examiners. These defects were included in the total of 247 defects that were found among the 334 children studied.

Unlike the report of the first study, the report of this study was not altogether clear about the ascertainment of the children who were examined, and despite Yankauer's belief that there was no "bias of selection which influenced the results," the 334 children may not have been representative of all of the children in grades 2–8 of the nine schools.

The main finding was that, among all 247 defects identified, only 22 had not been discovered by the schools. Provided the 334 children were a reasonable sample, this result was indeed good evidence of the Astoria plan's case finding efficiency, at least as regards school children in low income areas where most of the entrance examinations—on which the Astoria plan partly depends—are given by school physicians.
Yankauer noted the possibility that in higher income areas, where most of the entrance examinations were given by private physicians, the initial case finding might not have been good enough to permit the plan to function effectively. Jacobziner (1951) made a similar point, and said his experience indicated that the Astoria plan should be supplemented with one or more examinations, by school physicians, of the children examined by private physicians.

If the point which both authors made concerning private physicians' examinations is of consequence, it would seem desirable, in future studies using re-examination methods, to include separate tabulations according as the main examinations of the children are by private physicians or by school physicians.

**Study by Yankauer and Lawrence**

Of the various studies involving re-examination procedures, the most outstanding investigation so far published, and the one warranting the most detailed attention here, was the study conducted by Yankauer and Lawrence (1951) in Rochester, N. Y.

That city's program calls for examinations of all entering children, either by private or school physicians, as an entrance requirement. By re-examining an appropriate sample of first grade children who had been examined when they entered kindergarten, the investigators wished to learn whether enough new defects had accumulated during the one-year period to justify annual examining. Although Rochester and other large cities were exempted from New York State's law requiring annual examinations, the question posed by Yankauer and Lawrence was of obvious interest in connection with that law.

The re-examinations were performed by co-author Lawrence, who is a pediatrician. Her examining was conducted as independently as possible of the school records. She either interviewed the parents or had them complete a specially designed questionnaire on the child's medical history. Only those defects whose identification “required the professional time and skill of a physician” were counted in the study. The conditions to be considered “defects” were carefully defined, and the definitions constituted as close an approach to specifications for a “standardized” school health examination as the writer has noted in recent literature.

A sample of first-grade children was drawn by sorting the city's 70 elementary schools into three groups according to economic level, and then selecting 13 schools in such a way that the schools chosen from each economic grouping had about 15 percent of the first-graders in that grouping. It was not clear that random selection from the first grade population was insured in
this way, but the procedure gave a sample of 1,086 first-graders with proportionate representation from each economic grouping.

Lawrence was able to examine all but 30 of the 1,086 children. Of the 1,056 whom she examined, 59 children were later found to have had no entrance examination despite the school regulation. Chief interest therefore centered on the remaining group of 997 children who had been examined twice.

For 516 of the 997 children, the earlier examinations had been given by private physicians, while 370 had been examined by school physicians. There remained a group of 111 children whose school records showed that physicians had examined them, but did not clearly indicate whether the examiners were private or school physicians.

The report of the study does not give separate tabulations for the 516 children examined by private physicians and the 370 examined by school physicians. Nor does the report give, for the total group of 997 children, the full 2x2 scatter or cross tabulation of Lawrence's findings against the findings of the entrance examinations. Below, however, is the form of the scatter representing such a cross tabulation, and it shows, in the appropriate categories, the figures available from the published report.

<table>
<thead>
<tr>
<th>Lawrence's re-examinations</th>
<th>997</th>
<th>210</th>
<th>787</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children with defect</td>
<td>154</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Children without defect</td>
<td>56</td>
<td>?</td>
<td>?</td>
</tr>
</tbody>
</table>

The authors felt that the ratio of the 154 cases found by the school to the 56 cases not found by the school indicated that the second examination was not worth while. As will be noted later, there is reason to question the use of this ratio, even though the authors' conclusion was not unreasonable.

Yankauer and Lawrence went on to consider how often care was being received by the group of 56 children whose defects had not been found by the school. To learn this, the authors consulted the parents and, as necessary, the private physicians
or clinics concerned. It appeared that, among the 56 children, 35 were receiving adequate care and 21 were not. Further study of the 21 children who were not under care suggested that all but one of them could have been identified by a classroom teacher. This was seen as additional evidence that examinations conducted one year after the entrance examinations were “valueless from a case finding viewpoint.”

There is some question as to whether most of the 21 cases would actually have been caught by teacher observation even in schools where teachers are specially trained in that case finding procedure. And, as regards the 56 children whose cases were not known to the school, it is a question whether their care status—important though it was—should have been considered in relation to case finding as much as in relation to follow-up problems.

While agreeing with the authors’ conclusion that complete examinations given one year after the entrance examinations are not worth while, we should note that this conclusion left open the possibility that rapid and inexpensive examinations might be desirable on an annual basis. Rapid examinations could be expected to detect most of the cases like the 21 found to need care in this study, and, as regards the children whose defects were already known, the findings of rapid examinations could often be used to emphasize the need for care in those cases where the parents had not yet acted.

Moreover, if moderately complete examinations by school physicians were routinely scheduled one year after the entrance examinations, and if private physicians in the community were advised that this would be regular procedure, it might have the effect of a “quality control” procedure in helping to insure adequate examinations by private physicians. In the long run a continuous check of this kind might be simpler, less expensive, and more satisfactory all around than conducting re-examinations and an occasional campaign in the manner reported by Molner and Blanchard.

Adaptations for routine use

The eight studies reviewed above show that re-examination of the children concerned is an important and useful evaluative method. It remains to consider ways of making the method applicable to a wide range of school health programs.

We need only go a step further in the direction set by the work already done, particularly in the study by Yankauer and Lawrence. The adaptations of their methods sketched below are not offered as ideal solutions of the many technical problems which re-examining raises. The suggested procedures are never-
theless believed adequate for answering questions as to (a) how well a given program is finding the “defects that are actually present,” to use the phrase of New York City’s 1908 study; and (b) how well the given program is succeeding in securing the care that is requisite. These are questions which are frequently—and fairly enough—asked about the case finding and follow-up work of school health programs. The questions do not cover everything we would like to know about a particular program, but answers to them should provide basic information or “first facts” regarding any program’s effectiveness.

**Selection of re-examiners**

As an evaluative method, re-examination may be thought of as midway between judgmental evaluation and the experimentally controlled work to be considered in the next Section. Insofar as the results of the re-examination approach are reproducible, the method may be considered objective. Yet a good deal depends on the experts who are chosen to do the re-examining, and for that reason the method necessarily involves a subjective factor which enters the picture in the process of choosing physicians to conduct the study. From this viewpoint the selection of the experts to do the re-examining is quite as important as the selection of panel members in judgmental evaluation.

Although no flat rules can be laid down, it would seem very desirable that the content of the re-examinations include two quite different elements. One element would be the more recent and more carefully established research information available concerning children in the school-age range. Coverage of this could usually be insured by asking a professor of maternal and child health to be an examiner, or to recommend a pediatrician having knowledge and skills comparable to his own. The other important element to be considered is the type of medical experience which can only be gained from extensive service as a school physician. This qualification could well be sought in a physician who has been a supervisor in a program that has included at least two routine examinations of all children in the school system concerned.

If funds for an evaluative study are short, and if one physician having both qualifications is available, he might be asked to do all of the re-examining. But from the standpoints of both the reliability and the validity of the re-examination findings, it will ordinarily be best to seek the above qualifications in two different physicians.

We should digress to urge that, concurrently with future use of re-examining in evaluative studies, special studies will be made of the reliability of the work of a pair of examiners with
qualifications like those noted above. Such studies could be conducted by the American Academy of Pediatrics, the American Public Health Association, or the American School Health Association, perhaps in collaboration with a school of public health. The design of such work would resemble the procedure to be discussed below, except that two comparable pairs of physicians would give independent examinations to the same group of children. There should be alternation of the pair of physicians who give the first examination, and careful accounting should be kept of the time spent. The object would be to discover how well the examining done by one pair of specially qualified physicians correlates with that done by an equally qualified pair.

The magnitude of that correlation is important because it represents an upper limit of the correlations that can be expected in ordinary evaluative studies where one pair of experts is used to check a program's case finding efficiency. Suppose, for example, that the special studies have shown the correlation between the findings of two pairs of experts is approximately .75. Suppose also that ordinary evaluations of a number of different kinds of school health programs have shown that the correlation of a pair of experts' findings and the schools' previous findings ranges between .50 and .70 from one study to another. We would not measure these correlations against 1.00, but against .75, and we would probably consider that the correlations .50, .60, and .70 represented fair, good, and very good case finding, respectively.

Routine re-examining and sampling procedures

In ordinary evaluative studies using the re-examination method, the experts should be assisted by technicians of their own selection. In line with the general procedure used by Yankauer and Lawrence, the staff of the program being evaluated should cooperate with the technicians in providing whatever new or old laboratory findings and screening scores the expert re-examiners may require, so that the latter can spend practically all of their time on work that really requires their professional skills.

It is doubtful whether the experts need to, or should, interview more than a few of the parents in connection with the re-examinations. If such interviewing is not limited, it tends to make the re-examiners' findings less independent than they should be of the school's previous findings. In addition, the interviewing of parents often takes time which the re-examiners can spend as well or better on other diagnostic procedures. Perhaps the use of interviews in occasional cases, and the use of a questionnaire for the great majority of the cases, would be the best compromise. The choice of a suitable questionnaire should of course be made by the re-examiners, but we may mention that, in addition to
considering the forms developed by Singer-Brooks (1952) and Yankauer and Lawrence, some attention could well be given to selecting and adapting items from the latest available edition of the Cornell Medical Index. It may be recalled that use of this instrument, which has the merit that the validity of its materials has received substantial checking (see Brodman and others, 1949–51), was suggested in the report of the Pennsylvania study (Davis, 1955).

As regards the size of the sample of children to be re-examined, it may be said that an evaluative study would still be worth conducting if the experts had time to cover only some 200 children. However, that figure should be regarded as a minimum sample size.

Except where the school system's enrollment is sufficiently small that study of a 100 percent sample may be contemplated, the decision as to the maximum sample size that is desirable should not be affected by the school system's total enrollment. Instead, the determining factors should be the numbers and kinds of detailed breakdowns believed important enough to warrant special attention. Examples of breakdowns of possible interest are classifications of the children according as they have been examined chiefly by private physicians or by school physicians; according to the children's ages or the economic levels of their districts; and according to the number of years the children have been under the program. Before deciding that a particular breakdown is important enough to warrant special attention and the extra cost which that may involve, the investigators should consider whether that breakdown is likely to yield differences whose direction and magnitude are unknown, or only differences which could be predicted well enough on the basis of previous knowledge.

The kinds of breakdowns desired may also affect the type of sampling that is used, but if the breakdowns are not of major interest it will be admissible to take every nth child on the enrollment lists, where n is the total enrollment divided by the size of sample desired. If the enrollment lists are not in good order for the purpose, or if there is any other reason to think this procedure would not readily produce a truly random or "probability" sample, an expert on sampling should be consulted.

In case there is special interest in the children who only recently came to the school, and have therefore had little exposure to the program, it may be desirable to use a relatively high sampling ratio with them in order to bring a sufficient number of them into the sample. Ordinarily, however, it would be best to have the school authorities indicate on the enrollment lists who these children are, and to omit them from the sample by skipping over their names in the sampling process. If this is done, the
number of remaining children, not the total enrollment, should of course be used in obtaining \( n \).

When the expert physicians have completed their re-examinations of the children in the sample, each child should be classified as "with defect" or "without defect" according to the findings of the new examinations. The experts should then study the school's records of the same children, again classifying each child as either with or without defect, but only in accordance with what the school's records indicate. In each classification, the "with defect" group should include the children who have more than one defect apiece. So far as possible, the corrected as well as the uncorrected cases should be included in both classifications.

Cases that are hard to decide may arise with respect to either classification. Final decisions should of course rest with the expert re-examiners. Provided essential independence is maintained, however, there is no reason why the process of deciding doubtful cases should not include some consultation with the staff members of the program being evaluated.

The work that remains after completion of the tasks of examining and classifying the children may be outlined in three steps. The first two steps concern evaluation of the case finding and the third concerns evaluation of the follow-up work. Considering the fact that sound case finding is essential for effective follow-up work, the emphasis given here to case finding problems is believed to be in keeping with their importance.

**Case finding evaluation**

**Step 1.** As regards the program's case finding efficiency, the basic data consist of the scatter or cross tabulation of the two with-and-without defect classifications, as already indicated, for example, in connection with New York City's 1908 study and the findings of Yankauer and Lawrence. In addition to the scatter for the total sample, a separate scatter should be tabulated for each sub-group of children that is of special interest, although this should be attempted only for the sub-groups which have been sampled in adequate numbers.

To determine case finding efficiency, it is necessary to use some type of coefficient which takes account of the overall association represented in the scatter. The point correlation coefficient (which is also called "phi") is not the only coefficient that can be used, but is probably the most practicable one.\(^4\) The particular

\(^4\) Another index, called the tetrachoric correlation coefficient, is an alternative measure that is practicable if one has charts or tables to facilitate the computation of the coefficient's value. However, use of the tetrachoric coefficient involves the assumption that the defects concerned in the 2 x 2 scatter are normally distributed. We may grant that most defects are matters of degree.

79

Provided by the Maternal and Child Health Library, Georgetown University
coefficient that is used, however, is a much less important matter than insuring that the full scatter is provided in the report of the study, or at least that enough information is given so that readers can easily construct the scatter and compare it with similar scatters from other studies, applying whatever type of coefficient they may wish to employ for that purpose.

Reasons why use should be made of a coefficient that takes account of the scatter as a whole are given in most statistical texts dealing with correlative problems. We may here consider why, from a practical viewpoint, it is inadequate to employ a ratio, such as the one used by Yankauer and Lawrence. It will be recalled that they utilized the ratio of cases found by the school to the cases not found by the school, the total of those two groups being the children who, in Lawrence's re-examinations, were classified as "with defect."

To illustrate the inadequacy of such ratios (or percentages analogous to them), let us assume that a program's case finding has been evaluated by a method more or less like the one we have described, and that the scatter shown below represents the findings in a sample of 500 children. The correlation for this scatter happens to be .54.

<table>
<thead>
<tr>
<th>Re-examinations</th>
<th>( \text{Children with defect} )</th>
<th>( \text{Children without defect} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>School's findings</td>
<td>( 500 )</td>
<td>( 100 )</td>
</tr>
<tr>
<td>( \text{Children with defect} )</td>
<td>110</td>
<td>66</td>
</tr>
<tr>
<td>( \text{Children without defect} )</td>
<td>390</td>
<td>34</td>
</tr>
</tbody>
</table>

(as stressed by Palmer, 1934, and by Chenoweth and Selkirk, 1953, page 135). Yet it does not follow that defects can be regarded as normally distributed variables. If they are \( \text{not} \) normally distributed, the use of tetrachoric coefficients would overstate the extent of the underlying associations. It is also true that, for their part, point correlation coefficients tend to \text{understate} the extent of underlying associations, but the amount of understatement is probably not great, and it seems safer on the whole to underestimate than to overstate the associations concerned. Ideally, perhaps, one should compute and report both the point and the tetrachoric coefficients for each scatter, but that has hardly seemed requisite in this review. No tetrachoric coefficients have been used, and all correlations are point coefficients except where "ordinary" correlations are specified in connection with 4-way or 5-way ratings of nutritional status.
Suppose the experts did not calculate the correlation or consider the scatter as a whole; instead, they centered attention on the ratio 66:34 and indicated that the school physicians should improve the efficiency of their case finding by increasing that ratio to 75:25.

It might then occur to the school physicians that the 75:25 ratio could be obtained if they modified their views regarding the severity levels at which defects should be reported; this would permit the reporting of conditions which, although of the same general nature as those reported before, were relatively mild in degree. The school physicians might be led to do this in the best of faith, and the procedure would not be objectionable in principle, especially if notation were made of the fact that each of the new defects was a case of moderate or slight severity.

Suppose that a new group of defects of this nature were added to those already noted in the school records, and that the school's findings were again cross tabulated with the experts' findings. There is reason to believe that, except for chance fluctuations, the correlation would remain practically unchanged. We have therefore set up the scatter that would be expected on that assumption and the assumption that 25 defects were added to the school records in the manner just described. The effect is to change the figures in the scatter considerably, except of course for the top row of figures representing the experts' findings. The scatter, for which the correlation again computes as .54, is as follows:

<table>
<thead>
<tr>
<th>Re-examinations</th>
<th>Children with defect</th>
<th>Children without defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>School's findings</td>
<td>Children with defect</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>Children without defect</td>
<td>365</td>
</tr>
</tbody>
</table>

For purposes of convincing the skeptical, it might be worth while in connection with evaluative studies to test this proposition much in the manner that we have described. Since chance fluctuations can affect the correlational results perceptibly, tests in several different studies would be desirable. Another kind of data pertinent to the general problem will be reviewed later (Section 5) in connection with the screening study by Jess and Souther (1940).
It will be seen that the desired ratio of 75:25 has been achieved. Yet the scatter as a whole reflects the fact that, despite the shifts of frequencies in the various categories, there is still about as much disagreement as before between the ideas of the experts and the ideas of the school physicians concerning conditions that should be classified as defects.

This pair of scatters illustrates two points. One is that no ratio or percentage can adequately indicate overall efficiency of case finding. The other is that the level of severity of the case finding, although an important matter in itself, is practically independent of, and should not be confused with, questions regarding case finding efficiency. The latter questions are best discussed in terms of correlations or similar coefficients, and it is one of their advantages that they permit generalization of case finding efficiency without reference to severity levels.

Although levels of severity are not the key problems in case finding, and although the overall efficiency of case finding can be good or poor regardless of whether a high or low level of severity is used, it is of course desirable in an evaluative study to give attention to the severity level that has been employed by the program under consideration. One way of doing this is to compare the proportion of children indicated as having defects in the school records with the same proportion as found by the experts. Taking, for example, the first of the two scatters just reviewed, the figure 110 suggests that the severity level used by the school is only a little higher than that used by the experts, which is indicated by the figure 100. The fact that there is a relatively large difference between the analogous figures in the second scatter indicates, as we already know from the way this scatter was made up, that the school physicians shifted to a less stringent level of severity in selecting "with defect" cases.

However, it will ordinarily be best to deal with questions of severity level in connection with the discussions of specific cases to be described next.

**Step 2.** As the second and last step in evaluating the program's case finding, the experts should provide specific examples of changes that may be desirable in the concepts and definitions of defects used in the program. The examples of these changes should be developed by listing the cases for which the experts' and the school's findings were in marked disagreement, and taking up the cases one at a time with the program staff. As necessary, a few of the children should be called in for further checking or for demonstrating certain points.

For discussion purposes in the remainder of this Section, we may refer to the first of the two scatters cited above. The
cases which should be given special attention in this step of the study are the 34 children “missed” by the school and the 44 children “over-selected” by the school. (The group of 66 children is also likely to include some cases in which the defects identified by the experts differ from those found by the school, and such cases could be considered if there is time, but they are unlikely to prove much that will not be shown as well by studying the missed and over-selected cases.) The purpose of this part of the evaluation is to indicate why good case finding should have identified most of the cases in the group of 34 children, and should have included only a few of the cases among the 44 children.

We did not specify “all” of the 34 children and “none” of the 44 children because a few changes in the classification of the children will probably be found advisable at this time, due either to additional information supplied by the program staff or to additional findings made when the children are brought in for demonstration or further checking. Decisions on such cases should of course be made by the experts, although there is no reason why views of the program staff should not be weighed in the process of reaching the decisions. However, it will be best to leave unaltered the original scatter of the experts’ vs. the school’s findings. From the start of the study, it should be recognized that a certain amount of error is to be expected in any such scatter. The scatters of different studies will be more understandable and comparable if they represent the original cross tabulations without adjustments for later decisions.

As a result of the discussions and demonstrations the experts might decide, say, that 31 of the 34 “missed” children actually had defects which the school should have found, and that, among the 44 “over-selected” children, 4 had defects of consequence and 40 did not.

The 31 cases should be regarded as errors of omission, and the 40 cases should be considered as errors of commission. Information about the total group of 71 cases should be recorded in a manner which, though brief, will serve as guide material to the program staff and others concerned with case finding problems. The material could well be included in an appendix to the report of the study. Needless to say, the names of the children and details of the discussions of their cases should not be published.

Follow-up evaluation

Step 3. Evaluation of the program’s follow-up work should start with listing the children who were classified as “with defect” according to both the experts and the school records. These cases comprise the group of 66 children in the scatter which we have been using as an example of study findings. The experts should
add to this group the 31 children just verified as being cases of true defect among the "missed" group, and also those 4 children among the "over-selected" group who were finally judged to have defects of consequence.

The total of 101 children with well ascertained defects should be studied further with a view to deciding whether each child has received or is receiving adequate care. For a considerable number of children in the list, the adequacy or inadequacy of their care will have become evident enough in the previous work of the study. For the remainder the methods used by Yankauer and Lawrence should be followed. That is, the decisions regarding the children's care status should be reached by consulting, if necessary, the staff of the program, the parents of the children, and the private physicians or clinics concerned.

The care status of a child with an irremediable defect should be judged as adequate or inadequate according to whether provision has or has not been made for adjusting his school and home conditions in ways that seem best for his case. The care status of a child with two or more defects should be considered adequate only if he has received or is receiving adequate care for all of his defects. These suggested rules for classifying the more complex cases are arbitrary, but experience with similar problems in other fields indicates that such rules make for maximum comparability of one set of findings with another. Furthermore, the over-simplifications that result from applying the rules will not be serious if the report of the study includes brief narrative accounts of the more complex cases, along with recommendations on best ways of solving the problems which the individual cases present.

While classification of the 101 children into those whose care status is and is not adequate will be the main purpose of this step of the study, it will be important also to learn, so far as possible, how often the efforts of the school should be credited with securing the care for those who received it. To reach decisions about that, the experts, or the technicians representing them, should first consult the parents, inquiring both as to why the care was sought and when it was sought. The parents' report should then be checked against whatever evidence or indications can be found in the school records. When the two sources do not yield a consistent picture, considerable weight should be given to the views of the program staff.

If the care status or the school's role cannot be decided with certainty for a few of the cases, they should be mentioned in the report, but left out of account in what follows. Among the remaining children, the percentage comprised by each of the three groups indicated below should be computed. Each percent-
age may be considered a separate measure, although in a given community the three values are likely to have affected each other. We may designate these indexes as:

"Percentage a" for the children receiving inadequate care.

"Percentage b" for the children receiving adequate care in which the school's efforts were an important factor.

"Percentage c" for the children receiving adequate care without significant effort by the school.

We saw in Section 1 that serious difficulties beset "correction rates" as they are ordinarily compiled, and that much of the effort currently expended on them might be better spent on estimating the proportion of children who need attention and are not receiving it. A measure consistent with that viewpoint is the one we have designated as percentage a. Although this index is not of a sort that can be obtained routinely, its value could be ascertained periodically in connection with re-examination studies conducted, say, at five-year intervals.

Even though that procedure would be desirable, and even though it would probably be less costly than a good "correction rate" which a school might be able to develop on its own resources, we would not be measuring, with percentage a, what an ordinary correction rate is intended to measure, namely, the positive results of a program. Those results are best measured by percentage b.

Yankauer (1952) and others have rightly stressed that, regardless of how the positive results of a program are expressed, they should be judged in part against the professional personnel and facilities available for care in the community concerned. However, we do not have, and we probably will not have for a long time, any formula or other systematic way of allowing for community facilities or otherwise judging a school health program in relation to them.

About all we can say is that, although a good program should be able, in any event, to keep percentage a smaller than percentage b, the program should be able to do that more easily if the community's facilities are adequate than if they are inadequate. That is, if the community's facilities are adequate, they should tend to make percentage c large, and the effect of this would be to leave a relatively small gap for the school to close.

The viewpoint held by most private physicians regarding community facilities was ably stated by Neff (1939), who spoke as executive secretary of the medical society of Nassau County,
N. Y. He said: "Where the community resources are inadequate, the schools are in a strategic position to demonstrate such inadequacies and to bring pressure to bear upon the proper authorities to have these inadequacies overcome. But they must be willing to present positive evidence if they hope to accomplish their end."

In communities where facilities for the care of school-age children seem to be inadequate, the schools can find out whether that is true—and, if it is, present the positive evidence which Neff called for—by having experts re-examine a sample of the school children and bring attention to percentage a as compared particularly with percentage c, among those children whose defects are well verified.
EXPERIMENTAL APPROACHES

EVALUATIVE STUDIES of school health services have already profited to a considerable extent from the method of controlled experimentation. Yet, as this Section attempts to show, the gains achieved so far are small compared with those which could be made if more of the total effort now put into evaluative studies were invested in experiments, and if there were improvements in the selection of control groups and hypotheses.

A number of brief statements regarding the elements of experimental design, as they might well be applied in the field of school health services, are available in the discussions of Harte (1950), Stouffer (1950), Cochran (1955), Greenberg and Mattison (1955), and Albritton (1956).

We need not attempt to generalize the principles set forth by these authorities. But, for the reader who might have been led to suppose that the design of a modern experiment should be rather complicated if it is to be good, we should note Cochran's statement that "we are having to learn . . . how many different questions can be investigated in a single study . . . (and) the lesson seems to be not to be too ambitious. . . . Statisticians . . . may have oversold the power of statistical techniques to unscramble an omelet."

Cochran was referring to the alleged merits of relatively complex procedures, and he did not mean that the value of the classical model of the controlled experiment had been oversold. Indeed, he held that "the power of experimentation in speeding up progress is tremendous," and that "even if it sounds unrealistic," it is always worthwhile to ask "Why can't I do an experiment?"

Let us keep these views of Cochran in mind as we consider some of the more important experimental studies already con-
ducted in school health services and closely allied fields.

Health education

Turner (1928) was the author of the first of several experimental studies that have been conducted on the effects of giving health education to children. Under Turner’s supervision, special health instruction was given to fifth and sixth grade pupils in selected schools, while the children in the same grades of a comparable set of schools served as a control group. The children's gains in weight and height over a 20-month period were used as criteria of the effectiveness of the health instruction.

Turner found that the children in the experimental schools showed moderately greater gains in weight than the children in the control schools, and, in respect to gains in height, there was a considerably greater difference in the same direction. As a result, what was termed at that time the “underweight status” of the children (meaning their weight relative to their height) was worsened rather than improved. This was, of course, a reflection on the concept of “underweight status,” and did not, by itself, cast doubt on the effectiveness of health education. However, it would appear to be worth while to repeat the experiment using the same general design and criteria as were employed by Turner, if only because his finding of greater stimulation of height than of weight seems a little hard to believe.

Turner (1929) went on to express the opinion that schools should weigh all children frequently “as an educational means of interesting them in health and in health practices.” This general viewpoint has been echoed in several subsequent studies and discussions having to do with physical measurements. Apparently no one has attempted to find out, on either an experimental or a judgmental basis, whether children’s interest in health and health practices is really enhanced by taking, recording and discussing their weights or heights. Meanwhile it is possible that most of the efforts being made to use physical measurements in this way are giving children quite mistaken notions about the nature of health and its relation to physical growth.

A second experiment on effects of health education was conducted by Hardy and Hoefer (1936). Pediatric examinations were used as criteria of effectiveness of the experimental variable, which was intensive health instruction given to children in selected schools over a four-year period. A group of schools were designated as controls, but the children in the control schools did not have as good initial health status, according to the pediatric examinations, as the children in the experimental schools.

The results showed that the improvement in the children’s
health status, as indicated by the examinations, was considerably
greater for the experimental than for the control children. The
authors made the rather dubious assumption that the relatively
superior initial status of the children in the experimental group
was not too important so long as the gains made by that group
were relatively marked, and it was concluded that “improved phys-
ical condition was a definite resultant” of the special instruction.

A third important study of effects of health education was
conducted in Cattaraugus County, N. Y., by Grout and Pickup
(1938). Again a group of experimental children received compre-
hensive health instruction, while the children in two nearby coun-
ties, whose initial comparability with Cattaraugus County had
been established, were utilized as a control group. Three criteria
of effectiveness were used. One was an extensive questionnaire,
completed by each pupil, regarding his health habits and practices.
This instrument showed no important differences in the responses
of the experimental and control children. Yet some real, if rather
modest, differences between the two groups appeared in the other
two criteria of effectiveness. They were the pupils’ scores on a
separate test of health knowledge, and the children’s behavior in
respect to cleanliness and other health matters as judged by the
teachers and outside observers.

The possible use of pediatric examinations as a fourth
criterion of effectiveness was considered by the authors, but they
decided that such examinations were “unsuited” to their purpose
of testing the effects of health instruction on behavior. This was
not unreasonable, but there remains some question as to whether
improvements in children’s health knowledge and behavior really
mean much when they are induced as part of a special study, and
whether such improvements actually result in better health status
of the children concerned.

Since the education of parents in health matters is often
stressed in connection with school health services, one might sup-
pose that at least as much work had been done on that question
as on effects of giving health education to children. However, if
the effects of giving health education to parents have been suitably
tested, and if positive results have been obtained, the findings are
not commonly mentioned in discussions of the value of educating
parents in school health programs. Those discussions seem to be
based on logic and a priori assumptions.

In any event the importance of the problem warrants ex-
perimentation on a large scale. As one design that would be
feasible in any of our larger cities, two sets of schools with at
least four schools in each set could be selected in such a way as
to insure that the two sets were, as a whole, comparable. Staff
members having the same qualifications would conduct similar
programs in both sets of schools, except for a substantial differ-
ence between the two sets in respect to the amount of attention
devoted to educating parents. One way of providing for this
would be to assign extra staff members, preferably having the
same qualifications as the regular staff, to the schools who were
to give the larger amount of attention to educating parents.

However, the study would answer a more practical type of
question if the number of physicians and nurses, as well as the
time they spent on the program, were kept the same in both sets
of schools. Then, in one set, about half of the staff time would be
expended on educating parents (e.g., by requiring their attendance
at examinations and by other methods), while in the other schools
only some 10 percent of the staff time would be spent on parent
education, thus permitting a relatively large amount of attention
to be given to ordinary case finding and follow-up problems. The
measurements of effectiveness should be made one or two years
after the end of the test period, and should include (a) an index
of the amount of care actually received by the children, and (b)
parents' reactions to a standardized interview covering the experi-
mental variable without making direct reference to it (e.g., as
developed by Greenberg and others, 1952).

Nutrition studies

More experimental studies have been conducted on nutri-
tional and dietary problems than on any other single phase of
school health service. The first major study was conducted by
Kaiser and others (1926). The investigators wished to evaluate
the 20-week special "nutrition classes" which had long been part
of the school health program in Rochester, N. Y. The function
of the special classes was to see that underweight children re-
ceived medical care as necessary, rest and reduced activity in
and out of school, and suitable increases or changes in diet. The
authors pointed out that the key question which an evaluative
study should answer was: "Do other children who are as much
underweight, but do not receive the stimulation given in the nutri-
tion class, make satisfactory gains ultimately?" To answer this
question the investigators relied chiefly on data regarding the
gains in weight made by 632 children who entered the classes, as
compared with the gains made by a like number of underweight
children who did not enter the classes. Both groups were followed
during the 20-week period of the classes and for a year thereafter.

The authors' presentation of the statistical findings was
unsatisfactory, and the published figures need not be reproduced
here. So far as can be judged from the findings, however, they
indicated that little, if any, of the benefit conferred by the special
classes was retained at the end of the year. However, it appears from the report that the initial comparability of the experimental and control groups was not firmly established, and it is possible that this study involved a bias opposite to that in the study conducted by Hardy and Hoefer. That is, in this study the children who did not enter the special classes may have been somewhat superior, as a group, to the children who did enter the classes, and if so, the special classes might have accomplished more than the comparison of end-results indicated. The experiment is well worth repeating, with careful attention to controls, in several of the school systems which still maintain special classes for below-par children.

To study the effects of feeding school children supplementary breakfasts, Urbach and others (1948) divided 205 11-year-olds, all of whom were underweight, into three “matched” groups. The basis of the matching was not altogether clear. On arriving at school, one group (59 children) received an ordinary cereal, while another group (73 children) received an enriched cereal, and the third group (73 children) received no cereal. The test period was seven months, at the beginning and end of which numerous measurements and rating of all 205 children were made by persons who did not know the group to which any child belonged. In respect to gains in weight, no difference was found between the ordinary-cereal and enriched-cereal groups, although in both those groups the gains were greater than in the no-cereal group. The more important finding was that the children in the enriched-cereal group improved more than the other two groups of children in several respects, including skeletal maturity and oral conditions. It is to be hoped that several schools will repeat the comparison of enriched vs. ordinary cereals, using random assignments of children from matched pairs (see Tisdall and others, below).

Browe and Pierce (1950) selected 24 children with conjunctival symptoms, 21 with gum symptoms, and 19 with tongue symptoms. Over periods of one or two years these three groups received, respectively, vitamin A, ascorbic acid, and niacin. Control children receiving no vitamins were designated, and color photographs of the eyes, gums, and tongues of both the experimental and control groups were taken before and after the treatment period. Judgments as to whether the children showed improvement were made from the photographs by persons who did not know which children had received treatment. Browe and Pierce found statistically significant differences in the proportions of experimental and control children showing improvement, but when the work was extended to larger groups of children, the findings were apparently less clear cut. In a brief summary of the extended
study, Pierce and others (1953) said the results had been analyzed by two different statistical methods, both of which "showed the same general trend, but considerable variation in the incidence of significant differences." This conclusion was not elaborated upon.

Benjamin and Pirrie (1952) tested claims that vitamin B₁₂ helps school children whose rate of physical growth is considerably slower than average. Tablets containing that vitamin and tablets lacking it were given to random halves of 830 children whose poor physical condition had occasioned their assignment to London's open-air schools. The tablets were of different colors, but no one except the manufacturer knew which tablets contained the vitamin until the findings of the 8-week trial had been analyzed. The results showed only slight differences between the experimental and control children with respect to gains in height and weight. Although the findings tended on the whole to favor the children who had received the vitamin, the differences were not consistent from one age or sex group to another, and were not significant statistically.

In a study that was a model experiment in several respects, Tisdall and others (1952) tested the effects of giving children school lunches. As the first step, the school children in Toronto's least prosperous school district were thoroughly examined and tested in respect to all the physical and physiological variables ordinarily believed to be affected by nutrition. The investigators then set up 278 pairs of children who were matched, so far as possible, on the more important variables.

In each pair, assignment of the children to the experimental and control groups was decided by chance, except where purely random selection would have put siblings in opposite groups and perhaps caused family complications.

Over a 2-year period, the children in the experimental group were served a specially nutritious lunch, without charge, at a Red Cross center located near the school.

As anticipated, approximately 200 of the original 287 pairs completed the study; for purposes of the study "completion" meant that the experimental child of a pair ate at least 90 percent of the special lunches over the 2-year period. At the end of that period, the 400 children were then examined and tested again in respect to the same physical and physiological variables as at the start of the study.

Apart from small differences which were only to be expected in the serum levels of ascorbic acid, carotene, and vitamin A, the experimental children were slightly better off than the control children in respect to general physical condition and the condition of their teeth. However, the differences were not signif-
ificant statistically, nor were they of "practical significance" in the opinion of the investigators. At the same time, the investigators noted that the nutritional status of these 400 children had not been unsatisfactory initially, despite the fact that they were from low-income families. It was suggested that more positive results might be found in a test of the effects of special meals on children with poor nutritional status.

Plan for further experiment

To the reviewer it would seem especially desirable that such a test be conducted with undernourished preschool children, since first-grade children frequently appear to have received poor dietary care prior to entering school. The experimental variable would be nutritious food supplements, and they would be made available for all of the meals (not lunches alone) of the children in the experimental group.

In six or eight low-income communities which are some distance apart, it should be feasible to survey the diets of families having preschool children, and thus to identify at least 600 children aged 3 to 5 years whose diets are definitely inadequate.

The surveyed communities should be paired, and, in each pair of communities, matched pairs of children should be selected in such a way that one member of each pair of children was from each community. Methods like those used by Tisdall's group should be used for pairing the children, except that, in the matching process, consideration need not be given to more than two or three factors, and they should be chosen as the factors that are believed by experts to be most relevant to preschool nutrition. In this way it should be possible to set up, from the initial group of 600 children, some 250 pairs of children who were fairly well matched.

Coin tossing or some other random procedure should be used to decide, for each pair of communities, the one which would serve in the experimental group and the one which would be in the control group. This process would place 250 individual children in the group who would receive the dietary supplements, while leaving 250 closely comparable children in the control group.

Owing to changes of residence, unsatisfactory parental cooperation and other reasons, some 50 of the pairs of children would probably have to be dropped from the study during the course of a year. Whenever the investigators found it necessary or advisable to drop a child from either the experimental or the control group, the other child in that pair should also be excluded from further consideration.

The 200 pairs who remained in the study after one year, or those 150 who would probably remain after two years if the
study were continued that long, should be examined by methods like those which Tisdall's group used, and the value of the dietary supplements should be judged by comparing the examination findings in the experimental and control children.

Unless the communities in the study were some distance apart, a number of parents in the control communities would soon learn the details of the dietary supplements, and would tend to provide similar food for their own children—or at least that would happen considerably more often when the experimental and control groups were near to each other than when they were separated. Failure to reckon with this possibility might result in serious underestimation of the effectiveness of the dietary supplements.

It would not be necessary to ignore the needs of any child whose diet was found to be seriously deficient. A child in that category should be excluded from consideration in the study at the very start—and should be provided with proper food by the investigators or other authorities. The pairing of children for the experimental and control groups should be carried out with those remaining children whose diets were definitely inadequate, but not very seriously so.

It seems to the reviewer that the ethics of an experiment like the one sketched above are entirely defensible. It is too easy, or perhaps too convenient, for us to forget that wherever unverified assumptions are used as bases for operating programs or parts of programs, uncontrolled experiments are being conducted. For the most part, the results of uncontrolled experimentation have to be based on judgment, because the findings, by their very nature, cannot be conclusive. Moreover, even the judgmental findings of an uncontrolled experiment tend to be very slow in coming. These circumstances are scarcely defensible on ethical grounds. As an editorialist of the British Medical Association (1951) has aptly said, "a good experiment, well reported, may be more ethical and entail less shirking of duty than a poor one."

Hearing and speech

The difference between inadequately and adequately controlled tests is illustrated by certain studies of radium treatment for hearing loss in school children. In a first study Crowe, Guild and others (1942) recommended the insertion of radium applicators in the pharyngeal tissues of 239 children with impaired hearing, and the parents acceded to this treatment in 208 cases. Sometime later an additional 337 children with hearing loss were designated as controls. The authors admitted that there was doubt about the initial comparability of these children and the treated
children, but a comparison of the groups was made two years after the applicators had been used. The treatment was judged successful because hearing had become normal for 90 percent of the treated children, while that was true for only 46 percent of the untreated children. Later, however, co-author Guild (1950a and b) found reason to doubt the validity of the 1942 study, and indeed she found indications that the radium treatment might have some deleterious effect on hearing.

In view of the uncertainty about the value of the treatment, Bordley and Hardy (1955) conducted a conclusive test of the question. They selected 582 third-grade children whose hearing was poor and whose pharyngeal tissues seemed abnormal enough to make them candidates for the treatment. Applicators were inserted in the tissues of all 582 children, but only half the applicators contained radium. The other half contained an inert substance, and the physicians using the applicators did not know which were which at the time of inserting them.

Five years after the treatment period, the investigators were able to find and re-examine 193 of the treated children and 192 of the control children. It was revealed that the hearing of the two groups had improved to about the same extent. Thus Guild's doubts about the value of the applicators proved well founded, although it was not confirmed that the treatment had deleterious effects.

Studies of the effectiveness of the usual methods of treating impaired hearing have been reported by Gardner (1943) and by Bennett (1953). In each study the investigator sought to take advantage of the fact that, when treatment is recommended for a group of school children, some of their parents do, and some do not, see that treatment is obtained. These respective groups of children were regarded, in effect, as experimental and control groups. Re-examination of these groups showed, in each study, that substantially greater improvement had occurred among the treated children than among the untreated children.

The trouble with this type of study is that the value of the treatment is overstated if, on the average, the children who received treatment were less severe cases than the children who did not receive it; and conversely, the effectiveness of the treatment is understated if there was a tendency for the children receiving treatment to be relatively severe cases. The latter situation probably occurs more frequently than the former, because most parents do not seek treatment for their children immediately, and a parent is more likely to decide to seek treatment if the child's condition does not improve during the "wait-and-see" period. If this was typical of the circumstances in the studies of Gardner and Bennett, their findings underestimated the value
of the ordinary methods of treating children with poor hearing, but we will not know whether this is so until tests have been made with as good controls as those used by Bordley and Hardy.

Wilson (1954) conducted an interesting test of the value of having speech training given to kindergarten children by their regular teachers. To avoid the uncontrolled variations that might be involved if the experimental and control groups were under different teachers, Wilson used schools where the same teacher was responsible for two different groups of children. A chance basis was used to decide, for each teacher, which group of children was to receive the special training. Experimental and control groups, each totaling over 100 children, were set up in this way.

Before the experiment began, all of the children were measured in respect to their articulation of 18 consonants, and also in respect to their reading readiness. Then, over a three-month period the teachers gave the experimental children training in the articulation of 12 of the 18 consonants, while the control children were not specially trained in any sounds.

Retesting at the end of the training period showed that, although the training had not improved reading readiness, the experimental children's articulation was better for all 18 consonants, including the 6 which were not used in the training, than the performance of the control children on the same sounds.

The author noted that long-term experiments of this type would be very desirable. We need to find out, for example, whether speech training given to experimental children in the fifth or sixth grade produces better speech when they reach the eighth grade than is found at that time in control children who received no such training.

Posture and exercise

Ways of influencing school children's posture have been studied in two experiments. Schwartz and others (1928) gave a 4-month physical training program to 68 boys, while a comparable group of 50 boys were used as a control group. As criteria of posture, the investigators developed objective measures of bodily relations in standing and sitting positions. Comparisons of the before-and-after measurements of the experimental and control groups indicated that the exercise did not improve posture, nor did it change abdominal circumference, chest diameter, or chest expansion.

However, general bodily growth was somewhat greater in the experimental than in the control boys. Moreover, in various tests of physical strength which the authors gave to both groups of boys before and after the 4-month period, the experimental
group showed, on the average, about twice as much gain as the control group.

The boys were not measured later to see how well the greater gains made by the boys in the experimental group may have persisted. That question would be well worth testing in a repetition and extension of the experiment made by Schwartz' group, especially because the physiologist Tanner (1951) has expressed the opinion that muscles stimulated by exercise will "revert to their normal size for the child's age after the exercise ceases."

Clements and others (1950) conducted the other important experiment on posture. Experimental and control groups, each comprising 90 children aged 8–9 years, were set up "in such a way as to match initial posture, body types, and ages as nearly as possible." The authors did not specifically state that, after setting up the matched pairs, there was random assignment from each pair to experimental and control groups, but we may hope that this refinement was not overlooked. The experimental group participated in a 6-month program that included frequent gymnastics, posture training, dancing, games, and health talks, while the control children were not specially stimulated to engage in any of those activities.

As criteria of effects of the special program, attempts were made to use before-and-after photographs and ratings by a physician. However, the children in the experimental group became "extremely posture conscious," and this seemed to invalidate the use of photographs and the ratings made when the children knew they were being judged. Nevertheless, on the basis of supplementary ratings made when the children in the experimental group were "off guard," the physician found that their posture had improved markedly, and that the control children's posture had not changed.

Three months later, during which summer vacations occurred, the physician re-judged the children in the experimental group. He found considerable persistence of their gains, but he did not likewise re-judge the control children to see whether their posture, too, might have been improved by the summer vacation period. This was unfortunate, and it was also unfortunate that all of the children in this experiment were not observed and rated, outside of school, by several judges who did not know which children were in the experimental and control groups.

Respiratory diseases

As we saw in Section 1, over half the absence for illness in elementary school is due to respiratory conditions, and little
or no progress has been made in reducing those diseases. In view of that situation it would appear desirable to repeat, perhaps with variations to be worked out in consultation with epidemiologists, an experiment which was briefly reported by Kaiser (1941). In one of two comparable schools, he so arranged that special attention was given to identifying and promptly sending home children with respiratory infections, while in the other school “early recognition and prompt exclusion received no emphasis.” At the end of the school year the case rate (instances of absence) was relatively high in the experimental school, but the average duration of absence had been markedly shortened. As a result, the experimental school’s absence rate (days lost per pupil) was substantially lower than the absence rate in the control school.

The possibility that ultraviolet light might reduce respiratory infections was tested by Gelperin and others (1951) in an experiment whose design might be applicable to other specific proposals for cutting respiratory illness. The crucial part of the study was conducted in 6 schools, which were split in such a way that half the classrooms were exposed to genuine ultraviolet light, while the fixtures in the other classrooms, although similar in appearance, emitted a non-ultraviolet bluish light. Over the test period of four and one-half months, no difference could be demonstrated statistically between the children in the experimental and control classrooms with respect to absences for respiratory illnesses.

It is also of interest that respiratory diseases were involved in the findings of one of the earliest experiments reported in the field of school health. Bliss (1915, 1918) conducted a controlled test of the once popular practice of opening school windows at intervals throughout the day, in winter as well as summer. Children in classes following this practice and children in classes not following it were compared in respect to: (a) gains in height and weight; (b) performance on tests of fatigue; and (c) absence rates for illnesses. The first two of these criteria showed no important differences between the experimental and control classes, but in respect to the third criterion, the children in the experimental group had relatively high illness rates, particularly in respect to absence for colds, sore throats, and contagious diseases.

Apparently this experiment had the effect of stopping the practice of opening school windows regardless of season. That may have been just as well, and yet the results reported by Bliss are not altogether easy to understand, considering such other knowledge—including “negative” knowledge—as we have regarding respiratory conditions and contagious diseases. If only for “theoretical” reasons, it would seem worth while to repeat the experiment reported by Bliss.
Follow-up procedures

Since there have been relatively few experimental tests of follow-up procedures, it is gratifying to note that a well controlled study of that kind was conducted by Mather and others (1955). Schools in 20 Pennsylvania communities were selected and divided into 4 comparable sets, which we may designate as sets A, B, C, and D. The schools in set A were reserved for control purposes. In the schools of sets B, C, and D the nurses utilized a card-index system suggested by Gallagher and Gallagher (1952) to aid the follow-up work with children needing attention for defects. In schools C and D, the card-index system was used in combination with certain other devices, which need not be detailed here except for noting that one of them was general publicity.

Since it turned out that there were no significant differences among the results obtained for the schools of sets B, C, and D, the analysis dealt with them as though they were a single group of experimental schools. The important question, then, was the effectiveness of using the card-index system, as compared with the effectiveness of the usual follow-up routines which had been used in the control schools of set A.

In both the experimental and control schools the study was concerned with those third and fifth grade children who, according to the regular examinations of the school physicians, needed attention for uncorrected defects. Medical defects of that kind had been found among some 350 of the children in the experimental schools, and among 120 of the children in the control schools. After the card-index system had been used in the experimental schools for about three months, specially trained interviewers asked the parents of both groups of pupils whether physicians had been contacted regarding the children's need for care. As was pointed out in our earlier discussion of correction rates (Section 1), responses to this type of question tend to give an unduly favorable impression of the amount of care which children receive, and yet the correction rate obtained in this way is not invalid where the object of the program is simply to see that children needing care are brought to the attention of physicians.

In any event the study showed that the correction rate, so obtained, was 61 percent for the children in the experimental group and only 46 percent for the children in the control group. The study thus indicated that the effectiveness of follow-up work on medical defects was increased through the use of the card-index system. As regards dental defects, however, the card-index system seemed to have little or no value, since, for that group of defects, the analogous correction rate was 60 percent for the experimental children and 59 percent for the control children.
It is to be hoped many more studies will be conducted using a general design like the one employed by Mather and others. Further attention will be given to this type of experiment in connection with our review of dental studies later.

**Screening problems**

Although investigations of screening devices bulk large in school health literature, most of them tend to be quite similar in nature. Rather than attempting to review the findings of these studies, we may discuss: (1) a study design which, though common, is very questionable; (2) a particular study which was, in several respects, a model of the kind of controlled comparison that should be conducted oftener in the future; and (3) a plan for a needed study on ways of screening out or “pre-selecting” the children who should receive attention from school physicians.

**Questionable study design**

1. All of the investigators concerned with the problem have recognized that the findings of a screen should be checked against some kind of a criterion, such as the findings of a physician who is specially trained in diagnosing the functions which the screen is supposed to measure.

   Too often, however, an investigator has administered a screen to a group of children, and has asked the specialist physician to examine only those children who failed the screen. The children passing the screen have been ignored, along with the possibility that they may have included many children who, if examined by the specialist, would have been found to need quite as much attention as the cases selected by the screen.

   In this type of study the investigator has usually employed the number of children who failed the screen as the statistical base or denominator and, for the numerator, has used the number of children found to need care by the specialist physician. It has been assumed that the percentage or ratio obtained in this way is an adequate measure of the screen’s efficiency.

   As we attempted to bring out in connection with the Astoria study (Section 3) and the report of Yankauer and Lawrence (Section 4), percentages or ratios of this kind are affected by the severity levels used for the screen and criterion examinations. The results of studies reported in terms of such percentages or ratios are not directly comparable. Before any proper comparison of the findings could be made, one would have to ascertain the severity levels involved in the various studies and would have to use that information for adjusting the percentages or ratios that had been reported. This would be quite impractical,
if only because the severity levels involved in the criterion examinations are often difficult to estimate, and the process of making estimates would become an investigation in itself.

In order to secure data that will have practical value for comparative purposes, the investigator should see that both the screen and the criterion examinations are given to all of the children in the study group. The two sets of findings should be published, preferably, in the form of a 2x2 scatter. However, the use of some other form of presentation is admissible provided it includes enough of the data so that anyone wishing to set up the 2x2 scatter may do so from the figures given in the report. It will then be possible for readers to construct the scatter and compute, if the investigator has not already done so, the point correlation coefficient or some other measure of the statistical association that holds for the study findings. (The discussion of the study by Jenss and Souther, which will be reviewed shortly, includes data showing that correlations tend to be practically independent of the severity levels used.)

It is, of course, relatively costly to give the criterion examinations to all the children in the study group, and when one goes to the expense of doing so, it is important to see that the fullest possible use is made of the examinations. Thus, whenever it is planned to give criterion examinations to a sizeable group of children for the purpose of testing a screen, the investigator should consider the possibility of using the same examinations to test, not only the one screen, but several screens of the given kind. The additional screens could well include variations of the original screen, especially if there is uncertainty as to what variation of the screen may be most efficient.

The process of selecting screens to be tested is worth the expenditure of considerable professional time. The greater part of this time can well be spent in consulting specialists in the content of the screening problem at issue and experts in the field of tests and measurements.

In designing the study, it would be desirable to include plans for obtaining information on: (a) the cost of the equipment that will be needed for routine administration of each screen; (b) the time required for training staff members to administer the screen; and (c) the number of pupils that can be screened per day by one staff member. Finally, unless it is certain that a single type of criterion examination is sufficient, it will be desirable to utilize more than one criterion measure.

The study of Jenss and Souther

Some important aspects of screening research may be illustrated by a review of the investigation reported by
Jenss and Souther (1940). They wished to evaluate six different combinations of physical measurements which, at the time of their study, appeared to have some plausibility as screens for detecting children with poor "physical fitness." In using the latter term the authors referred, not to the child's athletic ability, but to his general physical condition, including especially his nutritional status.

The screens were tested on 713 New Haven school children, most of whom were in low-income families. The investigation may be regarded as a well controlled comparison of the screens, inasmuch as all children in the study group were examined with respect to all of the screens and with respect to all of the criteria as well. This procedure minimized the effects of the extraneous factors that tend to impair the comparability of results when other study designs are employed.

All the measurements used for the screens, and most of the measurements used for the criteria, were made when the children were 7 years old. A few of the measurements used for the criteria were taken when the children were age 6.

Without attempting to describe the screens in detail, we may identify them as:

1. Weight in relation to skeletal build.
4. The "ACH" index, in which the difference between arm girth and chest depth was obtained and related to hip width.
5. Pryor's index of weight relative to height and hip width.
6. The Baldwin-Wood tables of weight relative to height.

The first three of these screens were developed by Franzen (1929) as methods of identifying children who, in pediatric examinations, were likely to be rated low with respect to nutritional status. A few years later Franzen (1934a, 1934b, and 1935) devised the fourth screen listed above as a method of estimating the combined results of the first three screens. As a basis for constructing all four screens Franzen assumed that, in rating nutritional status, a pediatrician was really making a judgment of the child's soft tissues in relation to his skeletal build.

Franzen showed that, to whatever extent his assumption regarding pediatric ratings was sound, his screens were also
sound. However, he did not check the assumption by systematically comparing the findings of the screens with ratings of nutritional status actually made, independently of the screens, in regular pediatric examinations. Although Warner (1935) had reported a preliminary test of Franzen's assumption, her findings were inconclusive owing to difficulties, not with the screens, but with the pediatric ratings.

The last two of the six screens (Pryor's index and the Baldwin-Wood tables) were likewise measures which had not been adequately tested before, and some authorities believed that these screens might succeed where similar measures had failed in the past.6

As regards criteria, Jenss and Souther recognized that neither a pediatric rating of nutritional status nor any other single measure could be considered fully adequate as a criterion of physical fitness. The authors therefore utilized a number of different measures or indications of fitness, each of which could be considered sound as a partial criterion. The criterion measures were:

A. Gain in weight between ages 6 and 7, or during the year previous to the time of taking the screen measures.

B. Increase in muscle size between ages 6 and 7, as indicated by the difference between measurements of arm girth taken at those ages.

C. Ratings by a pediatrician of nutritional status at age 7.

D. Combination of the pediatrician's ratings of nutritional status at age 6 and at age 7.

E. This criterion was a combination of three measures: criterion B (gain in muscle size); criterion C (pediatrician's rating of nutritional status at age 7); and the

---

6 For the correlation between physicians' ratings and the early Wood-Woodbury tables, coefficients of .35 and .29 had been found, respectively, in the studies of Dublin and Gebhart (1923) and Clark and others (1925). The latter study had also dealt with Dreyer's index of weight relative to trunk length and chest circumference, and with Pirquet's index of the cube root of weight divided by sitting height. Although these two screens were not checked against a criterion, they were shown to correlate lowly with each other and with the Wood-Woodbury tables. The study of Jones (1938) had shown that, when 5-way ratings of nutritional status were used, the average value of the ordinary correlation coefficient was .50 for the ratings of four physicians vs. Tuxford's weight-for-height index. The relatively high value of .50 was to be expected, not only because ordinary correlation coefficients were used, but also because of other circumstances which were more or less unique to Jones' study.
findings of the pediatrician as to the child's need for medical and dental care at age 7.

The results of the study were reported in such a way that, for each screen in relation to each criterion, one could readily construct the 2x2 scatter and compute the point correlation coefficient. The reviewer has done this for the 6 screens versus the 5 criteria. The resulting 30 correlation coefficients are shown in the accompanying table.

Before discussing these findings and their significance for future work on physical fitness screens, let us consider an important by-product of the study having to do with the question of whether correlations are affected by the cutoff scores or levels of severity used in screening devices.

### CORRELATIONS BETWEEN SCREENS AND CRITERIA OF PHYSICAL FITNESS

*Data of Jenss and Souther, 1940*

<table>
<thead>
<tr>
<th>Screens:</th>
<th>( r_{AB} )</th>
<th>( r_{BC} )</th>
<th>( r_{CD} )</th>
<th>( r_{DE} )</th>
<th>( r_{EF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Weight (relative to skeletal build)</td>
<td>0.01</td>
<td>0.02</td>
<td>0.05</td>
<td>0.05</td>
<td>-0.01</td>
</tr>
<tr>
<td>2. Arm girth (relative to skeletal build)</td>
<td>-0.05</td>
<td>0.03</td>
<td>0.28</td>
<td>0.26</td>
<td>0.13</td>
</tr>
<tr>
<td>3. Subcutan. tissue (relative to skeletal build)</td>
<td>-0.04</td>
<td>0.01</td>
<td>0.03</td>
<td>0.00</td>
<td>0.08</td>
</tr>
<tr>
<td>4. ACH index (arm girth minus chest depth relative to hip width)</td>
<td>-0.02</td>
<td>-0.05</td>
<td>0.26</td>
<td>0.23</td>
<td>0.08</td>
</tr>
<tr>
<td>5. Pryor's index (weight relative to height and hip width)</td>
<td>0.17</td>
<td>0.01</td>
<td>0.13</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>6. Baldwin-Wood tables (weight relative to height)</td>
<td>0.03</td>
<td>0.00</td>
<td>0.32</td>
<td>0.24</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Jenss and Souther took the trouble to use two different cutoff scores to distinguish children "failing" each of the screens except the Baldwin-Wood tables. Thus, for each of five screens, one cutoff score was set to distinguish the lowest-standing 20 percent of the children in the group on which the screen was standardized, while another cutoff score was set to distinguish the lowest-standing 14 percent.

Since, according to statistical theory, there was no reason to expect that the coefficients for the 20 percent cutoffs would tend to be higher or lower than the coefficients for the 14 percent cutoffs, we have shown only the coefficients for the 20 percent cutoffs in the table. However, both sets of coefficients have been computed, and it is of interest to ask whether, as a matter of actual fact, the averages of the two sets are similar or substantially different.

In the table, the 25 coefficients based on the 20 percent cutoffs, or those which concern all screens except the Baldwin-Wood tables, yield an average value of .07. For the 25 analogous coefficients based on the 14 percent cutoffs (not shown in the table), an average value of .06 is found. The averages .07 and .06 happen to be rather more similar than we might have expected considering chance fluctuations, but the data provide a practical demonstration of the fact that correlation coefficients tend to be independent of cutoff scores or severity levels.

From the 30 coefficients in the table as a whole, it is clear that none of the screens correlated well with any criterion, since no coefficient higher than .32 was obtained, and most of the values were close to zero. This fact accords with the authors' conclusion that all of the screens tested in the study were unsatisfactory.

The fact that a wide range of proposed screens has been found unsatisfactory in the study of Jenss and Souther and in reports of earlier authors raises some doubt regarding the value of the combinations of physical measurements used in the grid of Wetzel (1941) and in the procedures suggested by Stuart and Meredith (1946) and Meredith (1949). The reviewer has been unable to find substantial evidence that these new measures are successful where the previous devices were not. It is of course possible that a future study will show that Wetzel's grid, Meredith's chart, or some other arrangement of physical measurements is valuable. But, for the present, a considerable burden of proof would seem to rest with those who recommend the use of such measures for screening purposes.

The investigation of Jenss and Souther was an excellent one for its time. The remarks which follow should be regarded less as criticism of their study than as considerations which may be relevant to future work.
The concepts of validity and reliability are worth brief discussion here, since they are frequently pertinent to screening studies, and they were involved in two problems given special attention by Jenss and Souther. For a full discussion of validity as distinct from reliability (or "replicability"), the reader should consult a text on measurement methods. As brief definitions, we may say that validity has to do with the extent to which a measure gets at what it is supposed to cover, while reliability has to do with the extent to which errors comprise or affect the measure.

It is not unusual for a measure to have high reliability and yet have little validity for its purpose. Conversely, a measure can have low reliability but substantial validity (although in practice a measure's effective validity is somewhat limited by low reliability, at least until its reliability is improved). While neither of these extreme combinations of conditions occurred in the study of Jenss and Souther, their report includes materials which exemplify, in moderate form, both kinds of situations.

Most authorities at the time of the study assumed that pediatric ratings of nutritional status (such as criteria C and D) were the most important criterion of fitness that was available. Accordingly, Jenss and Souther conducted a test of what they termed the stability of the ratings. By stability the authors meant essentially the same thing as is meant by reliability.

As one approach to the problem, a pediatrician was asked to rate the nutritional status of a group of 103 children, and, some 13 days later, to rate the same children again. A 4-way scale was used for both sets of ratings, and the second set of ratings was made as independently as possible of the first set. The ordinary correlation coefficient for the two sets of ratings was .73.

As a further approach, the authors asked three pediatricians to make independent ratings, within a week, of a group of 208 children. The ordinary correlation coefficients for the three sets of ratings obtained in this way were: .63 for the first versus the second pediatricians' ratings; .66 for the first versus the third pediatricians' ratings; and .66 for the second versus the third pediatricians' ratings.

Data on the reliability of commonly used measures are always instructive, and the above findings of Jenss and Souther are probably the best information that has been published on the reliability of pediatric ratings of nutritional status.¹

¹ Data yielding correlations of about the same size had been reported by Franzen (1929) and by Jones (1938), but no study had used a procedure which, for the purpose of studying the reliability of the ratings, was as well designed as the procedure employed by Jenss and Souther.
However, the reliability of the ratings was not basic to the question of the value of the ratings as a criterion measure. If reliability were the important problem, a future study could secure satisfactory ratings by simply following a procedure like the one used by Jenss and Souther in the second part of the test cited above. That is, an investigator could arrange to have three pediatricians rate all children in the study group. Then, by averaging the three ratings of each child (and thus canceling most of the errors in the ratings), a measure having very substantial reliability could be obtained.

This would not essentially improve the validity of the ratings, or the extent to which they got at nutritional status per se, and that is clearly a higher priority question than the problem of the reliability of the ratings. Medical science provides more or less continuous checks on other parts of the pediatric examination, but there seems to have been no adequate validation of that part of the examination in which the child's nutritional status is rated. We will consider ways of filling this gap after discussing another measure which was given special attention by Jenss and Souther. It provides an illustration of a measure whose validity was substantial on a priori grounds, but whose reliability was low.

In a special part of the study, the authors were at pains to show that, on the basis of the usual assumptions about factors affecting physical fitness, one could reasonably believe the children in the study group included a considerable number whose fitness was low. To make this point, the authors thoroughly investigated each child or his parents with respect to several variables which, for brevity, were termed "socioeconomic" measures. We need not discuss the details of this work nor the question of whether it was essential to the study's main purpose of testing the screens, but we should note that one of the measures obtained for each child was his consumption of "milk and leafy vegetables."

Unfortunately, many difficulties were encountered in the investigators' efforts to obtain reliable information concerning the children's diets. The report of the study stressed that the securing of reliable data on dietary intake should be considered an important part of any future study of physical fitness screens. However, Jenss and Souther felt that their own data on the consumption of milk and leafy vegetables were too unreliable for use as a criterion measure.

This was regrettable because, unless information on dietary intake is totally unreliable (i.e., consists of nothing but errors), it has obvious validity for research having to do with nutritional status. And, when a measure is known to have substantial validity but low reliability, it is worthwhile, despite the low reliability, to utilize the measure as a criterion in a study of screening
methods. If the investigator then finds that the measure correlates perceptibly with one or more of the screens, he can assume that its correlations with the same screens will be a good deal higher when steps are taken to improve the reliability of that criterion measure. It is ordinarily possible, in one way or another, to remedy the situation when low reliability is the problem, whereas if validity is low, little can be done to improve matters short of turning to some different type of measure.

It seems clear that ratings of nutritional status should be checked against a suitable measure of dietary intake. One approach would be a study of the “intensive survey” type, or a study using the procedure of recording, over a period of several years, detailed information on the diets of a representative group of children and relating that information to ratings of the children’s nutritional status. The ratings should of course be made in examinations conducted independently of any knowledge of the children’s diets.

The nutritionists King (1945) and Maynard (1950) have emphasized that until studies of that kind are reported we will not know the magnitude of the statistical relationship between diet and health. If it is feasible, for example, to apply such intensive survey methods to thousands of adults for the purpose of studying cardiovascular disease (see Dawber and others, 1951), it should not be too difficult to apply more or less similar methods to a few hundred children in order to learn a good deal more about their nutrition than we now know.

A more conclusive approach would be an investigation of the kind urged by Hill (1938). In commenting on the studies of physical fitness screens reported by Jones (1938) and other investigators, Hill recommended that future work be designed to answer the question of what actually happens to children as a result of changing their diets. Considerable information of that nature has been obtained in the studies of particular nutritional problems which we reviewed earlier in this Section. Yet it appears that a study comprehensive enough to answer the general question raised by Hill is still to be undertaken. A plan for one of the possible studies of that kind was sketched at the end of the earlier discussion of nutrition studies, page 93.

**Design for testing teacher versus nurse referral**

One of the concluding statements in the report of Jenss and Souther may be cited to introduce a general plan for a study of screens to select children for medical examinations.

In connection with the problem of ascertaining a child’s physical condition (as distinct from the more specific question of his nutritional status), Jenss and Souther declared that a sound screening method “must be found,” since the cost of full
pediatric examinations of all children in a school was "prohibitive for most communities." Our lack of well-tested screens for that purpose is as serious today as it was when Jenss and Souther commented on the problem in 1940.

A considerable number of school health authorities assume that the screening out or "pre-selection" of children to be examined by physicians is best carried out by the teacher, with help from the school nurse at certain stages of the process. As we saw in Section 3, the plan of giving the teacher primary responsibility for screening was first recommended in 1840 by William Alcott, but the procedure did not come into extensive use until reports of the Astoria study began to be published (see especially Wheatley, 1940).

Less often it is assumed that the screening of children for examinations should be carried out by the nurse, with the teacher helping in ways that do not greatly interfere with her other duties. This procedure was recommended by the eminent school physician Arthur Cabot in 1911, and was employed frequently for more than a decade thereafter. Although still in use (see, for example, the reports of Kahl, 1947; Snyder, 1953; and Buley, 1954), the procedure is apparently less commonly employed today than the plan of giving primary responsibility to the teacher.

While recognizing that, to some extent, both the teacher and the nurse are usually involved in each of these procedures, we may for brevity's sake refer to the respective screens described above as simply "teacher-referral" and "nurse-referral." It is assumed that neither teacher nor nurse attempts to make diagnoses, but that both procedures are simply screens in the sense pointed out by Dukelow (1956), who said the findings of a screen should be regarded as "presumptive evidence of disease, rather than a diagnosis on which a physician would feel justified in basing treatment."

No controlled comparison of teacher-referral and nurse-referral seems to have been conducted, and there is clearly a need for such work. If research funds were available on a large scale, it would be desirable to plan a whole series of tests in which the two screens were compared: (i) in schools where the health service budget was large and in schools where the budget was small; (ii) in schools where the initial qualifications of the teachers and nurses were very good and where those qualifications were relatively poor; (iii) under conditions where a substantial amount of in-service training was usually given teachers and nurses, and where such training was ordinarily slight; and (iv) in schools where children identified by a screen are usually referred to school physicians, and in schools where the referrals are ordinarily made directly to parents and private physicians.
If, however, adequate funds do not become available for such systematic variation of "background" factors, it would still be worth while for large school systems to compare the two screens under whatever conditions are feasible for the circumstances of the schools concerned. We will henceforth give attention to a general study design that could be used in a wide range of background conditions. It should be understood that, in reporting any study of this kind, the investigators should specify the amounts of money spent for each phase of the work, the initial qualifications of the teachers and nurses, the amounts of training given to each type of personnel in the course of the study, and whether the school ordinarily refers most of the children who seem to need attention to school physicians or directly to parents and private physicians. Eventually, then, it should be possible to form some opinion regarding the importance—or perhaps the unimportance—of the roles which background factors may play in respect to efficiency of the screens.

As a partial model we may briefly review the design used by Buck (1922) in a test which he conducted at the Rose school in Detroit. It is doubtful that he interpreted his results correctly, but he published enough of the findings to enable readers to make their own interpretations. Moreover, as a result of a reasonably diligent search, the reviewer is led to believe that no evidence better than Buck's has been published regarding the possible value of the teacher-referral screen. (In 1942, Miller reported a study of the teacher's ability to screen children for a limited number of defects, but his findings were given almost entirely in terms of percentages, and there seems to be no way to reconstruct the 2x2 tables which would show how much association there was between teachers' and physicians' findings. The data reported by Gudakunst, 1937, also seem to defy interpretation in terms of association tables.)

Buck believed nurse-referral was "somewhat more scientific" than teacher-referral, and he noted that a number of cities were using nurse-referral with apparent success. However, he felt it was less important to compare teacher-referral with nurse-referral, than to see how well teacher-referral compared with the findings of staff physicians, when both the teachers' and the staff physicians' findings were checked against the findings of expert physicians. That is, for purposes of his study, the teachers' findings and staff physicians' findings were regarded as screens, and the findings of expert physicians were used as the criterion measure.

Buck arranged to have 241 pupils inspected by teachers who had received special training in screening children for medical examinations. The same pupils were then examined by a squad of three physicians who were on the regular staff of the school.
health service. (At that time all school medical examining in Detroit was done by "triads" of physicians, with each physician responsible for conducting certain parts of the examination.) Finally, independent examinations were given the 241 children by a squad of three expert physicians who, as supervisors of the school health program, had outstanding qualifications and experience.

The results are shown in the accompanying 2x2 scatters.8

<p>| Findings of expert physicians (criterion measure) |</p>
<table>
<thead>
<tr>
<th>Children with defect</th>
<th>Children without defect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Findings of specially trained teachers:</td>
<td></td>
</tr>
<tr>
<td>Children with defect</td>
<td>156</td>
</tr>
<tr>
<td>Children without defect</td>
<td>85</td>
</tr>
<tr>
<td>Findings of staff physicians:</td>
<td></td>
</tr>
<tr>
<td>Children with defect</td>
<td>141</td>
</tr>
<tr>
<td>Children without defect</td>
<td>100</td>
</tr>
</tbody>
</table>

Simple inspection of them shows that the teachers' findings do not agree very well with the findings of the expert physicians, while substantial agreement exists between the findings of the staff physicians and the expert physicians. This impression is confirmed by the values of the point correlations, which compute as .24 for teachers versus expert physicians, and .67 for staff physicians versus expert physicians. Thus, contrary to what Buck believed on the basis of a percentage analysis of his results, the findings cast considerable doubt on the idea that teachers compare reasonably well with staff physicians in respect to ability to pick out children needing medical attention.

The study plan suggested below differs in several respects

---

8 Although the marginal totals of children with defect were not given as such by Buck, he provided the numbers of defects found by each group of personnel. To estimate the numbers of children with defect, it was only necessary to divide the numbers of defects by 1.3. Data in other parts of the report indicate that 1.3 is probably the best divisor to use, but the general results would be affected very little if 1.2 or 1.4 were used instead of 1.3.
from the design used by Buck. Nurses would of course be employed where Buck utilized staff physicians, and the study would extend over a full school year. For reasons which will be evident below, it would not be feasible to conduct the study, as Buck did, on a single group of children. Instead, use would be made of two different groups of children in schools which are some distance apart, but which are similar to each other in respect to parents' income levels, and particularly in respect to the children's health status as indicated by existing records. When reasonable comparability of the two groups has been established, a coin should be tossed to determine which group would be the "teacher-referral study group" and which the "nurse-referral study group."

As in Buck's study, the criterion measure would be independent examinations of all children by expert physicians, who, so far as possible, would have qualifications like those suggested earlier (Section 4) in connection with re-examination methods. The experts would be called in at the start of the study to give or supervise the training of the teachers and nurses, as well as being called at the end of the study to give the criterion examinations.

For an adequate test, nurses should be assigned to the teacher-referral study group as well as to the nurse-referral study group, although the nurses' duties would differ substantially in the two groups. The teachers who were already with the children would of course be kept with their respective groups, but the nurses could well be assigned to the two groups by some random procedure.

It would be simplest, and probably best, to give the same training to all four sets of personnel who would be concerned with the screening work—that is, to both teachers and nurses in the teacher-referral group and to both teachers and nurses in the nurse-referral group. The training of all teachers and nurses would be distributed throughout the first 6 weeks of the school year, and use could well be made of materials like those provided by Rogers (1945), Schneider and McNeely (1951), and by Wheatley and Hallock (1951).

During the 30 weeks following the training period, the teachers in the teacher-referral group would record their observations of the children in accordance with some up-to-date version of the procedures developed for the Astoria plan (Nyswander, 1942). The nurses in this study group would not inspect any children except individual cases believed by the teacher to warrant such inspections during teacher-nurse conferences, which would be held at about the eighteenth and twenty-eighth weeks of the 30-week period. At those times the nurses would not only inspect some of the children but would also give the teacher help in inter-
preting the observations that had accumulated in the teachers' records. After the second set of teacher-nurse conferences the teachers would make final decisions regarding the children who would and would not be considered referrals in that study group.

The nurses in the nurse-referral group would schedule their duties in such a way as to include a thorough inspection of each child during the first 8 or 10 weeks of the 30-week period, and, at about the twenty-eighth week, a second inspection of each child. The second inspections would be relatively brief for all children except those for whom the nurses saw reason to suspect some trouble. Just before or during each set of inspections, the nurses would ask the teachers to name any children who seemed to warrant special checking. However, in this study group the teachers would not be required to observe the children systematically, and they would be asked to keep only such notes as seemed to them especially pertinent in the light of the training they had received. Following the second set of inspections, the nurses would finally decide on the children who would and would not be considered referrals in this study group.

After the 30-week period, the remainder of the school year would be used for conducting criterion examinations of all children in both study groups, and for comparing the experts' findings with the decisions reached by teachers in the teacher-referral group and by nurses in the nurse-referral group. The comparative efficiencies of the two screening procedures would be ascertained from scatters and correlation coefficients analogous to those cited above from Buck's study.

A qualitative analysis of the cases "missed" and "over-referred" by each screen would indicate ways in which teacher-referral, nurse-referral, or both might be improved, although information of this nature should not be regarded as a better general guide to future policy than the statistical correlations. Finally, even though the relationship between the teacher-referral screen and the nurse-referral screen would not be of special interest from a practical viewpoint, it might be of some general as well as theoretical interest to publish the scatter for that association, together with a brief account of representative cases on which the teachers and nurses did and did not agree.

To make the general plan of the study easier to grasp, we have avoided mention, up to now, of a complication arising from the fact that the study would extend over a considerable period of time. This circumstance means that there will be a number of children who, so far as the teachers and nurses are concerned, should be referred immediately, or at least long before the criterion examinations are given.

This complication is not quite as serious as it may seem
because a majority of the children falling in this category are likely to be cases of acute respiratory or communicable disease and are therefore not of the kind that ordinarily come to issue in connection with screening for adverse conditions. At the same time, there will be certain other cases which are of the kind that screening procedures should detect, and which should be referred for attention without delay.

For cases in either category there should be consultation, if possible, between teacher and nurse on the need for the referral before it is made. Where time and opportunity permit such consultation, final decision on the case should rest with the teacher in the teacher-referral study group and with the nurse in the nurse-referral study group. Where consultation is not feasible, the referral should be made without delay by whichever staff member sees the child first. All such referrals should be made in accordance with whatever procedures are normal for the given school system. However, records should be kept of the referrals, showing who made them and to whom children were referred, while also distinguishing so far as possible between the acute and non-acute cases.

At the time of the criterion examinations, the experts or technicians under their direction should investigate all such referrals. The aims of this work would be to distinguish between cases which were and were not pertinent to the general screening problem; and, among the pertinent cases, to identify the cases referred by the teachers in the teacher-referral study group, and the cases referred by the nurses in the nurse-referral study group.

Only the last two categories, i.e., the referrals made by the persons mainly responsible for screening in the respective study groups (regardless of whether there was consultation or not) need to be considered when the study's main findings are analyzed. In the light of all available information on each "pertinent" referral, including the findings for the given child in the criterion examination of him, the experts would decide whether the case did or did not warrant medical attention.

Although it would probably make no great difference, it would be desirable to report the scatters and correlation coefficients both with and without the inclusion of the referred cases, which, when included in the scatters, would be entered in the same way as the other cases whom the teachers or nurses had identified as needing attention.

Dental programs

A number of authorities are in substantial agreement that there is a serious shortage of dental manpower relative to the true need for care; that the public does not realize the great im-
Importance of dental care; and that school dental programs may be considered sound in proportion as they educate the public to seek adequate care and thus create a firm basis for the training of additional dental manpower (see Wheatley, 1943; Klein, 1944; Bertrand and Hitt, 1948; American Dental Association, 1951; and Fulton, 1955). In a recent review of the overall picture, Russell (1955) has pointed out that even if optimum use is made of fluorides during the next decade or two, a large gap will still exist between the actual need for dental care and the manpower that will probably be available to fill the need.

It thus appears that, in the future as in the past, the primary aim of school dental programs should be the education of parents to seek regular care for their children. Experimental studies that could help to guide programs toward the achievement of that aim are not as numerous as one might hope and expect considering the seriousness of the problem. However, the reviewer has found reports of four such studies, and even though they are not models of experimental design, their findings are quite suggestive for future work.

**Morris' study**

Morris (1939) reported an extensive though uncontrolled experiment in 7 Michigan counties. The W. K. Kellogg Foundation paid for examinations given the school children in dentists' offices, as well as for a substantial part of the care that was provided and for certain courses in children's dentistry which were given the local dentists. The school teachers urged the children and their parents to take advantage of the free examinations and to secure whatever care was needed. Additional pressure on the parents was exerted by the local physicians and by a number of civic groups.

Measurement of the effectiveness of the project was attempted by asking the dentists to inform the county health departments regarding the numbers of children whose care needs were completed each year. The reporting by the dentists was apparently inadequate in some of the counties, for Morris was able to give data regarding only one county. In that county, among all children aged 6–10 the proportion receiving complete care rose from about one-fifth to two-thirds from the first to the third year of the project.

**Frankel's study**

Frankel (1940) conducted comparative tests of 3 different procedures. Each procedure was applied to a separate group of children in the first, third, and fourth grades. The test of each procedure extended over a period of 7 months.
One of the procedures was of special interest because it represented a plan frequently used in school dental programs today. Important elements of this plan had been worked out by Sutton (1925) in Atlanta, Ga., but the procedure was developed in detail by Turner and others (1937) in Malden, Mass., and is therefore termed the Malden plan. Under this plan the school or health department distributes a reply-form and note to the parent of each child; these materials urge that the child be taken to a dentist for examination and that the dentist be re-visited as often as necessary for completion of all care that the child needs. If the parent complies with these recommendations and secures complete care for the child, the dentist signs the form and it is returned to the school by the dentist, parent, or child. On occasion the teacher may be aided by the principal or a member of the health service staff, but the teacher does most of the follow-up work and she is usually expected to take steps in a prescribed sequence that is intended to insure the return of a maximum number of forms.

The effectiveness of this plan is often measured in terms of the percentage of children whom the dentists report as completed. Frankel, however, examined the children herself. Thus she was not only able to use her own findings to measure effectiveness, but she was also able to relate her findings to the reports of the dentists. The study’s results on the latter score will be reviewed later in this Section.

At the end of the 7-month test period, Frankel’s examinations showed that complete care had been received by 84 percent of the children exposed to the Malden plan. It is uncertain whether the same percentage would have been found if the children in the Malden-plan group had been fully comparable to the children in the other 2 groups at the start of the experiment; unfortunately, Frankel had concentrated rather too much on seeing that the B groups were of similar economic status, and she realized too late that the children in the Malden-plan group had received more dental care than those in the other study groups before the testing of the 3 procedures began.

In the second procedure tested by Frankel the same reply-form was used, but a dental hygienist distributed the form and did all of the follow-up work directed toward getting it returned. The methods used by the hygienist included all of the steps taken by the teacher. In addition, the hygienist talked to each child individually about the need for dental care, and she asked the parents who did not seek care within a few weeks to come to the school. There she pointed out the child’s cavities and gave the parent a detailed explanation as to why the child needed dental care at regular intervals.
In general, the amount of time which the hygienist spent on her work was so large that the findings for this procedure were of more theoretical than practical interest. For, in order to duplicate the hygienist's methods on a routine basis, a school would have to employ one hygienist for every 800 children, and this would often mean that most of the funds available for school health services would have to be spent on hygienists' salaries alone. In any event, Frankel's examinations showed that over the 7-month period, satisfactory care had been received by 42 percent of the children in the group who were followed up by the hygienist. Although this percentage was noticeably higher than the 24 percent found for the Malden plan, it would seem reasonable to hope that further work may discover some not-too-expensive ways of increasing the proportion of completed cases to even more than 42 percent.

The third procedure tested by Frankel was that of having the school nurse select, and, as time permitted, follow up the children whose physical examination records indicated that they especially needed dental care. Among the children exposed to this procedure, Frankel's re-examinations showed that 20 percent received satisfactory care during the period of the study. This result was scarcely surprising, yet it indicated what could be accomplished with a relatively small amount of effort and thus provided a background finding of some value.

**Nyswander's study**

The third study to be reviewed was part of the Astoria study directed by Nyswander (1942). She tested certain procedures that resembled to some extent the first 2 procedures tested by Frankel. Nyswander's test of the Malden plan was supplemented by an arrangement whereby the local dentists agreed to examine, without charge, the school children who were brought to their offices by parents. In combination with this arrangement Nyswander applied the Malden plan for a full school year to children in grades 1–8. The results were measured in terms of what the dentists reported on the reply-forms. At the end of the year these forms indicated that care had been completed for 30 percent of the children exposed to the plan.

A direct comparison of Nyswander's figure of 30 percent with the 34 percent which Frankel obtained for the Malden plan would be unsound because the 2 studies differed in respect to duration, the children's ages, and the measures of effectiveness used. It nevertheless seems a little surprising that Nyswander's test of the Malden plan did not show at least as high a percentage of completions as Frankel found. One is led to suspect that the free
examinations used in Nyswander's test did not help matters enough to justify the time and trouble which the school took to arrange for the examinations. However, no firm conclusion should be drawn regarding that point until clearer experimental evidence is available.

Nyswander also conducted a test of the procedure of having a hygienist conduct the follow-up work. This procedure was tested over a 6-month period, which was not much different from the 7-month period employed by Frankel. Otherwise, however, Nyswander's test differed markedly from Frankel's. In Nyswander's test the hygienist worked part-time for only 2 months, after which 4 months were allowed to elapse before the results from the replyforms were tallied. Moreover, Nyswander's test was conducted with upper-grade children, whose parents, according to Wheatley (1943), are somewhat easier to convince of the need for care than parents of the lower-grade children who were used in Frankel's test. Finally, Nyswander used the dentist's reports to measure effectiveness, where Frankel had relied on her own examinations. By chance, these differences between the studies cancelled each other's effects in the results, so that Nyswander, like Frankel, obtained 42 percent completions. One wonders what would be found if Nyswander's procedure were: (1) modified to extend over a school year; (2) applied to children of all grades; and (3) judged in terms of findings from examinations given by a school dentist or hygienist at the end of the year.

**Gold's study**

The last of the 4 experimental studies to be discussed was an investigation conducted by Gold (1945). Her study was complex, and a review of more than its main parts would not be useful here. She was chiefly interested in how much effect 2 procedures would have when they were applied, in combination, to eighth grade children. One procedure was a series of individual teacher-pupil conferences during which the child's personal need for dental care was explained and stressed. The other procedure consisted of giving an intensive dental health course in the children's regular class periods. Teachers administered both of these procedures to an experimental group of 255 children. A matched control group of the same size was exposed to only "incidental" discussions of dental health. The experiment extended over a period of 4 months.

Although it was useful to test the combined effect of the individual conferences and the intensive course, it was regrettable, as Gold herself noted, that her study design did not include tests to discover how effective each procedure was by itself. It was noteworthy, however, that certain subsidiary findings of the study
led Gold to consider the individual conferences as more effective than the course in dental health.

An important feature of the study was the fact that, as one of her measures of effectiveness, the author used the average number of teeth that were filled per child during the 4-month test period. Essentially, this measure was the gain in what is called the “F component” of the DMF rate, which, as will be noted later, is a particularly valuable index for experimental and evaluative purposes. In Gold’s study the data on gains were obtained in examinations given by a dental hygienist at the beginning and the end of the test period. The results showed that the children exposed to both the individual conferences and the health education course gained an average of 5.0 filled teeth per child.

This was a very large gain, but it was not too surprising in view of the intensive, if temporary, procedures to which the children in the experimental group were subjected. More surprising was the fact the control children, who had received only incidental dental health instruction, gained 1.8 filled teeth per child. Relative to the gain for the experimental group, the gain of 1.8 filled teeth appears small. It is nevertheless a high figure, and is well above any annual gain per child that school dental programs, including those with dental health instruction, have been able to achieve at reasonable cost.

It therefore seems likely that in Gold’s experiment some effect of the intensive procedures used with the children in the experimental group “spilled over” indirectly to the children in the control group. While Gold’s use of a sound method of measuring effectiveness has permitted us to cite her finding of 1.8 filled teeth per child as a probable example of this so-called “contamination factor,” it should not be supposed that the effect was at all unique in Gold’s study, or that it is not an equally serious problem in other experimental studies. The important point is that when this factor is not guarded against, a procedure used with an experimental group is likely to appear less effective than it actually is, since effectiveness is ordinarily judged against the findings for the control group. As indicated earlier in connection with the discussion of nutrition experiments, it is desirable to use experimental and control groups which are in different localities, and this is likely to require the cooperation of two or more school systems in a single experiment.

Further discussion of methods would not seem essential here because the experimental methods that are needed for studying dental programs are, in principle, identical with those which are appropriate to research in other phases of school health services. The writer will suggest a broad hypothesis that may merit testing in future experiments, and will then conclude this Section.
with a brief discussion of the use of DMF rates and reply-forms in evaluating programs.

The "many-person" hypothesis

In the process of examining the literature the reviewer has wondered whether more parents could be persuaded to take their children to dentists if the same amount of time which the teacher spends under the full Malden plan were spent by several different people, or by having a different person communicate briefly with the parent each time an additional step is required in the follow-up sequence.

The idea of bringing pressure to bear on parents by having several different people stress the need for regular visits to dentists is far from new, as the reviewer recalls seeing this idea discussed in reports of the international congresses on school health that were held from 1904 to 1913. However, something really new would be contributed if school authorities could arrange suitable tests of the "many-person" procedure in comparison with the procedure of having one person do all or most of the follow-up work.

An essential feature of such tests would be the equalization, so far as possible, of time and cost factors for the 2 procedures. In applying the Malden plan it should not be assumed, as some writers have apparently assumed, that the time spent by teachers and principals or other supervisory instructional staff is "free"; instead the cost of such time should be carefully estimated with consideration of the salaries of the staff members involved. (In this connection see the critique of Rast, 1954, regarding Astoria and Malden plans.)

The arrangements used in the many-person procedure should provide for having one person keep account of all communications with the parents of a given group of children. Whereas in the Malden plan a teacher ordinarily does this for the children in each grade group, in the many-person procedure the records could perhaps be kept for a whole school by a volunteer assistant or by part-time help contributed by a civic group. Thus most of the time involved in the many-person procedure could be spent on contacting the parents through whatever visits, notes, or telephone calls may be appropriate for particular cases.

The occupations of the individuals who contact the parents would probably have to differ somewhat from one set of tests to another, depending on the background conditions that already exist in the areas where the experiments are conducted. For purposes of initial tests, it would not seem unreasonable to assume that the occupations of the persons who contact parents do not make a great deal of difference, so long as those persons hold
responsible positions, and, as necessary, the nature of their posi-
tions is indicated to the parents. If and when it should be found that the general hypothesis is supported by the tests, there will be time enough to find out how much difference the occupations of persons who contact parents may make, i.e., whether or not contacts made by family physicians, for example, tend to be more effective than those made by PTA officers.

So much for problems of experimental design and hypoth-
theses. What can be inferred from the literature regarding ways of evaluating dental programs on a routine basis?

**DMF counts**

The epidemiological value of counts of permanent teeth that are decayed, missing, and filled was shown in the studies of Collins (1931), Stoughton and Meaker (1931–32), and Munblatt (1933). The full significance of such counts for public health dentistry, and the importance of combining them into the “DMF rate,” was pointed out by Klein and Palmer (1937). The letters in the rate refer to the numbers of permanent teeth that are (D) decayed and unfilled; (M) missing or needing extraction; and (F) satisfactorily filled. Since the numbers of these teeth rise rapidly with age, it is very important that there be accurate determination and clear reporting of the ages of the children concerned in DMF data.

In the past 2 decades the use of DMF rates in special studies has increased markedly, but application of such rates for routine program evaluation has not been at all commensurate with their potential value for that purpose. This point has been stressed by a number of authorities, including Baker (1951) and Gerlach (1953). DMF rates or measures consistent with them have been used to evaluate programs in at least two States (New Jersey, by Wisan and Chilton, 1948; and Pennsylvania, by Grace, 1952–55). Yet it appears that few administrators at the local level realize the feasibility of using DMF rates for “built-in” evaluation of their dental programs.

Although chief interest usually centers, as in Gold’s study, on the F component, it is easy and very desirable to obtain all 3 components of the DMF rate during inspections of the children’s teeth by a dental hygienist. Thus a school system could well employ a hygienist for a few days every fall to make such inspections in a sample of at least 300 of the children concerned in the school dental program.

The sample should be chosen by a random method like the one described earlier in connection with medical re-examination procedures (Section 4). The children’s dates of birth should be shown on the enrollment list used for the sampling, and the sample...
should omit children who, at the time of the inspections, have not reached age 6 or are more than a year older than the normal graduation age for the school.

The $D$, $M$, and $F$ counts for each child should be recorded on a separate card showing his age at last birthday. The cards should then be sorted by age. For each single-age group of children, the total number of $D$ teeth, the total number of $M$ teeth, and the total number of $F$ teeth should be obtained. Each of these numbers should be divided by the number of children in the given single-age group. The quotients so obtained are the components of the DMF rate for the given age group, and are simply added together to obtain that rate.

The DMF rates and their component parts for the successive age groups should be charted with age shown along the horizontal axis. After a complete chart has been made, it could be re-plotted in simplified form as a classroom exercise in the upper grades. Some of the children could help with the plotting, and the activity could thus be made a part of the regular instruction in dental health given to upper-grade children.

The data should be plotted so that the $D$ components of the successive ages appear as the topmost band of the chart. This band should be lightly shaded. The $F$ components should be shown as the middle band, and should be left unshaded. The $M$ components, which are quite small in elementary school children, should be shown as a darkly shaded band at the bottom of the chart.

The band of $F$ teeth in the middle of the chart can then be represented as a barrier which, if kept wide, will prevent the $D$ teeth in the upper band from going into the band of $M$ teeth below. It can also be pointed out that the width of middle band of $F$ teeth is a measure of the program's success in what is termed its "reparative" aspect. The effectiveness of what are distinguished as "preventive" efforts (particularly through water fluoridation and topical application of fluorides to the children's teeth), is measured by the extent to which the total area made up by the $D$, $M$, and $F$ bands is decreased from year to year.

The lines drawn through the plotted points to distinguish the three bands can be fitted simply "by eye," and in so doing it is admissible to give relatively little weight to occasional values which, in the light of DMF data based on large groups, appear to represent chance deviations from expected values. It should be remembered that the object of having an appropriate sample of the children inspected is not to learn the precise value of each component for each age group, but is rather to obtain enough age-specific data so that one can estimate the width of each com-
ponent band and of the total area on the chart which the three bands comprise.

While it is sound to chart the data as described above, it is important to record in a separate table (and perhaps also on the back of the chart) the values of all age-specific components as they were actually obtained from the inspections, and also the number of children inspected in each age group. This will permit use of the data for statistical studies of trends or for comparing the results of a given program with those of other programs.

When the inspections are conducted at the start of the next year and each year thereafter, one chart should be made as described above for all the children in the sample. A second chart should be made showing only those children who were enrolled in the school at the start of the previous year, and who were therefore exposed to the program for at least a year. If the given program has not been much more effective than the programs in the areas from which the new pupils came, the 2 charts will be essentially similar. However, if the given program has been markedly successful, the second chart will show a more favorable —and a fairer—picture than the first one. Even if only one of the charts is made, it is important that 2 tables be prepared, with the first showing the age-specific component rates for all of the children sampled, and with the second giving the corresponding information for only those children who were exposed to the program for at least a year.

**Dental reply-forms**

At the present time the reply-forms used in the Malden and Astoria plans are of uncertain usefulness for evaluative purposes. Frankel, in the study reviewed above, provided data on 179 children whom dentists said they had completed during the 7-month period of the experiment. Frankel's before-and-after records of these children showed that 77 percent of them had received no care. However, this finding was not as convincing of the unreliability of dentists' reports as Frankel believed. The children concerned were in the age range when many of the deciduous teeth are still present. Whereas Frankel assumed that care of these teeth was as necessary as care of the permanent teeth, many dentists have disagreed with that viewpoint (see the report of the Oral Hygiene Committee of Greater New York, 1930; and Brekhus and others, 1944). To such dentists it would not be inadmissible to report as "completed" any child who had cavities in the deciduous teeth only.

Strusser and Sandler (1948) attempted to check the accuracy of reply-forms by having a hygienist examine the teeth of 64 high school boys 6 months after private dentists had reported
them as completed. The results were compared with those found for boys of the same age who had received care 6 months earlier in clinics, and for whom it was certain that all care was completed at that time. The hygienist's examinations showed no important difference between the 2 groups of boys.

So far as it went this finding was evidence of the validity of the reply-forms. It was nevertheless rather slender evidence, and better evidence is clearly needed. Such evidence could be obtained through special studies in several elementary schools that are now using the Malden plan. Schools in rural as well as urban districts should be included in the study. Dental hygienists should examine the children in all of the grades at the start of the year. Then, whenever a reply-form is returned indicating that a child's care has been completed, the hygienist should re-examine the child without delay. Children not reported as completed should be re-examined at the end of the year and investigated to see whether they are under treatment at that time. The usefulness of the reply-forms could then be judged by comparing the before-and-after DMF counts of: (1) the children reported as completed; (2) the children still under treatment at the end of the year; and (3) the remaining children.

Annual DMF counts like those described earlier will always be desirable, if only because the results of preventive efforts can scarcely be assessed without making such counts. However, as long as the shortage of dental hygienists lasts, it will be very difficult to have DMF counts made in many rural school districts. Those districts could nevertheless evaluate the success of their "reparative" efforts through the use of reply-forms, provided special studies show that the presently-used forms, or perhaps certain revisions of them, yield valid data.
REFERENCES

American Academy of Pediatrics. 1949c. Special tabulations of unpublished data from the Child Health Study; table 19 of volume titled “U. S. and Regional Tables.” Available at Children’s Bureau, Washington, D. C.
American Dental Association. 1951. The Maintown dental health project; how civic groups and dental health societies can plan for better dental health. The Association, Chicago.

Provided by the Maternal and Child Health Library, Georgetown University


Bordley, John E., and Hardy, Wm. G. 1955. The efficacy of nasopharyngeal irradiation for the prevention of deafness in children; joint research project of School of Medicine and School of Hygiene and Public Health, Johns Hopkins Univ., with PHS grant B–19, C–4. Mimeographed; available from authors. (The study design, without the results, was reported with same main title as above by Hardy & Bordley 1951, Am. J. Pub. Health 41 (Aug. Suppl.):57–61.)


126


Greer, I. G. 1948. Personal communication to Davis and Arena (q.v.) regarding the military records of graduates of five North Carolina orphanages. J. Pediat. 33:85.

Health Information Foundation. 1954. Rx for healthier children; how we can improve our poor record for correction of health impairments found in school physical exams. The Foundation, N. Y. 22 pp. (See also Mather and others, 1955.)


Keene, Chas. H. 1929. The physical welfare of the school child; a textbook in school hygiene and health work in the schools for normal schools and colleges. Houghton Mifflin, N. Y. 50 pp.

Kerr, James. 1897. School hygiene in its mental, moral and physical as-


Mace, Henry F. 1944. Physical condition of soldiers at induction; what the schools might do to overcome the defects shown. J. School Health 14:213–217.


Mather, Wm. G., Whitman, Lauris B., Samson, A'Delbert P., and Ayers, Mary E. 1955. Nurse home visits boost corrective action rate. Pub. Health Rep. 70:188–189. (This is a further brief report on the study which was also reported by Health Information Foundation, 1954, q.v.).


New York State Education Department (University of the State of N. Y.) 1952. Analyses of statistical data from annual reports of cities, villages and central, union free, and elementary schools, 1951-52. The Department, Bureau of Health Service, Albany. 19 pp.


Palmer, George T. 1934. Evaluating present practices; a study of physical defects and their follow-up. Pages 76-82 and 250-255 in: Advances in health education, as evidenced at the 7th health education conference of the American Child Health Association, University of Michigan, June 20-24, 1922. The Association, N. Y.


Reason, Paul L., Foster, Emery M., and Will, Robert F. 1953. The common core of State educational information; State educational records...


Warner, Estella Ford. 1935. The use of the nutritional status indices developed by the ACHA. Pages 16-24 in Am. Child Health Assn., Principles and practices in school health education; from the 8th Health Education Conference etc. The Association, N. Y.

Warner, Francis. 1893-96. I. Results of an inquiry as to the physical and mental condition of 50,000 children seen in 106 schools. II. Mental and physical conditions among 50,000 children seen 1892-94 and the methods of studying recorded observations, with special reference to the determination of the causes of mental dullness and other defects. J. Roy Stat. Soc. 56:71-100 and 59:125-168.


Wilson, Charles C. (Ed.) 1953. School health services; a report of the Joint Committee on Health Problems in Education of the National Education Association and the American Medical Association, with the cooperation of contributors and consultants. National Education Association, Washington, D. C. 486 pp.


This page is blank in the original document.
<table>
<thead>
<tr>
<th>Term/Category</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abdominal circumference</td>
<td>96</td>
</tr>
<tr>
<td>Absence rate for illness</td>
<td>3, 98</td>
</tr>
<tr>
<td>Academy study: see American Academy of Pediatrics</td>
<td></td>
</tr>
<tr>
<td>Accidents</td>
<td>1, 4-6, 8-9, 57</td>
</tr>
<tr>
<td>ACH index</td>
<td>102, 104</td>
</tr>
<tr>
<td>Acute conditions</td>
<td>114</td>
</tr>
<tr>
<td>Administration of school health services: see Jurisdictional problem</td>
<td></td>
</tr>
<tr>
<td>Administrative analysis</td>
<td>51, 53</td>
</tr>
<tr>
<td>Age, importance of, on DMF rates</td>
<td>121-122</td>
</tr>
<tr>
<td>Akerly, H. E.</td>
<td>36, 38, 128</td>
</tr>
<tr>
<td>Albritton, E. C.</td>
<td>87, 125</td>
</tr>
<tr>
<td>Aleott, W. A.</td>
<td>44-45, 55, 109, 125</td>
</tr>
<tr>
<td>Ambiguity of illness and accident rates</td>
<td>7-9</td>
</tr>
<tr>
<td>American Academy of Pediatrics</td>
<td></td>
</tr>
<tr>
<td>consultation from</td>
<td>59, 77</td>
</tr>
<tr>
<td>study by, 26-33</td>
<td></td>
</tr>
<tr>
<td>hours of dental service, 32-33</td>
<td></td>
</tr>
<tr>
<td>hours of medical service, 28-29</td>
<td></td>
</tr>
<tr>
<td>numbers of physicians and nurses, 26-28</td>
<td></td>
</tr>
<tr>
<td>American Association for Health, Physical Education and Recreation</td>
<td>59</td>
</tr>
<tr>
<td>American Dental Association</td>
<td></td>
</tr>
<tr>
<td>consultation from</td>
<td>59</td>
</tr>
<tr>
<td>survey by, 31-32</td>
<td></td>
</tr>
<tr>
<td>viewpoint of, 115</td>
<td></td>
</tr>
<tr>
<td>American Child Health Association</td>
<td></td>
</tr>
<tr>
<td>study by Franzen, 53</td>
<td></td>
</tr>
<tr>
<td>survey of cities, 34</td>
<td></td>
</tr>
<tr>
<td>American Medical Association</td>
<td></td>
</tr>
<tr>
<td>consultation from</td>
<td>60</td>
</tr>
<tr>
<td>survey by, 26</td>
<td></td>
</tr>
<tr>
<td>viewpoint on Selective Service findings, 11-12</td>
<td></td>
</tr>
<tr>
<td>American Nurses Association</td>
<td>60</td>
</tr>
<tr>
<td>American Public Health Association</td>
<td></td>
</tr>
<tr>
<td>consultation from</td>
<td>60, 77</td>
</tr>
<tr>
<td>Evaluation Schedule, 34-35</td>
<td></td>
</tr>
<tr>
<td>Health Practice Indexes, 34-35</td>
<td></td>
</tr>
<tr>
<td>American School Health Association</td>
<td></td>
</tr>
<tr>
<td>60, 77</td>
<td></td>
</tr>
<tr>
<td>Analytical studies of Selective Service data</td>
<td>12-13</td>
</tr>
<tr>
<td>Arena, J. M.</td>
<td>12, 128</td>
</tr>
<tr>
<td>Arm girth</td>
<td>102-104</td>
</tr>
<tr>
<td>Articulation, experiment on, 96</td>
<td></td>
</tr>
<tr>
<td>Association of School Business Officials</td>
<td>36</td>
</tr>
<tr>
<td>Association tables: see Scatters</td>
<td></td>
</tr>
<tr>
<td>(2x2), examples of</td>
<td></td>
</tr>
<tr>
<td>Asthma</td>
<td>14</td>
</tr>
<tr>
<td>Astoria study and plan</td>
<td></td>
</tr>
<tr>
<td>critique by Rast, 120</td>
<td></td>
</tr>
<tr>
<td>dental aspects, 117-118</td>
<td></td>
</tr>
<tr>
<td>evaluation of, 71-73</td>
<td></td>
</tr>
<tr>
<td>findings on teacher referral, 55-56</td>
<td></td>
</tr>
<tr>
<td>parents' attendance at examinations, 56-57</td>
<td></td>
</tr>
<tr>
<td>and a test of teacher referral, 109, 112</td>
<td></td>
</tr>
<tr>
<td>Athletics, and definition of school physici ans</td>
<td></td>
</tr>
<tr>
<td>Atlanta, Ga., 116</td>
<td></td>
</tr>
<tr>
<td>Attendance at examinations: see</td>
<td></td>
</tr>
<tr>
<td>Parent attendance at examinations</td>
<td></td>
</tr>
<tr>
<td>Average daily attendance, 37</td>
<td></td>
</tr>
<tr>
<td>Ayres, L. F.</td>
<td>24, 51, 125, 129</td>
</tr>
<tr>
<td>Ayers, M.</td>
<td>131</td>
</tr>
<tr>
<td>Background factors</td>
<td></td>
</tr>
<tr>
<td>in school health costs</td>
<td></td>
</tr>
<tr>
<td>expenditures for all schooling, 41-42</td>
<td></td>
</tr>
<tr>
<td>expenditures for public health, 39-41</td>
<td></td>
</tr>
<tr>
<td>in study of teacher vs. nurse referral, 110</td>
<td></td>
</tr>
<tr>
<td>Bain, K., 130</td>
<td></td>
</tr>
<tr>
<td>Baker, N. H., 121, 125</td>
<td></td>
</tr>
<tr>
<td>Baldwin-Wood tables, 102, 103, 104, 105</td>
<td></td>
</tr>
<tr>
<td>Bauer, W. W., 12, 125</td>
<td></td>
</tr>
<tr>
<td>Baumgartner, L., 125</td>
<td></td>
</tr>
<tr>
<td>Behavior, children's, in health matters, 89</td>
<td></td>
</tr>
</tbody>
</table>

Provided by the Maternal and Child Health Library, Georgetown University
Bell, E., 6, 126
Below-par children, 53–54; see also
Nutritional status
Benjamin, B., 92, 126
Bennett, M., 95, 126
Bertrand, A. L., 115, 125
Bias
in illness and accident reporting,
7–9, 22
in expenditure data, 36–39
in high school data, 69
Biennial survey of Office of Education,
36–39, 41–42
Bierman, J. M., 125
Blanchard, V. S., 67, 132
Bliss, D. C., 98, 126
Bordley, J. E., 95, 126
Braille classes, 70
Breakdowns of data, effect on sample
size, 78
Breakfasts, supplementary, 91
Brekus, P. J., 123, 126
British Medical Association, 94, 126
British solution of jurisdictional
problem, 41
Britten, R. H., 12, 126, 134
Broadbent, T. D., 126
Brodman, K., 49, 78, 126
Brouha, L., 135
Browe, J. H., 91, 126, 133
Brown, W. P., 2, 21, 131
Brownell, C. L., 186
Buck, C. E., 19, 110–111, 126
Budget for school health services, 41,
46, 109
Built-in evaluation of dental pro-
grams, 121–123
Buley, H. C., 109, 126
Burdiick, E. D., 3, 127

Cabot, A. T., 109, 126
California
illness data, 4
study of State program, 46–47
Cameron, W. R., 6, 126
Cancer, 1
Card-index system, in follow-up, 99
Cardiovascular disease: see Heart
conditions
Care status of children, follow-up on,
84
Caries, dental, 11, 15–16, 31–33, 114–
124
Case finding evaluation
clinical conferences with staff, 82
desirability of reliability study,
76–77
inadequacy of percentages, 80–82
need for use of full scatter, 80
study of Walker and Randolph, 20
Case, J. D., 136
Case rate for illness, 3, 98
Cattaraugus County, N. Y., 89
Caziarc, D. R., 125
Cereals, experiment with, 91
Check-list, in Evaluation Schedule,
34–35
Chenoweth, L. B., 80, 126
Chest dimensions, 96, 102–103
Chilton, N. W., 121, 136
Chipman, S. S., 129
Chorea, 44
Cincinnati, O., 67
Ciocco, A., 6–7, 15, 126–127
City programs, judgmental evalua-
tions of, 50–61
Franzen's 1933 report, 52–53
Nyswander's Astoria study, 54–58
Phair's study, 51–52
Raper's study, 50–51
study by Ayres and Clark, 51
studies by Wheatley and others,
58–59
Woodruff's report, 53–54
Civic groups, in dental follow-up,
115, 120
Clark, J. W., 127
Clark, T., 3, 51, 127
Clement, E., 16, 97, 127
Clements, F. W., 16, 97, 127
Cleveland, O., 51
Clinical conferences with staff, 82
Cochran, W. G., 87, 127
Cohn, H., 44, 127
Collins, S. D., 3, 7, 8, 121, 127
Communicable diseases, 1–8, 47, 52
Community facilities, relevance of,
85–86
Community health services, expendi-
tures for:
see Public health expenditures
Conferences with staff on specific
cases, 82
Congenital malformations, 1
Conjunctival symptoms, 91
Contagious diseases: see Communicable
diseases

Provided by the Maternal and Child Health Library, Georgetown University
Contamination factor, in experiments, 94, 119
Continuity of nursing service, 30
Control groups, selection of, 87-124
Cooperation of parents and private physicians, 34
Coordination of effort, 48, 55
Cornell Medical Index, 49, 78
Cornell, W. S., 62, 127
Correction rates
evaluative studies of, 20-22
example of contrasting rates, 18-20
in follow-up evaluation, 85-86
in Mather's study, 99
in Phair's study, 51-52
in Tennessee study, 45-46
Correlation coefficients, discussion, vii, 79-80
Cost
of full pediatric examinations, 108-109
of screening equipment, 101
of school health services: see Expenditures
of teacher's work in Astoria and Malden plans, 120
Crippled children's programs, 40
Criteria, in screening studies, 100-108, 112-113
Critique of Astoria and Malden plans, 120
Crowe, S. J., 94-95, 127
Culbert, R. W., 57, 130
Cut-off scores, 65, 104-105; see also Severity levels of defects
Davens, E., 127
Davis, G. W., 18, 21, 48-49, 78, 127
Davis, W. H., Jr., 12, 128
Dawber, T. R., 108, 128
Deciduous teeth, questionable importance of care for, 123
Defects, prevalence of
Collins' data of 1931, reference to, 127
data of Maxwell and Brown, 21
Rapeer's findings, 50-51
Selective Service data, 9-12
Wilzbach's data, 68
Degrees of defect, 49, 79
Dental clinics, services of, 32-33
Dental conditions in Selective Service data, 11-12, 15
Dental health education
experiment on, 118
with DMF charts, 122
Dental hygienists, 25, 31, 115-119, 121-124
Dental inspections, data on, 31
Dental manpower, shortage of, 114-115, 124
Dental programs
DMF counts, 121-123
follow-up experiments, 99-100, 115-120
many-person hypothesis, 120-121
survey of American Dental Association, 81-82
Dental reply-forms, 116-118, 122-124
Dental service, hours of, 32-33
Dentists
in private practice, 33
in school programs, 25, 31-33, 114-124
Derryberry, C. M., 138
Design for studies: see Plans for new studies
Detroit, Mich.
re-examination study, 67
teacher-referral study, 110
Diagnoses, medical, 47, 61, 109
Diet, children's
data on intake, 107-108
suggested experiment on, 93-94
Digestive disorders, 4-5, 44
Diphtheria, 1, 7
Disagreement among experts, value of noting, 60
DMF counts and rates, 119, 121-124
charting of, 122-123
definition of, 121
effect of age on, 121
in Gold's study, 119
in program evaluation
at local level, 121
of preventive aspects, 122, 124
of reparative aspects, 122
of reply-forms, 124
at State level, 121
Downes, J., 8, 128
Drake, T. G. H., 135
Dreyer's index, 103
Dublin, L. I., 103, 128
Durhbow, A. B., 133
Dukelow, D. A., 26, 109, 128-129
Duration of illness, 3-7
Early literature, vii, 24, 44

Provided by the Maternal and Child Health Library, Georgetown University
Elementary vs. secondary schools, vi, 35–36
Ellis, R. G., 135
Epilepsy, 20, 44
Equalization of time and cost factors, 90, 120
Erdmann, A. J., Jr., 126
Errors of omission and commission, 83
Ethics of controlled experiments, 94
Evaluation Schedule (APHA), 34-35, 48
Evaluative methods
  dental programs, 121–125
  DMF counts, 121–123
  follow-up evaluation, 115–124
  reply-forms, 124–125
  medical services
    experimental evaluation, 90–114
    judgmental evaluation, 43–61
    re-examining and sampling, 62–86
    case finding evaluation, 79–83
    follow-up evaluation, 83–86
    statistical rates, 1–23
Exercise, studies of, 44, 96–97
Expenditures
  for instruction, 41–42
  for physical education, 37–39
  for public health, 36–37, 39–41
  in Pennsylvania, 49
  for school health, 36–38
Experience factor, importance of
  for school nurses, 30
  for school physicians, 29–30
Experimental method, vi, 87, 119
Experimental studies, 87–124
  Dental programs, 114–124
  Follow-up procedures, 99–100
  Health education, 88–90
  Hearing and speech, 94–96
  Nutrition, 90–94
  Posture and exercise, 96–97
  Respiratory diseases, 97–98
  Screening problems, 100–114
Expert judgment, use of: see Judgmental evaluation
Family physicians: see Private physicians
Fatigue tests, 98
Fluorides
  and DMF counts, 122
  effect on need for care, 115
  topical treatments in schools, 31
Folk, O. H., 134
Follow-up evaluation
  dental services, 99–100, 115–124
  medical services, 83–86, 99–100
  classification of findings, 84–85
  general method, 83–84
  relation to community facilities, 85–86
  rules for irremediable and multiple-defect cases, 84
Foster, E. M., 36, 38, 128, 133
Fowler, E. P., 13, 128
Fowler, H. M., 135
Frankel, R. R., 115–118, 123, 128
Franzen, R. H., 52–53, 64–67, 102, 106, 123, 133
Free dental examinations, 117–118
Free treatment, 51
Frequencies of defects: see Defects, prevalence of
Fulton, J. T., 115, 123
Gall, E., 135
Gallagher, C. D., 99, 128
Gallagher, J. R., 99, 128
Gardner, W. H., 95, 128
Gebhart, J. C., 108, 128
Gelperin, A., 98, 128, 131
General practitioners, data on, 27–29
Gerlach, L. A., 121, 128
German, J. W., 20, 129
Germany, early study, 44
Giese, H., 127
Goititre, 44
Gold, L., 118–119, 121, 129
Goldstein, M. S., 10, 129
Gordon, K., 127
Gossett, H. M., 135
Grace, L. G., 121, 129
Granoff, M. A., 128, 131
Greenberg, B. G., 27, 90, 129
Greer, I. G., 13, 129
Grid of Wetzel, 105
Grout, R. E., 39, 129
Gudakunst, D. W., 110, 129
Guide to a Community Health Study (APHA), 35
Guild, S. E., 94–95, 127, 129
Gullick, L. H., 24, 129
Gum symptoms, 91
Hagen, T. L., 58–136
Hagerstown, Md.
  illness data, 6
Selective Service study, 15
Haldeman, J. C., 40, 132
Hallock, G. T., 112, 136
Hardy, M. C., 88, 129
Hardy, W. G., 95, 126
Harper, F., 58, 136
Harris, M. E., 129
Harte, R. A., 87, 129
Health classes, studies of, 53–54, 90
Health department expenditures, 36–37, 39–41
Health department jurisdiction, problem of
British solution of, 41
Moss'survey, 46
Rapeer's study, 50
statistics on, 25
Health education experiments, 88–90
by Grout and Pickup, 89
by Hardy and Hoefer, 88–89
by Turner, 88
Health examiners, in Williams' study, 47
Health Information Foundation, 129
Health officers, data on, 27–28
Health Practice Indexes (APHA), 34–35
Health services in schools, definition of
in this review, vi
in survey data, 24
Hearing defects, 10–11, 15, 94–96
Heart conditions, 1, 10–11, 15, 70, 108
Height, 88–90, 101–108
Helm, F. E., 26, 128, 129
Henderson, F., 41, 129, 130
Herlihy, L., 91, 130
Hernias, 10–12
High school graduates, Selective Service data on, 14
High school students and evaluative studies, 67–70
Hill, B., 108, 130
Hip width, 102
Hipsey, E. H., 127
Hirschfeld, G., 8, 130
Hiscock, I. V., 47–48, 130, 136
Hitt, H. L., 115, 125
Hobson, C. J., 26, 87, 41, 134
Hoefer, C. H., 88, 120
Hoenig, L., 135
Home safety, education for, 57
Home visits by nurses, 46, 51
Hood, T. R., 9, 130

Hours of service
dental, 32–33, 49
medical, 28–29, 49
Hubbard, F. W., 26, 38, 130
Hubbard, J. P., 27, 130
Hypotheses
selection of, for experiment, 87–124
suggested test of “many-person” hypothesis, 119–120

Illness in schools, 6–8
absence rate, 3
case rate, 3
discussion, 5–6, 7–8
“reverse” effects, 7
trend studies, 6–7
in Hagerstown, 6–7
in New Haven, 6
study by Mason, 4–5
study by Metropolitan Life Insurance Co., 4

Immunization, 1–2, 7, 48
Incidence of defects: see Defects, prevalence of
Independent examinations, importance of, 16–17, 62–86

Influenza, 1
In-service training of teachers and nurses, 109
Instructional expenditures, 41–42
Intensive survey of dietary intake, 108
Interviewing, 29, 43, 45, 71, 73, 77, 90
Irradiation for hearing loss, 94–95
Irremediable cases, 12–13, 20, 84
Ivy poisoning, 5

Jackson, S. H., 135
Jacobziner, H., 57, 73, 130
Jenss, R. M., 8, 101–108, 130
Joint education-and-health jurisdiction, 25, 41, 46–47
Joint State Government Commission, Pennsylvania, 18–21, 48–50
Jones, R. H., 103, 106, 108, 130
Judgmental evaluation, 43–61
recommendations for, 59–61
studies of city programs, 50–61; see detail under City programs, judgmental evaluations of studies of State programs, 43–50; see detail under State programs, judgmental evaluations of

143
Jurisdictional problem
background effect on cost, 40-41
British solution, 41
judgmental evaluation of, 60-61
Moss's survey, 46
Rapeer's study, 50
statistics on, 24-25
Kahl, J. A., 109, 130
Kaiser, A. D., 16, 90, 98, 130
Kansas, accident study, 9
Keene, C. H., 1, 130
Kellogg, W. K., Foundation, 115
Kerr, J., vii, 44, 130, 131
 Kilander, H. F., 25, 131
Klein, H., 15, 33, 115, 121, 127, 131
Knutson, J. W., 131
Krause, R. F., 133
Laboratory findings, 77; see also
Physiological variables
Lanciano, R. C., 13, 131
Langer, 8127
Lawrence, R. A., 73-75, 80, 137
Lesnik, M. J., 131
Levels of severity: see Severity levels of defects
Levine, E., 30, 135
Linde, J. I., 6, 128, 131
Linking of Selective Service records with other data: see Record-linking studies of Selective Service data
Literature, early, vii, 44
Loch, W. E., 127
London, England, 92
Long, J. A., 135
Lorge, I., 126
Lorinser, C. I., 44, 131
Lossef, S., 135
Losty, M., 155
Lunches, school, experiment with, 92-93
Lyon, G. R., 13, 14, 68, 131
Mace, H. F., 12, 131
Mack, P. B., 135
MacKinnon, F. C., 129
Malden (Mass.) plan critique by Rast, 120
designs for new studies, 120-121, 124
Frankel's study, 116-117
Nyswander's study, 117-118
Mann, H., 43
Manny, F. A., 131
Manpower, dental, shortage of, 114-115, 124
"Many-person" hypothesis, 120-121
Marshall, F. G., 135
Mason, R., 4-5, 8, 131
Massachusetts, 43-45
Matching, in experimental studies, 91, 92, 93, 97
Maternal and child health funds, 40, 68
Mather, W. G., 18, 99, 131
Mattrison, B. F., 87, 129
Maxwell, C. H., 2, 21, 25, 131
McPetridge, E., 58
McGill, K. H., 134
MCH funds, 40, 58
McNeely, S. A., 112, 134
McQuarrie, L., 126
Meadors, G. F., 128
Meaker, V. T., 134
Measles, 1
Medical service, hours of, 28-29
Memory factor, in surveys, 28-29
Menzenr, L. F., 32, 132
Mental disorders, 10-11, 14
Meredith, H. V., 105, 132, 135
Merrow, S. B., 133
Metropolitan Life Insurance Co., 4, 8, 132
Midgley, A. L., 126
Miller, B. W., 110, 132
Minneapolis, Minn., 51
Missed cases, in 2x2 scatter, 83-84, 113
Mitchell, H. H., 132, 133
Model studies, 51-52, 58, 101-108, 110-111
Moen, B. D., 31-32, 132
Mohler, J. G., 67, 132
Moore, F. E., Jr., 128
Morbidity: see Illness in schools
Morris, E. W., 115, 132
Mortality rates and trends, 1-3
Moss, B., 41, 46-47, 132
Motivation, as study goal, 48
Mountin, J. W., 40, 132
Movement of families, 19, 69, 123
Multiple-defect cases, 11, 17, 84
Munblatt, M. A., 121, 132
Muscle size, gain in, 103
Musculoskeletal defects, 10-13; see also Skeletal maturity

Provided by the Maternal and Child Health Library, Georgetown University
Mustard, H. D., 132
Myopia, 44

Narrative accounts of complex cases, 84
Nassau County, N. Y., 85
National Council for Chief State School Officers, 36
National Education Association, 38
National Safety Council, 8–9, 132
Neff, J. L., 86–86, 132
Nephritis, 1
Newhall, C. A., 133
New Haven, Conn.
illness trends, 6
physical fitness study, 102
New Jersey, 121
New Orleans, La., 58–59
New York City
Astoria study, 55–57, 117–118,
re-examination studies, 62–67,
70–73
study of open-air classes, 53–54
New York City Department of Health, 62–64, 76, 132
New York State
evaluation of program, 45
correction rates, 21
mortality rates, 2
New York State Education Department, 7, 22, 132
Non-public schools, 35, 40, 58
North Carolina, orphanage study, 13
Norton, H. J., 130
Norwich, N. Y., 14
Nuisance diseases, 5, 47
Nurse referral
in California, 47
design for testing, 108–114
in Phair's study, 52
in Rapeer's study, 50
Nurses in schools
numbers and rates, 25, 27–28,
30–31
training of, 30
Nursing services
continuity of, 30
need for data on, 30–31
Nutrition experiments, 90–94
on special classes, 90–91
on cereals, 91
on vitamins, etc., 91–92
on school lunches, 92–93
on preschool children, plan for,
93–94

Nutritional status
Nyswander's finding, 54
relation to Selective Service find-
ings, 15–16
reliability of ratings, 106
screening study, 101–108
validity of ratings, 107–108
Nyswander, D. B., 55, 54–58, 117–
118, 133
Odd categories of 2x2 scatter, 65, 83
Office of Education, data of
instructional expenditures, 41–42
school health expenditures, 36–38
special surveys, 24–26
Oklahoma, 47–48
Ontario cities, 51–52
Open-air classes
evaluation of, 53, 54
vitamin study with, 92
Oral Hygiene Committee of Greater New York, 123, 133
Orphanage study, North Carolina, 13
Orthopedic handicaps, 70
Over-referrals or over-selected cases,
83–84, 113
Palmer, C. E., 15, 33, 121, 127, 131
Palmer, G. T., 34, 80, 133
Palmer, M., 135
Panels of experts, in judgmental studies, vi, 60–61
Parent attendance at examinations
in Astoria study, 55–56
in Phair's study, 51–52
in Tennessee study, 46
Parents, health education of, 89–90
Parochial schools: see Non-public schools
Part-time personnel
judgmental evaluation of, 61
physicians and nurses, 28–30
"Pathway" study, 64–67
Pediatric examinations and ratings,
88–89, 102–107
Pediatricians
in private practice, 28–29
in schools, 27
Pennell, M. Y., 130
Pennsylvania
correction rates, 18–19
evaluative studies, 20–21, 48–50
follow-up study, 99–100
mortality rates, 3
Pennsylvania Joint State Govern-
ment Commission, 18–20, 48–50

Provided by the Maternal and Child Health Library, Georgetown University
Percentages, inadequacy of, in case finding evaluation, 80-82
Periodicals, need for reporting studies in, 60
Per-pupil expenditure for instruction, 41-42
for school health services, 36-39
Perrott, G. St. J., 12, 126, 133
Pertinent referrals, in a screening study, 114
Phair, J. T., 51-52, 57, 133
Phi correlation coefficient, 79-80
Philadelphia, Pa., 18-19
Phillips, A. J., 135
Photographs, in experiments, 91, 97
Physical education and definition of health services, vi
expenditures for, 37-39
relation to Selective Service findings, 13-14
studies of exercise, 96-97
Physical fitness early studies of, 43-45
experiments in, 96-97
screening study of, 101-108
Selective Service findings, 9-16
Woodruff's report, 53-54
Physical measurements as screens, 101-105
Physical training: see Physical education
Physicians in schools definition of, 27
numbers and rates, 26-28
Physicians, private or family: see Private physicians
Physiological variables, 92; see also Laboratory findings
Pickup, E. G., 89, 129
Pierce, H. B., 91-92, 126, 133
Pirquet's index, 103
Pirrie, G. D., 92, 126
Plans for new studies annual DMF counts, 121-123
dental reply-forms, 123-124
health education of parents, 89-90
nutrition experiment, 93-94
pertinence of Schmidt's proposal, 13, 16-17
survey of physician and nurse services, 29-30
teacher vs. nurse referral, 108-114
test of "many-person" dental follow-up, 120-121
Platt, P. S., 133
Plotting of DMF rates, 122
Pneumonia, 1
Point correlation coefficient, 15, 79-80, 101, 104
Poison ivy, illness due to, 5
Poliomyelitis, 1
Positive results, measure of, 85
Posture studies, 15-16, 96-97
Preschool ages mortality rate, 2
plan for nutrition test, 93-94
supervision of, 46
Pre-selection of children for examination, 44-45, 47, 54-58, 109-114
Prevalence of defects: see Defects, prevalence of
Preventive aspects of dental program, 122
Primary causes of Selective Service rejection, 10
Priority of severe defects, 49-51, 58-59, 81
Private physicians in Academy study, 28-29
and correction rates, 18-20
in dental follow-up, 115, 121
in Detroit, 67
in Franzen's 1934 study, 67
gaining cooperation of, 55
recommendation of Wheatley's group, 59
responsibility for care, vi
in sampling children, 78
on Selective Service findings, 11-12
in teacher vs. nurse referral study, 109-110
and underprivileged children, 26
viewpoint of, 85-86
in Wilzbach's study, 68
in Yankauer's studies, 73-75
Private schools: see Non-public schools
Probability sample, 78
Profile surveys, 34-35
Proration of salaries of physical education instructors, 38
Pryor's index, 102
PTA officers, in dental follow-up, 121
Public health, definition of, 39-40
Public health expenditures, 36-37, 39-41
Public Health Service, 30, 133
Publicity, as a variable in follow-up study, 99

Qualifications
of school physicians, 29
of experts in re-examination studies, 76
of panelists in judgmental studies, 59
of nurses, 31, 109
of teachers, 109
Quality control of examinations, 75
Questionable design in screening studies, 100-101
Questionnaires in re-examination studies, 77

Radium treatment for hearing loss, 94-95
Randolph, C. R., 20, 45, 49, 135
Randomization, 33, 78, 93-94, 97, 121
Rapeer, L. W., 1, 19, 50-51, 133
Rapid examinations, 75
Rast, G. E., 120, 123
Reading readiness, experiment on, 96
Reason, P. L., 36, 38, 133
Record-linking studies of Selective Service data, 14-16
Red Cross center, in school lunch experiment, 92
Redwood City, Cal., 4-5
Re-examination method
adaptations for routine use, vi, 77-85
case finding evaluation, 79-83
follow-up evaluation, 83-85
re-examining procedure, 77-78
sampling procedure, 78-79
selection of re-examiners, 76
severity-level problem, 80-82
Re-examination studies
Franzen's “pathway” study, 64-65
Moner and Blanchard, 67
New York City Health Department, 62-63
Walker and Randolph, 20
Wallace's group, 70
Wilzbach's study, 67-69
Yankauer's studies, 71-72
Yankauer and Lawrence, 73-75
Released time, for dental care, 32
Reliability, concept of, and examples, 106-108

Reliability of examinations
data of Walker and Randolph, 29
for nutritional status, 106-108
proposed study of, 76-77
for visual acuity, 63-66
Reparative aspects of dental program, 122-124
Replicability, concept of, 106-108
Reply forms
dental, 116-118, 123-124
medical, 62
Report cards, for uncorrected defects, 47
Respiratory diseases
data on, 3-6
studies of, 97-98
trends in, 6-8
Response rates in surveys, 24-33
Reverse effects in illness rates, 7
Rheumatic fever, 1, 14
Rich, H., 135
Robbins, M. H., 127
Robertson, E. C., 135
Rochester, N. Y.
nutrition study, 90
re-examination study, 73
Rogers, J. F., 24, 112, 134
Ross, A. R., 135
Rowell, H. C., vii, 137
Rowntree, L. G., 10-11, 134
Rural schools
coverage of, in surveys, 26-28
and shortage of dental hygienists, 124
Russell, A. L., 115, 134
Russell Sage Foundation, 24

Sampling procedures
for medical examinations, 78-79
for dental inspections, 121-122
Samson, A. P., 131
Sandler, H. C., 123-124, 134
Scarlet fever, 1
Scatters (2x2), examples of, 56, 63, 65, 72, 74, 80-82, 111
Schachat, W. S., 135
Schloss, S., 26, 37, 41, 134
Schmidt, W. H., 13, 17, 134
Schneider, E., 112, 134
Scholz, G. C., 131
School health services, definition of, vi
School lunches, experiment with, 92-93
School physicians: see Physicians in schools
Schwartz, L., 16, 96, 134
Scoliosis, 44
Scope of this review, vi-vii
Screening problems, 100-114
design for testing teacher vs.
nurse referral, 108-114
physical fitness, 101-108
questionable study design, 100-101
visual acuity, 64-67
Secondary schools, vi, 35-37, 41
Selection of panels for judgmental studies, 59-60
Selection of examiners for re-examination studies, 76
Selective Service examinations, 9-17, 22-23
analytical studies, 12-13
main findings
causes of rejection, 10
prevalence of defects, 11
possible future studies, 16-17
record-linking studies, 13-16
Selkirk, T. K., 80, 126
“Separates,” undesirability of, 60
Sequence of steps in dental follow-up, 116
Severity levels of defects, 49-51, 58-59, 64, 80-82, 100-101, 104-105; see also Cut-off scores
Shortages
of dentists, 114-115
of dental hygienists, 114, 124
Siffert, R. F., 135
Sight-conservation classes, 70
Singer-Brooks, C., 78, 134
Size of sample in re-examination studies, 78
Sitting height, 103
Skeletal build, 102
Skeletal maturity, 91; see also Musculoskeletal defects
Skin conditions, 4-5, 8, 11
Smallpox, 1, 7
Smith, L. M., 2, 26, 134
Snellen chart, 65-66
Snyder, G. W., 109, 134
Social Security Administration, 39, 134
Socioeconomic measures, in fitness study, 107
Soft tissues, 102
Souther, S. P., 81, 101-108, 130
Speech, experiment in, 96
Stability of pediatric ratings, 106
Standardized examinations, 73
State education departments, 36
State programs, judgmental evaluations of, 43-50
California, 46-47
Massachusetts, 43-45
New York State, 45
Oklahoma, 47-48
Pennsylvania, 48-50
Tennessee, 45-46
Washington State, 47
Statistical rates as criteria, 1-23
accident rates, 8-9
correction rates, 17-22
DMF counts, 121-123
Hiscock’s views, 48
illness rates and trends, 3-8
mortality rates and trends, 1-3
Selective Service findings, 9-17
Stevenson, R., 135
Stokes, J., Jr., 135
Stouffer, S. A., 87, 134
Stoughton, A. L., 121, 134
Strength, physical, 96
Strow, C. W., 3, 130
Strusser, H., 123-124, 134
Stuart, H. C., 105, 135
Study design: see Plans for new studies
Subjective factor, in re-examination method, 76
Sunburn, 5
Subcutaneous tissue, 102
Tax-supported dental clinics, 32
Teacher-nurse conferences in Astoria plan, 55
in a test of teacher-referral, 112
Teacher observation: see Teacher referral
Teacher-pupil conferences in dental follow-up, 118-119
Teacher referral
Alcott’s recommendation, 44-45
Astoria study, 55-56
Buck’s study, 110
design for testing, 108-114
Franzen’s finding, 53
Technicians in re-examination studies, 77
Teeth, defective, in Selective Service data, 11-12
Tennessee evaluation of State program, 45-46
re-examination findings, 20
Terman, L. M., 135
Tetrachoric correlation coefficient, 79-80
Thompson, L. R., 134
Tibbits, H. G., 30, 135
Tisdall, F. F., 92-94, 135
Tongue symptoms, 91
Tonsillectomies
Kaiser's experiment, 16
relation to Selective Service findings, 15-16
 relevance for health supervision, 8
Toronto, Canada, 92
Training of school physicians, 29
Transfer of forms, methods for, 55
Trend data
in illnesses, 6-7
in mortality, 1-3
Triads of physicians, 111
Trial-and-error methods, 55, 94
Trunk length, 103
Tuberculosis, 12, 68
Turner, C. E., 88, 116, 135
Tuxford's index, 103

Ultraviolet light, experiment with, 98
Uncontrolled experiments: see
Trial-and-error methods
Uncorrected defects, 19, 21, 46, 85-86
Underweight, 11-12, 88, 90-91
Urbach, C., 91, 135
Urban-rural differences, 27-28

Validity, concept of, 106-108
Validity of Snellen test, 66
Venereal diseases, 10-12, 68
Ventilation, 44, 98

Virchow, R., 44, 135
Visual conditions, 10-12, 15, 62-67, 70
Vitamins, experiments with, 91-92
Voluntary dental clinics, 32
Volunteer assistance, 120

Wait-and-see period, as used by parents, 95
Walker, E. M., 130
Walker, W. F., 20, 45, 49, 133, 135
Wallace, H. M., 70, 135
Warner, E. F., 103, 138
Warner, F., 44, 136
Washington State, 47
Watkins, J. H., 136
Weaver, R. H., 26, 136
Weighing children, as an educational device, 88
Weight, 11, 15-16, 88, 90-91, 101-108
Wetzel, N. C., 105, 136
Wheatley, G. M., 2, 58, 109, 112, 115, 118, 136
Whitman, L. B., 131
Whitney, A., 133
Whooping cough, 1
Will, R. F., 133
Williams, J. F., 47, 136
Wilson, B. A., 96, 136
Wilson, C. C., 6, 136
Wilson, G. W., 126
Wilson, L., 127
Wilzbach, C. A., 67-70, 136
Winslow, C.-E. A., 21, 45, 136
Wisan, J. M., 121, 136
Wisconsin, 3
Wolff, H. G., 126
Wood, T. D., vii, 137
Woodruff, O. I., 53-54, 137
Wood-Woodbury tables, 103
Wrightstone, J. W., 135

Yankauer, A., 71-75, 80, 137
Yu, H., 135

☆ U. S. GOVERNMENT PRINTING OFFICE: 1957—429997
This page is blank in the original document.