Effect of Incomplete Information on Estimating Prevalence of Disease

By ALBERT P. ISKRANT, M.A., and QUENTIN R. REMEIN, B.A.

One of the most difficult problems for the statistician who is attempting to measure the incidence or prevalence of abnormal physical conditions in groups or samples of persons is the effect on his estimates of the "not observed"—the nonrespondents in the sample group to be studied.

This problem occurs in many fields of investigation. The statistician may wish to estimate the prevalence of syphilis in males of certain age groups, on the basis of the results of blood tests made as part of their selective service examinations. How about persons deferred for nonmedical reasons, volunteers, or those rejected for physical reasons before being blood tested? Does the omission of blood-test results on these persons bias the estimate and vitiate our generalizations?

Perhaps an estimate is needed of the prevalence of certain diseases in a community, based on the results of a screening examination—a mass blood-testing, mass X-ray, or multiple screening program. Will the nonrespondents show the same results as those who are examined—the volunteers or those examined as members of groups (as in industrial programs)? Do those who know they have the disease in question stay away from the clinic or come in for a check-up? Even if a population sample is obtained, what is the effect on the survey results of those who will not allow themselves to be examined because they know they have the disease, don't want to "know for sure," or are ill from other causes? And even if they do allow themselves to be examined, when an attempt is made to make findings of the screening survey more specific by securing diagnoses on suspects, what is the effect of the physician's hesitancy to make a diagnosis or his unwillingness to report it? Is the physician more or less willing to report a diagnosis if he knows that the person being screened has previously been diagnosed as having the condition which is being studied?

Another area in which lack of information about part of a group may influence results is in investigations of the effectiveness of a particular type of treatment on a selected group of persons. In this group will be some patients for whom post-treatment observation is not complete because of the patient's unwillingness to be examined, because he moved out of the area of the study, or because of death from other or unknown causes.

We have used data from the Richmond, Va., and Atlanta, Ga., multiple screening projects, from the Mississippi blood-testing projects, and from a special follow-up study of the Division of Venereal Disease to analyze the effect of nonresponse of patients on prevalence estimates and also on case finding and evaluation of treatment.

Mr. Iskrant, chief of program analysis in the Division of Chronic Disease and Tuberculosis, Public Health Service, was formerly a statistician in the Division of Venereal Disease. Mr. Remein is a statistician in the latter division.
This paper was presented at the annual meeting of the American Public Health Association and the Biometrics Society, November 1, 1951.

Richmond Project

The Richmond multiple screening project was designed to discover syphilis and other diseases detectable by blood testing and other diagnostic methods, and to educate the public regarding these conditions. Suspects—persons screening above or below specified levels or with abnormal readings—were referred to physicians, who were asked to report their diagnoses to the health department on forms provided for the purpose. To evaluate the effectiveness of this project as a case-finding program and to estimate the prevalence of certain diseases in the Richmond population, information was needed on (*a*) who responded to the appeal and who did not and (*b*) whether physicians examined all suspects and reported the results of all examinations to the health department or whether the recorded results were biased.

The Richmond multiple screening survey was intended to reach persons 15 years of age and over. Of the total 37,609 persons who were screened, 37,498 were known to be 15 years old or over, of known sex, and either Negro or white. Comparison of the respondents in the survey with the 1950 census of Richmond showed that the survey population had a more than proportionate share of white females, and therefore was not in proportion to the census as to race and sex.

About one-fourth of the white female respondents, all aged 25 to 34, were tested, whereas examinations were made of only one-seventh of the total population of the city. Negroes, both male and female, were poorly represented in all age groups. The young and the old of both races did not participate as well as did the middle-aged.

If the prevalence of diseases were related only to sex, race, and age, the findings for each group in the survey could be expanded to the total population of Richmond. But perhaps prevalence of disease is in part dependent on socioeconomic status or on quality of disease prevention or medical care, which also may be based on socioeconomic status. Education is usually considered a good index of this status, and records were available for persons tested in Richmond.

Persons going through the screening line were questioned regarding their education. Data on schooling from the 1950 census for Richmond are not yet available but, unless there has been a great change since 1940, it would appear that the screenees are better educated than the general population of the city. This is true for all four race-sex groups. Although the population of Richmond in 1950, by sex, race, age, and schooling, is not the same as in 1940, the percentages of 1940 groups going through the screening line are presented in table 1. While these percentages obviously are not valid for 1950, their relationship may be. The group with the highest representation in the screening line was composed of Negro female college graduates, followed by white female college graduates. Next came white females with 1 to 3 years of college, and then Negro male college graduates. The white college man was the poorest in attendance of all the college groups.

Table 1. Percent of 1940 population over 25 years of age going through screening line, by years of school completed, race, and sex, Richmond, Va., January–July 1950

School years completed	White		Negro	
	Male	Female	Male	Female
None	1.5	2.5	1.8	0.7
Grade school:				
1–4	10.7	11.6	7.0	3.4
5–6	11.6	18.2	8.7	5.8
7–8	13.2	26.4	16.0	11.7
High school:				
1–3	19.0	39.8	22.8	19.6
4	32.5	57.2	33.3	33.4
College:				
1–3	36.4	62.6	54.1	43.0
4	35.7	67.2	58.2	76.4
Total college	36.0	64.3	56.2	54.7
Not reported	43.0	64.1	63.5	48.7
Total	21.4	41.2	13.2	11.7

In each group studied, more white females attended the clinic than white males. In each group other than college graduates, the white female attended better than the Negro female. In all but two groups (high school and college graduates), the Negro male attended better than the Negro female, and, in all groups above the high school level, the Negro male attended better than the white male. For all sex-race groups, increase in attendance was associated with increase in education. Thus, the poor at-

tendance of Negroes at all ages may be a result of lack of education. Obviously, then, no generalizations for the entire population of Richmond can be made from results of the screening survey, since the attendance of respondents was biased in favor of whites, females, and the better educated.

Other characteristics of the respondents and nonrespondents in Richmond not revealed by the survey records are reported by the Health Information Foundation, which conducted an analysis of the multitest clinic (*1*). The foundation reports that the respondents usually come from the middle-income group rather than from the high- or the low-income group, from households with a low average number of persons, and from households where some member is covered by health insurance.

The next problem considered was the nonresponse or incomplete diagnosis by physicians, due either to the screenee's unwillingness to visit the physician or to the physician's hesitancy to report his findings. Physicians whose surnames began with the letter B were chosen as a sample, and the forms for all persons who had been referred to these physicians were withdrawn from the files. There were 93 physicians in this group, to whom 1,102 suspects had been referred. Hypertension was selected as a test to determine if the high proportion of previously known cases reported might be due to a tendency of physicians to return the report form on previously known cases and to withhold it in cases where a new diagnosis had to be established. Although the percentage of forms returned increased with the number of suspects referred to the physician, there was no relationship between the percentage of returns and the percentage of "condition present" (table 2). Neither was any relationship found in the percentages having hypertension either newly diagnosed or previously known. Thus, it would seem that in estimating the number of conditions diagnosed from the partial returns by physicians no obvious bias is introduced.

Table 2. Comparison of number and percent of "conditions" identified by physicians with number and percent of forms returned, Richmond Multiple Screening Survey

Percentage of forms returned	Physicians reporting	Suspects reported	Forms returned	Condition identified as present	
				Number	Percent
0–69	30	324	99	47	47.5
70–89	5	333	270	122	45.2
90–100	17	401	388	189	48.7

NOTE: More than one "suspect" referred to each physician in group whose surnames began with letter B.

Atlanta Program

The Atlanta multiple screening program was designed to blood test and X-ray large numbers of persons to find syphilis, tuberculosis, and other suspected conditions, through examination of the blood and the chest. Because of the large-scale nature of this endeavor, it was decided to concentrate on case finding and not to attempt to obtain data extraneous to that objective. Data on each individual responding to the appeal to come to the clinic were limited to age, race, and sex; no information was requested regarding schooling, income, or other socioeconomic factors.

Large numbers of persons were tested in Atlanta and large numbers with previously unknown conditions were discovered. But how successful was the project in terms of its potential? In making both case-finding and prevalence estimates, can it be assumed that the proportion of the population tested is the same as the proportion of the existing conditions found? In other words, as far as the condition being sought is concerned, do people come to the clinic at random? The 1950 census figures will indicate whether the young or the old, males or females, whites or nonwhites came in.

Unlike Richmond, Negroes are heavily represented among the respondents, with better representation than white persons for both sexes and at all ages. Moreover, the groups 15 to 34 years of age responded most. This is true of both races and both sexes. But how about the rich or the poor, the well educated or the not so well educated? Schooling was not checked on the Atlanta forms. The Bureau of the Census is currently preparing a tabulation for us which compares the education and income of the head of the house of the respondents and of the nonrespondents to the survey.

But how about biases inherent in the conditions themselves? Do the people who have syphilis or tuberculosis or diabetes come in to the clinic or stay out? More particularly, do the people who have been previously diagnosed as having a disease stay out or come in? The only way to find out is to examine either all or a sample of the nonrespondents.

Ideally, a sample should be selected before a screening project begins. After the project is completed, an attempt should be made to find and to examine those who did not respond to the public appeal. Since this had not been done, it was hoped that it would be possible to select a sample of the Atlanta population immediately after the project, check the responses against the survey records, and then attempt to examine the nonrespondents. However, funds were not available for a project of this type, so arrangements were made for analysis of a small sample of Negroes. This sample was selected by us and interviewed by staff members of Atlanta University, both for social-anthropological characteristics and for their participation in the Atlanta multiple screening program. The records were matched against the multiple screening records, and the sample of persons, all of whom were 15 years of age and over, was divided into two groups: those who responded to the appeal to come to the clinic for multiple screening; and those who did not. Letters were sent to all those not identified as respondents, inviting them to come to the city clinic for a chest X-ray and blood tests for syphilis, diabetes, and anemia. Few of the letters were returned as nondeliverable, and many of those in the sample selected came in for an examination.

One difficulty encountered was that this project was considered as research, for which insufficient funds were available. However, through the courtesy of the Georgia Division of Venereal Disease Control, two investigators were lent to us for 2 weeks, and an additional group of persons in the sample was induced to take the examination. At the end of 2 weeks the investigators returned to their usual tasks, leaving 560 of the 1,208 in the sample unexamined.

Because of the high percentage of positive blood tests (approximately 30 percent), it was decided that further follow-up would be profitable. After the investigators had done their best to get the nonrespondents to come in for examination, teams went into homes and offices in an attempt to obtain blood from the most uncooperative cases. Because of technical difficulties, this test was limited to the blood test for syphilis.

Altogether, there were 2,449 Negroes 15 years of age or over in the sample examined. Of this number, slightly more than half responded to the survey appeal; of these, 84 percent were eventually tested for syphilis. The remaining 16 percent were listed as "uncooperative," "moved," "died," and "could not locate."

The results of the different phases of follow-up are given in table 3. The percentage of positive blood tests at all ages is much higher in the nonrespondents than in the original respondents to the survey. Moreover, the percentage is higher in the second phase of the follow-up of the nonrespondents in the sample than in the first phase, indicating an unwillingness to respond on the part of those with a high positivity rate. We cannot explain the lower rate in the most resistant cases—those which had necessitated a blood test at home. However, this rate is still much higher than the rate for the original respondents. Presumably then, in Atlanta, there were more cases of syphilis among those Negroes who stayed away from the survey than among those who came in.

X-rays for tuberculosis also showed a slightly higher percentage of positives among the nonrespondents (2.2 vs. 1.7) than among the respondents. However, for anemia and diabetes the opposite was true: the respondents showed higher abnormal rates than the nonrespondents (anemia, 9.6 vs. 8.2; and diabetes, 5.4 vs. 3.1).

Mississippi Project

A special syphilis case-finding project in Mississippi also provides data on nonrespondents. This project included an appeal to the public to have blood tests. At the end of a designated period a team went into homes and attempted to get samples of blood from those who did not respond to the appeal. In 9 out of 11 counties the percentage of positive blood tests among those tested at home was higher than among

Table 3. Results of serologic tests for syphilis for Negroes tested during the Atlanta multiple screening survey and as a result of follow-up

Age (years)	Respondents		Nonrespondents					
	Number	Percent	First phase		Second phase		Tested at home	
			Number	Percent	Number	Percent	Number	Percent
15–24	279	9. 7	75	10. 7	27	11. 1	13	15. 4
25–34	311	21. 2	132	31. 8	38	36. 8	29	41. 4
35–44	290	26. 6	109	42. 2	42	45. 2	18	16. 7
45–54	199	28. 1	93	34. 4	34	44. 1	12	50. 0
55–64	102	25. 5	80	30. 0	14	50. 0	13	23. 1
65 plus	60	10. 0	51	27. 5	19	31. 6	6	16. 7
Total [1]	1, 241	20. 8	541	30. 7	174	36. 8	91	29. 7

[1] Includes one case, age unknown.

those who volunteered. In the 11 counties, the syphilis rate for nonrespondents was 11.8 percent compared to 8.0 for the respondents.

Georgia Case-Finding Project

A special case-finding project was carried out in a town in Georgia where, after the initial survey, an attempt was made to test the whole population. A slightly higher percentage of positives and a considerably higher rate of previously unknown syphilis was found in the nonrespondents than in the respondents (*2*).

Effect on Therapy Evaluation

Therapy evaluation is another area in which nonresponse affects finding of relapsed cases and estimates of prevalence of diseases needing re-treatment. Here we start with a group of persons who are known to have, and who know they have, a condition requiring medical treatment. They are treated and, whenever possible, all or part of the group are examined periodically, in an attempt to identify the persons who relapse or progress, and to determine failure or re-treatment rates. As in the usual type of case-finding project, the purpose of the follow-up is twofold: to find the failures and offer them further treatment; and to calculate failure or success rates. To induce treated persons to return for post-treatment observation, reliance can be placed on patient education and the patient's interest in his own health; on letters urging him to protect his health by coming in for an examination; on follow-up by a worker who visits him and urges him to return to the clinic; and on visits to his home or place of business, where he can be at least partially examined.

Division of Venereal Disease Study

A special intensive follow-up study made by the Division of Venereal Disease affords opportunity to observe differences in re-treatment rates among patients followed up by special "research investigators" and patients followed routinely—by patient education or by letter (*3*). Both groups of patients were treated by the same clinics, during the same period, and by the same methods of treatment, the only difference being in the method of follow-up. Of the group followed routinely, 42 percent were observed for 2 years; of the intensively followed group, 92 percent. No sizable differences between the failure rates were noted after the first year (table 4). Throughout the first year, the cumulative failure rate is higher in the group that was observed more completely, possibly due to earlier detection of relapses. In the early months of observation, the seronegativity rate in the intensively followed group was higher, again undoubtedly due to more complete examinations. However, at the end of 2 years there was no appreciable difference between the re-treatment rates of both groups.

Table 4. Cumulative re-treatment rates of secondary syphilis patients in a group followed intensively and in a group followed routinely

Observation period (months after treatment)	Intensive follow-up (250 patients)	Routine follow-up (1,856 patients)
3	1. 6	0. 7
6	6. 4	4. 8
9	10. 9	7. 9
12	11. 7	11. 1
15	12. 9	13. 2
18	13. 7	13. 9
21	15. 0	14. 8
24	15. 0	15. 8

Summary

1. Failure to respond to an appeal to "know for sure," when addressed to persons who are apparently well or who are being treated for various conditions, presents problems for both the case finder and the statistician.

2. Evidence from Richmond, Va., and Atlanta, Ga., suggests that the general public does not respond to such appeals at random, but that certain segments of the population respond better than others, because of the type of appeal or because of other factors.

3. These differences in response affect the observed prevalence of some diseases in the groups responding, when the presence of these conditions is affected by age, race, sex, or socioeconomic or other factors.

4. Knowledge of the presence or absence of the conditions being studied may affect the person's decision to respond to the appeal to have an examination.

5. Data from both Atlanta and Mississippi show higher syphilis prevalence in nonrespondents. In Atlanta, somewhat higher tuberculosis rates, but lower anemia and diabetes rates, are found among nonrespondents than among respondents.

6. Data from the Richmond multiple screening survey show no differences in the proportion of morbidity reported by physicians for suspects referred to them, regardless of the percentage of suspects reported on.

7. Data from the Division of Venereal Disease of the Public Health Service show no differences at the end of 2 years in cumulative re-treatment rates between a group of patients followed routinely (42 percent) and a group followed intensively (92 percent).

REFERENCES

(*1*) Boek, Walter E.: An analysis of the multi-test clinic of Richmond, Virginia. New York, Health Information Foundation, 1951, pp. 138–142.

(*2*) Warner, W. Lloyd, Hill, Mozell E., Bowdoin, C. D., Rion, J. Wallace, and McCall, Bevode: Syphilis prevalence and community structure. J. Ven. Dis. Inform. 32: 157–166 (1951).

(*3*) Bauer, Theodore J.: Evaluation of antisyphilitic therapy with intensive follow-up. I. The plan. J. Ven. Dis. Inform. 32: 355–359 (1951).

Recommended Reading About World Public Health

In the American Journal of Public Health for December 1951: "International Health—A Symposium;" including—1. Introductory Remarks, C.-E. A. Winslow; 2. The Role of WHO, Past, Present, and Future, Brock Chisholm; 3. Some Aspects of the WHO's Programs in the Americas, Fred L. Soper; 4. Application of WHO Programs and Policies in a Region, Chandra Mani; 5. Bilateral International Health Programs of the United States, Henry Van Zile Hyde; 6. Our Stake in World Health, Frank G. Boudreau. (This symposium summarized also in *Public Health Reports* for February 1952.)

In The Annals of the American Academy of Political and Social Science for November 1951: "The Search for National Security," including—Public Health and Foreign Policy, Leonard A. Scheele.