# Use of Death Rates in Evaluating Multiple Screening

CHARLES M. WYLIE, M.D., Dr.P.H.

ILLNESS can be controlled or cured only when the patient or his physician recognizes the need for treatment. The physician's help is usually sought, however, only when symptoms or signs occur that are not culturally accepted as normal. Since much long-term illness passes through a prolonged asymptomatic phase, there is a great unmet need for medical treatment in the population of the United States. Multiple screening has been devised to separate off persons with this need, without requiring the total adult population to be examined periodically by physicians.

Multiple screening uses two or more tests to sort out persons who probably have abnormalities from those who probably do not. The immediate aim of multiple screening is to refer for medical care those who have positive test results. Its ultimate aim is to reduce illness, disability, and death in the population by the early detection and treatment of disease.

Even when multiple screening is completely effective in detecting disease in the early stages, screenees will continue to die for various reasons. Screening tests do not cover all asymptomatic disease, and medical care is not sufficiently advanced to control all diagnosed conditions. Moreover, some conditions existing in the positive screenees may not be diagnosed, and new conditions may arise after screening, to remain undetected for some time.

Since persons with positive screening tests have more disease than those with negative test results, we may expect that death rates also will differ in these two groups. Unless medical care is much more effective than is currently assumed, persons with positive tests will have the higher death rates. Therefore, death rates may form a useful basis for evaluating screening tests, the more effective test separating off groups with high death rates.

The purpose of this paper is to present mortality figures for 2,298 residents of Baltimore who took multiple screening tests in 1954, to compare death rates in persons with positive and with negative test results, and to describe methods for using death figures to evaluate multiple screening.

## Background

In the last 3 months of 1954, the Commission on Chronic Illness invited 6,967 selected residents of Baltimore, aged 17 years and older, to attend a clinic for multiple screening. The residents were part of a random sample of the city population who had been reported free from serious health problems on household interview some months previously (*1*).

Of those invited, 2,023 came for screening. An additional 275 persons, who desired to attend the clinic after reading newspaper descriptions of the project, took the tests. Of the total screened, 35 percent were under 35 years of age, 39 percent were between 35 and 49 years, and 26 percent were 50 years of age and older.

*Dr. Wylie is assistant professor of public health administration, Johns Hopkins University School of Hygiene and Public Health, Baltimore, Md.* *Rose Mary Jacobs and Janet Hare contributed significantly to the study. The study was financed in part by grant 1G–6791A from the National Institutes of Health, Public Health Service.*

Seventy-nine percent were white, and 45 percent were males.

For descriptive purposes, the screening tests may be divided into two groups: (*a*) major tests, in which persons were referred to their physicians if the abnormal result was previously unknown; and (*b*) minor tests, for which no referrals were made. Major tests consisted of blood pressure measurement, six-lead electrocardiogram, 70-mm. chest X-ray, blood and urine sugar tests 1 hour after a glucose drink, hemoglobin level, urine albumin, serologic test for syphilis (STS), and two questions about discomfort on exertion. The minor tests included hearing, vision, dental examinations, and height and weight measurement (*2*).

The followup procedures, to determine the status of each person on December 31, 1959, have been fully described elsewhere (*3*). By searching the latest available city and suburban directories for Baltimore and its surrounding area, we obtained for a part of the study group more recent addresses and telephone numbers than appeared in our records. We then mailed a mimeographed letter and a questionnaire to each individual; 52 percent completed and returned the questionnaire. For an additional 23 percent, answers were obtained by telephone. Finally, we searched the 1955–59 death certificate files for the names of all persons who had not been contacted by letter or telephone and to confirm deaths reported in the returned questionnaires.

**Table 2. Number of deaths and crude death rate in screenees by major test result, according to cause of death, 1955–59**

| Cause of death [1] | Number of deaths with major test— | | Crude death rate per 1,000 with major test— | |
|---|---|---|---|---|
| | Positive | Negative | Positive | Negative |
| All causes | 78 | 27 | 73. 3 | 21. 9 |
| Hypertensive heart disease (440–447) | 10 | 3 | 9. 4 | 2. 4 |
| Vascular lesions affecting central nervous system (330–334) | 5 | 0 | 4. 7 | . 0 |
| Other cardiovascular disease (400–434, 450–468) | 28 | 4 | 26. 3 | 3. 2 |
| Malignant neoplasms (140–205) | 12 | 10 | 11. 3 | 8. 1 |
| Other | 23 | 10 | 21. 6 | 8. 1 |

[1] Figures in parentheses refer to International List numbers.

The followup program clarified the current status of 2,031 screenees. For 267 individuals, 12 percent of all screenees, information was incomplete. We believed that they were not then resident in the Baltimore metropolitan area, and we knew that they had not died in

**Table 1. Number persons screened in 1954, number dying during 1955–59, and number of deaths per 1,000 screenees for each test group, by age in 1954**

| Test results | All ages: Screened [1] | All ages: Deaths | All ages: Deaths per 1,000 | Under 35: Screened | Under 35: Deaths | Under 35: Deaths per 1,000 | 35–49: Screened | 35–49: Deaths | 35–49: Deaths per 1,000 | 50 and over: Screened | 50 and over: Deaths | 50 and over: Deaths per 1,000 | Age-adjusted death rate per 1,000 [2] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| All tests | 2, 298 | 105 | 45. 7 | 790 | 5 | 6. 3 | 862 | 20 | 23. 2 | 631 | 80 | 126. 8 | 52. 1 |
| Positive: | | | | | | | | | | | | | |
| Major | 964 | 78 | 80. 9 | 165 | 2 | 12. 1 | 373 | 12 | 32. 2 | 417 | 64 | 153. 5 | 65. 9 |
| Minor | 719 | 24 | 33. 3 | 178 | 1 | 5. 6 | 327 | 7 | 21. 4 | 209 | 16 | 76. 6 | 34. 5 |
| Negative | 615 | 3 | 4. 9 | 447 | 2 | 4. 5 | 162 | 1 | 6. 2 | 5 | 0 | . 0 | 3. 6 |

[1] Includes 15 persons of unknown age.
[2] Calculated by the direct method for a population containing equal numbers of persons under age 35, 35–49, and 50 years and over.

Maryland in the 5 years following screening. In this paper, all of this group are assumed to be alive.

## Mortality

Table 1 distributes the 2,298 persons and the 105 known deaths by age when screened and by the group test results. In each age range, death rates were markedly higher for persons with positive major tests than for persons with positive minor tests or with no positive tests. Since those with positive major tests included the greater proportion of older persons, we have made some adjustment for age to describe the overall mortality experience (footnote 2, table 1). The age-adjusted death rate for persons with positive major tests was twice that for persons with minor positive results and 18 times higher than for those with negative results.

The distribution of causes of death differed considerably between screenees with and screenees without positive major tests. Table 2 shows that 78 of the 105 deaths (74 percent) occurred in persons with positive tests results; 43 of the 50 deaths from cardiovascular causes (86 percent) occurred in that group. Crude death rates for the various disease classes

**Table 3. Number persons screened in 1954, number dying during 1955–59, and deaths per 1,000 screenees for each test, by age in 1954**

| Test result | Age (years) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | All ages | | | Under 35 | | | 35–49 | | | 50 and over | | |
| | Screened | Deaths | Deaths per 1,000 | Screened | Deaths | Deaths per 1,000 | Screened | Deaths | Deaths per 1,000 | Screened | Deaths | Deaths per 1,000 |
| Chest X-ray: | | | | | | | | | | | | |
| Positive | 394 | 49 | 124. 4 | 60 | 2 | 33. 3 | 106 | 4 | 37. 7 | 228 | 43 | 188. 6 |
| Negative | 1, 642 | 39 | 23. 8 | 669 | 3 | 4. 5 | 665 | 13 | 19. 5 | 308 | 23 | 74. 7 |
| Electrocardiogram: | | | | | | | | | | | | |
| Positive | 315 | 51 | 161. 9 | 42 | 1 | 23. 8 | 86 | 3 | 34. 9 | 187 | 47 | 251. 3 |
| Negative | 1, 977 | 54 | 27. 3 | 748 | 4 | 5. 4 | 775 | 17 | 21. 9 | 454 | 33 | 72. 7 |
| Blood pressure: | | | | | | | | | | | | |
| Positive | 167 | 25 | 149. 7 | 13 | 0 | . 0 | 59 | 4 | 67. 8 | 95 | 21 | 221. 0 |
| Negative | 2, 125 | 80 | 37. 6 | 777 | 5 | 6. 4 | 802 | 16 | 20. 0 | 546 | 59 | 108. 1 |
| Questionnaire: [1] | | | | | | | | | | | | |
| Positive | 147 | 24 | 163. 3 | 14 | 1 | 71. 4 | 44 | 1 | 22. 7 | 89 | 22 | 247. 1 |
| Negative | 2, 014 | 73 | 36. 2 | 739 | 3 | 4. 1 | 766 | 18 | 23. 5 | 496 | 52 | 104. 8 |
| Blood sugar: [2] | | | | | | | | | | | | |
| Positive | 80 | 12 | 150. 0 | 11 | 0 | . 0 | 26 | 1 | 38. 5 | 43 | 11 | 255. 8 |
| Negative | 2, 187 | 92 | 42. 1 | 773 | 5 | 6. 5 | 828 | 19 | 22. 9 | 586 | 68 | 116. 0 |
| Urine sugar: [2] | | | | | | | | | | | | |
| Positive | 62 | 11 | 177. 4 | 14 | 0 | . 0 | 19 | 0 | . 0 | 29 | 11 | 379. 3 |
| Negative | 2, 141 | 93 | 43. 4 | 725 | 5 | 6. 9 | 810 | 19 | 23. 5 | 606 | 69 | 113. 9 |
| Urine albumin: | | | | | | | | | | | | |
| Positive | 97 | 20 | 206. 2 | 33 | 0 | . 0 | 24 | 3 | 125. 0 | 40 | 17 | 425. 0 |
| Negative | 2, 111 | 84 | 39. 8 | 709 | 5 | 7. 1 | 804 | 16 | 19. 9 | 598 | 63 | 105. 4 |
| Serologic test for syphilis: | | | | | | | | | | | | |
| Positive | 49 | 5 | 102. 0 | 6 | 0 | . 0 | 21 | 1 | 47. 6 | 22 | 4 | 181. 8 |
| Negative | 2, 163 | 97 | 44. 8 | 759 | 5 | 6. 6 | 813 | 19 | 23. 4 | 591 | 73 | 123. 5 |
| Hemoglobin: | | | | | | | | | | | | |
| Positive | 34 | 0 | . 0 | 20 | 0 | . 0 | 11 | 0 | . 0 | 3 | 0 | . 0 |
| Negative | 2, 213 | 103 | 46. 5 | 759 | 5 | 6. 6 | 834 | 20 | 24. 0 | 620 | 78 | 125. 8 |
| Height and weight: [3] | | | | | | | | | | | | |
| Positive | 581 | 37 | 63. 7 | 126 | 2 | 15. 9 | 223 | 7 | 31. 4 | 232 | 28 | 120. 7 |
| Negative | 1, 712 | 67 | 39. 1 | 664 | 3 | 4. 5 | 636 | 12 | 18. 9 | 412 | 52 | 126. 2 |

[1] Symptoms of cardiac disease.
[2] 1 hour after drinking 50 gm. of glucose.
[3] 30 percent or more above central weight for persons of medium frame, according to Metropolitan Life Insurance Co. tables.

were higher for persons with positive tests, but not significantly so for neoplasms, for which there was no specific screening test.

Some persons, including a few who eventually died, did not take the complete series of tests. Known diabetics, for example, were not asked to take the blood and urine sugar tests, and, due to mechanical failure, some 400 chest X-rays were unsatisfactory. Table 3 shows the number of persons who completed each test and the number who later died, correlated with age and with test result. The findings for the hearing, vision, and dental examinations were not analyzed.

For all tests except hemoglobin level, persons with positive test results had higher death rates than those with negative tests. This finding was consistent in all age groups for the chest X-ray and electrocardiogram. The higher death rates for those with positive height and weight tests occurred only among persons under 50 years of age. In the remaining tests, the higher death rates for positives occurred in the older groups. In the 34 screenees with positive hemoglobin tests, no deaths occurred. As we would expect, mortality was higher among the older screenees, whether test results were positive or negative.

### Indices of Effectiveness

Most studies of multiple screening have measured test performance by effectiveness in separating off persons with specific diseases. In this study, we desired to evaluate each test by its efficiency in separating off those who died in the 5 years following screening.

Table 4, based on figures in table 3, presents five indices which have some value in measuring this performance. Death sensitivity, the percentage of all dead persons classified as positive by each test, is listed in column A. This index can be considered only along with death specificity, which is the percentage of all 5-year survivors classified as negative by each test (column B). Four tests—urine sugar, STS, hemoglobin, and height and weight—give poor results when these indices are considered jointly.

Column C shows the age-adjusted death rates for persons with positive tests. Tests with the highest rates are the urine albumin, urine sugar, questionnaire, and electrocardiogram. The death rate in positives is an inadequate sole criterion for measuring test performance, since it gives a high rank to the urine sugar test, previously evaluated as poor.

Age-adjusted death rates in persons with

**Table 4. Indices of screening effectiveness based on deaths of screenees during 1955–59**

| Test | Death sensitivity [1] (A) | Death specificity [2] (B) | Age-adjusted death rates per 1,000 | | Age-adjusted mortality ratio [3] (E) | Average rank (F) |
|---|---|---|---|---|---|---|
| | | | Positive (C) | Negative (D) | | |
| Chest X-ray | 55. 7 | 82. 3 | 86. 6 | 32. 9 | 2. 6 | 4 |
| Electrocardiogram | 48. 6 | 87. 9 | 103. 3 | 33. 3 | 3. 1 | 2 |
| Blood pressure | 23. 8 | 93. 5 | 96. 3 | 44. 8 | 2. 1 | 6 |
| Questionnaire [4] | 24. 7 | 94. 0 | 113. 7 | 44. 1 | 2. 6 | 3 |
| Blood sugar [5] | 11. 5 | 96. 9 | 98. 1 | 48. 5 | 2. 0 | 7 |
| Urine sugar [5] | 10. 6 | 97. 6 | 126. 4 | 48. 1 | 2. 6 | 5 |
| Urine albumin | 19. 2 | 96. 3 | 183. 3 | 44. 1 | 4. 2 | 1 |
| Serologic test for syphilis | 4. 9 | 97. 9 | 76. 5 | 51. 2 | 1. 5 | 8 |
| Hemoglobin | . 0 | 98. 4 | . 0 | 52. 1 | . 0 | 10 |
| Height and weight [6] | 35. 6 | 75. 1 | 56. 0 | 49. 9 | 1. 1 | 9 |

[1] $\frac{\text{Deaths classified positive}}{\text{All deaths}} \times 100$

[2] $\frac{\text{5-year survivors classified negative}}{\text{All 5-year survivors}} \times 100$

[3] Death rate in positives/death rate in negatives.

[4] For symptoms of cardiovascular disease.

[5] 1 hour after drinking 50 gm. of glucose.

[6] 30 percent or more above central weight for persons of medium frame, according to Metropolitan Life Insurance Co. tables.

negative tests are given in column D. The lower this death rate, the more effective was the test. Ranking high in efficiency are the chest X-ray and electrocardiogram, while in lowest ranks are the STS and hemoglobin tests. The death rate in the negatives is a criterion which correlates fairly closely with the joint consideration of sensitivity and specificity.

Finally, column E gives the mortality ratio, the death rate in positives divided by the death rate in negatives. The urine albumin, electrocardiogram, and chest X-ray tests rank high, and the STS, hemoglobin, and height and weight tests rank low by this criterion.

At present, we can say only that each criterion contributes something to the evaluation of the tests; none seems outstandingly good or bad, and none is adequate for sole consideration. We have therefore produced an average ranking for each test (column F), giving equal weight to each of the five indices. This average ranking suggests that the STS, hemoglobin, and height and weight tests performed so poorly in separating off high mortality groups that they might well be abandoned. Highly effective were the urine albumin, electrocardiogram, chest X-ray, and questionnaire.

## Discussion

As a method of evaluating multiple screening tests, the use of diagnoses made as a result of screening has many defects. In few studies are persons with negative tests examined to determine whether test results are false negatives, and there is considerable evidence that many examinations do not provide sufficient information for making a diagnosis. Medical schools do not stress the diagnosis of asymptomatic disease; therefore, many general practitioners have difficulty in deciding what labels to attach to conditions found in asymptomatic patients with positive diagnostic tests. For example, some physicians may make a diagnosis of heart disease based only on an electrocardiographic abnormality found on screening; other physicians might make a negative diagnosis on the basis of the same finding.

In contrast to medical diagnoses, deaths do not depend on evaluation by physicians. Since groups with high death rates need medical care more urgently than groups with low death rates, screening tests perform a useful function if they refer high mortality groups for medical care. The use of death rates to evaluate screening test performance has, therefore, some validity.

The degree of validity is not perfect, however. Some screening tests aim at conditions which rarely cause death; the hemoglobin level for anemias and intraocular pressure for glaucoma are examples of this group. Mortality studies can underevaluate the success of such tests. Also, some screening tests, such as the chest X-ray for tuberculosis, detect conditions for which medical care is highly effective. A comparison of death rates in persons with both positive and negative test results would again underevaluate the effectiveness of such tests. However, the majority of tests are probably treated fairly by mortality data, since the conditions detected are frequent causes of death and are only moderately affected by presently available medical knowledge.

Should each test be evaluated only in terms of deaths from conditions which should be detected by the test? The answer is probably "No." If the electrocardiogram, for example, successfully separates off persons with noncardiovascular disease as well as cardiovascular cases, that test should be given due credit for this additional yield. Furthermore, a test such as height and weight, not aimed at a particular disease condition, can be evaluated only in terms of deaths from all causes. This is probably the optimum procedure for all tests.

One pioneering study of multiple screening has published mortality data (*4*). Many of the tests used in that study of longshoremen in San Francisco were similar to the tests used in Baltimore. Despite the great differences in the populations tested, the ranking of the effectiveness of the tests is similar. An unexpected finding, both in Baltimore and in San Francisco, has been the superior performance of the urine albumin test. Albuminuria in older groups of the asymptomatic population seems to have a prognosis more grave than was previously suspected.

This followup study in Baltimore suffers from the defect of incomplete information for

12 percent of all screenees. The 267 incompletely traced persons included more persons in the younger age groups, more nonwhites, and more persons with negative tests than those whose current status is accurately known. This defect has probably had little effect on death figures for each test and for the ranking of tests, but it may have produced an artificially low death rate among screenees with negative tests. Thus, the annual crude death rate per 1,000 persons with all tests negative was 1.2 in Baltimore, compared with 4.3 for longshoremen in San Francisco where losses to followup were less severe.

The second major problem in the Baltimore study is the relatively small number of persons screened, and the small number of deaths on which the analysis must be based. Age groups had, therefore, to be broad in this presentation, and even the "age-adjusted" death rates were rather crude adjustments.

Finally, we would briefly mention one theoretical problem involved in using death rates to evaluate screening test performance. A screening test, able to separate off an older group of positives, will produce a group with a higher mortality because of age alone. Only when death rates are compared for positive and negative groups of similar age range will the effectiveness of a test in screening for disease be determined. The age-adjusted mortality ratio discussed in this paper minimizes the effect of separating off the older age groups.

Are we being unfair to the test in making this age adjustment, however? Should a test be given credit for screening for both age and disease, since both conditions result in greater need for medical care? We believe that the answer to both questions is "No," but realize that opinions will differ.

## Summary and Conclusions

To evaluate the effectiveness of multiple screening, this study has used 5-year mortality figures for 2,298 residents of Baltimore who took multiple screening tests in 1954. The study has shown that screening tests separate off high mortality groups who can reasonably be given priority in medical care. This finding gives a sound basis for multiple screening and weakens the contention that all adults should be seen periodically by physicians, not just those with positive screening tests.

Five indices which use death rates in evaluating screening tests have been presented. Each index has good and bad points, and no one index is adequate to form the sole basis for evaluation. The indices suggest that tests with adequate performance were 70-mm. chest X-ray, six-lead electrocardiogram, blood pressure measurement, questionnaire for symptoms of cardiovascular disease, blood and urine sugar, and urine albumin. Tests with poor performance were the serologic test for syphilis, hemoglobin level, and height and weight test.

Little is known of how much persons with positive screening test results are benefited by the early detection of their diseases. It is possible that multiple screening may fail in its ultimate aim to reduce illness, disability, and death in the population, because medical care may not control the diagnosed conditions. Only when the benefits of early detection are clearly established can multiple screening be encouraged widely in the United States.

### REFERENCES

(1) Commission on Chronic Illness: Chronic illness in a large city. Harvard University Press, Cambridge, Mass., 1957.

(2) Roberts, D. W., and Wylie, C. M.: Multiple screening in the Baltimore study of chronic illness. J.A.M.A. 161: 1442–1446, Aug. 11, 1956.

(3) Wylie, C. M.: Participation in a multiple screening clinic with five-year followup. Pub. Health Rep. 76: 596–602, July 1961.

(4) Buechley, R. W., Drake, R. M., and Breslow, L.: Height, weight, and mortality in a population of longshoremen. J. Chronic Dis. 7: 363–378, May 1958.