Some Proposed "Comparability Areas" for U.S. Statistics on Cause of Death

DAVID HEWITT, M.A., JEAN MILNER, M.S.A., and ADELE CSIMA, M.A.

THE EPIDEMIOLOGIST or public health worker who finds some striking geographic contrast in the mortality attributed to a particular disease is often so inhibited by doubts about the uniformity of diagnostic practices that he dare not draw any firm conclusion. As a rule he has neither the resources nor, perhaps, the legal right to pursue his inquiry beyond this inconclusive stage. A highly instructive exception to this rule, however, was provided by Anderson's recent study of geographic variation in deaths due to bronchitis and emphysema in Canada (*1*). At an earlier stage of Anderson's work it had appeared that, in comparison with Canada as a whole and with neighboring Ontario in particular, the province of Manitoba had very high death rates for "emphysema without mention of bronchitis"— I.C.D. 527.1 (*2*) and low rates for chronic bronchitis—I.C.D. 502.0 (*2*). He was therefore faced with the question, "Do physicians in Manitoba . . . fail to diagnose and mention chronic bronchitis in . . . patients who die from emphysema while their confreres in Ontario nearly always do; or do men in Manitoba . . . develop a 'dry' form of emphysema . . . while men in Ontario develop a 'wet' form characterized by bronchitis?"

The authors are with the department of epidemiology and biometrics, University of Toronto, Canada. Mr. Hewitt is an associate professor, Mrs. Milner is a fellow, and Mrs. Csima is an assistant professor.

Not content with leaving the question open, Anderson traced and questioned individual physicians who had prepared certificates assigning death to various chronic respiratory conditions. Their replies revealed that much of Manitoba's apparent excess mortality from emphysema was indeed due to a preference on the part of Manitoba physicians for diagnostic terms less commonly used elsewhere. Moreover, the reason for this preference, as many of the physicians perceived, was that at the University of Manitoba "undergraduate teaching for many years was such as to discourage a physician from making a clinical diagnosis of chronic bronchitis" (*3*). Because of the rapid rise in mortality attributed to emphysema, epidemiologists had, of course, already been alerted to the likelihood that fashion in diagnosis was an important influence on the statistics of death from this cause. A parallel situation is that of the mortality attributed to pulmonary and other venous embolism, and here again the rate for Manitoba, based on A86 of the Intermediate List (*2*), deviates further from the Canadian norm than that for any other Province (*4*).

It might at first seem unlikely that imprinting of a particular viewpoint on undergraduates at one medical school could have a determining influence on the mortality statistics for an entire Province, but more than half of the physicians in Manitoba were trained at this one school. In a random sample of 155 Manitoba physicians drawn from the medical directory

(5), 92—59.4 percent—were found to have graduated from the University of Manitoba. The situation in Manitoba is therefore extreme, but by no means unique. In Arkansas, Indiana, Kentucky, South Carolina, and Vermont, more than 50 percent of all physicians in each State were products of a single medical school in 1959 (6). A high degree of professional self-sufficiency is as likely to give rise to spurious differences between the death rates of States as of Canadian Provinces, and cause-specific rates for these five States should therefore be regarded with special caution. On the other hand, there are certainly States in which the medical corps has been recruited in a much more eclectic fashion, so that less than 10 percent of the physicians come from any one school (as in California, New York, and seven other States) and, with regard to diagnostic practices, these States should be reasonably comparable with one another and with the United States as a whole. There are in addition certain groups (for example, five of the six New England States) within which the distributions by school of origin are similar, even while they differ markedly from the national aggregate.

These considerations suggest that it would sometimes be wise, as an insurance against lack of comparability, to restrict interstate mortality comparisons to groups of States whose physicians exhibit some minimum degree of measured resemblance in respect to their medical schools of origin. In this paper, we suggest several such groups, which we propose to call "comparability areas," by analogy with the registration areas devised by the Bureau of the Census (7), which are still used for some purposes by the National Center for Health Statistics. These comparability areas parallel the registration areas in that: (a) comparisons within one such area are expected to have greater validity, on the average, than comparisons involving jurisdictions outside the areas; and (b) with the passage of time and without any change in the conditions for admission, these areas may be expected to include progressively more States. There is, however, an important distinction. Admission of a State to the registration areas betokens the attainment of some standard of completeness not yet reached by States outside the area. No such connotation, however, is involved in the admission of any State to, or its exclusion from, a comparability area; no judgment of the "quality" of diagnosis is at any time intended or implied. We do not claim to have devised even a partial solution to the vexing problem of comparability in epidemiologic studies, but we hope that our proposal will stimulate further discussion and study.

Calculation of Comparability Index

Although a particular State may legitimately be said to have a high or low average level of comparability, no single axis exists along which States could be ranged in the order of some supposed characteristic of general comparability. Rather, interstate comparability is a property of particular pairs of States, and any measure of it will, in the first instance, have to be calculated separately for every possible pair. Such a measure is conveniently expressed in the form of an index taking values between zero and unity; a value of zero would mean that no medical school is represented in both States; a value of unity would mean that the two States contain an identical "mix" of medical graduates. Between these extremes the comparability index previously suggested elsewhere (4) is defined as

$$C_{12}=\sum(P_{i1}P_{i2})/\sqrt{(\sum P_{i1}^2)(\sum P_{i2}^2)},$$

in which the subscripts 1 and 2 distinguish the States to be compared and P_i is the proportion of physicians in either State who have graduated from the ith medical school. Either term in the denominator tends to be small (and the index value large) when the quantities P_i are all of similar magnitude, as they are, for example, in California but not in Indiana. The index also tends to be large when the larger values of P_{i1} in the numerator coincide with the larger values of P_{i2} (as when Virginia and West Virginia are taken together).

It seems likely that there are varying degrees of resemblance and contrast between particular medical schools. However, for lack of any usable information on this point, we have been obliged in our formulation to treat each school as unique. Consequently a low value of C for a pair of States means only that spurious mortality differences could be important, not that they are positively to be expected. On the other hand, a high value of C does mean that interstate comparison is unlikely to be affected by bias of the type under consideration.

All the figures used for the calculations reported in our paper have been derived from tables in the Health Manpower Source Book No. 11, Medical School Alumni (*6*). These tables show, by State of residence, the number of graduates from each of 78 individual, currently active medical schools; graduates of these 78 schools accounted for 87.2 percent of the physicians located in mid-1959 in the United States, its territories, or on temporary foreign assignment. We were obliged to treat all those who graduated in Canada (2.3 percent) as coming from a single school, likewise those from foreign countries (6.3 percent), those from "extinct" schools (4.4 percent), and those whose school of origin was not ascertained (0.1 percent). Figures were available and used for residents of each of the 50 States, the District of Columbia, Puerto Rico, and the U.S. Territories, but we did not use the figures for Federal physicians. Thus the data for our calculation occupied a table of 82 rows (corresponding to

Membership of complete comparability areas with average internal comparability value of 0.5

State	Area number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Alaska	X	X	X	X	X	X	X	X	X					
Arizona	X	X	X	X	X	X	X	X	X	X				
California	X	X	X	X	X	X	X	X	X	X				
Florida	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Hawaii	X	X	X	X	X	X	X	X	X	X				
Idaho	X	X	X	X	X	X	X	X	X					
Illinois	X	X	X	X	X	X	X	X						
Montana	X	X	X	X	X	X	X	X	X					
Nevada	X	X	X	X	X	X	X	X	X	X				
New Jersey	X	X	X	X	X	X	X	X	X	X				
New Mexico	X	X	X	X	X	X	X	X	X	X	X	X		
North Dakota	X	X	X	X	X	X	X	X	X					
South Dakota	X	X	X	X	X	X	X	X	X					
Utah	X	X	X	X	X	X	X							
Washington	X	X	X	X	X	X	X	X	X					
Wyoming	X	X	X	X	X	X	X	X	X					
Puerto Rico	X	X	X		X	X				X				
U. S. Territories	X	X	X	X	X	X	X	X		X				
Colorado	X													
Minnesota		X												
Missouri			X											
Nebraska				X										
New York					X					X				
Oregon						X								
Texas							X						X	
Connecticut										X				
Delaware								X	X					
Maine										X				
Massachusetts										X				
New Hampshire										X				
North Carolina									X				X	
Pennsylvania								X	X					
Rhode Island										X				
Alabama											X	X		
Georgia														X
Louisiana												X		
Mississippi											X	X		
Tennessee											X			
Virginia												X		
West Virginia												X		

the physicians' schools of origin) and 53 columns (corresponding to their current residence). The output obtained from the computer consisted of a 53 by 53 symmetric matrix of comparability values.

The highest value of C was 0.9058, for the comparison of Maine with Massachusetts; the lowest was 0.0086 for Nebraska with South Carolina. All four States with the highest average values of C (unweighted arithmetic means) were all without any medical school of their own (Arizona, Hawaii, New Mexico, and Nevada) while the fifth (Florida) was also well known as a heavy net importer of physicians from all over the country. At the other extreme, the lowest average values of C were those for South Carolina, Vermont, Arkansas, Georgia, and Indiana. As already mentioned, four of these States received more than 50 percent of their physicians from a single source; two medical schools in Georgia had graduated 60 percent of the State's complement of physicians.

Membership of Comparability Areas

A knowledge of the individual values of C could be directly useful in the planning of a study when the investigator wishes either (*a*) to identify States acceptable for comparison with a particular State in which he is interested (if he were, for example, a public health official in that State) or (*b*) to select groups of States whose mortality statistics may safely be used together in testing some hypothesis, for example, as representatives of mainly agricultural versus industrial populations or of seaboard versus inland regions. However, most geographic studies of mortality are frankly exploratory, seeking to generate rather than to test hypotheses, so that the investigator has no prior interest in comparisons involving particular States. In such a situation, it is desirable to base the study as broadly as possible, but to avoid including regions too disparate in nosology, diagnostic practice, or medical vocabulary. These opposing requirements may best be reconciled in a complete comparability area, which we define as follows:

A comparability area comprises a number of States so selected that the average value of the comparability index for all interstate comparisons within the area exceeds some chosen level; such an area will be called complete if no further State can be introduced without depressing the average internal comparability below the chosen level.

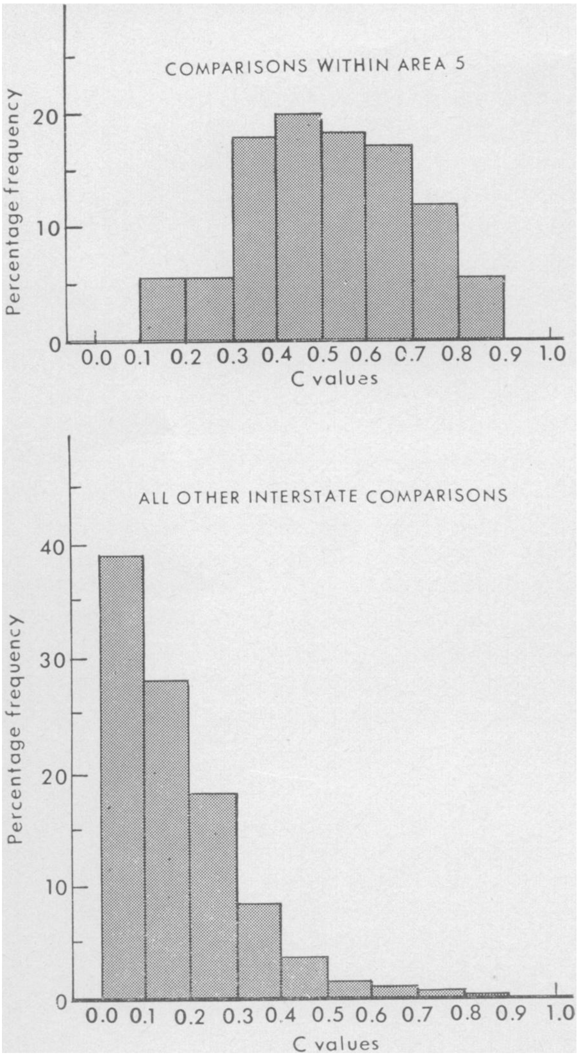
It should be noted that this definition refers only to comparisons between entire States: smaller subdivisions of these areas, such as counties, are likely to have much lower levels of comparability with one another.

A routine for selecting the members of a complete comparability area which has proved amenable to computer treatment is as follows:

1. Select any pair of States for which C exceeds the chosen level.
2. Find the third State in such a way that the average of the three values of C linking members of the first pair to the third State and to one another is as high as possible.
3. Test whether this average value of C exceeds the chosen level and, if it does, find a fourth State in such a way that the six relevant values of C will have as large an average as possible.
4. Repeat the test and, if necessary, continue in the same way to recruit a fifth, sixth, *n*th . . . member until the average value of C does fall below the chosen level.
5. The *n*th State is then disqualified, and the preceding (n—1) are held to constitute a complete comparability area.
6. Select a new starting point and repeat steps 2 to 5.
7. Repeat steps 1 to 6 until all possible starting points have been exhausted.

We have followed this routine, using a chosen level of 0.5 for the average value of C within comparability areas. Some areas with identical membership were reached by different routes (that is, with members recruited in a different sequence). After elimination of such duplications, 14 distinguishable areas remained, ranging in size from two to 19 States (see table). Areas 1 through 7, consisting of 18 or 19 members each, have a common core of 17 members. Only three of these core members (Florida, Illinois, and New Jersey) and only two of the alternate members (Puerto Rico and New York) lie wholly east of the Mississippi River. Presumably this over-representation of western States in the main comparability areas reflects the mingling together in the West of many streams of migrants, including physicians, who

Frequency distribution of a measure of comparability between States, based on the mix of medical graduates in each State



have come from older and more separated communities in the East. Areas 8, 9, and 10 also contain many western States from the core membership of areas 1 to 7, but with the addition of some representation from the northeastern seaboard. In particular, area 10 contains five of the six New England States together with New York. Areas 11 to 14 are smaller groups with some representation of the Southeast. Thirteen States do not appear in the table because they did not qualify for membership in any comparability area at the chosen level of 0.5.

The figure shows the frequency distribution of C for 171 comparisons internal to area 5 and the very different distribution for the balance of all other interstate comparisons. These data give the impression that a mortality study extending over about one-third of the States, if suitably chosen, will be much safer than one in which States are compared indiscriminately. Of course, this safety has to be bought at a price. Deciding to use strictly defined comparability areas would in some ways be like adopting a more stringent criterion of statistical significance; in each instance the object would be to reduce the risk of what statisticians call type 1 error (*8*). In the present context, the type 1 error is one that is made when we give credence to a spurious difference in mortality rates. The price to be paid for a reduction of this risk is some increase in the risk of type 2 errors—those errors made when we fail to detect a genuine difference in mortality. States with the highest degree of mutual comparability, as here defined, also tend to be similar in respect to various factors influencing disease risk. Hence the safest comparisons are not necessarily the most fruitful. Even so, they can yield results of interest. Thus, the most highly comparable pair of States (Maine and Massachusetts) can readily be shown to differ significantly. For example, age-specific comparisons restricted to white males show that the proportion of deaths in Massachusetts attributed to cirrhosis of the liver is about double the corresponding proportion in Maine (*9*).

Discussion

Concern over the effects on mortality statistics of variations in diagnostic fashion inclines some persons to discount all such data as valueless. This judgment, however, seems to us to be overly fastidious. At the other extreme, there may be some persons who believe that problems of comparability become serious only at the level of international comparisons. In support of the latter view, one could point to the likelihood that many of the differences between individual physicians will be canceled out in the statistics for any moderately large population and could also note that evidence of bias arising from differences between groups of physicians is largely anecdotal. Even if such bias were an important factor, it could only be brought to light by extensive cross-tabulation of the causes of death against the characteristics of the physicians signing the death certificates. If this cross-tabulation were ever to be undertaken on a large scale, the results would still be difficult to interpret because the characteristics of the physicians might well be confounded with the real differences in the proportions of patients dying from particular causes. In other words, the uneven distribution by region and by type of practice of physicians from different backgrounds of training and experience has a double consequence. First, it makes the nature and amount of diagnostic variation difficult to measure. Second, it makes the statistical effects of any such variation more serious for epidemologic studies than they would otherwise be.

We have attempted to set up a rationally based position somewhere between the extremes mentioned. In order not to waste expensively acquired data and at the same time not to run unnecessary risks of being misled, we propose, in effect, to grade the data and distinguish those parts which may be misleading from those which deserve greater credence. This grading is not an attempt to rate diagnostic performance on some scale of merit or to adjudicate between differing usages in disease nomenclature. What are graded are the actual building blocks of a statistical mortality study—the comparisons between pairs of rates.

Other proposals for reducing the risk of mistaken inference from mortality statistics have generally been concerned with the reliability of individual rates, or even with the reliability of the individual death certificates contributing to these rates. A number of authors have pointed out that diagnosis tends to be more clear-cut in young patients than in the elderly. Moreover, according to one view (*10*), it may be prudent not to use data on causes of death of persons beyond the age of 65 (a procedure which would exclude 60 percent or more of all deaths). Another approach is to restrict attention to rates based on rather broad cause-of-death categories, so that differences in allocation will occur mainly within, rather than between, categories (*11*). In the method we propose there need be no ban on the study of elderly groups nor on the use of fairly fine diagnostic categories.

The procedure we have outlined could be varied in a number of ways. The chosen level of C could be set lower so as to obtain larger comparability areas, but at some cost to the validity of the results expected. This variation could be achieved with the same source material and the same computer programs that we have already used. Also with the same material, but with some modification of the program, comparability areas could be constituted so that all internal comparisons, not just their average, would have a C value above the chosen level. Again with material from the same source (*6*), but data not used here, differences in physicians' ages could be taken into account, possibly by treating fellow alumni of a medical school as coming from different schools unless they graduated in the same decade. Finally, as H. I. Sauer, supervisory statistician, Heart Disease Control Branch, Public Health Service, suggested in a personal communication to us in 1968, "the hospital in which a physician serves as intern may influence his vocabulary in certifying cause of death even more than does his medical school." It would, of course, be easy to substitute a classification by internships for the classification by undergraduate schools used here—provided that the basic information were available.

Summary

After becoming qualified at a particular medical school, physicians do not disperse uniformly all over the United States but tend to take up practice in circumscribed regions. Because of variations in diagnostic preferences and in the medical vocabulary among medical schools, and consequently among their graduates, these geographic patterns of physician settlement can give rise to spurious differences between States in statistics on causes of death. An index is therefore proposed for measuring the degree of comparability between any pair of States, together with a method for building up "comparability areas" in which interstate comparisons will have some assurance of validity. Fourteen comparability areas are proposed, based on the known geographic distributions of medical school alumni in 1959. All but 13 States have a place in one or more of these areas.

REFERENCES

(1) Anderson, D. O.: Geographic variations in deaths due to emphysema and bronchitis in Canada. Canad Med Assoc J 98: 231–241 (1968).

(2) World Health Organization: Manual of the international statistical classification of diseases, injuries, and causes of death. Revision 7. Geneva, 1957.

(3) Anderson, D. O.: Observations on the classification and distribution of pulmonary emphysema in Canada. Canad Med Assoc J 89: 709–716 (1963).

(4) Hewitt, D.: Mortality from cardiovascular-renal diseases in Ontario and elsewhere. 1. Descriptive. 2. Analytic. J Chronic Dis 21: 323–340 (1968).

(5) Feasby, W. R., editor: Tenth annual Canadian medical directory. Seccombe House, Toronto, 1964.

(6) Stewart, W. H., and Pennell, M. Y.: Health manpower source book. Medical school alumni. PHS publication No. 263, sec. 11. U.S. Government Printing Office, Washington, D.C., 1961.

(7) Linder, F. E., and Grove, R. D.: Vital statistics rates in the United States 1900–1940. U.S. Government Printing Office, Washington, D.C., 1943, pp. 95–100.

(8) Hoel, P. G.: Introduction to mathematical statistics. Ed. 3. John Wiley & Sons, Inc., New York, 1962.

(9) National Office of Vital Statistics: Vital statistics of the United States for the years 1961–1965. U.S. Government Printing Office, Washington, D.C., 1963–1967.

(10) Sigurjonsson, J.: Index rates for comparing the importance of arteriosclerotic and degenerative heart diseases as a cause of death. Amer J Med Sci 250: 395–401 (1965).

(11) Reid, D. D., and Rose, G. A.: Assessing the comparability of mortality statistics. Brit Med J No. 5422: 1437–1439, Dec. 5, 1964.

Tearsheet Requests

Dr. David Hewitt, Department of Epidemiology and Biometrics, School of Hygiene, University of Toronto, Toronto 5, Canada

Public Health Service Staff Appointment

Dr. Raymond T. Moore has been appointed associate commissioner of the Environmental Control Administration, an element of the Consumer Protection and Environmental Health Service of the Public Health Service.

Dr. Moore was acting director of the Bureau of Radiological Health since 1968 and deputy director of the radiological health program since January 1967.

Dr. Moore practiced general medicine in Sequin, Tex., from 1949 to 1958. He has had broad experience in occupational health, industrial medicine, and radiological health. Following studies under an Atomic Energy Commission fellowship in industrial medicine at the University of Rochester, Rochester, N.Y., he joined the Public Health Service as a commissioned officer.

His early assignments in the Service included studying the effects of low-level radiation on radiation workers, lecturing on occupational health at the University of Pittsburgh, and advising on medical aspects of radioactive fallout.

Between 1963 and 1967 Dr. Moore served as medical officer at the Nevada Test Site, project officer for two research grants in medical diagnostic radiology, national coordinator for the Medical Liaison Officers Network, and program director in radiological health for Public Health Service Region VII in Dallas, Tex. He also represented the Public Health Service in planning and implementing the health program for the NS *Savannah.*

Dr. Moore was born and educated in Arkansas. He received his B.S. degree from Arkansas State Teachers College at Conway in 1939 and his M.D. from the University of Arkansas in 1944. He served his internship at the Baptist Memorial Hospital in San Antonio, Tex., in 1945. Dr. Moore also received a master of industrial science degree from the University of Rochester in 1959. He is a member of the American Academy of Occupational Medicine, the American Medical Association, the American Public Health Association, and the Industrial Medical Association.