frequency of face-to-face interaction, preceded by the increasing structural differentiation and the separation of various functional activities. Furthermore, the association is confronted with the problem of enforcing the pertinent features of its program through a relatively expensive outlay. Thus with an increase of controls there is a corresponding increase of staff and administrative expenditures. This phenomenon of increased controls, efforts to improve communication, and the use of additional professional help can be found emerging not only when the membership declines, but also when the membership increases at a rapid rate.

(9) Increase of expenditures can also be explained in terms of increased capital outlay to improve facilities in order to maintain organizational prestige.

(10) With a decline in the membership of an organization, there is no immediate or actual decline in income. This phenomenon is due primarily to the greater efficiency of the organization in collecting dues and carrying out financial drives.

(11) Material property will increase over a period of time, and this increase is closely related to the expenditures for staff and upkeep. Unless the material property is withdrawn from use or permitted to deteriorate, these service expenditures cannot be reduced below a certain level.

A METHODOLOGICAL ANALYSIS OF SEGREGATION INDEXES *

OTIS DUDLEY DUNCAN AND BEVERLY DUNCAN
University of Chicago

THERE have been proposed in the literature several alternative indexes of the degree of residential segregation of the nonwhite population of a city.[1] This paper shows that all of these can be regarded as functions of a single geometrical construct, the "segregation curve." From this there are developed several important implications: (1) The proposed indexes of segregation have a number of hitherto unnoticed interrelationships which can be mathematically demonstrated. (2) Some of them have mathematical properties of which their proponents were unaware, and which lead to difficulties of interpretation. (3) As a consequence, the status of the empirical work already done with segregation indexes is questionable, and their validity for further research is undetermined.

This paper consists of a summary of the mathematical analysis made of segregation indexes and of a documentation of the conclusions listed. The problem of validating segregation indexes is viewed as one of some importance, not only in its own right, but also as an illustration of the difficulties in finding an adequate rationale for much sociological research using index numbers.

THE SEGREGATION CURVE

Consider the k census tracts of a city. The i^{th} tract contains N_i nonwhites and W_i whites, totalling to $N_i + W_i = T_i$. Summing over i, $\sum_1^k N_i = N$, $\sum_1^k W_i = W$, and $\sum_1^k T_i = T$. For each tract compute the nonwhite pro-

* Revision of paper read at a meeting of the Midwest Sociological Society, April, 1954. The clerical assistance of Florence Sugeno, Richard W. Redick and Robert Glassburg is gratefully acknowledged, as is the financial assistance of the Social Science Research Committee of the University of Chicago. This research was supported in part by the U. S. Air Force under Contract Number AF 33 (038)-25630, monitored by the Human Resources Research Institute. Permission is granted for reproduction, translation, publication and disposal in whole and in part by or for the U. S. Government.

[1] Donald O. Cowgill and Mary S. Cowgill, "An Index of Segregation Based on Block Statistics," *American Sociological Review*, 16 (December, 1951), pp. 825–831; Julius A. Jahn, "The Measurement of Ecological Segregation: Derivation of an Index Based on the Criterion of Reproducibility," *American Sociological Review*, 15 (February, 1950), pp. 100–104; Julius A. Jahn, Calvin F. Schmid, and Clarence Schrag, "The Measurement of Ecological Segregation," *American Sociological Review*, 12 (June, 1947), pp. 293–303.

portion, $q_i = N_i/T_i$, and array the tracts in ranks 1 to k in order of magnitude of q_i. With this ordering compute tract by tract the cumulative proportions of nonwhites and whites, letting the cumulative proportion of nonwhites through the i^{th} tract be X_i and the cumulative proportion of whites be Y_i, e.g., $X_2 = (N_1 + N_2)/N$, $Y_2 = (W_1 + W_2)/W$. The segregation curve is the function $Y_i = f(X_i)$, as graphed in Figure 1. The observed segregation curve, together with the nonwhite proportion in the entire city, $q = N/T$, contains all the information involved in the calculation of any of the segregation indexes suggested in the literature. As is suggested below, research on segregation is not likely to progress far with the research problem limited to the study of this information alone.

DEFINITION OF INDEXES

We here define, with reference to the segregation curve, the several indexes proposed in the literature, omitting the proofs of the equivalence of our definitions and those originally given. In all cases these proofs involve only elementary algebra and geometry.

The "Gini Index," *Gi*,[2] is the area between the segregation curve and the diagonal of Figure 1, expressed as a proportion of the total area under the diagonal. It can also be defined as the "mean cost rating" of the cost-utility curve with Y = cost and X = utility;[3] or as the weighted mean difference with repetition,[4] of the tract nonwhite proportions, q_i, divided by the mean difference, $2pq$, of the binomial variable of color, for the total city population, scoring each white person unity and each nonwhite zero (where $q = N/T$, $p = 1 - q$). The simplest formula for computing *Gi* is $\sum_1^k X_{i\text{-}1}Y_i - \sum_1^k X_iY_{i\text{-}1}$, keeping the tracts in the order established for constructing the segregation curve.

The "Nonwhite Section Index,"[5] here denoted D, for dissimilarity or displacement,[6] is the maximum vertical distance between the diagonal and the curve in Figure 1, i.e., the maximum of the k differences $(X_i - Y_i)$. Alternatively, suppose there are s tracts for which $q_i \geqq q$; then $D = X_s - Y_s$. If x_i and y_i are the *uncumulated* proportions of the city's nonwhites and whites, i.e., $x_i = N_i/N$, $y_i = W_i/W$, then $D = \frac{1}{2} \sum_1^k |x_i - y_i|$. Furthermore, D is the weighted mean deviation from q of the tract proportions, q_i, divided by the mean deviation, $2pq$, for the total population. It may be interpreted as the proportion of nonwhites who would have to change their tract of residence to make $q_i = q$ for all i (hence the term, displacement).

Our interest in the Cowgills' index[7] is confined to the mathematical form of the index, without regard to the important but logically distinct issue of the appropriate size of area units. The general form of the index is the ratio of the number of areas occupied exclusively by whites to the maximum number of areas which could be so occupied. To obtain a relationship to the segregation curve we have considered a slight further generalization: the ratio of the number of *persons* living in exclusively white areas to the total whites in the city. The generalized Cowgill Index (*Co*) is then the length of that segment of the curve, if any, which coincides with the vertical drawn from (1,0) to (1,1) (see Figure 2).

It may be noted that the foregoing indexes can be described as measuring directly the degree of departure of the segregation curve from the diagonal, which is the norm of even distribution. Other such indexes of "unevenness" could doubtless be suggested.[8] The remaining indexes proposed in the literature can be related to the segregation curve only by explicitly introducing the city non-

[2] Jahn, Schmid, and Schrag, *op. cit.*, Index #3.

[3] Otis Dudley Duncan, "Urbanization and Retail Specialization," *Social Forces,* 30 (March, 1952), pp. 267–271; Otis Dudley Duncan, Lloyd E. Ohlin, Albert J. Reiss, Jr., and Howard R. Stanton, "Formal Devices for Making Selection Decisions," *American Journal of Sociology,* 58 (May, 1953), pp. 573–584.

[4] Maurice G. Kendall, *The Advanced Theory of Statistics,* London: Griffin and Co., 3d ed., 1947, Vol. I, Ch. 2.

[5] Jahn, Schmid, and Schrag, *op. cit.,* Index #4; Josephine J. Williams, "Another Commentary on So-Called Segregation Indices," *American Sociological Review,* 13 (June, 1948), pp. 298–303.

[6] Donald J. Bogue, *The Structure of the Metropolitan Community,* Ann Arbor: University of Michigan, 1949; Edgar M. Hoover, Jr., "Interstate Redistribution of Population, 1850–1940," *Journal of Economic History,* 1 (November, 1941), pp. 199–205.

[7] Cowgill and Cowgill, *op. cit.*

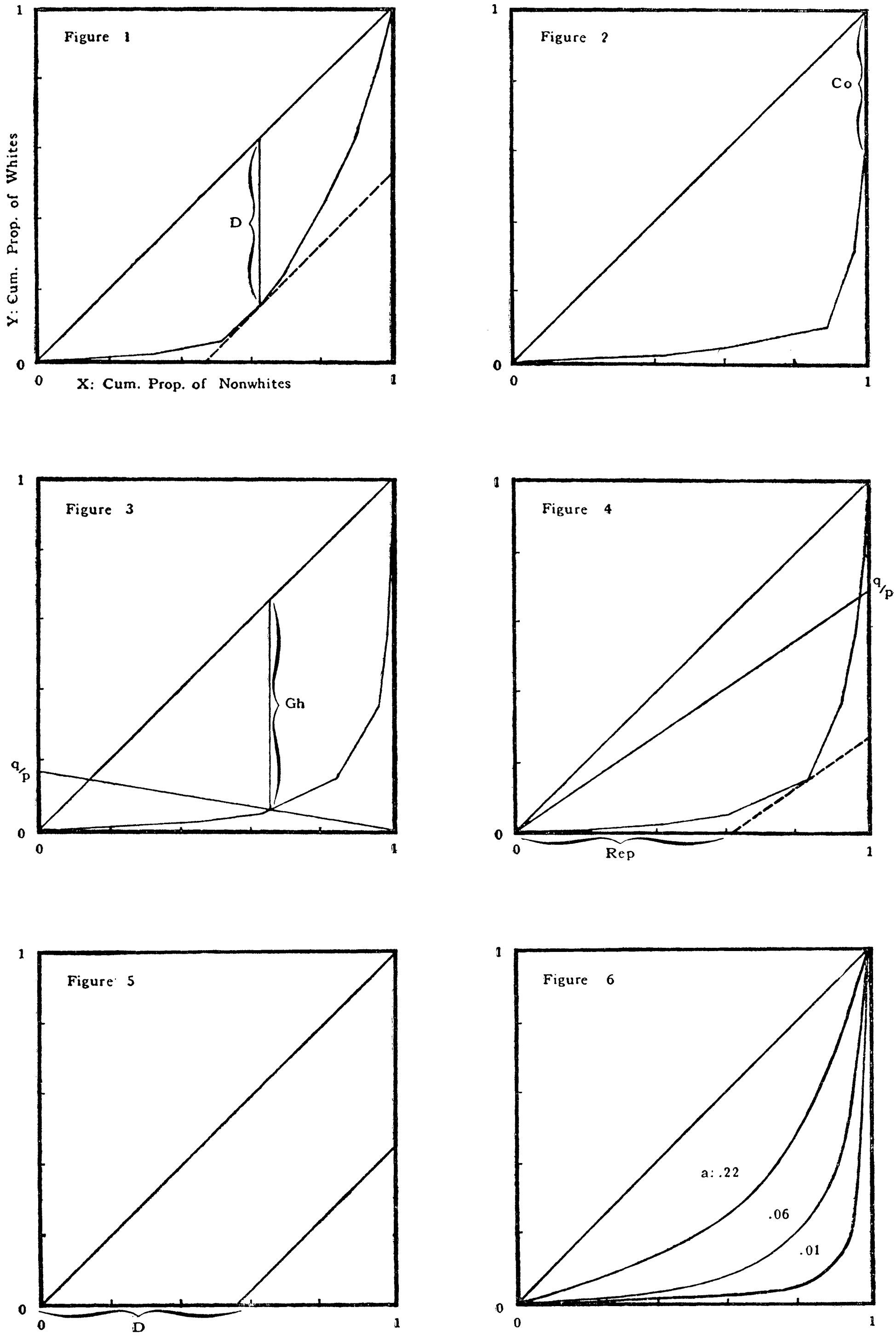

FIGURE 1. "Section" index in relation to segregation curve (curve for Macon, Ga., 1940 $D = .47$).
FIGURE 2. Generalized Cowgill index (Syracuse, N. Y., $Co = .42$).
FIGURE 3. "Ghetto" index (Louisville, Ky., $Gh = .60$).
FIGURE 4. "Reproducibility" index (Birmingham, Ala., $Rep. = .62$).
FIGURE 5. Williams' model of the segregation curve, with $D = .56$.
FIGURE 6. Hyperbola model, $Y = aX/(1 - bX)$, for selected values of a.

white proportion, q. We assume throughout that $q \leqq .5$.

The "Nonwhite Ghetto Index," Gh,[9] is found graphically by plotting the line $Y = q(1 - X)/p$. The index value is then $(X_g - Y_g)$, denoting by (X_g, Y_g) the point where this line intersects the segregation curve (see Fig. 3).

The "Reproducibility Index," Rep,[10] is formally identical with the index of efficiency used in prediction work.[11] To obtain Rep graphically construct the line parallel to $Y = qX/p$ which is "tangent" to the segregation curve, i.e., which intersects but one point of the curve or which coincides with that segment (if any) which has a slope q/p. Then the value of Rep is the X-intercept of the auxiliary line (see Figure 4).

The correlation ratio of the binomial variable, color, on tract is by definition the square root of the variance between tract proportions divided by the total variance of the population; i.e.,

$$eta = \sqrt{\frac{\sum_1^k T_i q_i^2}{Tpq} - \frac{q}{p}}.$$

In the case of a binomial variable, *eta* is identical with the mean square contingency coefficient, *phi*,[12] and is equal within a very close approximation to the intraclass correlation.[13] It is, therefore, a well known statistic, appearing, e.g., in Robinson's formula for ecological correlation as a measure of "clustering by area."[14] The "revised index of isolation" recently suggested by Bell[15] is identical with the square of *eta*. This term, as well as the term "segregation score" used by Jahn *et al.*,[16] seems somewhat superfluous. It also seems undesirable to restrict the use of *eta* to the case of tracts of equal size, as the latter authors do. Unlike the other indexes, *eta* involves a squared term, and no simple geometric relationship of *eta* to the segregation curve has been found. That a relationship exists is, however, indicated below.

INTERRELATIONS OF INDEXES

Previous work has made it clear that the foregoing indexes are not independent. Jahn *et al.*[17] found moderate to high empirical correlations among four of them. Hornseth[18] demonstrated that the same four could all be expressed in formulas of one general type. Williams found relationships among them for a segregation curve of a specified type.[19]

In fact, there are definite relationships among the several indexes which hold irrespective of the form of the segregation curve, and which can be derived formally without reference to data. For example, the minimum value of Gi is D, and the maximum is $2D - D^2$. Some of the other relationships which have been found are the following:

(1) $qD/(1 - pD) \leqq Gh \leqq D$

(2) $1 - p(1 - D)/q \leqq Rep \leqq D$, unless the term on the left is negative, in which case $Rep = 0$.

(3) $Gh \leqq eta \leqq \sqrt{Gh}$

(4) $1 - 2p(1 - Gh) \leqq Rep \leqq Gh/(p + qGh)$

(5) Co is the minimum value of Gi and D; $qCo/(1 - pCo)$ is the minimum of Gh; and if $Co > 1 - q/p$, then $Rep \geqq 1 - p(1 - Co)/q$.

(Some values of *eta* and *Gh* reported by Jahn *et al.* which are inconsistent with the third relationship are to be attributed to their use of a formula for *eta* not weighted for tract size, or to computational errors.)

From the above it is clear that there is almost necessarily a high correlation between D and Gi, as well as between *eta* and Gh. On the other hand the correlation between

[8] Leo A. Goodman, "On Urbanization Indices," *Social Forces*, 31 (May, 1953), pp. 360–362.

[9] Jahn, Schmid, and Schrag, *op. cit.*, Index #1; Williams, *op. cit.*, p. 301.

[10] Jahn, *op. cit.*

[11] Lloyd E. Ohlin and Otis Dudley Duncan, "The Efficiency of Prediction in Criminology," *American Journal of Sociology*, 54 (March, 1949), pp. 441–451.

[12] Williams, *op. cit.*

[13] Leslie Kish, "On the Differentiation of Ecological Units," Ph.D. dissertation, University of Michigan, 1952.

[14] W. S. Robinson, "Ecological Correlations and the Behavior of Individuals," *American Sociological Review*, 15 (June, 1950), pp. 351–357.

[15] Wendell Bell, "A Probability Model for the Measurement of Ecological Segregation," *Social Forces*, 32 (May, 1954), pp. 357–364.

[16] Jahn, Schmid, and Schrag, *op. cit.*, Index #2.

[17] Jahn, Schmid, and Schrag, *op. cit.*

[18] Richard A. Hornseth, "A Note on 'The Measurement of Ecological Segregation' by Julius Jahn, Calvin F. Schmid, and Clarence Schrag," *American Sociological Review*, 12 (October, 1947), pp. 603–604.

[19] Williams, *op. cit.*, p. 302.

D and *Gh* need not be high if there is considerable variation in *q*. These mathematical relationships, therefore, appear to account satisfactorily for the empirical intercorrelations of the index values reported by Jahn *et al.*

A MODEL FOR THE SEGREGATION CURVE

Williams' model of the segregation curve, referred to above, is represented geometrically by the line parallel to the diagonal of the graph which includes the points $(D, 0)$ and $(1, 1 - D)$ (see Figure 5). For such a curve Williams showed that $Gi = 2D - D^2$, $Gh = D$, and $eta = \sqrt{D}$. It can also be shown that $Co = D$ and $Rep = D$.

The Williams model, although a useful analytical construct, does not serve well to describe empirical segregation curves. We have worked with an alternative model which assumes that the segregation curve has the form of a hyperbola, $Y = aX/(1 - bX)$, where a and b are both nonnegative, and $a + b = 1$. Figure 6 shows selected curves of this form. It can be seen that the smaller the value of a, the greater is the degree of segregation, as measured by *Gi* or *D*. For a segregation curve of the form described, it can be shown that the segregation indexes have the following formulas:

(1) $Gi = 1 + 2a(b + \log_e a)/b^2$.

(2) $D = (1 - \sqrt{a})/(1 + \sqrt{a})$

(3) $Co = 0$

(4) $Rep = (\sqrt{ap} - \sqrt{q})^2/bq$, when $q/p \geqq a$, and $Rep = 0$, when $q/p < a$.

(5) $$Gh = 1 - \frac{\sqrt{4abpq - a^2} - a}{2bpq}$$

No exact formula for *eta* has been found, but in view of the relationship between *eta* and *Gh* stated in the preceding section, the following is suggested as a close approximation:

(6) $eta = (Gh + \sqrt{Gh})/2$, where the value of *Gh* is taken from the formula just given.

To fit the hyperbola to census tract data for a city, *D* was calculated from the data, and the parameters of the curve were computed from the formulas $b = 1 - a$, and $a = (1 - D)^2/(1 + D)^2$, the latter being the solution for a of the second formula in this section.

A hyperbola was fitted to the census tract data for each of the 60 tracted cities of the United States in 1940. Then, predicted values of each of the indexes were computed, entering these parameters and the corresponding observed q's in the foregoing formulas. The results summarized in Table 1 make it clear that most of the segregation indexes can be predicted rather closely, given *D, q,* and the assumption that the segregation curve conforms to the hyperbola model. The largest errors occur for *Co,* since the hyperbola model requires this index to be zero. Even so, the model does not sacrifice a great deal of information. There are 18 cities whose *Co* index is actually zero, and another 13 have values of *Co* under .05. An index of this type is evidently of little use in comparing cities on the basis of census tract data.

TABLE 1. SELECTED MEASURES OF CLOSENESS OF FIT OF ACTUAL SEGREGATION INDEX VALUES TO VALUES CALCULATED FROM THE HYPERBOLA, FOR 60 TRACTED CITIES: 1940

Segregation Index	Mean Arithmetic Error	Mean Absolute Error	Root-Mean-Square Error	Correlation Index [1]
D	0.0	0.0	0.0	1.0
Gi	—0.006	0.015	0.019	.980
Gh	—0.003	0.044	0.066	.945
Rep	0.028	0.065	0.102	.908
Eta	—0.016	0.039	0.054	.958
Co	0.078	0.078	0.124	...

[1] Correlation index $= \sqrt{1 - (\Sigma d^2)/k \cdot Var(I)}$, where the d's are differences between corresponding actual and calculated values of a segregation index, and $Var(I)$ is the variance of the actual values. The correlation indexes for this problem are slightly lower than the Pearsonian correlations between actual and observed values.

Two important conclusions may be drawn from the experience with the hyperbola model. First, there appears to be a characteristic form for the segregation curves of most large American cities, despite the considerable variation among them in degree of segregation. Second, for this universe of cities, there is little information in any of the indexes beyond that contained in the index, *D*, and the city nonwhite proportion, *q*. Each of the other indexes can be obtained to a close approximation given *D, q,* and the assumption of the hyperbolic form of the segregation curve. This conclusion might require modification if area units other than tracts were employed; and it has not been checked for any date except 1940.

EMPIRICAL CONSEQUENCES OF THE CHOICE OF AN INDEX

A number of criteria have been offered for the choice of an index formula,[20] with no consensus on the matter having been reached. In our judgment it has been insufficiently emphasized that the empirical results obtained with an index may be strongly affected by its mathematical properties.

For example, Jahn *et al.* report several correlations between *Gh* and other variables,[21] of which three are large enough to be statistically significant and possibly of theoretical importance. We have reworked these correlations with a somewhat different sample of 46 cities, calculating segregation indexes for the nonwhite, rather than the Negro, population.

The correlation of *Gh* with the 1939–40 crude death rate from tuberculosis for these cities was .58. A much higher correlation, .83, was found between the city nonwhite proportion, q, and the tuberculosis death rate. Further, since q and *Gh* correlated .51, the partial correlation between *Gh* and the tuberculosis death rate, with q held constant, dropped to .32, which is on the borderline of statistical significance at the .05 level. Or, when the tuberculosis death rate was standardized for color, the correlation with *Gh* dropped to .29, which is of doubtful significance. In addition it was found that *Gi* correlated only —.02 and .06, respectively, with the crude and standardized tuberculosis death rates. It appears, therefore, that the tuberculosis death rate responds primarily to the proportion nonwhite, rather than to the degree of nonwhite segregation, and that the correlation obtained with *Gh* reflects primarily the correlation of *Gh* with q. The difference between the results obtained with *Gh* and *Gi* follows from the fact that *Gi* depends only on the segregation curve, while *Gh* involves not only the curve but also the nonwhite proportion, q.

It was also found that the correlation between *Gh* and the per cent of overcrowded housing, .40, dropped to .02 with q partialled out, and that the correlation with *Gi* was a nonsignificant —.13, with q being a better predictor than either *Gh* or *Gi,* correlating .76 with per cent of overcrowding. Similarly the r of —.39, between *Gh* and Thorndike's "G" was clearly attributable to the high correlation, —.87, between q and "G," and even changed in sign to an r of .11 when q was partialled out. The correlation of *Gi* with "G" was a nonsignificant .18.

These results indicate that the ability of a segregation index to predict other variables is an insufficient criterion of its worth. If we wished to predict the tuberculosis death rate, the percentage of overcrowding, or Thorndike's "G," we would use none of the segregation indexes, but instead, q, as the predictor. Yet the theoretically interesting question would remain, e.g., is the tuberculosis death rate associated with the segregation of nonwhites? To investigate this question we require a measure of segregation whose validity is established independently of its correlation with the death rate.

INADEQUACIES OF SEGREGATION INDEXES

The literature on segregation indexes contains references to a number of difficulties in their use and interpretation. Some of these require additional comment.

All of the segregation indexes have in common the assumption that segregation can be measured without regard to the spatial patterns of white and nonwhite residence in a city. Yet it is common knowledge that in some cities—e.g. Chicago—the nonwhite population is predominantly clustered in a "Black Belt," whereas in other cities nonwhite occupancy takes the form of scattered "islands" or "pockets." Surely whatever variables of ecological organization and change are related to the degree of segregation must also be affected by the spatial pattern of segregation. We have found that in 1940, in 50 of the 51 nonsuburban tracted cities studied, the nonwhite population was more concentrated toward the ecological center of the city than the white population. (This study involved the use of an "index of centralization," which has not yet been described in pub-

[20] Cowgill and Cowgill, *op. cit.*; Hornseth, *op. cit.*; Jahn, *op. cit.*; Jahn, Schmid, and Schrag, *op. cit.*; Julius A. Jahn, Calvin F. Schmid, and Clarence Schrag, "Rejoinder to Dr. Hornseth's Note on 'The Measurement of Ecological Segregation,'" *American Sociological Review,* 13 (April, 1948), pp. 216–217; Kish, *op. cit.*; and Williams, *op. cit.*

[21] "The Measurement of Ecological Segregation," *op. cit.*, pp. 302–303.

lished work, but which is in some respects similar to the Gini segregation index.) Other aspects of the spatial pattern of segregation need to be studied as well. It seems unlikely that any single index of segregation will be found sufficient for the purposes of such research.

In none of the literature on segregation indexes is there a suggestion about how to use them to study the *process* of segregation or change in the segregation pattern. As a first step in this direction we have experimented with an adaptation of the method of expected cases to determine (1) how much of the segregation of nonwhites can be attributed to differences between whites and nonwhites in income, occupational status, and rentals paid, and (2) whether changes in degree of segregation over a ten-year period are related to changes in these variables. It has been found, for example, that there is a marked difference between southern and northern cities in the influence on residential segregation of white-nonwhite differentials in labor force and occupational status. When this variable is held constant, the Gini index of segregation is reduced by 12 to 22 per cent in most southern cities, but by only two to nine per cent in most of the cities of the North. Such preliminary findings indicate the advisability of taking account of socio-economic factors in analyzing differences in residential segregation. There is, further, a need for research to develop mathematical and empirical bases for anticipating the effects on measures of segregation of such changes as a marked increase in the nonwhite proportion, an improvement of the nonwhite's relative socio-economic status, or an invasion-succession sequence.

The problem of the appropriate areal unit for research on segregation has been forcefully stated by the Cowgills.[22] As they imply, it is easy to gerrymander tract boundaries to increase or decrease the apparent degree of segregation. However, the problem cannot be solved merely by reducing the size of areal units, e.g., to blocks. The objections made to the census tract basis apply also, *mutatis mutandis,* to blocks. For example, if all nonwhites resided on alleyways and all whites in street-front structures, then even a block index would fail to reveal the high degree of segregation. The most complete discussion of the problem of area unit has been given by Wright,[23] who has indicated the formidable difficulties in the way of finding segregation measures which are not relative to the system of area units used.

IMPLICATIONS FOR INDEX THEORY

In this paper we have not sought to formulate a comprehensive set of criteria for determining the validity of a segregation index. Probably such an attempt would be premature in the present stage of empirical investigation and conceptualization of the phenomenon of segregation. In our judgment the criteria thus far suggested in the literature fall short of comprehensiveness, and not all of them are likely to be generally accepted. However, we feel that the work reported here is relevant to the problem of validating segregation indexes, and possibly suggestive for the general problem of index development in social research.

Specifically, we have established the following: (1) There was a lack of clarity and consistency in the specifications for a segregation index originally proposed. Jahn *et al.* suggested that "a satisfactory measure of ecological segregation should . . . not be distorted by . . . the proportion of Negroes."[24] Though it has never been made clear what would constitute "distortion" in this respect, it is apparent from our analysis that the nonwhite proportion, q, does enter into the formulas for such indexes as *Gi, D, Gh,* and *Rep,* and that q is involved in different ways in the several formulas. For example, if the segregation curve remains constant, but q changes, then the values of *Gh, Rep,* and *eta* will be affected, but not those of *Gi, D,* and *Co*. As yet there is no criterion to determine which of these is the more desirable property for an index number. Lacking such a criterion it is perhaps doubtful whether a meaningful comparison can be made of the degrees of segregation of two cities with greatly different q's. (2) The empirical correlations among alternative indexes are clarified by determining the mathematical relationships holding among them—either in general, or under the assumption of a par-

[22] Cowgill and Cowgill, *op. cit.*

[23] John K. Wright, "Some Measures of Distribution," *Annals of the Association of American Geographers,* 27 (December, 1937), pp. 177–211.

[24] "The Measurement of Ecological Segregation," *op. cit.*, p. 294.

ticular model of the segregation curve. Thus the assessment of segregation indexes is carried a step beyond the speculation of Jahn *et al.* that their four indexes comprise "two sets of independently discriminative measures," [25] and Hornseth's erroneous judgment that these four indexes "are for practical purposes identical measures." [26] (3) The mathematical analysis of segregation index formulas discloses the areas of redundancy and ambiguity among them, i.e., it permits a conclusion as to the circumstances in which two indexes will give interchangeable results and in which they will give incompatible results. Hence it goes beyond the truism that the empirical results obtained with an index are in part a function of the mathematical form of the index, to indicate what specific property of the index is responsible for the kind of results obtained. Thus, for example, the finding that Southern cities are more segregated than Northern can be properly qualified if it is known that the index being used responds in a certain way to variations in q, and that Southern cities generally have higher nonwhite proportions than Northern.

A difficult problem of validation is faced by the proponent of a segregation index formula. The concept of "segregation" in the literature of human ecology is complex and somewhat fuzzy, i.e., the concept involves a number of analytically distinguishable elements, none of which is yet capable of completely operational description. Yet it is a concept rich in theoretical suggestiveness and of unquestionable heuristic value. Clearly we would not wish to sacrifice the capital of theorization and observation already invested in the concept. Yet this is what is involved in the solution offered by naive operationism, in more or less arbitrarily matching some convenient numerical procedure with the verbal concept of segregation. The problem must be faced of considering a variety of possible selections of data and operations on these data in an effort to capture methodologically what is valuable in the work done with the concept prior to the formulation of an index. As we have suggested, it may be that no single index will be sufficient, because of the complexity of the notion of segregation, involving as it does considerations of spatial pattern, unevenness of distribution, relative size of the segregated group, and homogeneity of sub-areas, among others. In short, we are emphasizing the distinction between the problems of (a) working from a limited set of data to a mathematically convenient summary index, and (b) working from a theoretically problematic situation to a rationale for selecting and manipulating data.

Lazarsfeld and his co-workers [27] have taken the lead in a much needed effort to codify the procedures by which concepts are so specified that index construction may profitably be undertaken and to rationalize the decisions involved in formulating an index for research use. Sociologists will learn, as economists have, that there is no way to devise adequate indexes which avoids dealing with theoretical issues. Incidentally, one lesson to be learned from the relatively unproductive experience with segregation indexes to date is that similar problems are often dealt with under different headings. Most of the issues which have come up in the literature on segregation indexes since 1947 had already been encountered in the methodological work on measures of inequality, spatial distribution, and localization in geography and economics.[28]

---

[25] *Ibid.*, p. 299.

[26] Hornseth, *op. cit.*, p. 604.

[27] Paul F. Lazarsfeld and Allen H. Barton, "Qualitative Measurement in the Social Sciences: Classification, Typologies, and Indices," in *The Policy Sciences,* edited by Daniel Lerner and Harold D. Lasswell, Stanford: Stanford University Press, 1951.

[28] P. Sargant Florence, W. G. Fritz, and R. C. Gilles, "Measures of Industrial Distribution," Ch. 5 in National Resources Planning Board, *Industrial Location and National Resources,* Washington: Government Printing Office, 1943; Edgar M. Hoover, Jr., "The Measurement of Industrial Localization," *Review of Economic Statistics,* 18 (November, 1936), pp. 162–171; Wright, *op. cit.*; John K. Wright, "Certain Changes in Population Distribution in the United States," *Geographical Review,* 31 (July, 1941), pp. 488–490; Dwight B. Yntema, "Measures of the Inequality in the Personal Distribution of Wealth or Income," *Journal of the American Statistical Association,* 28 (December, 1933), pp. 423–433. See also Federal Housing Administration, *The Structure and Growth of Residential Neighborhoods in American Cities,* by Homer Hoyt, Washington: Government Printing Office, 1939, Ch. 5; P. Sargant Florence, *Investment, Location, and Size of Plant,* Cambridge: Cambridge University Press, 1948; and George C. Smith, Jr., "Lorenz Curve Analysis of Industrial Decentralization," *Journal of the American Statistical Association,* 42 (December, 1947), pp. 591–596.