

Chapter 14

Index Bias and Current Practices

Standard versions of indices of uneven distribution take their minimum value of zero only under the condition of exact even distribution. Most segregation researchers and consumers of segregation studies are habituated to accepting this benchmark for social integration. On reflection, however, it is an unusual point of reference for assessing segregation. For one thing, exact even distribution usually is not logically possible because individuals, families, and households cannot be distributed in fractional parts as almost always is needed to achieve exact even distribution. The resulting departure from uneven distribution is likely to be negligible when segregation is being assessed for broad group comparison using relatively large spatial units such as census tracts. But it will be non-negligible when measuring segregation for small groups and/or when using small spatial units such as blocks.

A second reason for viewing exact even distribution as an unusual reference point is that it does not correspond to the notion that race (or more generally “group membership”) is statistically unrelated to neighborhood of residence in keeping with the usual “baseline” null hypothesis adopted in studies seeking to assess quantitative group disparities on socioeconomic outcomes. To the contrary, *exact even distribution is an unexpected outcome under a model of random distribution wherein race and neighborhood are statistically independent*. Thus the occurrence of exact even distribution can signal that race is systematically associated with residence through some kind of structured social dynamic (e.g., a group quota allocation process).

As a consequence of these two factors, scores for all popular indices of uneven distribution are inherently subject to upward bias in the following sense; they have positive expected values when residential distributions of individuals and households are random and thus standard indices will signal that segregation exists even when there is no significant statistical association between group membership (e.g., race) and residential location.

Index bias is a concern for several reasons. One is that, while bias is sometimes negligible and can safely be ignored; bias can be and often is non-negligible. When

this is the case, bias can distort index scores and result in misleading assessments of the level of segregation in a particular case as well as misleading assessments of how the case in question compares with other cases including the same city at another point in time. A second reason for concern about bias is that it varies in complex ways that can make it difficult for researchers to diagnose its presence and deal with its undesirable consequences. A third reason for concern is that, because researchers are aware that bias can render index scores untrustworthy, they guard against it by foregoing many kinds of segregation studies that they would otherwise undertake if index scores could be trusted.

The current state of affairs presents difficult challenges to researchers. They want to view index bias as negligible for all cases in a given study so they can set aside concerns that assessments of segregation are untrustworthy when examining values of individual cases at a point in time, or when comparing values for a case over time, or when comparing values across cases. Unfortunately, it is not always safe to assume that scores can be trusted. In response to this situation, researchers routinely adopt multiple ad hoc strategies with the goal of avoiding and/or “dealing with” the undesirable consequences of index bias.

A few methodological studies have advocated dealing with bias directly at the point of measurement by adjusting observed scores to remove the impact of bias and obtain unbiased scores (e.g., Winship 1977; Carrington and Troske 1997; Allen et al. 2009; Mazza and Punzo 2015). To date, however, few researchers have embraced such strategies. The main reason for this appears to be that the resulting index scores are complicated to explain and interpret and the best approaches to implementing the adjustments are technically and computationally demanding.

What most researchers do instead is adopt “indirect” rather than “direct” approaches to dealing with index bias. That is, they measure segregation using “standard” (i.e., biased) versions of indices and then they adopt a variety of strategies to cope with the problem that scores may be differentially distorted by bias. Unfortunately, the strategies researchers use are a patchwork of informal, ad hoc practices. They are well-intentioned, but they are subject to criticism on multiple counts. The most important criticism is that the prevailing practices do not directly deal with index bias at the point of measurement for individual cases. Consequently, index scores for individual cases that are suspect of being distorted by bias are never “corrected” and in most studies these cases are not even identified. Consequently, index scores for individual cases affected by bias remain untrustworthy and cannot be safely used for even elementary descriptive tasks such as: assessing the level of segregation for individual cities on a case-by-case basis, making direct comparisons of segregation between any two cases, assessing differences in segregation between different group comparisons for a single city, or following a single case over time.

There is no sugar-coating the current situation. *Prevailing practices for dealing with bias do not yield trustworthy segregation index scores for individual cases.* At one level this is not surprising because the strategies researchers use to cope with index bias do not adopt the goal of obtaining trustworthy index scores for individual cases that have only a negligible amount of bias (e.g., less than 2 points). They instead employ a two-pronged strategy. They first try to screen out cases most likely

to be distorted by severe levels of bias. They then try to “work around” the problem of moderate levels of bias for many of the “surviving” cases. The main strategies researchers use in pursuing this approach are informal “rule-of-thumb” practices for screening cases from the analysis and/or minimizing the undesirable consequences of cases where bias is likely to be a non-trivial concern. Common strategies for dealing with bias include the following:

- assess segregation using larger spatial units such as census tracts instead of smaller units such as blocks;
- focus on comparisons of broad group populations and avoid comparisons involving smaller subgroups within populations – for example, compare all Whites with all Blacks instead of comparing low-income Whites with low-income Blacks;
- apply a variety of ad hoc sample restrictions to exclude potentially problematic cases in the full data set from the subset of cases used for the final analyses; and
- weight cases in the analysis data set differentially in hopes of minimizing the influence that potentially problematic cases may exert on results.

These strategies and ones similar to them are widely used primarily because they are easy to implement. More rigorous alternative approaches are available but are rarely adopted due partly because they are less well known but also because they are more complex and demanding. I view the current state of affairs with concern. First, as I noted above, the practices researchers use do not improve the measurement of index scores at the level of individual cases. Second, the “protective” practices are applied inconsistently and in patchwork fashion. Third, there is little formal methodological work to show that the practices being used are in fact effective in eliminating and/or minimizing the undesirable impact of untrustworthy index scores.

Finally, and perhaps most importantly, I worry that the “cures” adopted for dealing with index bias have undesirable side effects that in some cases may be “as bad as the disease.” In particular, prevailing practices restrict the scope of segregation studies and constrain research designs in nonrandom and ultimately undesirable ways. They shift study designs toward investigating a narrower set of questions that can be addressed using a smaller subset of cases and group comparisons where standard index scores are viewed as more trustworthy.

Obviously, this is not the situation researchers want. They would prefer to have trustworthy index scores for as many cases as possible and for as wide a range of group comparisons and research situations as possible. Happily, the difference of means framework I introduce in this monograph makes it possible to take a major step toward this goal. Working from within this framework I am able to develop refined versions of widely used indices of uneven distribution to correct the problem of index bias directly at the point of measurement. The new measures are attractive on several counts. First, they are not exotic or dramatically different. They are refined versions of popular indices and researchers do not have to adopt unfamiliar approaches to measuring uneven distribution. Second, the refinements that yield unbiased versions of indices involve minor adjustments in index calculations that are simple and easy to implement but yet very effective in providing robust protection

against index bias over a broad range of conditions and group comparisons. Third, the technical basis for achieving unbiased index scores allows researchers to continue to invoke familiar substantive interpretations of popular indices with only subtle changes. Finally, the new measures can be used at little cost or risk. When bias in fact is negligible, as sometimes is the case, scores of unbiased versions of indices track scores of standard versions very closely and the two versions will yield essentially identical results. The scores for standard and unbiased versions of indices differ only when bias is non-negligible and scores for standard versions of indices do not yield trustworthy assessments of uneven distribution.

Based on these points I suggest that the unbiased versions of indices that I introduce in this monograph provide valuable new alternatives for research. They can be used interchangeably with standard versions of indices in any situation where standard index scores can be trusted and results will be the same. But, more importantly, the unbiased versions can be used in many additional situations where standard indices cannot be safely used. Thus, the unbiased versions of index scores I introduce here expand the potential scope of segregation studies to include group comparisons and study situations that researchers currently would avoid.

I devote the remainder of this chapter to the task of “setting the stage” for introducing the unbiased versions of popular indices. I serve this goal by first reviewing the general problem of index bias. I then review the prevailing practices researchers use to try to minimize the undesirable effects of index bias and note my concerns about these practices focusing on technical questions of their efficacy considered narrowly and also on the insidious impact of these practices on segregation research more broadly. Finally, I review options that have been previously suggested for how bias might be addressed directly at the point of measurement and consider why they have not gained wider adoption. The existence of this chapter indicates that I believe it is worthwhile to review these topics in some detail. However, I will not be surprised and will take no offense if some readers choose to skip forward to Chap. 15 where I outline the basis for formulating unbiased versions of popular indices and Chap. 16 where I review their behavior in empirical applications. I turn now to reviewing basic issues and current practices.

14.1 Overview of the Issue of Index Bias

The dissimilarity index (D) is the most widely used measure of uneven distribution so it comes as no surprise that it has received especially close scrutiny on the issue of index bias. Taeuber and Taeuber (1965) provided a thoughtful early discussion of the issue in their appendix chapter reviewing issues in segregation measurement. They noted that zero, the value of D that signals integration conceived as exact even distribution, does not obtain under random distribution and furthermore is usually logically impossible even under strategic, purposive assignment because

individuals and households cannot be assigned in fractional parts (1965: 231–235).¹ Later methodological studies characterized D 's positive expected value under random assignment (i.e., $E[D] > 0$) as “bias” and raised awareness that bias in D varies in complex ways that can make scores for D problematic in many situations (e.g., Cortese et al. 1976; Winship 1977). The issue has now received regular attention for four decades and a large literature has grown with contributions from many methodological studies that have considered the nature of index bias, its practical consequences, and possible approaches for diagnosing and dealing with it (e.g., Taeuber and Taeuber 1976; Cortese et al. 1976, 1978; Blau 1977; Winship 1977, 1978; Massey 1978; Falk et al. 1978; Farley and Johnson 1985; Boisso et al. 1994; Carrington and Troske 1997; Ransom 2000; Allen et al. 2009; Mazza and Punzo 2015).

Consensus exists on many important points relating to certain technical aspects of index bias. Several key understandings trace to Winship's (1977) influential early analysis of the bias behavior of D and S . Of particular note, Winship introduced two analytic formulas for calculating the expected value of D (denoted by $E[D]$) under random distribution. Both formulas are based on a formal model of random distribution of households from two groups over areas of constant population size (t_i). He termed one formula “exact” because it implements detailed calculations based on the binomial probability distribution and can be applied at both small and large values of area population size. He termed the other formula an “approximation” because it draws on simpler calculations that yield satisfactory results when area population size is not small (i.e., when $t_i \geq 25$). Examining the approximation formula, $E[D] = 1 / \sqrt{2\pi t_i P Q}$, clarifies how $E[D]$ varies over study design and demographic conditions. Specifically, it reveals that two terms – area population size (t_i) and the relative size of the reference group (P) – determine how the value of $E[D]$ varies with city racial composition and with study design (i.e., the size of spatial units used in assessing segregation).

The first term, the area pairwise population count (t_i), has an inverse relationship with $E[D]$; all else equal, $E[D]$ declines as t_i increases. This relationship can provide a rationale for why research moved from once commonly assessing segregation using small areas such as blocks to more often using larger areas such as census tracts. It also provides a rationale for avoiding group comparisons which involve small combined populations. The practice provides a measure of protection against index bias, but this comes with substantial costs. It eliminates the option of investigating segregation in smaller cities and communities where tracts are too big to capture segregation patterns. It also eliminates the option of studying segregation involving small groups and subpopulations.

¹ The latter point is not widely appreciated but deserves greater attention because individuals typically are imbedded in a family or household that sociologically cannot be viewed as divisible and in most cases will be racially homogenous. Accordingly, Winship (1977) advocates assessing segregation for households rather than individuals because individuals within households are not “statistically independent”.

The second key term in Winship's approximation formula for $E[D]$ is PQ , the product of group population proportions. The value of this term is controlled by P – the pairwise proportion of the reference group in the combined city-wide population of the two groups. P in turn determines Q , the pairwise proportion of the comparison group, based on $Q=1-P$, and so also determines the value of PQ . The value of PQ has an inverse relationship with $E[D]$; all else equal, $E[D]$ is lower when PQ is higher. The maximum for PQ occurs when the two groups are equal in size ($P=Q=0.5$). So bias in D ($E[D]$) grows larger as groups become more imbalanced in size (i.e., as P departs from 0.5). This relationship can provide a rationale for excluding cases from analysis when one group in the segregation comparison is small in relative size. Again, the practice provides protection against index bias, but it comes at a cost; it eliminates the option of investigating segregation in communities where groups are imbalanced in size. Thus, for example, it precludes the possibility of investigating segregation in the initial stages of a new group's entry into a residential system since group size will in most cases be highly imbalanced.

Winship assessed the impact of area population size (t_i) and city racial composition (P) on index bias ($E[D]$) by tabulating the values of $E[D]$ obtained from analytic formulas over varying combinations of t_i and P . The results he reported showed that area size and city racial composition have complex, non-linear, non-additive effects on $E[D]$. Later studies confirm his findings with similar results obtained by analytic and simulation exercises investigating the issue of index bias (e.g., Carrington and Troske 1997; Allen et al. 2009; Mazza and Punzo 2015). I summarize the most important findings of these studies as follows.

- D is subject to bias under all conditions; that is, the expected value of D under random distribution always is greater than zero (e.g., $E[D] > 0$).

Significantly, the positive value of $E[D]$ truncates the range of D in empirical analyses by setting a “floor” for the minimum value below which D is unlikely to fall in the absence of exceptional circumstances (e.g., assignment of individuals and households by quota and in fractional parts).

- In many situations the index value of 0, which obtains only under exact even distribution, is not logically possible due to the integer nature of population counts and the non-independence of individuals in families and households.²
- The magnitude of bias for D varies inversely with the population size of areal units (t_i).

Other things equal, $E[D]$ grows smaller as area population size grows larger; it moves toward being negligible when area population size is very large.

- The magnitude of bias varies inversely with pairwise balance in city racial composition.

²By non-independence of individuals in families and households I mean that attaining exact even distribution would require that individuals living together within families and households would be separated and distributed independently across areas.

- Other things equal, $E[D]$ grows larger as city racial composition becomes more imbalanced. More exactly, $E[D]$ is lowest when $P = Q = 0.5$ and increases at an increasing rate as P departs further from 0.5.
- The joint impact of area population size (t_i) and city racial composition (P) on the magnitude of bias is complex. Specifically, the effects of each factor are non-additive and nonlinear such that a bias-promoting change in one factor amplifies the other factor's impact on bias.

The most important conclusion to be drawn from these studies is more general and deserves to be separated from the others.

Bias can be non-trivial in magnitude in many cases and it can vary greatly in magnitude from case to case including different cities, different group comparisons, or a given city-group comparison tracked over time.

Consequently, bias can complicate measurement and potentially lead researchers to draw incorrect conclusions about the levels and patterns of variation in uneven distribution across group comparison, across cities, and over time.

Significantly, all of the points just listed apply to all popular indices of uneven distribution except one. More specifically, the points listed above apply to the gini index (G), the Atkinson index (A), the Hutchens square root index (R), and the Theil entropy index (H). One popular measure – the separation index (S) – is an exception; index bias is less of a problem for this index than for any other widely used index of uneven distribution.

Bias for S is smaller in magnitude than for any other popular index. In addition, variation in bias for S across cases is less complicated than for any other popular index. The major reason for this is that bias for S is determined by just one factor – area population size (t_i) – with $E[S]$ being given by the simple calculation $E[S] = 1/t_i$ (Winship 1977). Thus, in contrast to other indices, bias for S does not vary with city racial composition (P). Accordingly, analyses reported in Chap. 16 show that the separation index (S) exhibits a lower level of bias than other indices under all conditions and especially when city racial composition is imbalanced. Indeed, the levels of bias for the separation index (S) are so much lower and so much less complicated than for other indices, this alone could be a compelling reason to always consider using S in empirical analyses. That said, $E[S]$ is never zero and bias can render scores for S problematic in some extreme circumstances. Consequently, while using S to measure uneven distribution can go a long way to protecting against the potential distorting impact of index bias, using S cannot in itself guarantee that bias does not adversely affect index scores.

14.1.1 *Effective Neighborhood Size (ENS): A Further Complication*

Previous methodological studies provide valuable insights about the nature of index bias. Unfortunately, however, these insights do not necessarily provide an adequate basis for diagnosing the presence of bias in empirical studies. The reason for this is that the expected values of index scores (i.e., $E[\bullet]$) under random assignment are more complicated in empirical studies than in analytic studies. Three factors pose difficulties for researchers seeking to assess and deal with index bias in empirical studies.

- Neighborhood size often varies substantially across spatial units.
- The non-negligible presence of other groups not included in the segregation comparison often varies markedly across cases.
- The extent to which other groups not included in the segregation comparison co-reside with the two groups in the comparison often varies across cases.

Each of these three factors complicates bias because they affect the value of t_i which, as noted above, plays a central role in determining the expected values of indices under random assignment (i.e., $E[\bullet]$). In empirical studies area population size (t_i) can be highly variable and this makes its impact on $E[\bullet]$ more difficult to establish. As a rule of thumb, t_i varies in predictable ways across the kind of areal units used in measuring segregation. For example, t_i is lower when using census blocks compared to census tracts. So, all else equal, one can safely expect bias will be a greater concern for blocks than for tracts. But there is a further complication in empirical studies; the population size of the areal unit used (e.g., tracts) can vary considerably across units.

The exact impact of variation in area size (t_i) on bias can be complicated to assess for a given case. But it is easy to grasp that it can be important because empirical distributions of population counts for areas often span a wide range and tend to be skewed right with unusual outliers. Variation in area population size occurs for many reasons including: differences between areas with high-density apartment buildings vs. areas with low-density, single-family detached housing; the presence of non-institutional group quarters such as work camps, college dorms, and military barracks, convents, etc.; and the presence of institutional group quarters such as prisons, facilities for the elderly and disabled, and other institutions. As a result, it can be inappropriate to use a single value of area population size (t_i) when estimating $E[D]$ by analytic formulas. As an alternative, one could extend the formulas for $E[D]$ to take account of variation in area size. Another alternative is to adopt computation-intensive methods such as estimating the sampling distribution of $E[D]$ under random distribution using city- and comparison-specific bootstrap simulations as advocated by Carrington and Troske (1997) and Allen et al. (2009).

Unfortunately, all of these options introduce complexity and substantial computational burdens and so are unlikely to be widely adopted by researchers.³

The next complication arises when other groups not in the segregation comparison are present in the city population. To see this, first note that, strictly speaking, it is not area population size per se that is relevant to index bias; it is the “pairwise” population count in the area. In view of this I introduce the term “effective neighborhood size” (ENS) to refer to the value of the combined population counts for the two groups in the comparison in the areal unit. The value of effective neighborhood size (ENS) sometimes corresponds to the value of area population size, but ENS is conceptually distinct and can depart from overall area population size. Indeed, ENS can take dramatically different values from overall area population size when the combined relative size of other groups not in the segregation comparison is large.⁴

Effective neighborhood size (ENS) equals area population size (t) only when the city population consists of just the two groups in the segregation comparison. This situation is often assumed in methodological studies to simplify analysis, but the assumption is untenable in empirical studies where the presence of other groups in the population can cause the value of effective neighborhood size (ENS) to depart dramatically from overall area population size. Under random distribution for *all* groups ENS will be smaller than area population size and estimates of index bias based on overall area population size will be too low. This can cause commonly used “rules-of-thumb” for protecting against bias to fail. For example, researchers may use census tracts as the spatial units for assessing segregation in hopes that bias will be negligible because tract populations are large. But ENS can still be low even when using census tracts if the two groups in the segregation comparison are both small. For example, this might occur when investigating the segregation of Asian subgroups (e.g., the Chinese and Korean subpopulations) or when investigating segregation across income subgroups (e.g., Whites and Blacks in the top quintile or decile of the distribution of household income).

In simple situations one could replace the value of area population size with a smaller value of ENS by multiplying average area population size by the proportionate representation of the two groups in the comparison in the total population.⁵ Unfortunately, this is inadequate because the value of effective neighborhood size (ENS) is affected by another complicating factor; namely, the extent to which the other groups in the city population co-reside with the two groups in the segregation comparison. If the other groups co-reside extensively with the two groups in the comparison (as would be the case under random distribution of all groups), ENS will be smaller than area population size (t_i) and approach its minimum possible value. All else equal, index bias would then be higher. But, if the other groups in the

³ Analytic techniques advocated by Mazza and Punto (2015) may help reduce the computational burden. But the task will still likely be too complex for wide adoption in empirical studies.

⁴ The impact of this factor has not previously been carefully studied. I provide analyses in the next chapter that show how it can have important impacts on E[D] and how these impacts can make previous strategies for dealing with index bias (e.g., Winship 1977) less effective.

⁵ For the moment I set aside the factor of variation in area size.

population are completely segregated from the two groups in the comparison, the two groups of interest will be the only groups present in the areas where they reside. In this situation the value of ENS then will take its maximum possible value and match area population size. All else equal, index bias would then be lower. The “correct” value of ENS in empirical analyses will typically fall somewhere between these minimum and maximum values depending on whether the other groups in the city population are weakly or strongly segregated from one or both of the groups in the segregation comparison. Since multi-group distributions vary widely in real cities, this issue carries complex implications for index bias and greatly complicates the assessments of index bias across group comparisons and cities.

In sum, the distinction between overall area size and effective neighborhood size (ENS) and the other complications noted above can have important practical implications for assessing index bias. When ENS is known with precision, analytic formulas for calculating expected values of bias (e.g., $E[D]$) can potentially provide a reasonable guide to identifying when bias is negligible or problematic. When one or more of the complications noted in the above discussion are present, the same formulas can yield incorrect expected values of bias. Previous methodological studies have not recognized this problem. As a result, strategies for dealing with bias that rely on estimating expected values of index scores under random distribution ($E[\bullet]$) can perform poorly in empirical studies.

14.1.2 The Practical Relevance of Variation in Effective Neighborhood Size

In the face of these complications, one option is to estimate values of $E[\bullet]$ by bootstrap simulation methods (per Carrington and Troske 1997; Allen et al. 2009, and Mazza and Punzo 2015). In principle, applying these methods with observed residential distributions can yield superior results of $E[\bullet]$ because the estimates do not depend on simplifying assumptions about the value of effective neighborhood size (ENS).

I explored using this option by examining expected values the dissimilarity index (D) for block-level segregation between Whites and Blacks for CBSAs in 2000. For this analysis I computed values of $E[D]$ by three methods. First I computed two values of $E[D]$ using Winship’s (1977) “approximation” and “exact” formulas. To establish the value of ENS to use in the formulas, I calculated the median value of ENS over blocks in the CBSA that had nonzero counts for the combined White and Black population. I additionally computed values of $E[D]$ based on bootstrap simulations that do not make simplifying assumptions about ENS.⁶

⁶Specifically, I estimated $E[D]$ by distributing White and Black individuals randomly over areas where Whites and Blacks were found in the observed residential distribution for the city and calculating D from the resulting random distribution. I repeated the exercise 1000 times for each CBSA and took the average of the obtained values of D as the estimate of $E[D]$ for the CBSA.

I found that the values of $E[D]$ based on the three methods were highly correlated ($r^2 \geq 0.95$). But, importantly, they were not exact substitutes for one another and the differences often had important consequences. First, while values of $E[D]$ correlated across methods, the average values of $E[D]$ varied by method. Values of $E[D]$ based on Winship's approximation formula were much higher than those based on the exact formula (consistent with results presented in Winship (1977)). Second, values of $E[D]$ based on bootstrap simulation methods were lower than values obtained using analytic formulas. Specifically, estimates of $E[D]$ based on the Winship's exact formula were on average 40% higher than estimates from bootstrap simulations. This indicates that assessment of bias using analytic formulas will be too high and adjustments of index scores using estimates of bias based on analytic formulas would tend to significantly "over-correct" and yield estimates of unbiased segregation that are too low.

I conducted similar exercises for other popular indices of uneven distribution; specifically, G, R, H, and S. The results for these indices were similar to what I just described for D. Estimates of bias obtained by analytic formula were higher than estimates based on bootstrap simulation methods. The key point for present concerns is that the magnitude of estimates of $E[\bullet]$ varies by method. This indicates that estimating expected values of index scores under random distribution is not a simple task in empirical studies. For now it appears that the most accurate alternative is to use the computationally demanding method of bootstrapping (per Carrington and Troske (1997) and Allen et al. (2009)) to obtain estimates of expected values ($E[\bullet]$) of measures of uneven distribution. The estimates are superior because they do not rely on strong assumptions (i.e., that areas are all the same size and that effective neighborhood size is constant across areas) but instead directly incorporate the observed variation in ENS across areas. Unfortunately, the practical burdens associated with this approach will deter most researchers from adopting the methods.

14.1.3 Random Distribution Is a Valid, Useful, and Conceptually Desirable Reference Point

The literature on segregation measurement includes many statements noting that random distribution can serve as a valid and desirable reference point for assessing segregation (e.g., Jahn et al. 1947; Reiner 1972; Zelder 1972; Cortese et al. 1976; Winship 1977; Blau 1977; Boisso et al. 1994; Carrington and Troske 1995; Carrington and Troske 1997; Ransom 2000; Allen et al. 2009; Mazza and Punzo 2015). For example, Cortese, Falk, and Cohen offer the succinct argument that it is "natural" to "construct an index which takes a value of zero when the distribution is random" (1976: 631). The unbiased measures suggested by Winship (1977), Carrington and Troske 1995, Carrington and Troske (1997), Allen et al. (2009), and Mazza and Punzo (2015) all have this property. The measures I introduce in Chap. 15 also have this property.

One obvious benefit is that when indices have this property the value of zero can then serve as the reference point for evaluating whether the index value obtained indicates that race or other group membership plays a role in segregation over and above the consequences of chance. Using indices with this quality would bring segregation research into conformity with long-standing convention in the study of group disparities in socioeconomic outcomes. Inequality research in all domains except the study of residential segregation evaluate group disparities on socioeconomic outcomes (e.g., education, occupational status, income, etc.) based on comparisons of group means that take expected values of zero when group membership (i.e., race) has no statistical association with the stratification outcome in question.

No significant objection has been or can be raised against the goal of seeking “unbiased” segregation indices with these properties. Taeuber and Taeuber (1976) and Winship (1977) have correctly noted that segregation resulting from random factors can be substantively meaningful in its own right. But this of course does not undercut the desirability of having unbiased indices whose scores provide a trustworthy signal that segregation departs from levels expected under random distribution. Winship argues that measures possessing this quality are especially desirable when interest is focused on the *causes* of segregation rather than its *consequences* (1977: 1065). Moreover, even when one is interested in the consequences of segregation, it can be valuable to know whether the segregation involved reflects systematic social dynamics, stochastic variation in residential distributions, or artifactual components of index values.

14.2 Prevailing Practices for Avoiding Complications Associated with Index Bias

I noted at the beginning of this chapter that most segregation researchers are aware of the problem of index bias and based on concern about this potential problem they routinely adopt strategies to minimize its undesirable consequences. This represents a practical compromise between the ideal of assessing and dealing with bias directly at the point of measurement – which until now has not been possible – and foregoing segregation research altogether. Researchers thus face the dilemma that segregation is an important social phenomenon that warrants sustained investigation but methodological studies establish that bias can distort segregation index scores and have adverse impacts on results and findings. Because direct solutions to this problem have not been available, researchers have adopted two general approaches for coping with concerns about index bias. One is to identify and avoid using especially problematic cases. The other is to differentially weight cases to try to minimize the impact of problematic cases.

Surprisingly, researchers almost never use direct methods of assessing bias to identify potentially problematic cases. This is difficult to understand and raises the question of why researchers use inferior proxy approaches instead of more rigorous

methods. Computation intensive bootstrap methods – which arguably yield the best estimates of $E[\bullet]$ – are relatively new and arguably are too demanding for general use. But analytic methods for assessing $E[\bullet]$ set forth in Winship (1977) have rigorous foundations and are easy to implement. It would seem that these methods provide an obvious and compelling option for identifying segregation comparisons that are most likely to be distorted by bias. Nevertheless, researchers instead rely on informal “rules of thumb” to screen cases. These informal methods tend to be crude and imprecise in comparison to available analytic methods for directly evaluating $E[\bullet]$. Common examples include the following practices.

- Restrict segregation studies to comparisons involving broad population groups; avoid comparisons involving small populations or subgroups within broader populations.
- Assess segregation using larger spatial units such as census tracts; avoid smaller spatial units such as census blocks or census block groups.
- Restrict segregation studies to only comparisons where group ratios are relatively balanced and avoid comparisons where group ratios are highly unbalanced.
- Assess segregation using full count (100 %) data; avoid sample data.
- Weight cases differentially – discounting cases presumed to be distorted by bias – when performing statistical analyses assessing variation in segregation over time or across groupings of cases and when performing regression analyses investigating cross-area variation in segregation.

The practices just listed are not necessarily all implemented in every study and the individual practices are not always implemented in exactly the same way. But almost all empirical studies adopt some combination of multiple practices similar to the ones listed above. The best justification one can offer for these “rule-of-thumb” practices for dealing with index bias is that, while they are not necessarily optimal, they are easy to implement and may be useful.

14.2.1 Unwelcome Consequences of Prevailing Practices

Researchers adopt the practices just described with the best of intentions and the practices probably do provide a measure of protection from situations where undesirable consequences of index bias are especially great. My concern is that segregation studies rely too heavily and uncritically on these informal practices. One basis for my concern can be expressed in the simple question, “Is there compelling evidence to indicate that the practices are effective in accomplishing the intended goal of eliminating undesirable impacts of index bias?” Unfortunately, the answer is “no, not really.” The practices are appropriately characterized as rough-and-ready “rules-of-thumb” whose efficacy has not been established by rigorous methodological studies.

I comment on these issues further in the next section to explain the points more carefully. But I should note here that I see these issues as secondary because it is easy to imagine substituting better practices. The more serious concern is that *even if these prevailing practices for dealing with the problems associated with index bias are refined to work as well as possible they still have the undesirable consequence of restricting the scope of segregation studies*. This issue is insidious because it is less obviously “visible.” But its impact on segregation research is substantial and far reaching.

Importantly, this undesirable consequence is not reduced when one adopts more rigorous practices for diagnosing situations where index bias is likely to be problematic. The practices researchers adopt to avoid problems associated with index bias make it impossible to conduct many studies that researchers would otherwise undertake if index bias were not a concern. The following is a list of research topics that are of clear scientific interest but currently are “off limits” because prevailing practices for dealing with index bias will preclude analyses that could address questions relating to these topics.

- studying segregation at finer levels of neighborhood resolution such as using small spatial units such as census blocks,
- studying segregation in smaller metropolitan areas and non-metropolitan areas (because segregation in these areas can only be captured well using smaller spatial units such as blocks),
- studying segregation involving populations that are small in absolute size such as Asian and Latino subgroups (e.g., Vietnamese or Salvadoran) or “first settler” and early arriving” Latino and Asian populations in new destination communities,
- studying segregation between population subgroups based on social characteristics such as education, income, family/household type, or other similar characteristics, especially considered in combination, and
- studying segregation involving groups that differ substantially in relative size.

As the situation currently stands, these and many other kinds of studies are precluded due to researchers’ concerns that index scores obtained for the comparisons involved cannot be trusted. The undesirable consequence of this is that the research literature is severely skewed toward examining a narrow subset of segregation comparisons that survive a gauntlet of restrictions placed on group comparisons, analysis samples, and study design (e.g., size of spatial unit). Accordingly, most empirical studies of segregation in the contemporary literature focus on tract-level segregation for large metropolitan areas and on group comparisons involving minority populations that are large in terms of both absolute and relative group size. Of course these cases are important and sociologically interesting in their own right. But researchers should not lose sight of the fact that this is a narrow subset of cases and is not representative of the full range of situations and group comparisons that research would consider if study designs were not narrowly restricted to reduce concerns about index bias.

This raises the concern that our understanding of segregation patterns is based on a particular subset of cases and comparisons chosen for practical, not theoretical and substantive, reasons. Equally importantly, it raises the related concern that researchers cannot undertake studies of segregation in many situations that have potentially important value for understanding segregation dynamics. For example, it is of obvious scientific interest to study the trajectory of segregation over time for new immigrant populations. But this currently is not possible because prevailing restrictions on study designs preclude the possibility of assessing segregation in the early stages of this process when the group is small in both absolute and relative size.

In some areas of inquiry the impact of concerns about index bias on the scope of segregation studies is pervasive and near-total. One example of this is the near total disappearance from the literature of studies that assess segregation at smaller spatial scales. Analysis of segregation based on block-level data once was common (Taeuber 1964; Taeuber and Taeuber 1965; Sorenson et al. 1975; Schnore and Evenson 1966; Farley and Taeuber 1968, 1974; Roof and Van Valey 1972; Van Valey and Roof 1976). Nowadays it is rare.

This change in the literature is not based on theoretical or substantive concerns. To the contrary, assessing segregation at small spatial scales has obvious substantive value because it can potentially detect segregation that might otherwise be missed. Accordingly, block-level analysis is better suited for studying the emergence of segregation patterns for newly arriving migrant or immigrant populations because patterns of segregation during their initial settlement would not be evident if segregation is measured using larger units such as census tracts.⁷ Similarly, block data are relevant for nonmetropolitan areas and non-core counties where census tracts are too large to sustain meaningful segregation analysis. But contemporary empirical studies rarely investigate segregation using block data. It is not because segregation in these settings just mentioned is substantively unimportant or scientifically uninteresting. Instead, it is because *segregation study designs have “retreated” to supposedly safer ground to avoid the complications of index bias* that arise when measuring segregation based on small areas. The unfortunate byproduct of this is that it has inhibited the investigation of segregation in smaller cities and communities.

Another closely related example is that empirical segregation studies systematically avoid examining segregation in metropolitan areas where one of the populations in the analysis is a relatively small proportion of the population or is small in absolute population size. For example, Farley and Frey’s (1994) influential study of trends in segregation from Whites for Blacks, Latinos, and Asians restricted its analysis to metropolitan areas where the minority group in the comparison either reached 20,000 in overall population or represented 3% or more of the city population. As a result, out of 318 total metropolitan areas, their analysis included only 232 areas for

⁷Lichter et al. (2010) used block-level data to study White-Latino segregation in new destinations. They offered compelling arguments for why block level data was necessary. But they did not address the problem of index bias.

White-Black segregation, only 153 areas for White-Latino segregation, and only 66 areas for White-Asian segregation.

The metropolitan areas excluded from comparison were those for which the minority group was small in relative and/or absolute size. Many of the excluded cases have non-negligible populations for the groups in question ideally would be included in studies investigating how segregation varies with basic factors such as size of city, relative group size, and trends in absolute and relative group size. However, since prevailing practices exclude cases over key ranges of these variables, many interesting research questions cannot be addressed.

Similar consequences are seen in studies of segregation among subgroups within various minority populations. For example, in a study of segregation patterns for five Asian-origin groups (Chinese, Japanese, Korean, Vietnamese, and Asian Indian), Massey and Denton (1992) restricted their analysis to metropolitan areas where the size of the Asian-origin group in question was 5,000 or higher. This limited the scope of their analysis to no more than 11 metropolitan areas for any single group. In addition, they reported segregation scores only for group comparisons where both groups in the segregation comparison had 5,000 persons and this eliminated 20–30% of possible comparisons involving other Asian-origin groups. They explicitly justified these restrictions in terms of concerns about index bias stating “Since the index of dissimilarity is inflated by random variation when group sizes get small (Massey 1978), we only compute indices when the group size in the SMSA exceeds 5,000” (Massey and Denton 1992: 171). Massey and Denton are clear that they did not adopt these restrictions on study design based on theoretical interest or other substantive concern but rather adopted the restrictions solely as a means of guarding against adverse consequences of index bias.

A final example I note is the impact on research examining racial segregation between racial groups after they have been secondarily grouped on socioeconomic status or other social characteristics relevant for group differences in residential distributions. Empirical investigations of this type routinely limit their analyses to a handful of very large cities. Furthermore, to proceed with analysis in this small subsample of cities they collapse the detailed data on socioeconomic characteristics (e.g., income) into a small number of broad groupings (e.g., 3–5 categories). Again, these restrictions in study design are adopted primarily to avoid complications associated with index bias. Evidence of this is found in the following statements from two important studies investigating racial-ethnic segregation across socioeconomic standing.

Since the number of minority members is small in some socioeconomic categories, particularly those at the upper end of the socioeconomic spectrum, we focus attention on three sets of 20 SMSAs that have the largest numbers of blacks, Hispanics, and Asians ... Focusing on the top 20 SMSAs for each group maximizes the number of minority members within each socioeconomic category and increases the stability of the segregation indices. (Denton and Massey 1988: 799–800)

Since dissimilarity indices become unreliable and difficult to interpret when the number of minority members is very small (Massey 1978), we only compute figures for those metropolitan areas where the minority population reached 5,000. Massey and Fischer (1999: 318)

The several examples reviewed above illustrate that empirical studies of segregation routinely adopt restrictions on study designs to avoid situations where index bias can complicate assessments of the level of segregation and its variation across cases. In the absence of better alternatives for dealing with index bias, these practices can perhaps be seen as necessary precautions. Nevertheless, it is important to recognize that the practices have many unwelcome consequences and it would be more desirable to have unbiased versions of indices of uneven distribution so the current restrictions on the scope of segregation studies can be relaxed.

14.2.2 Efficacy of Prevailing Practices: Screening Cases on Minority Population Size

In the ideal, the practices researchers adopt to minimize complications associated with index bias would have clear rationales and be established as effective by rigorous methodological studies. One approach would be to identify potentially problematic cases by using either analytic formulas (Winship 1977) or bootstrap methods (e.g., Carrington and Troske 1997; Allen et al. 2009; Mazza and Punzo 2015). For example, one might require that expected values of $E[D]$ be below some value deemed “acceptable” – say 3–5 points. But empirical studies of segregation do not screen cases this way nor do they report the levels and ranges of $E[D]$ for the cases in the analysis sample.

Instead, empirical studies rely on informal practices such as screening cases based on “thresholds” on absolute and relative group size. The potential concern is that this is an imprecise way to screen problem cases. I explored the issue empirically using a data set with observations on White-Minority segregation for CBSAs in 1990, 2000, and 2010. I screened cases requiring that each case have at least 2,500 persons in both groups in the decade of observation and with the smaller group in the comparison comprising at least 3% of the combined group total.⁸ Screening criteria similar to these are routine in empirical studies. Their application here yielded an analysis data set with 3,570 cases.

This result itself deserves comment. Relaxing the case selection criteria to require cases to have only 500 persons and for the smaller group in the comparison to comprise only at least one-half of one percent of the combined group total would yield 6,655 cases. The additional 3,085 cases would be highly relevant for assessing how segregation compares in smaller communities and communities where one

⁸The data set included White-Black, White-Latino, and White-Asian comparisons.

group in the comparison is small in relative size. This could apply, for example, to establishing “baselines” for White-Latino segregation in micropolitan areas and non-core counties of the Midwest and South that emerged as new destination communities experiencing Latino population growth during the period 1980–2000. Current practices do not permit these cases to be considered. The unbiased indices I introduce in Chap. 15 make it possible for researchers to focus on these communities using spatial units as small as blocks (instead of tracts) if they wish to do so.

For each segregation comparison I calculated the value of D and estimates of bias based on values of $E[D]$ obtained using both Winship’s analytic formulas and also by bootstrap methods. The question I address is whether the restrictions on the study design and analysis sample yield an analysis data set where concern about bias is negligible. The main conclusions are the same whether using either set of estimates of $E[D]$ so I report results for $E[D]$ computed by formula because few researchers are likely to compute bootstrap estimates in empirical studies. I first consider results when segregation is assessed using tract-level data, the most conservative choice for minimizing potential bias. Here the mean for $E[D]$ was 7.36. Equally and perhaps more importantly, its values displayed considerable variation across cases with an inter-decile range of 8.86 with 10% of cases at or below 3.74 and 10% of cases at or above 12.60. So the first takeaway point is that the screening criteria did not reduce the typical potential for bias to negligible levels. A second takeaway point is that screening cases did not yield an analysis data set where the potential for bias is uniform across cases. This is not surprisingly because relative group size is an important determinant of $E[D]$ and it varies widely across cities even after screening out cases where percent minority is below 3%.

Another finding is that the level of underlying potential for bias in D varies across group comparisons. The mean for $E[D]$ is 6.10 for White-Black segregation, 7.02 for White-Latino segregation, and 10.86 for White-Asian segregation. The cross-group variation traces to the fact that, on average, the relative size of the minority population is smaller for the comparisons involving Latinos and even more so for comparisons involving Asians. This raises concerns that bias might distort cross-group comparisons on segregation. The means on D are 48.48 for the White-Black comparisons, 35.13 for the White-Latino comparisons, and 39.21 for the White Asian comparisons. It is interesting to observe that the difference of 3.84 between the White-Asian and White-Latino averages for $E[D]$ is almost as large as the difference of 4.08 between the White-Asian and White-Latino averages for D .

The important point here is that the conventional approach to screening cases does not do away with nagging concerns about the potential role of bias. Furthermore, these results only get worse when segregation is measured using data at lower levels of geography such as for block-groups and blocks. For example, when calculated using block-level data, the means for $E[D]$ are 21.98 for White-Black segregation, 35.13 for White-Latino segregation, and 39.21 for White-Asian segregation. The results for $E[D]$ also varied considerably across areas and across group comparisons as observed for $E[D]$ computed using tract-level data.

14.2.3 Efficacy of Prevailing Practices: Weighting Cases by Minority Population Size

Researchers often are aware of concerns that index bias can distort results even after applying sample restrictions aimed at excluding the most problematic cases. In many studies researchers address this concern by weighting cases by minority population size for the city when performing statistical analyses such as computing summary statistics (e.g., means) for groups of cases or estimating regression equations. Unfortunately, the efficacy of this strategy is not rigorously established.

The practice is sometimes described as being an appropriate way to deal with “unreliable” cases but this rationale is open to question. Cases with biased index scores are not “unreliable” in the usual statistical sense of that term. To the contrary, biased index scores are highly reliable in the sense of yielding consistent results under given study conditions. The problem is not that the scores are inconsistent; the problem is that they are consistently high; that is, they are reliable but still untrustworthy because they are biased upward.

Weighting cases by minority population size does not “correct” the higher and potentially misleading index scores that may result from bias for some cases. So what does the practice accomplish? One clear consequence is to strongly skew analysis results in the direction of reflecting segregation patterns found in cities that have large minority populations. In most studies this means that a relatively small subset of cases will receive larger weights and have a disproportionate influence on results of statistical analyses. In contrast, a larger number of remaining cities will receive smaller weights and have modest-to-negligible influence on results. This amounts to reducing the “nominal” sample size for the macro units (usually cities) as results will be similar those obtained when excluding cases with small minority populations.

Minority population size is at best only a crude proxy for bias potential (i.e., $E[D]$). Accordingly, screening and weighting on this item can introduce at least two kinds of distortions to results. Holding relative group size constant, many smaller cities will be discounted or excluded from the analysis altogether when more careful diagnostic analysis would show that their index scores are as trustworthy as those for larger cities (because bias is intrinsically related to relative group size, not to absolute size). The practical result is that weighting cases to protect against bias will tend to be “hit and miss” in effectiveness but the practice will definitely skew results to more closely reflect segregation patterns for cities with large minority populations.

The main point is that current approach of guarding against undesirable consequences of bias based on using informal proxy criteria is open to question. Moreover, even if problematic cases were identified more carefully (e.g., using bootstrap methods to estimate $E[D]$), an important underlying problem would remain; current practices do not correct flawed scores so the cases can be trusted and used in the analysis. Instead, the cases that are impacted by index bias are excluded or discounted and analysis results thus reflect segregation patterns observed for a small

subset of cases that are not adversely impacted by bias. This is hardly an ideal study design. These cases, while important in their own right, are not necessarily representative. So one is left hoping, but not knowing, that “true” segregation patterns in the large fraction of cases that are excluded or discounted do not differ from the segregation patterns in the smaller subset of cases that dominate the analysis results.

14.2.4 An Aside on Weighting Cases by Minority Population Size

Statistical theory provides a different and potentially defensible rationale for case weighting when performing statistical analyses of variation in segregation across cities and communities. It is that the dependent variable (i.e., the index score) exhibits differential variability across cities. The relevant statistical issue is heteroskedasticity – a violation of the ordinary least squares (OLS) regression assumption that error variance is constant across cases. This issue is distinct and separate from index bias. Index bias is systematic with regard to the direction of its impact on index scores; biased cases have consistently inflated values for index scores. In contrast, heteroskedasticity does not involve bias; it involves greater volatility in scores around the model-predicted average and the volatility reflects scores that are below the predicted average as well as scores that are above the predicted average. When heteroskedasticity is present, estimates of means and regression coefficients are unbiased but significance tests in OLS regression may be questioned because the assumptions underlying the tests are not met.

One strategy for dealing with heteroskedasticity in aggregate-level regressions is to perform weighted least squares (WLS) regression using case weights (w) that are proportional to the inverse of each case’s expected error variance (Hanushek and Jackson 1977). Statistical theory indicates the appropriate weight (w) would be the reciprocal of the expected error variance of D . This can be calculated directly.⁹ But some might view absolute size of the minority population as a potentially acceptable proxy and defend weighting cases by population size on this count.

This would perhaps be justified if variation in index scores was greater when minority population size is small. But empirical analysis suggests this is not the case. This is due to two reasons, one simple and one complex. I explored the issue by examining the empirical associations among three variables – the score for D , predisposition for bias measured by $E[D]$, and minority population size – using the data set and measures introduced and described in the previous section. The simple

⁹ D is the White-Black difference of proportions ($p_1 - p_2$) residing in areas where proportion White for the area equals or exceeds that for the city as a whole. The expected variance (σ^2) of a difference of proportions is obtained by squaring the standard error of the difference of proportions – $\sigma(p_1 - p_2)$ – given by $\sqrt{p_1q_1/n_1 + p_2q_2/n_2}$ or, alternatively, $\sqrt{pq(1/n_1 + 1/n_2)}$ if using the pooled calculation of $p = (p_1n_1 + p_2n_2)/(n_1 + n_2)$.

part of the story is that values of D do not display heteroskedasticity in relation to minority population size. More specifically, dispersion in the values of D around the mean is relatively constant across minority population size so there is no obvious empirical basis for weighting cases by minority population size to compensate for heteroskedasticity.

The complex part of the story is that predisposition for bias (i.e., $E[D]$) is moderately and inversely correlated with minority population size.¹⁰ This might lead one to expect that dispersion in residuals would be larger when minority population size is small. Instead, however, the dispersion in residuals for D is lower, not higher, when $E[D]$ is high. This is because index bias raises the “floor” for D since bias precludes low scores. This then truncates the range of variation in D in comparison to the range of variation in D when $E[D]$ is low.

Since the argument for weighting cases by minority population size to deal with the statistical issue of heteroskedasticity is weak, it is appropriate to ask whether the practice is warranted on any basis. The best one can say in defense of the practice is that it may tend to reduce the influence of cases that on average have higher levels of bias (i.e., higher values on $E[D]$). But this purpose could be better served by establishing weights based on direct assessments of bias. However, even if case weights were well-calibrated to reflect bias, the practice of down-weighting cases proportional to bias is a weakly justified ad hoc procedure. It does not “repair” or “correct” inflated index values for individual cases. Misleading cases remain misleading. What the practice does accomplish is to minimize the influence of potentially misleading scores when they are averaged in with other scores that are viewed as less misleading.

If the rationale for case-weighting is not particularly strong, is it at least benign? This question is hard to answer. One thing is clear; weighting by minority population size skews results toward patterns of segregation observed in cities with large minority populations. This is definitely a non-representative subset of cities disproportionately including large cities and medium-sized cities where percent minority is higher. Whether this influences findings in undesirable ways or not is unclear and may depend on the question being addressed. If one is investigating patterns and variation in segregation *for all cities* – that is, to understand how segregation varies across cities based on urban-ecological factors (e.g., population size, racial composition, population growth, etc.) equal weighting of all cases is more appropriate. Weighting cases by minority population shifts the focus away from outcomes for all cities and toward outcomes for minority individuals residing in cities with large minority populations. Skewing results in this way may be tolerable for some research questions. But it would be best for researchers who use these practices to acknowledge the issue and reflect on how findings might be affected.

¹⁰This is because absolute minority size has a moderate association with relative group size which is intrinsically related to index bias for D and all other indices of uneven distribution except S .

14.2.5 Summing Up Comments on Prevailing Practices

In this section I have argued that the research designs of empirical studies of residential segregation are shaped in important ways by researchers' concerns about the possible undesirable consequences of index bias. Motivated by these concerns, and with the best of intentions, segregation researchers routinely adopt a variety of informal practices such as restricting analysis samples to exclude cases where they suspect bias may render index scores untrustworthy and differentially weighting remaining cases when conducting statistical analyses. The goal is to minimize the potentially undesirable impacts of bias on index scores for cases that are included in the analysis sample.

I raised concerns that the efficacy of this patchwork of informal practices is open to question on various counts not the least of which being that bias is "flagged" by crude proxies instead of by using best available direct approaches for diagnosing the potential for bias. In the final analysis, I argued that the greater concern is that, even if these prevailing practices for dealing with index bias are refined and improved, they would continue to have an important but largely unappreciated undesirable consequence. This is that the practices narrow the scope of segregation studies in two important ways. First, they restrict empirical analysis to a subset of potentially non-representative cases and group comparisons where index scores are presumed to be less problematic. Second, they eliminate the possibility of investigating many important research questions that involve situations where standard indices are viewed as prone to non-negligible bias.

Based on this I argue that the most desirable strategy all around is to deal with bias at the point of measurement and obtain index scores that are not distorted by index bias. Having unbiased index scores would make it possible to use individual cases "as is". It would eliminate the need to screen and exclude cases due to concerns about bias. It would eliminate the need to use weighting procedures to minimize the influence of cases with biased scores on results of statistical analyses. The attractiveness of this kind of solution has not been overlooked. But past efforts to deal directly with index bias at the point of measurement have not gained acceptance. I review the reasons for this in the next section.

14.3 Limitations of Previous Approaches for Dealing Directly with Index Bias

The potential benefits of dealing directly with the index bias at the point of measurement have not gone unrecognized and a variety of suggestions for developing unbiased versions of segregation indices have been offered over the decades. To this point, however, none of these suggestions has gained wide acceptance in empirical research. The kind of approach proposed most often is to adjust scores of standard versions of index scores downward to eliminate the impact of upward bias

associated with their expected values under a baseline model of random distribution (e.g., Cortese et al. 1976; Winship 1977; Farley and Johnson 1985; Carrington and Troske 1997; Allen et al. 2009; Mazza and Punzo 2015). For example, Winship (1977) and Carrington and Troske (1997) have proposed a relatively simple “norming” adjustment that has intuitive appeal.¹¹ They propose calculating “unbiased” or “bias adjusted” scores for D , designated here as D^* , based on the following calculation.

$$D^* = (D - E[D]) / (1 - E[D])$$

The justification for the calculation is that the value obtained indicates the degree to which observed departure from uneven distribution (D) exceeds the departure expected under a baseline model of random distribution (i.e., $E[D]$). In principle this adjustment can be applied to any index of uneven distribution for which the expected value under random distribution ($E[\bullet]$) can be estimated.

Unfortunately, conceptual and practical issues have worked against wide adoption of this procedure. Regarding conceptual issues, the interpretation of D^* is more technical and abstract than the interpretation of the conventional version of D . For example, negative values are possible and, while this is a valid result under the procedure, it is unsettling to many researchers. This negates one of the appealing aspects of D ; namely, the ease with which its interpretation can be conveyed to broad audiences as well as professional audiences. Regarding practical issues, the method requires estimating $E[D]$ as part of the analysis. In principle this can be accomplished using either analytic formulas or bootstrap simulation methods. But so far these options have not been embraced by segregation researchers due at least in part to the technical and computational burdens associated with estimating $E[D]$.

Prospects for adoption of this approach in the future are poor. One reason for this is the formula-based methods for estimating $E[D]$ that are most easily implemented can perform poorly when the full population of the city includes groups other than the two groups in the segregation comparison. Unfortunately, this condition is common in many research situations. It undercuts the potential value of using simple formula-based approaches to estimating $E[D]$ because estimated values tend to be too high and in turn can lead values of D^* to be too low because the adjustment to remove the impact of index bias is too aggressive. Until now this problem has gone unnoticed in the literature. In principle, the problem can be overcome by drawing on refined versions of formula-based estimates of $E[D]$ or using estimates based on bootstrap simulation methods, but complexity and increased computational burden associated with these superior approaches to estimating $E[D]$ makes it unlikely researchers will adopt these options.

¹¹ Allen and colleagues (2009) also suggest a similar strategy.

14.4 Summary

In this chapter I pointed out that empirical studies of residential segregation are strongly influenced by concerns about index bias. These concerns are reflected in the study designs researchers adopt and in the methods of statistical analyses researchers use. One important consequence of this is that researchers carefully avoid studying segregation in situations where they suspect bias will render scores of standard versions of indices of uneven distribution untrustworthy. Accordingly, they avoid studying group comparisons involving small groups; they avoid studying group comparisons where groups are imbalanced in size; they avoid measuring segregation using smaller spatial units such as census blocks; and they avoid examining segregation in smaller communities. Even after adopting these restrictions on study design, researchers continue to have concerns that bias makes some index scores untrustworthy. Analysis reviewed in the chapter shows their concern is well justified. Motivated by these concerns researchers routinely weight cases differentially based on minority population size when performing statistical analyses on the assumption that this will minimize the impact cases with scores inflated by bias will have on results. In a very real sense this has the practical effect of reducing the sample size even further and skewing it toward a non-random subset of cases. Taken collectively, these several practices limit the scope of segregation studies so attention is focused disproportionately on patterns of segregation for large metropolitan areas with minority populations that are large in absolute and relative terms. And even among this subset of cases, results of statistical analyses disproportionately reflect segregation patterns for cities with larger minority populations.

The adoption of these practices is well intentioned. But the current state of affairs is far from ideal. As things currently stand, even after restricting study designs to avoid problematic cases, researchers remain less than confident about scores for the individual cases in their studies and routinely weight cases differentially when performing statistical analysis to minimize the impact of index bias. This concern complicates elementary tasks in segregation analysis such as being confident about the index score for a given case, or comparing scores for two cases, or following the score for a single case over time. More importantly, concern about index bias leads researchers away from investigating segregation in a wide range of situations that would be theoretically relevant and sociologically interesting if index scores could be trusted.

The better alternative is to deal with problem of index bias directly at the point of measurement. Previous suggestions for accomplishing this task have involved applying after the fact adjustments to standard versions of index scores. These “bias adjusted” indices have never gained wide usage. In part this is because they have involved complex and often computationally demanding procedures. In addition many researchers find the resulting measures to be unfamiliar and therefore more difficult to interpret and explain to nontechnical audiences. Finally, researchers simply have not yet been convinced that the approach of applying corrective adjustments

to standard index scores yields robust and effective results over the wide range of situations encountered in “real world” empirical studies.

In the next chapter I introduce a new solution for moving beyond the current unsatisfactory situation. By drawing on the difference of means formulation of indices of uneven distribution, I identify new insights about the nature of index bias that make it possible to address index bias at the point of measurement. The insight is that, when segregation is cast as a group difference on average levels of scaled group contact, bias can be traced to a relatively simple source; namely, how group contact with the reference group is impacted by self-contact which inherently differs for the reference group and the comparison group. Eliminating self-contact from index calculations by assessing group contact based on “neighbors” instead of “area population” eliminates this inherent source of bias in index scores. Chapter 15 reviews the basis for establishing unbiased versions of popular indices. Chapter 16 reviews the performance of the “unbiased” versions of popular indices to establish that, as desired, they have expected values of zero under random assignment. It also makes the case that the new measures allow researchers to use familiar indices with greater confidence and dispense with most of the ad hoc practices that currently restrict the scope of segregation studies.

References

- Allen, R., Burgess, S., & Windjeijer, F. (2009). *More reliable inferences for segregation indices, Working Paper Series no. 09/2016*. Bristol: Centre for Market and Public Organisation, Bristol Institute of Public Affairs, University of Bristol.
- Blau, F. (1977). *Equal pay in the office*. Lexington: Lexington Books.
- Boisso, D., Hayes, K., Hirschberg, J., & Silber, J. (1994). Occupational segregation in the multidimensional case: Decomposition and tests of significance. *Journal of Econometrics*, *61*, 161–171.
- Carrington, W. J., & Troske, K. R. (1995). Gender segregation in small firms. *Journal of Human Resources*, *30*(3), 503–533.
- Carrington, W. J., & Troske, K. R. (1997). On measuring segregation in samples with small units. *Journal of Business & Economic Statistics*, *15*(4), 402–409.
- Cortese, C., Frank Falk, R., & Cohen, J. (1976). Further considerations on the methodological analysis of segregation indices. *American Sociological Review*, *41*, 630–637.
- Cortese, C., Frank Falk, R., & Cohen, J. (1978). Understanding the standardized index of dissimilarity: Reply to Massey. *American Sociological Review*, *43*, 590–592.
- Denton, N., & Massey, D. S. (1988). Residential segregation of Blacks, Hispanics, and Asians by socioeconomic status and generation. *Social Science Quarterly*, *69*, 797–817.
- Falk, R. F., Cortese, C., & Cohen, J. (1978). Using standardized indices of dissimilarity: Comment on Winship. *Social Forces*, *57*, 713–716.
- Farley, R., & Frey, W. H. (1994). Changes in the segregation of whites from blacks during the 1980s: Small steps toward a more integrated society. *American Sociological Review*, *59*, 23–45.
- Farley, R., & Johnson, R. (1985). On the statistical significance of the index of dissimilarity. In *The proceedings of the social statistics section of the American Statistical Association* (pp. 415–420). Washington, DC: American Statistical Association.
- Farley, R., & Taeuber, K. E. (1968). Population trends and residential segregation since 1960. *Science*, *159*, 953–956.

- Farley, R., & Taeuber, A. F. (1974). Racial segregation in the public schools. *American Journal of Sociology*, 79, 888–905.
- Hanushek, E. A., & Jackson, J. E. (1977). *Statistical Methods for Social Scientists*. Orlando: Academic Press.
- Jahn, J., Schmid, C. F., & Schrag, C. (1947). The measurement of ecological segregation. *American Sociological Review*, 12, 293–303.
- Lichter, D. T., Parisi, D., Taquino, M. C., & Grice, S. M. (2010). Residential segregation in new Hispanic destinations: Cities, suburbs, and rural communities compared. *Social Science Research*, 39, 215–230.
- Massey, D. S. (1978). On the measurement of segregation as a random variable. *American Sociological Review*, 43, 587–590.
- Massey, D. S., & Denton, N. A. (1992). Residential segregation of Asian-origin groups in U.S. metropolitan areas. *Sociology and Social Research*, 76, 170–177.
- Massey, D. S., & Fischer, M. J. (1999). Does rising income bring integration? New results for Blacks, Hispanics, and Asians in 1990. *Social Science Research*, 28, 316–326.
- Mazza, A., & Punzo, A. (2015). On the upward bias of the dissimilarity index and its corrections. *Sociological Methods and Research*, 44, 80–107.
- Ransom, M. R. (2000). Sampling distributions of segregation indices. *Sociological Methods and Research*, 28, 454–475.
- Reiner, T. A. (1972). Racial segregation: A comment. *Journal of Regional Science*, 19, 137–148.
- Roof, W. C., & Van Valey, T. L. (1972). Residential segregation and social differentiation in american urban areas. *Social Forces*, 51, 87–91.
- Schnore, L. F., & Evenson, P. C. (1966). Segregation in southern cities. *American Journal of Sociology*, 72, 58–67.
- Sorenson, A., Taeuber, K. E., & Hollingsworth, L., Jr. (1975). Indices of racial residential segregation for 109 cities in the United States 1940 to 1970. *Sociological Focus*, 8, 125–142.
- Taeuber, K. (1964). Negro residential segregation: Trends and measurement. *Social Problems*, 12, 42–50.
- Taeuber, K., & Taeuber, A. (1965). *Negroes in cities: Racial segregation and neighborhood change*. Chicago: Aldine Publishing Company.
- Taeuber, K., & Taeuber, A. (1976). A practitioner's perspective on the index of dissimilarity: Comment on Cortese, Falk, and Cohen. *American Sociological Review*, 41, 884–889.
- Valey, V., Thomas, L., & Roof, W. C. (1976). Measuring segregation in American cities: Problems of intercity comparison. *Urban Affairs Quarterly*, 11(4), 453–468.
- Winship, C. (1977). A re-evaluation of indices of residential segregation. *Social Forces*, 55, 1058–1066.
- Winship, C. (1978). The desirability of using the index of dissimilarity or any adjustment of it for measuring segregation: Reply to Falk, Cortese, and Cohen. *Social Forces*, 57, 717–721.
- Zelder, R. E. (1972). Racial segregation: A reply. *Journal of Regional Science*, 19, 149–153.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 2.5 International License (<http://creativecommons.org/licenses/by-nc/2.5/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

