

A New Twist on an Old Approach: A Random-Interaction Approach for Estimating Rates of Inter-Group Interaction

John R. Hipp · George E. Tita · Lyndsay N. Boggess

Published online: 3 February 2010

© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract There are numerous instances in which researchers wish to measure the rate of intra- or inter-group interactions (whether positive or negative). When computing such measures as rates there is great uncertainty regarding the appropriate denominator: we analytically illustrate how the choice of the denominator when calculating such rates is not trivial and that some existing strategies create a built-in relationship between the computed rate and the group composition within the entity. Another strand of prior work only focused on the relative occurrence of intra- versus inter-group events, which does not account for the important theoretical possibility that both types of events might increase in certain social contexts. Our approach provides an advance over these earlier strategies as it allows one to take into account the relative frequency of interaction between members of different groups, but then translates this into per capita rates. We also provide an empirical example using data on inter- and intra-group robbery and aggravated assault events for block groups in a section of the city of Los Angeles to illustrate how our procedure works and to illustrate how other approaches can lead to dramatically different conclusions.

Keywords Inter-racial crime · Intra-racial crime · Racial/ethnic heterogeneity · Neighborhoods · Crime · Random probability of interaction

J. R. Hipp (✉)

Department of Criminology, Law and Society, University of California, Irvine,
3311 Social Ecology II, Irvine, CA 92697, USA
e-mail: john.hipp@UCI.edu

G. E. Tita

Department of Criminology, Law and Society, University of California, Irvine,
2307 Social Ecology II, Irvine, CA 92697, USA

L. N. Boggess

Department of Criminology, University of South Florida, Tampa, FL, USA

J. R. Hipp

Department of Sociology, University of California, Irvine,
3311 Social Ecology II, Irvine, CA 92697, USA

Introduction

Scholars have long been interested in the rates of inter- and intra-group events. Though groups are able to form along many different dimensions of social life, race and ethnicity represent two of the more salient dimensions for fostering group identity. Patterns of friendship networks among adolescents (Hallinan and Williams 1989; Moody 2001; Moody and White 2003; Mouw and Entwisle 2006; Quillian and Campbell 2003), the formation of work teams within organizations (Hinds et al. 2000; Ruef et al. 2003), and inter-group marriage (Alba and Golden 1986; Gray 1987; Jones 1991; South and Messner 1986) have all been examined with a focus on the rate of intra- and inter-racial interactions. With violence being a ubiquitous feature of urban centers, and given the significant change in the demographic composition of cities over the last half century, it is not surprising that rates and patterns of intra- and inter-personal violence have also garnered much attention.

Sociologists and criminologists have sought to gain a better understanding of two issues related to inter-group events. An early body of scholarship focused on the question of the *relative* rates of inter- and intra-group crime (O'Brien 1987; Sampson 1984; South and Messner 1986); these studies focused on testing whether inter-group crime events occurred more frequently than expected by chance. A more recent body of scholarship has asked what structural determinants lead to higher rates of inter- or intra-group crime (Jacobs and Wood 1999; Messner and South 1992; Parker and McCall 1999; Wadsworth and Kubrin 2004). These more recent studies have adopted a strategy of computing *rates* of inter-group crime, and then testing whether certain characteristics of (usually) large cities or metropolitan areas are associated with such higher rates. However, actually measuring rates of inter- and intra-group interactions is a more methodologically challenging question than it may appear on the surface.

When computing inter- and intra-group crime rates, there are three key methodological issues: (1) the physical closeness (propinquity) of other group members; (2) the relative size of each group; (3) the preference for interaction with fellow group members. First, propinquity is simply the notion that those who are closest in physical space will be most likely to interact, and measuring it requires defining and then taking into account the local context in which such interactions occur. Second, the relative size of each group can be illustrated in an extreme example: a racially homogenous community will experience no inter-group crime—not because of a preference for within-group crime, but simply because there are no opportunities given the lack of other groups in the community. Third, the preference for committing crime events against members of the same group is what we are frequently interested in estimating.

Studies have attempted to account for propinquity effects when viewing inter-group crime either by using relatively small geographic units of analysis, or by including a measure of segregation when the analysis is conducted at the level of the city or metropolitan area. Sometimes missing from these efforts, however, is a careful consideration of the important implications that the relative sizes of the groups play in studies wishing to estimate *rates* of inter- or intra-group crime. A key challenge when computing these rates is the need to use an appropriate denominator. Some research has proceeded by including the population of the offender's or victim's group in the denominator (Jacobs and Wood 1999; Parker and McCall 1999), whereas other research has included the total population (Wadsworth and Kubrin 2004). As we will analytically illustrate below, these strategies do not assume a baseline rate of inter-group crime that would be of theoretical interest, and in fact create built-in relationships with the racial/ethnic composition of the unit of analysis. Our empirical example below highlights that ignoring the relative sizes of the groups in the

population when defining rates yields spurious results that are often dramatically different for the racial/ethnic composition measures than those obtained when appropriately accounting for the relative sizes of the groups.

In this paper, we propose a general method for estimating inter- or intra-group interactions that can be applied to a variety of social interactions (e.g., violence, friendship, marriage). Our method carefully accounts for the relative size of the groups under study, and then under a baseline assumption of random interaction computes the relative probability of interaction within and between group members. By testing the extent to which such interactions deviate from this baseline we are able to measure the relative preference for across-group interactions. We then propose a solution to translate these proportions into per-capita rates that can be analyzed, thereby providing a major advancement over methods that compare the *relative* occurrence of inter- and intra-group events. Thus, we are measuring the number of intra- or inter-group crimes relative to the expected number of interactions between members of those groups under an assumption of random interaction. We conclude by illustrating our approach on a study of violence in South Los Angeles, and contrast our findings with those using different intra- or inter-group crime “rates.”

Estimating Inter-Group and Intra-Group Events

Baseline Models

For much scientific inquiry, “baseline models” are used as a form of comparison. That is, baselines can be used to standardize different observations in a study to make them comparable. Per capita crime rates represent one of the simplest and most common baseline models within criminology. In order to make meaningful comparisons in terms of the aggregate amount of crime in a geographic unit (such as cities, counties, or even neighborhoods) one might assume that the amount of crime is linearly related to the number of persons in an area. Given this assumption, the number of crime events is divided by the population to create a per capita rate.

To assess the appropriateness of a baseline model, it is important to understand what is being assumed about the nature of social interactions. In the case of per capita rates, one is adopting a theoretical model in which an implicit cost function limits the number of social interactions an individual can maintain. For example, the number of friendships an individual can maintain is limited by the amount of time needed to sustain each ongoing interaction. Likewise, potential offenders are limited in the number of fellow citizens they can assault or rob. If the model did not place such a limit on interactions, then as the number of persons increase in any given context (organization, neighborhood, city, etc.) the number of social interactions would increase exponentially. Indeed, Mayhew and Levinger (1977) posed the possibility of such a model and argued that per capita rates may not be the ideal choice for certain social phenomena. They pointed out that scholars need an explicit theory of the social interaction process when considering an appropriate rate, and suggested that the nonlinear effects of dyadic interaction may be important to consider. The model using the population size as a denominator is implicitly positing that it is unreasonable to suppose that the residents of a large city such as New York City have the time to interact with any more than a small proportion of their fellow residents. Given this, the number of possible interactions will remain relatively constant over various population sizes, and therefore the linear relationship between population size and crime events is a reasonable baseline. Of course, there may be some degenerate instances at extremely small

population values in which the number of interactions will be dependent on the size of the population: i.e., a solitary individual cannot interact with anyone, two persons may weary of constant interaction, etc. We argue that such size effects will very quickly level off and become dominated by the constraint of time in which the total number of interactions changes linearly with the population size.

Alternative baseline models could be proposed: for instance, Felson (2002) argues that computing the number of motor vehicle thefts per population does not really capture the true level of exposure; rather, a more appropriate measure is the number of motor vehicle thefts per capita *automobiles*. Thus, the baseline assumption here is that the number of motor vehicle thefts will change linearly with the number of automobiles present. In the extreme, a city with no automobiles will have no automobile theft.

Focusing on the crime events committed by members of a specific group within a population presents little methodological challenge as we can propose a baseline model in which the number of crime events in a city committed by members of group A is linearly related to the number of group A members residing in the city. Indeed, this is the approach taken by virtually all studies focusing on race-specific crime in using the population size of the offender's group as the denominator to compute such rates.

However, things become considerably more complex when we turn to the question of crime events committed by members of one group against members of a specific other group (or members of their own group). The baseline model to choose in this instance is not entirely clear. One approach assumes that the number of such events increases linearly with the number of members of the offenders' group: thus, the number of crimes committed by members of group A against members of group B will increase linearly with the number of members of group A. For example, one study tested structural theories emphasizing that racial competition, racial segregation, racial composition, and economic and labor market opportunity would increase levels of intra- and inter-group homicides in large U.S. cities (Parker and McCall 1999). Another study focusing on structural characteristics asked whether different forms of social control—both economic and political—affected the level of inter-group homicides in large U.S. cities beyond the effects of the racial composition and minority group threat (Jacobs and Wood 1999). Essentially, a baseline model constructed using the size of the offender's group assumes a constant number (a quota) of crime events committed by group A members against members of group B. It is this "quota" that constrains individual actions rather than the opportunity set (i.e., group B population) and crimes against a particular group will be foregone once one's quota is achieved. In the extreme this will be problematic, as a group A member living in a city with no group B members will be unable to fulfill this quota. In such cities, this baseline model still assumes that the same number of crime events per capita will be committed by members of group A against members of group B, but this is simply not possible. Think of the city where there are very few members of group B and many members of group A: if each A fulfills their "quota" of crime events against members of group B, the consequence will be that members of group B will experience repeat victimization on an ongoing basis.

A second approach to creating a baseline model when comparing the amount of inter-group crime between geographic units uses the total population as the denominator (Wadsworth and Kubrin 2004). For example, a study asked whether such structural factors as racial inequality and the racial composition led to higher levels of inter-group homicide (Wadsworth and Kubrin 2004). This baseline model assumes that the number of crime events committed by members of group A against members of group B will increase linearly with the total size of the population, which leads to some rather peculiar

assumptions about the social world. To illustrate, consider a city with a relatively small proportion of group A members (say, 5 percent) and the rest group B members. This baseline model assumes that if we then double the number of group A members along with an equal reduction in group B members we expect to observe the *same number* of inter-group crimes committed by group A members against group B members. This implies that each group A member will commit half as many such inter-group crimes despite the fact that the number of nearby group B members has only reduced slightly. It is not clear what theoretical model would propose such a response, and therefore what we should conclude if such a baseline model was rejected. Similarly, if the city experiences an increase in the number of group B members with no decrease in the number of group A members, this is posited to linearly increase the number of crime events committed by group A members against them. This is neither consistent with the “quota” assumption described above, or the “opportunity” assumption that we describe shortly, but instead assumes a linear relationship. We are aware of no individual-level theory that would give rise to such behavior; therefore, it is difficult to know what to conclude if this baseline model is rejected.

Using the total population in the denominator presents an additional theoretical quandary: Why should one expect that crimes between groups A and B will increase linearly as the population of group C increases? For a given population, as the number of members of group A leave the area and are replaced by members of group C, this baseline model assumes that the total number of crimes committed by members of group A against members of group B will stay the same—that is, each remaining group A member will increase their quota against group B members in order to offset the events no longer committed by those departing group A members. Also, if group B members leave and are replaced by group C members, group A members are assumed to commit the same total number of crimes against group B members—i.e., each group B member will experience repeated victimization by group A members.

Note that each of these first two baseline models start with what we argue are some rather unconventional assumptions and then tests the extent to which the social world deviates from these assumptions. However, it is not clear the utility of this given that these theoretical baselines do not conform to theoretical or empirical expectations.

We instead build on a long line of literature and propose a more reasonable baseline model that assumes crime events are not explicitly committed against members of specific groups, but rather are based on randomness within a particular opportunity structure. That is, we assume that no preferences exist for within-group or between-group contact and therefore interaction probabilities are determined solely by the relative sizes of the populations. Given this baseline model, we are then able to test the extent to which behavior *deviates* from random interactions. The underlying intuition of this approach comes out of the statistics literature and is embodied in the logic of assumed randomness in the construction of cross-classified tables. As examples of this, some prior research has assumed that crime events occur randomly between members of different groups, and then tested the extent to which the observed data deviated from randomness (Messner and South 1992; O’Brien 1987; Sampson 1984; South and Messner 1986). These studies focus on *relative* rates of inter- and intra-group crime and test whether inter-group crime events occurred more frequently than expected by chance. Although this is an appropriate test of the randomness assumption, a limitation is not providing an intuitive sense of the *magnitude* of this deviation from randomness.

The segregation literature has a similar baseline model based upon randomness in which residents are randomly distributed across the neighborhoods within a city (Grannis 2002; Massey and Denton 1993). Then, the observed distribution of households across the

neighborhoods of the city is compared with this baseline to assess the degree of segregation. Note that such an approach is very similar to ours here: in each instance, the baseline assumption is one of no preference. Of course, it is not posited that this is how the world works, but rather it posits how the world would appear if there were no such preferences, and then compares the observed world with this baseline to get a sense of the degree of segregation (in the segregation literature) or the degree of inter-group violence (in the criminological literature). For instance, the index of dissimilarity measure of segregation (Duncan and Duncan 1955) assumes a random probability of settlement in the neighborhoods within an urban area: thus, the baseline model is that the geographic subunits within a larger area will have the same group proportions as the entire area. The measure then computes the extent to which the group composition in each of the subareas differs from that of the larger area and sums these totals. Such an approach is therefore similar at least in spirit to our approach for estimating inter-group crime rates.

We next describe how we compute our baseline.

Random Probability of Interaction Baseline

We begin with an assumption of no preference for within- or between-group contact. To understand the issue of computing the probability of inter- and intra-racial crime events under the assumption of no preference for within- or between-group contact, consider a simple stylized example of a single small town with 100 residents. This town is small enough geographically that all residents can potentially commit a violent act against any other person (in larger cities, it is important to account for propinquity). Suppose as well that this town is isolated from any other towns (in instances of adjacent cities or neighborhoods, the model is complicated by the possibility that residents of adjacent cities or neighborhoods can also commit violent acts against residents). This simple example allows isolating the key issues at stake, and will thus be illuminating.¹

In this simple example, the baseline probability of any violent event is simply a function of the probability of general interaction. If we assume no preferences in whom individuals interact with—hence randomness—each individual has an equal chance of interacting with each of the other 99 residents. Thus, the total number of possible interactions in this community is calculated as

$$(N)(N - 1) \quad (1)$$

where $N = 100$ is the size of the community. The reason for the second expression $(N - 1)$ is that we assume residents cannot interact with themselves. Therefore, in this example there are 9,900 possible pairs of interactions among the population.²

Now consider that this community contains members of two distinct groups. To begin, we again assume that the probability of interaction among all residents in the community is random—that is, there is no in-group bias present in which residents prefer to interact with members of the same group and this lack of in-group bias does not differ over groups.

¹ Given that our focus here is on accurately computing inter- and intra-group crime rates, we set aside such additional complications of handling propinquity issues in larger cities given that they are outside the scope of this study, can be handled as an extension of the model, and do not affect any of the results presented here.

² Note that we could complicate this further by considering the frequency of interaction. Such an extension, however, would simply attach weights to each of these potential interactions, and would thus be a straightforward extension to our model described here. We thus keep the example simple to aid in understanding the issue.

We are thus establishing a baseline assumption of no within-group preferences, and then testing the extent to which the data deviate from this. We can now compute the probability that a member of group A will initiate an interaction with another member of group A, the probability that a member of group B will initiate an interaction with another member of group B, the probability that a member of group A will initiate an interaction with a member of group B, and the probability that a member of group B will initiate an interaction with a member of group A. Suppose group A has 40 members, and group B has 60 members in this hypothetical community. The total number of expected interactions between two members of group A is

$$(N_A)(N_A - 1) \tag{2}$$

where $N_A = 40$ is the size of group A in the community. The total number of expected interactions between two members of group B is

$$(N_B)(N_B - 1) \tag{3}$$

where $N_B = 60$ is the size of group B in the community. Finally, the total number of expected interactions initiated by a member of group A with a member of group B is

$$(N_A)(N_B) \tag{4}$$

and the total number of expected interactions initiated by a member of group B with a member of group A is

$$(N_B)(N_A) \tag{5}$$

where N_A and N_B are the sizes of the two groups, as defined before. Note that these two equations allow for asymmetric relations: that is, $(N_A)(N_B)$ is the possible number of instances in which a member of group A initiates contact with a member of group B, and $(N_B)(N_A)$ is the possible number of instances in which a member of group B initiates contact with a member of group A. These two values are equal in our baseline model.

Given the total number of possible interactions in the community as defined above, we can then translate each of these measures of the total number of interactions of a particular type into a rate for the community by dividing by the total number of possible interactions. Thus, given that an interaction has occurred, the probability that it involved two members of group A is:

$$[(N_A)(N_A - 1)]/[(N)(N - 1)] \tag{6}$$

where all terms are defined as before. Given an interaction, the probability that it involved two members of group B is:

$$[(N_B)(N_B - 1)]/[(N)(N - 1)] \tag{7}$$

where all terms are defined as before. Given that an interaction occurred, the probability that the interaction was initiated by a member of group A with a member of group B, or by a member of group B with a member of group A, is:

$$[(N_A)(N_B)]/[(N)(N - 1)] \tag{8}$$

where all terms are defined as before.

For instance, in a hypothetical community in which 40 residents are members of group A and 60 residents are members of group B, having observed an interaction between two people, the probability of that interaction having been initiated by someone of group A

Table 1 Probability of interaction between racial groups for hypothetical communities with different relative group compositions

Community	Group percentages		Probabilities of interaction			
	A	B	A with A	B with B	A with B	B with A
1	0	100	0.000	1.000	0.000	0.000
2	10	90	0.009	0.809	0.091	0.091
3	20	80	0.038	0.638	0.162	0.162
4	30	70	0.088	0.488	0.212	0.212
5	40	60	0.158	0.358	0.242	0.242
6	50	50	0.247	0.247	0.253	0.253
7	60	40	0.358	0.158	0.242	0.242
8	70	30	0.488	0.088	0.212	0.212
9	80	20	0.638	0.038	0.162	0.162
10	90	10	0.809	0.009	0.091	0.091
11	100	0	1.000	0.000	0.000	0.000

Note: There are two groups: A and B. A with A indicates the probability of interaction between two group A members, etc

with another member of group A is .1576, the probability of an interaction occurring between two members of group B is .3576, the probability that the interaction was initiated by a member of group A against a member of group B is .2424, and the probability a member of group B initiated interaction with a member of group A is .2424. Note that the probability of the two types of inter-group interaction is equal. That is, minority-majority interactions are equally as likely as majority-minority interactions under these random conditions.³

Table 1 illustrates these properties by listing the probability of these four types of interaction given a range of possible community compositions. A crucial point to highlight regarding Table 1 is that the probabilities of these various types of interaction *do not change linearly* as the proportions of the various groups change. In a community in which group A constitutes just 10 percent of the population (community 2), just 0.9 percent of the interactions will be between two members of group A under randomness; however, this share of interactions increases to 3.8 percent when group A constitutes 20 percent of the population (community 3), and 8.8 percent, 15.8 percent, and 24.7 percent as their share of the population increases to 30, 40 and 50 percent, respectively. Likewise, the proportion of interactions which will occur between two members of different groups changes nonlinearly as these group shares change: the expected share of inter-group interactions when group A constitutes just 10 percent of the population (community 2) is 18.2 percent (since it is the sum of the A on B and B on A interaction probabilities). However, this increases to 32.3, 42.4, 48.5, and 50.5 percent as group A's share increases to 20, 30, 40, and 50 percent, respectively. This illustrates the nonlinearity of the baseline expected interaction rate between members of different groups given a particular group distribution in a

³ This differs from Blau's (1987) focus on the effect such proportions have on the *individual's* likelihood of interacting with members of the other group. Whereas his approach focused on the fact that in such instances individual members of the smaller group will have many more intergroup interactions than will individual members of the larger group, our focus here is on the aggregated number of such interactions.

community. This nonlinearity is generally not appreciated in prior work, though it has quite dramatic effects on estimated models as we illustrate below.

Our approach defines inter- and intra-group crime rates conditional on a random interaction assumption. We follow the long line of literature estimating per capita crime rates that implicitly posits that social interactions (or crime events) require a certain amount of time and energy. Therefore, such events are a linear function of the number of persons of the two groups in the population after taking into account this random probability of interaction. Therefore, we compute the number of intra- or inter-group crimes relative to the expected number of interactions between members of those groups under an assumption of random interaction, which implies two steps: (1) compute the conditional probability of each type of inter- or intra-group interaction given the group compositions in the community (as just described), and (2) multiply these proportions by the “community” population to translate into a per capita rate.⁴ If the community contains only these two groups, these conditional probabilities will sum to one. Intuitively, the conditional probabilities are telling us how many crime events of this type we should expect by chance given this group composition, and multiplying by the community population places these into the familiar metric of per capita crimes.

Stylized Example

We next illustrate the consequences of measuring inter- and intra-group crime rates using these different baseline models with our simple, stylized example. We will illustrate that the alternative baselines we described above yield values with built-in relationships to the racial/ethnic composition, which is frequently of theoretical interest. Consider a hypothetical instance in which not only is the probability of interaction between various residents in the community random, but that the likelihood they will commit a violent act upon a fellow resident is also random (conditional on an interaction). In this setup, the probability of a violent act between or within groups will be the same as the probability of simple interaction.⁵ There is therefore no tendency towards or against inter-group violence. Suppose that in this hypothetical community of 100 residents, 100 violent events occurred in the prior year; therefore, the per capita crime rate is 1 (as 100 crime events occurred in a community of 100 persons). To compute the number of crime events of each type, we simply multiply the total number of crime events (100) by the probabilities of interaction given in Table 1 above: this is shown in columns 4 through 7 in Table 2 for a series of hypothetical communities with different group compositions. To translate these inter- and intra-group crime events into crime *rates*, we then divide these values for crime events by the chosen population denominator.

There is general uncertainty in the literature as to which “population” is appropriate to use in the denominator when calculating intra- and inter-group crime rates, and these differing approaches use the differing baseline model assumptions we have described. We show the varying “rates” produced for hypothetical communities of differing group compositions using these alternative denominators in columns 8 to 12 in Table 2. For example, some scholars have proposed dividing intra-group crime events among members of group A by the population of group A; if this approach were taken here, the computed

⁴ The “community” can be whatever unit of analysis is employed in the particular study: cities, neighborhoods, census tracts, block groups, schools, classrooms, etc.

⁵ Of course, we could change this probability of a crime event given an interaction from 1 to .1 or .01, or any other multiple and the substantive results would be unchanged.

Table 2 Showing the relationship between various rates of inter- and intra-group crime and the proportion of groups in the community and the group heterogeneity

Community	(1)	(2)	(3)	(4) (5) (6) (7)				(8)	(9)	(10)	(11)	(12)			
				Crime events by racial type									Various rates for inter- and intra-group crime		
Group percentages		Ethnic heterogeneity		A on A crime		B on B crime		A on A crime per offender group		B on B crime per offender group		A on B crime per victim group		A on B crime per total population	
A	B	A	B	A on A crime	B on B crime	A on A crime	B on B crime	A on A crime per offender group	B on B crime per offender group	A on B crime per offender group	A on B crime per victim group	A on B crime per total population			
1	0	100	0	0.00	100.00	0.00	0.00	–	1.000	–	0.000	0.000			
2	10	90	0.18	0.91	80.91	9.09	9.09	0.091	0.899	0.909	0.101	0.091			
3	20	80	0.32	3.84	63.84	16.16	16.16	0.192	0.798	0.808	0.202	0.162			
4	30	70	0.42	8.79	48.79	21.21	21.21	0.293	0.697	0.707	0.303	0.212			
5	40	60	0.48	15.76	35.76	24.24	24.24	0.394	0.596	0.606	0.404	0.242			
6	50	50	0.5	24.75	24.75	25.25	25.25	0.495	0.495	0.505	0.505	0.253			
7	60	40	0.48	35.76	15.76	24.24	24.24	0.596	0.394	0.404	0.606	0.242			
8	70	30	0.42	48.79	8.79	21.21	21.21	0.697	0.293	0.303	0.707	0.212			
9	80	20	0.32	63.84	3.84	16.16	16.16	0.798	0.192	0.202	0.808	0.162			
10	90	10	0.18	80.91	0.91	9.09	9.09	0.899	0.091	0.101	0.909	0.091			
11	100	0	0	100.00	0.00	0.00	0.00	1.000	–	0.000	–	0.000			

Note: There are two groups: A and B. Ethnic heterogeneity is a Herfindahl index measure of heterogeneity of the two groups. Columns 8 and 10 use group A population as the denominator, columns 9 and 11 use group B population, column 12 uses the total community population as the denominator

rates would be those shown in column 8. If we were to take a similar approach for computing the rate of crime events among members of group B, we would divide events among group B members by the population of group B, as shown in column 9. For computing rates of inter-group violence, some researchers have suggested using the population of the victimized group or the population of the offending group in the denominator, and we show these computations in columns 10 and 11, respectively. Other scholars have suggested using the total population for computing inter-group violence rates, and column 12 shows these calculations for our simple example.

In comparing the different “rates” obtained by employing different denominators, we highlight some interesting results. Recall that since this example assumes random interaction among residents, “rates” should not differ as the population composition changes. Whereas our approach performs perfectly in that the “rate” of inter- and intra-group crime events holds constant at 1.0 as the relative proportions of the groups change (not shown), the other measures show some troubling results. We see that the “rate” of intra-group crime when using the size of the group’s population as the denominator is a function of the relative proportion of the community’s population constituted by the particular group in columns 8 and 9. Thus, in community 1, which is entirely composed of members of group B, we see in column 9 that the group B intra-group crime rate is indeed 1, the accurate value. However, in community 2 in which 10% of the residents are from group A and 90% are from group B, the intra-group B crime rate is now .899. This occurs despite the fact that we are still defining crime as a random event. This highlights that although the numerator (the probability of a crime event) changes nonlinearly, the denominator (the proportion of the population constituted by the group) changes linearly. Therefore, the popular approach adopted by studies of using the group population to compute a crime rate is creating a built-in relationship into the equation being estimated between the relative size of the groups and the estimate of the intra-group crime rate. This bias only gets worse as group B’s proportion of the population diminishes. Thus, in community 3 in which group B constitutes 80% of the population, their rate of intra-group crime is .798. In community 10, in which group B constitutes only 10% of the population, their rate of intra-group crime is just .091. Again, we emphasize that these rates should remain constant over these hypothesized communities since the probability of such crime events remains the same; thus, this is quite severe bias. Failing to take into account this nonlinear change in the probability of interaction as the composition of groups change quite clearly poses problems for studies simply treating this change linearly.

There are similar problems when computing the rate of inter-group crime. Whether using the population of the offender’s group as the denominator, or the population of the victim’s group as the denominator, the implications of Table 2 are that neither is a satisfactory solution. If we use the offender’s group as the denominator (column 10), we see that the rate of inter-group crime committed by group A members moves monotonically in the opposite direction of the proportion of the population constituted by group A. In community 2, in which group A constitutes 10% of the population, the A on B inter-group crime rate is .909. This falls to .808 when group A constitutes 20% of the population, and decreases to .101 when group A constitutes 90% of the population. This is quite severe bias given that the actual rate in each case should be 1. The effect moves in exactly the same way (though opposite) if we use the victim’s group in the denominator (column 11). Note a troubling implication: in a community comprised of 90% group A members and 10% group B members, using the offender’s group in the denominator will conclude that the “rate” of B on A crimes is 9 times greater than the rate of A on B crimes (.909 versus .101) even though they are in fact identical. Furthermore, the conclusion would be the exact

opposite if the researcher instead chose to use the victim's group in the denominator. This highlights that the relationship between the group composition of the community and this inter-group crime "rate" will be a monotonic function of the size of the two groups in the community—not a particularly desirable characteristic given that prior studies often are interested in *testing* the relationship between the composition of the groups in the community and the inter-group crime rate.

Similar problems arise using the total population as the denominator for constructing the rate of inter-group crime (see column 12 of Table 2.) One notable feature of this approach is that whereas the A on B inter-group crime rate increases as the proportion of the population constituted by group B goes from 0 to 50%, it then falls as this proportion goes from 50 to 100%. Thus, this is not a simple monotonic relationship, but instead is a nonlinear relationship. This induced relationship is not a trivial one, but can be rather substantial, as we will illustrate down below in our empirical example. Furthermore, the fact that these "rates" differ over these communities with differing compositions is troubling given that they should in fact be identical given our hypothetical model of randomness.

Consider one more issue: since this inter-group crime rate is a function of the squared relationship between the two groups, one might suppose that taking into account the group heterogeneity in the community would resolve this problem. The intuition is straightforward: the most common measure of the heterogeneity in a community is simply a sum of squares of the proportions of the groups in the community. To address this question, column 3 in Table 2 shows the measure of heterogeneity for each of these hypothetical communities. In the case of the inter-group crime rate using total population in the denominator, we see a nearly perfect positive correlation between this measure and the ethnic heterogeneity of the community. In fact, if we divide this rate by the ethnic heterogeneity value we obtain a constant value over each of these different community compositions. This is a reassuring finding in that it suggests that one strategy is to estimate a model in which the outcome is an inter-group crime rate using the total population as the denominator, and one of the predictors is the ethnic heterogeneity of the community *with its coefficient constrained to one*. There is no need to estimate this coefficient given that this is simply a correction for this otherwise built-in relationship. However, such an approach is problematic given that studies frequently purport to *test* the effect of racial/ethnic heterogeneity (Wadsworth and Kubrin 2004) rather than simply including it as a correction to a fundamental misspecification. In such studies, what needs to be tested is not the extent to which this parameter differs from zero, but rather the extent to which it differs from *one*: values significantly greater than one suggest heterogeneity increases inter-group violence, whereas values significantly less than one actually imply *decreased* inter-group violence.

For the measures using either the offender population or the victim population in the denominator, taking into account ethnic heterogeneity does not help. Dividing these "rates" by the ethnic heterogeneity of these hypothetical communities does not yield a constant value over these different community compositions. Thus, a relationship remains between the group compositions and these particular measures of inter-group rates even after controlling for the level of ethnic heterogeneity. Furthermore, there will frequently be a built-in relationship between the measure of ethnic heterogeneity and these "rates" of inter-group crime. Although there will be no correlation between these measures and the community ethnic heterogeneity *as long as all of these community proportions are equally plausible* in the population from which the sample derives, this non-relationship will break down if the communities tend to be clustered at one end of

this spectrum in the particular sample employed. Since minority group members constitute a relatively smaller proportion of the total population, communities tend to cluster at the lower end of the spectrum and imply a nonzero correlation between ethnic heterogeneity and the inter-group crime rate. The correlation between ethnic heterogeneity and the inter-group crime rate using the offender's population in the denominator will be nearly -1 if all the communities are composed of fewer than 50% of offender group members; this correlation will be nearly 1 using the victim's population in the denominator. It is not clear what substantive conclusions can be attached to the coefficient for ethnic heterogeneity as a covariate given the mathematical relationship between heterogeneity and such inter-group "rates."

When Might Other *Rates* be Useful?

We point out that there are some instances in which these alternative rates of inter-group events might be of interest. For example, suppose a researcher is interested in capturing the average experience with inter-group crime of residents within a particular geographic unit. In this case, dividing the number of crime events committed by group B members against group A members by the number of group A members would provide a measure of the average experience with such inter-group crime by a group A member. This might be useful if one is interested in understanding how the average experience of individual group A members with inter-group crime might impact their perceptions towards other racial/ethnic groups. Likewise, if one wished an estimate of the average number of intergroup violent events committed by members of group B against members of group A, dividing these events by the group B population would provide this information. Or, suppose a researcher is interested in capturing the average level of intra- versus inter-group violent victimization experienced by group A residents in a neighborhood. Using group A members in the denominator will yield (1) the average number of intra-group events they experience, and (2) the average number of group B on group A events experienced by group A members. Though these approaches do provide a measure of individual experience from aggregated data regarding intra- and inter-group violence (acknowledging the caveats regarding the risk of committing the ecological fallacy with such an approach), we caution that because they do not account for the population composition these approaches would not provide information about the relative tendency towards such violence given structural constraints.⁶

We emphasize that it is always crucial that the researcher first consider what exactly the theoretical construct of interest is. If one is interested in relative rates of intra- and inter-group violence in an area, we have emphasized that it is crucial to account for the relative group compositions of the area as highlighted by Blau (1977, 1987), and to employ our approach. If one is interested in certain other phenomena, other approaches can be useful. Nonetheless, the researcher must always carefully specify the construct of interest.

⁶ As an extreme example to illustrate our point, such approaches would conclude that there is little to no Eskimo on Latino violence in Los Angeles; while this would accurately reflect that Latinos in Los Angeles are not concerned about violence perpetrated by Eskimos because they do not experience it, this is trivially true given that there are essentially no Eskimos in the environment.

Empirical Example

Data

We next turn to an empirical example for a vivid demonstration of the effect such decisions on computing inter- and intra-group crime “rates” can have on substantive conclusions. Given the mathematical relationship between the group composition and the computed rates of inter- and intra-group crime we analytically showed above, we posit that these effects will be non-trivial.

For these examples, we combined data obtained from the Los Angeles Police Department (LAPD) with data from the U.S. Census on the socio-demographic characteristics of census block groups. The police data contained all UCR reports for robberies and aggravated assaults for the entire South Bureau for the period 2000–2002. We pool the data across the 3 years and construct an average count for each crime type. In addition to the race and ethnicity of the offender and victim (when known), the data also included the address of each crime, which permitted us to geocode to street addresses and then aggregate to 2000 census block group boundaries.⁷ South Bureau of Los Angeles in 2000 had a population of about 661,000 persons in 494 block groups, and was nearly 50% Latino, 37% black, 9% white, 4% Asian, and 2% other races. The average median income in these census block groups was \$30,279 with a 29 percent poverty rate.

We classified each crime based on the race/ethnicity of the offender and the victim. We coded all events for which the race/ethnicity was known for both offenders and victims. This information was known for approximately 87.6 percent of the victims of robberies and 99.8 percent of the victims of aggravated assaults; it was known for approximately 97.5 percent of the offenders of robberies and 93.7 percent of the offenders of aggravated assaults. Since we are focusing on Latinos and blacks, there are four possible types of crime⁸:

- a) black on black (c_{bb})
- b) Latino on Latino (c_{ll})
- c) black on Latino (c_{bl})
- d) Latino on black (c_{lb})

We created estimates of the conditional probability of interaction between groups based on the equations above. Thus, there is the conditional probability of within group interaction for blacks (i_{bb}) and Latinos (i_{ll}) based on Eqs. 6 and 7 above. And the conditional probability of across group interaction is based on Eq. 8 above. Because the probability of interaction between two group members is the same regardless who initiates the interaction, we only need one conditional probability interaction to handle these two possible crime types: i_{bl} for black-Latino interactions. For each of these conditional probabilities, we multiplied it by the block group population and included it in the equations we estimate. Thus, multiplying by the block group population places these into the familiar metric of per capita crimes.

⁷ Additional analyses showed the results to behave similarly when using tracts as proxies for neighborhoods.

⁸ We do not view crimes involving whites or other racial/ethnic groups because the low residential population of these groups results in unstable estimates.

Exogenous Neighborhood Measures

We included several characteristics of block groups that might be important for explaining inter- and intra-group crime rates. These measures all come from the 2000 U.S. Census. To take into account the racial/ethnic composition of the block group, we included measures of the percent Latino and percent black. We included a measure of racial/ethnic heterogeneity by using a Herfindahl index (Gibbs and Martin 1962: 670) of five racial/ethnic groups (white, Asian, Latino, black, and other race), which takes the following form:

$$H = 1 - \sum_{j=1}^J G_j^2 \tag{9}$$

where G represents the proportion of the population of ethnic group j out of J ethnic groups. To capture the effect of neighborhood economic resources, we included the block group median income. Given that broken families might reduce oversight capability, we included the percent single parent households in the block group. We measured residential stability with the average length of residence of households in the block group in 2000. We capture crowding conditions with the percentage of households classified as crowded (defined as more than one person per room). We also included two measures of group inequality: in the within race equations we included a measure of the average education achievement of Latinos or blacks (for the appropriate outcome variable), and a measure of Latino or black income inequality as measured by the Gini Index.⁹ For the inter-group crime models, we substitute for these measures *relative* group comparisons: the absolute value of the ratio of Latino to black education achievement, and the absolute value of the ratio of Latino to black logged household income. We list the summary statistics of the variables used in the analyses in the “Appendix”. Our diagnostics showed no multicollinearity problems among the variables in the analyses: the variance inflation factors did not show values larger than 8, the condition index ranged from 5.91 to 6.78, and systematically excluding particular variables from the models showed the results to be very robust.

Methods: Negative Binomial Models

We estimated negative binomial regression models for these four separate count outcomes. Thus, the count outcome is estimated with a Poisson distribution, with the additional parameter with an assumed gamma distribution that accounts for the non-independence of the crime events. For instance, the black on black violent crime equation is:

$$c_{bb} = \alpha + i_{bb} \tag{10}$$

where c_{bb} is the average number of crime events for the 2000–2002 period, i_{bb} is the conditional probability of an interaction between two black residents multiplied by the block group population size, as defined above in Eq. 4 (logged, with a coefficient

⁹ The Gini coefficient is defined as:

$$G = \frac{2}{\mu^2} \sum_{i=1}^n ix_i - \frac{n+1}{n}$$

where x_i is the household’s value of income, μ is the mean income value, the households are arranged in ascending values indexed by i , up to n households in the block group. To account for the binning in the Census data, we utilized the Pareto-linear procedure (Aigner and Goldberger 1970; Kakwani and Podder 1976), which Nielsen and Alderson (1997) implemented in the prln04.exe program provided by Francois Nielsen at <http://www.unc.edu/~nielsen/data/data.htm>.

constrained to 1), α is an intercept. To translate these count outcomes into rates, we include as an offset variable the preferred denominator, natural log transformed, with a coefficient constrained to one. We included our conditional probabilities (multiplied by the block group population as described above) as the offset measures. For instance, in the inter-group crime models we included the conditional probability from Eq. 8 (multiplied by the block group population) in the model with a coefficient constrained to one. In these models, the exponentiated value of α gives the expected number of crime events (given this random interaction assumption) per 1,000 population (when multiplied by 1,000).

We later generalize this model by including our exogenous measures of interest:

$$c_{bb} = \alpha + i_{bb} + BX \quad (11)$$

where all terms are as defined above, X is the matrix of the exogenous measures described above and B is a vector of their effects on the outcome.

To increase the efficiency of our estimates, we estimated each of these four equations separately and then combined the covariance matrices of results into a single covariance matrix before computing the standard errors. Such an approach is implemented in the *suest* command in Stata. This procedure combines the results for these models, and allows for significance testing across equations.

Results

Models Without Covariates

We begin by computing the average rates of intra- and inter-group aggravated assault rates using our baseline model, as well as the other baseline models. Given that each of these three baselines generates “rates” in different metrics, it is not informative to directly compare the raw values they produce. However, it is instructive to view the *relative* crime rates between the different intra- and inter-group crime types, given that researchers are frequently interested in comparing the relative frequency of these different types of crime. We see in Table 3 that the most frequent type of aggravated assault event, by far, is black on black aggravated assault, regardless of the denominator used. However, the relative frequency of this type of crime compared to the others differs based on the denominator: whereas our random interaction approach finds that the black on black aggravated assault rate is 7.3 times larger than the Latino on Latino aggravated assault rate, this gap is considerably narrower if the offender group population is used in the denominator (4.5) or the total population is used in the denominator (3.4). If we were to ask the degree to which blacks assault other blacks rather than Latinos, we would mistakenly conclude that this tendency is much weaker when using the population of blacks as the denominator (4.3 times as likely) than when using our random interactions approach (nearly 6 times as likely). Conversely, Latinos’ within group tendency for assaults is overestimated when using the total population or the Latino population in the denominator: it is about 4 times as likely as Latino on black assaults with these approaches, whereas it is just twice as likely with our approach. Likewise, it appears that Latinos are more likely to be assaulted by Latinos than blacks when using the total population as the denominator (about 50% more likely), whereas no such tendency is evident using our approach, as the rate of within group assaults is actually slightly lower than black on Latino assaults. Finally, although black on Latino assaults are 2.5 times as likely as Latino on black assaults using our random interaction approach, they appear 4 times as likely when using the offender population as

the denominator. Thus, not properly accounting for the probability of interaction between the group members can lead to estimates that are greatly over- or under-estimated.

Although we see less evidence for a within-group preference for robberies, we again see some sharp differences based on how the rate is computed. Whereas black intra-group robberies are nearly five times more likely than Latino intra-group robberies when using our random interaction approach, this tendency appears much smaller when using as the denominator the offender group population (3.4) or the total population (2.6). The relative tendency for blacks to be robbed by blacks rather than Latinos is overestimated when using the offender group as the denominator rather than our approach. Furthermore, the relative tendency of Latinos to rob fellow Latinos rather than blacks is overestimated when using the total population or the Latino population as the denominator (about 5 times as likely) compared to our approach (2.65 times as likely). Finally, although blacks are much more likely to rob Latinos than are Latinos to rob blacks in our approach (over 14 times as likely), this tendency is greatly overestimated when using the offender population as the denominator (23.7 times as likely).

We point out that these relative rates of crime are unsurprising when one considers the relative size of these two groups in our study area, along with the analytics we presented in Table 2. For example, given that blacks represent a relatively smaller proportion of the population, Table 2 implies that dividing black intra-group violence by the black population will result in a relative underestimate of the amount of this type of crime, and that black on Latino violence divided by the black population will be relatively overestimated (i.e., look at A on A, and A on B, crime per offender group when group A is a small proportion). And the opposite will be the case for the analogous crime rates for Latinos given that they are the larger population. Indeed, Table 3 shows that when using the population of the offender group as the denominator, black intra-group violence is relatively underestimated compared to Latino intra-group violence, and that Latino intra-group violence is overestimated relative to Latino on black violence. It also shows that black intra-group violence is underestimated relative to black on Latino violence, and that black on Latino violence is overestimated relative to Latino on black violence. All of these results are anticipated by the Table 2 analytics.

Models Including Covariates

Of particular theoretical interest is including covariates in the model to determine which characteristics of neighborhoods are associated with higher rates of inter- or intra-group crime in 2000–2002. For these analyses, we focus on four types of aggravated assaults: within group for blacks and Latinos, and the two across group types.¹⁰ To illustrate the sensitivity of the results to the denominator used when calculating these inter- and intra-group crime “rates”, we estimated four models for each crime type using the different offset variables we have discussed.

We begin by focusing on the black on black aggravated assault models, shown in Table 4. Model 1 shows that the presence of more Latinos increases black on black aggravated assaults relative to what we would expect under a random interaction assumption, whereas the presence of more blacks or racial/ethnic heterogeneity decreases

¹⁰ The pattern of results using robbery as the crime type showed similar effects. Given this, and for space constraints, we focus on the aggravated assault results here.

Table 3 Comparing three different denominators when computing intra- and inter-group aggravated assault and robbery rates from 2000 to 2002

	Random probability of interaction	Population of offender group	Total population
Computing various crime “rates”			
<i>Aggravated assault rates per 1,000 population</i>			
Black on black	57.53	19.87	7.54
Latino on Latino	7.84	4.37	2.22
Black on Latino	9.68	4.64	1.46
Latino on black	3.91	1.14	0.56
<i>Robbery rates per 1,000 population</i>			
Black on black	25.73	9.81	3.76
Latino on Latino	5.24	2.92	1.46
Black on Latino	28.11	13.62	4.17
Latino on black	1.97	0.58	0.29
Ratio of various crime “rates”			
<i>Ratios of various aggravated assault rates</i>			
Ratio of black on black to Latino on Latino	7.34	4.55	3.40
Ratio of black on black to black on Latino	5.94	4.28	5.16
Ratio of black on black to Latino on black	14.70	17.50	13.58
Ratio of Latino on Latino to black on Latino	0.81	0.94	1.52
Ratio of Latino on Latino to Latino on black	2.00	3.84	3.99
Ratio of black on Latino to Latino on black	2.47	4.09	2.63
<i>Ratios of various robbery rates</i>			
Ratio of black on black to Latino on Latino	4.92	3.36	2.58
Ratio of black on black to black on Latino	0.92	0.72	0.90
Ratio of black on black to Latino on black	13.04	17.04	13.09
Ratio of Latino on Latino to black on Latino	0.19	0.21	0.35
Ratio of Latino on Latino to Latino on black	2.65	5.07	5.08
Ratio of black on Latino to Latino on black	14.24	23.66	14.53

Note: Outcome is average crime events over the 2000–2002 period

them, controlling for the other variables in the model.¹¹ However, the results differ dramatically depending on the denominator used in calculating the rate of black on black aggravated assault. For example, the effect of percent Latinos is about 60 percent larger

¹¹ We also estimated these models without the racial/ethnic heterogeneity measure. The results were substantively similar, suggesting the absence of estimation difficulties being encountered due to simultaneously including both this measure along with the racial composition measures.

Table 4 Rate of black on black aggravated assaults, averaged from 2000 to 2002, comparing various baseline models

	Random mixing assumption	Offender's race	Total population
Latino	0.013* (0.007)	0.021** (0.007)	0.036** (0.008)
African American	-0.012* (0.006)	0.024** (0.006)	0.070** (0.007)
Ethnic heterogeneity	-0.010* (0.004)	0.007 [†] (0.004)	0.026** (0.004)
Median income	-0.020* (0.008)	-0.017* (0.008)	-0.012 (0.008)
Single parent households	0.005 (0.005)	0.005 (0.005)	0.005 (0.005)
Residential stability	0.032* (0.016)	0.032 [†] (0.016)	0.028 (0.017)
Percent crowded households	0.013* -0.005	0.014** (0.006)	0.016** (0.006)
Within-race education achievement	-0.07 (0.081)	-0.091 (0.079)	-0.132 [†] (0.080)
Within-race income inequality	0.008 (0.006)	0.013* (0.006)	0.020** (0.006)

Standard errors in parentheses. *N* = 468 block groups. Negative binomial regression models using various offset measures. All models include an intercept

** *p* < .01(two-tail test)

* *p* < .05 (two-tail test)

[†] *p* < .05 (one-tail test)

when using the offender's race (model 2) as the denominator, and about 180% larger when using the total population as the denominator (model 3). We also see that the effects of more blacks or more racial/ethnic heterogeneity appear to have *positive* effects when using one of these other denominators, in direct contrast to the negative effects detected using our approach. These findings mimic our analyses above, which highlighted that these alternative baselines for computing intra-group crime rates create erroneous results because of the built-in mathematical relationships with the group composition of the neighborhood.

The results are more consistent across baselines for the non-race covariates in the models. Although our analytics above highlighted that other techniques for calculating intra- and inter-group crime "rates" will induce mathematical relationships with the neighborhood group composition, the degree to which other covariates in the model are biased due to this mis-specification will depend on their degree of correlation with the outcome measure and these racial composition measures. For example, although median income generally reduces black on black aggravated assaults, the size of this effect is 16% smaller when using the offender's race as the offset, and 38% smaller (and non-significant) when using the total population as the offset compared to our approach. Likewise, although the effects of residential stability are generally positive, the size of this effect is 12 percent smaller (and non-significant) in the model using the total block group population as the offset. In contrast, the effect for the percent crowded households is 28% larger when using the total population as the offset compared to our approach. The effects are even more dramatic for the measure of household income inequality among blacks: compared to our random probability of interaction model with a nonsignificant finding, the effect size increases about 60% when using the offenders' race as the offset (and is now significant) and about 240% when using the total population as the offset. This highlights that whereas

Table 5 Rate of Latino on Latino aggravated assaults, averaged from 2000–2002, comparing various baseline models

	Random mixing assumption	Offender's race	Total population
Latino	−0.030** (0.006)	−0.004 (0.006)	0.027** (0.006)
African American	−0.001 (0.005)	−0.001 (0.005)	0.002 (0.005)
Ethnic heterogeneity	−0.017** (0.004)	−0.003 (0.004)	0.011** (0.004)
Median income	−0.023** (0.009)	−0.023** (0.008)	−0.023** (0.008)
Single parent households	0.010 [†] (0.006)	0.010 [†] (0.006)	0.011 [†] (0.006)
Residential stability	0.003 (0.024)	0.009 (0.023)	0.014 (0.023)
Percent crowded households	0.004 (0.005)	0.004 (0.005)	0.004 (0.005)
Within-race education achievement	0.024 (0.056)	0.013 (0.054)	0.020 (0.051)
Within-race income inequality	−0.004 (0.006)	−0.004 (0.006)	−0.002 (0.006)

Standard errors in parentheses.
N = 479 block groups. Negative binomial regression models using various offset measures. All models include an intercept

** *p* < .01 (two-tail test)
* *p* < .05 (two-tail test)
[†] *p* < .05 (one-tail test)

the covariates in the model other than the racial composition measures generally are minimally affected by the choice of a denominator for the intra-group crime rate, sometimes there can be quite strong differences depending on the correlations among the variables in the model.

We see equally dramatic effects when viewing the rate of Latino intra-group aggravated assaults in Table 5. Using our random interaction assumption as the offset, the number of such assaults decreases as the percentage Latino in the neighborhood increases. This suggests that ethnic enclaves will result in lower rates of within group assaults among Latinos, which is certainly a plausible finding. On the other hand, this effect is essentially zero when using the Latino population as the offset measure, and actually flips sign when using the total block group population as the offset. Likewise, we see that although the ethnic heterogeneity in the neighborhood reduces the number of Latino intra-group assaults using our random interaction assumption, this effect is non-significant when using the Latino population as the offset, and again reverses sign when using the total population as the offset. The non-race measures in the model have relatively similar effects regardless which “rate” is used as the outcome, as they generally show weak effects.

Turning to the two inter-group crime types, we again see dramatic differences based on how these rates are calculated. For black on Latino aggravated assaults (shown in Table 6), an increase in the percentage Latino modestly decreases this type of crime using our random interaction assumption. However, if we use the offender's race as the offset the sign reverses, and this sign reversal shows an even stronger effect when using the total population as the offset. Whereas our approach shows that increasing the percent blacks has no effect on black on Latino assaults, there appears to be a strong positive effect when using the Latino population or the total block group population as the offset. And the effect of ethnic heterogeneity is dramatically different when employing our offset rather than the

Table 6 Rate of black on Latino aggravated assaults, averaged from 2000–2002, comparing various baseline models

	Random mixing assumption	Offender's race	Victim's race	Total population
Latino	-0.015 [†] (0.009)	0.017* (0.008)	0.002 (0.010)	0.039** (0.010)
African American	-0.007 (0.008)	-0.004 (0.008)	0.043** (0.008)	0.050** (0.009)
Ethnic heterogeneity	-0.019** (0.006)	0.003 (0.006)	0.006 (0.007)	0.027** (0.006)
Median income	-0.028** (0.009)	-0.030** (0.009)	-0.031** (0.010)	-0.032** (0.010)
Single parent households	0.014 [†] (0.007)	0.014* (0.007)	0.012 (0.007)	0.012 [†] (0.007)
Residential stability	0.048 (0.030)	0.055 [†] (0.029)	0.043 (0.032)	0.050 [†] (0.030)
Percent crowded households	0.017** (0.006)	0.017** (0.006)	0.021** (0.006)	0.020** (0.006)
Absolute value of Latino to black education ratio	0.577 (0.488)	0.386 (0.399)	0.526 (0.532)	0.359 (0.431)
Absolute value of Latino to black income ratio	0.226 (0.165)	0.185 (0.160)	0.231 (0.163)	0.199 (0.155)

Standard errors in parentheses. *N* = 455 block groups. Negative binomial regression models using various offset measures. All models include an intercept

** *p* < .01(two-tail test)

* *p* < .05 (two-tail test)

[†] *p* < .05 (one-tail test)

total population as the offset. The non-race measures in the model are generally similar regardless which denominator we use. One exception is that the effect of crowded households on inter-group assaults is about 20 percent larger when using the number of Latinos or the total population as the offset rather than our approach.

Finally, we point out that the results are equally inconsistent when measuring Latino on black aggravated assault rates, as seen in Table 7. Using our random interaction assumption, an increase in the percentage Latinos in the block group, or an increase in racial/ethnic heterogeneity, increases Latino on black crime. Using the size of the victim's or offenders' racial group as the offset, however, yields non-significant results for these two measures, and using total population as the offset yields opposite conclusions. We also see that our approach finds a negative effect of an increasing black population for Latino on black aggravated assaults, whereas using the black population or the total block group population as the offset yield a strongly *positive* effect. Among the non-race variables, the effect of the Latino to black education ratio differs considerably when using these various baselines: compared to our random interaction assumption, the size of the effect is about 20% smaller when using the black population or the Latino population as the offset, and 36% smaller (and non-significant) when using the total population as the offset. Likewise, the size of the effect for the Latino to black income ratio is about 15% smaller when using the Latino population or the total population as the offset compared to our approach. This

Table 7 Rate of Latino on black aggravated assaults, averaged from 2000–2002, comparing various baseline models

	Random mixing assumption	Offender's race	Victim's race	Total population
Latino	−0.021* (0.009)	−0.003 (0.011)	0.009 (0.009)	0.030** (0.010)
African American	−0.029** (0.008)	0.022* (0.009)	−0.027** (0.008)	0.027** (0.009)
Ethnic heterogeneity	−0.023* (0.009)	0.005 (0.009)	−0.004 (0.009)	0.022* (0.009)
Median income	0.003 (0.013)	0.002 (0.014)	0.001 (0.014)	0.001 (0.014)
Single parent households	0.015 [†] (0.008)	0.014 (0.008)	0.015 [†] (0.008)	0.014 [†] (0.008)
Residential stability	−0.016 (0.028)	−0.021 (0.030)	−0.009 (0.028)	−0.015 (0.030)
Percent crowded households	0.006 (0.009)	0.011 (0.009)	0.006 (0.009)	0.010 (0.009)
Absolute value of Latino to black education ratio	0.867* (0.377)	0.684 [†] (0.400)	0.719* (0.332)	0.556 (0.349)
Absolute value of Latino to black income ratio	0.401* (0.197)	0.343 [†] (0.179)	0.377 [†] (0.201)	0.331 [†] (0.177)

Standard errors in parentheses. $N = 455$ block groups. Negative binomial regression models using various offset measures. All models include an intercept

** $p < .01$ (two-tail test)

* $p < .05$ (two-tail test)

[†] $p < .05$ (one-tail test)

again highlights that whereas there will generally be few differences for the other measures in the model, this is not a certainty and is dependent on the correlations among the variables in the model.

Conclusion

Given the considerable theoretical interest in assessing the sources of inter- and intra-group social interactions, it is crucial that researchers appropriately translate these events into rates. We proposed a measure that accounts for the number of intra- or inter-group crimes relative to the expected number of interactions between members of those groups under an assumption of random interaction. Our procedure (1) appropriately accounts for the probability of interaction within and across groups; and (2) converts this into per capita rates. This latter feature allows researchers to test the relationship between hypothesized covariates and these types of inter- and intra-group relationships. It also allows viewing how each of these inter- and intra-group rates changes over time. We are aware of no solutions that address each of these issues. Whereas prior work has appropriately accounted for such a random interaction probability in testing whether inter-group crime events occurred more frequently than expected by chance (O'Brien 1987; Sampson 1984),

such a statistical test does not lend itself to testing the effects of covariates on inter- and intra-group crime separately. Given that there may be reason to test for the actual *rate* of inter-group crime, rather than its prevalence relative to intra-group crime, we suggest that our proposed procedure has considerable practical use for researchers.

We have demonstrated that the choice of denominator when computing inter- and intra-group rates is crucially important. Our analytic results highlighted that using the offender's population, the victim's population, or the total population as the denominator in such rates will induce built-in relationships between the racial composition and these rates. Given the interest in testing whether the racial composition of a community affects the rate of inter-group crime, this is a serious failing. As a consequence, some of the results of prior research testing the effect of certain community characteristics on rates of inter-group crime may need to be revisited (Jacobs and Wood 1999; Parker and McCall 1999; Wadsworth and Kubrin 2004). Our examples using a dataset of crime events in block groups in one area of the city of Los Angeles dramatized these effects. It should be emphasized that these were not small bias effects. Instead, we saw on numerous occasions that the conclusions of such models could be completely reversed based on the decision of how to compute the intra- or inter-group crime rate. For example, in a model predicting Latino on black aggravated assaults, the effect of percent black was completely reversed from a negative to a positive coefficient based on the choice of denominator. Likewise, the effect of racial/ethnic heterogeneity was completely reversed from a negative to a positive effect depending on the chosen denominator for both black on Latino, and Latino on black, aggravated assaults. And the conclusion that Latinos engage in less intra-group violence when living in Latino-dominated barrios was completely reversed when using a different denominator, leading to the mistaken conclusion that they engage in *more* intra-group violence in such scenarios. Clearly, these are not trivial decisions.

Of some solace was the finding that other variables in such models will generally be less dramatically affected by the denominator used when constructing the intra- or inter-group crime rate. In many instances, the size of the effect for these other covariates will be similar. Nonetheless, it must be emphasized that this need not always be the case. In fact, this will depend on the degree to which such measures correlate with the racial composition measures and the outcome measure. If these additional variables have a relatively substantial correlation with the racial composition measures, it is quite possible that the estimate of their multivariate effect on inter- and intra-group crime rates will be substantially affected by how the crime rate is computed. In some instances, the bivariate correlation of these measures with the outcome will be substantially changed based on the rate used. Whereas our analytics along with the findings from this empirical example are a caution to researchers that the choice of a denominator when creating intra- or inter-group crime rates requires careful consideration, our proposed solution provides a useful resource for scholars wishing to test the effects of various measures on either inter- and intra-group crime rates in future work.

We also emphasize that although we have focused on inter- and intra-group crime rates in this example, it is a straightforward generalization to utilize our procedure in other instances in which researchers wish to measure inter- or intra-group events. For instance, researchers wishing to study the rates of inter- or intra-group school friendships, rather than simply their relative occurrences, can adopt this procedure. Given that some contextual characteristics may affect *both* intra- and inter-group events similarly, approaches that simply compare the *relative* rates of inter- and intra-group interactions alone are inadequate. That is, a characteristic that affects both inter- and intra-group crime similarly will not be detected in a study that focuses only on their relative occurrence. Our procedure

allows researchers to test the effect of contextual characteristics on rates of intra- and inter-group interactions *separately*.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

Appendix

See Table 8.

Table 8 Summary statistics for variables used in the analyses

	Mean	SD
Latino	48.30	25.00
African American	36.62	7.60
Ethnic heterogeneity	44.92	14.47
Median income	30.28	13.84
Single parent households	24.88	11.93
Residential stability	10.91	3.67
Percent crowded households	31.30	17.16
Within-race education achievement: Latinos	9.50	1.81
Within-race income inequality: Latinos	36.89	10.79
Within-race education achievement: African Americans	12.63	1.51
Within-race income inequality: African Americans	41.73	13.55
Absolute value of Latino to black education ratio	2.37	1.20
Absolute value of Latino to black income ratio	1.49	1.43

Note $N = 494$ block groups

References

- Aigner DJ, Goldberger AS (1970) Estimation of Pareto's law from grouped observations. *J Am Stat Assoc* 65:712–723
- Alba RD, Golden RM (1986) Patterns of ethnic marriage in the United States. *Social Forces* 65:202–223
- Blau PM (1977) A macrosociological theory of social structure. *Am J Sociol* 83:26–54
- Blau PM (1987) Inequality and heterogeneity: a primitive theory of social structure. Free Press, New York
- Duncan B, Duncan OD (1955) A methodological analysis of segregation indexes. *Am Sociol Rev* 20: 210–217
- Felson M (2002) *Crime and everyday life*. Sage, Thousand Oaks
- Gibbs JP, Martin WT (1962) Urbanization, technology, and the division of labor: international patterns. *Am Sociol Rev* 27:667–677
- Grannis R (2002) Discussion: segregation indices and their functional inputs. *Sociol Methodol* 32:69–84
- Gray A (1987) Intermarriage: opportunity and preference. *Popul Stud* 41:365–379
- Hallinan MT, Williams RA (1989) Interracial friendship choices in secondary schools. *Am Sociol Rev* 54:67–78
- Hinds PJ, Carley KM, Krackhardt D, Wholey DR (2000) Choosing work group members: balancing similarity, competence, and familiarity. *Organ Behav Hum Decis Process* 81:226–261
- Jacobs D, Wood K (1999) Interracial conflict and interracial homicide: do political and economic rivalries explain white killings of blacks or black killings of whites? *Am J Sociol* 105:157–190
- Jones FL (1991) Ethnic intermarriage in Australia, 1950–52 to 1980–82: models or indices? *Popul Stud* 45:27–42
- Kakwani NC, Podder N (1976) Efficient estimation of the Lorenz curve and associated inequality measures from grouped observations. *Econometrica* 44:137–148

- Massey DS, Denton NA (1993) *American apartheid: segregation and the making of the underclass*. Harvard, Cambridge
- Mayhew BH, Levinger RL (1977) Size and the density of interaction in human aggregates. *Am J Sociol* 82:86–110
- Messner SF, South SJ (1992) Interracial homicide: a macrostructural-opportunity perspective. *Social Forum* 7:517–536
- Moody J (2001) Race, school integration, and friendship segregation in America. *Am J Sociol* 107:679–716
- Moody J, White DR (2003) Social cohesion and embeddedness: a hierarchical conception of social groups. *Am Sociol Rev* 68:103–127
- Mouw T, Entwisle B (2006) Residential segregation and interracial friendship in schools. *Am J Sociol* 112:394–441
- Nielsen F, Alderson AS (1997) The Kuznets curve and the great U-turn: income inequality in U.S. counties, 1970 to 1990. *Am Sociol Rev* 62:12–33
- O'Brien RM (1987) The interracial nature of violent crimes: a reexamination. *Am J Sociol* 92:817–835
- Parker KF, McCall PL (1999) Structural conditions and racial homicide patterns: a look at the multiple disadvantages in urban areas. *Criminology* 37:447–477
- Quillian L, Campbell ME (2003) Beyond black and white: the present and future of multiracial friendship segregation. *Am Sociol Rev* 68:540–566
- Ruef M, Aldrich HE, Carter NM (2003) The structure of founding teams: homophily, strong ties, and isolation among US entrepreneur. *Am Sociol Rev* 68:195–222
- Sampson RJ (1984) Group size, heterogeneity, and intergroup conflict: a test of Blau's inequality and heterogeneity. *Social Forces* 62:618–639
- South SJ, Messner SF (1986) Structural determinants of intergroup association: interracial marriage and crime. *Am J Sociol* 91:1409–1430
- Wadsworth T, Kubrin CE (2004) Structural factors and black interracial homicide: a new examination of the causal process. *Criminology* 42:647–672